

REGRESSÃO COM ERROS DE MEDIDA E
PONTOS DE MUDANÇA UTILIZANDO
METODOLOGIA BAYESIANA

Daisy Gomes de Souza Tu

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO GRAU
DE
DOUTOR EM ESTATÍSTICA

Área de Concentração: Estatística

Orientador: Prof. Dr. Heleno Bolfarine

- São Paulo, dezembro de 2003 -

REGRESSÃO COM ERROS DE MEDIDA E
PONTOS DE MUDANÇA UTILIZANDO
METODOLOGIA BAYESIANA

A tese foi defendida por Daisy Gomes de Souza Tu e aprovada pela Comissão Examinadora em 14 de maio de 2004. Este exemplar corresponde à redação final, devidamente corrigido segundo as recomendações dessa Comissão.

Comissão Julgadora:

- Prof. Dr. Heleno Bolfarine (orientador) - IME-USP
- Prof. Dra. Pilar L.Iglesias - PUC-CHILE
- Prof. Dr. Hélio Migon - IM-UFRJ
- Prof. Dr. Filidor E. V. Labra - IMECC-UNICAMP
- Prof. Dra. Marcia Branco - IME-USP

*Dedico este trabalho ao meu
Senhor e Salvador Jesus, que
me amou e deu a sua vida por
mim.*

Resumo

O objetivo principal desse estudo foi analisar os modelos de regressão linear estrutural Normal e t -Student com erros nas variáveis, com mudança abrupta em um ou mais parâmetros após o k -ésimo ponto desconhecido (ponto de mudança) de uma seqüência finita de observações, utilizando métodos bayesianos de análise de dados.

A análise inferencial envolveu o problema de detectar a existência de mudança nos parâmetros, estimar o ponto de mudança k e os demais parâmetros, e também fazer a análise preditiva.

O modelo Normal com erros de medida e sem pontos de mudança, foi analisado sob diferentes escolhas de distribuições à priori incluindo também prioris impróprias.

Foram demonstrados teoremas que estabelecem condições para a existência das distribuições posterioris nos modelos Normal e t -Student com erros dependentes (com e sem ponto de mudança) sob algumas escolhas específicas de prioris impróprias.

Alguns algoritmos do tipo MCMC (“Markov Chain Monte Carlo”) foram utilizados com o objetivo de amostrar das posterioris, como os algoritmos de “Gibbs”, “Grouped Gibbs”, “Modified and Collapsed Gibbs” e “Metropolis-Hastings em Gibbs”. Foi feita uma análise de sensibilidade a variações nos valores dos seguintes hiperparâmetros: a razão das variâncias residuais (nos modelos identificáveis) e o número de graus de liberdade da distribuição t (quando este é assumido conhecido).

A qualidade do ajuste dos modelos foi avaliada utilizando resíduos bayesianos de validação cruzada.

Abstract

The main purpose of this study was to analyze the Normal and Student-t linear regression with measurement error models with an abrupt change in a finite sequence of observations, using bayesian methods for data analysis.

The inferential analysis considered the problem of detecting the existence of parameter changes, estimating the change point k and the remaining parameters, and also doing predictive analysis.

The Normal measurement error model without change point was analyzed under different prior distributions including improper priors.

Conditions for the existence of the posterior distribution under some specific improper prior distributions were demonstrated for the Normal and Student-t with dependent error models (with and without change point).

Some MCMC ("Markov Chain Monte Carlo") were used with the purpose of sampling from the posteriors like for instance: Gibbs, Grouped Gibbs, Modified and Collapsed Gibbs and Metropolis-Hastings in Gibbs.

A sensitivity analysis to different values of the following hyperparameters was done: the ratio of the residual variances (on the identifiable models) and the number of degrees of freedom of the Student-t distribution (when it was assumed known).

To asses model adequacy, bayesian cross-validation residuals were used.

Agradecimentos

A Deus, por ter me sustentado nos momentos difíceis da doença de meu marido e durante toda a trajetória desse trabalho. A Ele sejam dadas a Honra, a Glória e o Poder para sempre.

Ao meu orientador, o Professor Heleno Bolfarine, realmente pelo privilégio que tive de trabalhar sob sua orientação e pelo respeito, confiança, compreensão e sua enorme dedicação durante todo o desenvolvimento desse trabalho.

Ao Professor Chen Ming-Hui, pela orientação que possibilitou o desenvolvimento dos teoremas sobre a existência da posteriori do modelo estrutural Normal do Capítulo 2, e também por checar meus primeiros programas computacionais durante o semestre em que estive na “Worcester Polytechnic Institute – U. S. A.”.

À Professora Pilar Iglesias, eu realmente agradeço pelas sugestões e esclarecimentos nos vários encontros que tivemos.

Ao Professor Alexandre Kawano da Escola Politécnica, pelas dicas na fase de demonstração de um dos teoremas sobre a existência da posteriori no modelo t -dependente.

Ao Walter eu agradeço imensamente por seu trabalho notável e eficiente na edição dessa tese, e também por conseguir sempre arranjar um tempo para esse trabalho, e não somente isso, mas por sua enorme paciência.

Aos meus queridos amigos Marcia, Dione, Luiz Gustavo, Arthur, Denise, Oswaldo e Cátia, pela amizade, boa vontade e encorajamento com que sempre me presentearam.

À CAPES e ao CNPq, respectivamente, pela bolsa de doutorado no país e a bolsa sanduíche (PDEE).

Finalmente, em especial, à minha mãe, por suas orações e sua força e também

por eu poder sempre contar com seu amor incondicional. Também ao meu marido, que sempre desejou e procurou fazer o melhor por mim.

Sumário

1	Introdução	1
1.1	Caracterização geral dos problemas de ponto de mudança	1
1.2	O problema alvo desse estudo	3
1.3	Levantamento bibliográfico	6
2	Regressão linear Normal com erros nas variáveis	11
	Introdução	11
2.1	O modelo estrutural e o modelo funcional com erros Normais na estatística clássica	12
2.2	O modelo estrutural e o modelo funcional na estatística bayesiana	15
2.2.1	O modelo bayesiano estrutural	16
2.2.2	O modelo bayesiano funcional	23
2.3	O modelo estrutural identificável com particular priori de componentes independentes	25
2.4	Métodos computacionais descritos na literatura	28
2.5	Implementação de algoritmos para o modelo da seção 2.3	29
2.5.1	Distribuições condicionais completas à posteriori baseadas nos dados observados	29
2.5.2	Distribuições condicionais completas à posteriori baseadas nos dados completos (D)	32
2.5.3	Implementação do algoritmo original de Gibbs	32
2.5.4	Implementação do algoritmo de Gibbs em blocos	34

2.5.5	Implementação do algoritmo “modified collapsed Gibbs”	36
2.5.6	Implementação do algoritmo híbrido de Metropolis-Hastings em Gibbs	38
2.6	O modelo estrutural não identificável com particular priori de componentes independentes	42
2.7	O modelo estrutural com particular priori própria de componentes condicionalmente independentes	46
2.7.1	O modelo com restrição de identificabilidade	46
2.7.2	O modelo sem restrições de identificabilidade	49
2.8	Estabelecimento das condições para a existência da distribuição à posteriori sob diferentes escolhas de prioris impróprias	52
2.8.1	Condições necessárias e suficientes no caso do modelo estrutural identificável	53
2.8.2	Condições necessárias e suficientes no caso do modelo não identificável	76
2.9	Estimação do modelo	81
2.10	Avaliação do ajuste e análise de sensibilidade	84
2.10.1	Sensibilidade do modelo a diferentes valores da razão das variâncias residuais	85
2.10.2	Avaliação da qualidade do ajuste	85
2.11	Algumas medidas de comparação de modelos	98
2.12	EXEMPLOS	103
2.12.1	Análise do modelo estrutural normal com restrição de identificabilidade	103
2.12.2	Análise do modelo estrutural normal sem restrição de identificabilidade	118
2.12.3	O modelo estrutural com restrição de identificabilidade e priori imprópria	126

2.12.4	O modelo estrutural sem restrição de identificabilidade e priori imprópria	136
2.12.5	Resumo dos resultados:	143
3	O modelo linear estrutural identificável t-Student com erros nas variáveis	146
	Introdução	146
3.1	O modelo estrutural t -independente com o número de graus de liberdade conhecido	149
3.1.1	Definição do modelo	149
3.1.2	Implementação de algoritmos MCMC no modelo com particular priori própria de componentes independentes	155
3.1.3	Distribuição preditiva	157
3.1.4	Avaliação do ajuste	158
3.2	O modelo estrutural t -independente com graus de liberdade desconhecidos	163
3.2.1	Definição do modelo	164
3.2.2	Implementação de algoritmos MCMC no modelo t -independente para duas particulares escolhas de priori própria	165
3.3	O modelo estrutural t -dependente com o número de graus de liberdade conhecido	166
3.3.1	Definição do modelo	166
3.3.2	Implementação de algoritmos MCMC no modelo com particular priori própria de componentes independentes	170
3.3.3	Distribuição preditiva	171
3.3.4	Avaliação do ajuste	172
3.4	O modelo estrutural t -dependente com graus de liberdade desconhecidos	172

3.5	Condições para a existência da posteriori no modelo estrutural t -Student com erros dependentes e número de graus de liberdade conhecido	173
3.6	Sensibilidade do modelo a diferentes valores do número de graus de liberdade	182
3.7	Algumas medidas de comparação de modelos	183
4	Uma análise do modelo estrutural Normal com erros nas variáveis com um ponto de mudança desconhecido	187
	Introdução	187
4.1	O modelo estrutural com um ponto de mudança e priori com componentes independentes	189
4.2	A análise do modelo que permite mudança em todos os parâmetros com particular priori de componentes independentes	194
4.2.1	Distribuição à posteriori	195
4.2.2	Distribuições condicionais completas à posteriori baseadas nos dados completos	195
4.2.3	Estimação do modelo que permite mudança em todos os parâmetros	197
4.3	Implementação do algoritmo de Gibbs no modelo que permite mudança em um subconjunto fixado de parâmetros	199
4.4	Detecção de mudança	202
4.4.1	Procedimento 1 – Um teste baseado na distribuição à posteriori de k	203
4.4.2	Procedimento 2 – Uso da distribuição à posteriori da diferença (ou quociente) entre os parâmetros.	208
4.4.3	Procedimento 3 – Uso de critérios de seleção ou escolha de modelos	208
4.5	Distribuição preditiva e avaliação do ajuste	218
4.5.1	Distribuição preditiva	219

4.5.2	Densidade condicional preditiva de validação cruzada	220
4.5.3	Resíduos de validação cruzada	223
4.6	Estabelecimento das condições para a existência da posteriori sob algumas escolhas de prioris impróprias	225
4.6.1	Condições necessárias e suficientes no modelo que permite mu- dança em todos os parâmetros	225
4.6.2	Condições suficientes no modelo que permite mudança apenas em σ_x^2	228
4.7	EXEMPLOS	231
4.7.1	Análise do modelo estrutural normal identificável	231
4.7.2	Análise do modelo estrutural não identificável	245
4.7.3	Análise do procedimento 1 para detetar mudança baseada na distribuição à posteriori de k	259
4.7.4	Análise do procedimento 2 para detetar mudança baseada na distribuição à posteriori das diferenças (ou quocientes)	269
4.7.5	Análise do procedimento 3 para detetar mudança baseada nos critérios de seleção de modelos.	275
4.7.6	Resumo dos resultados	278
5	Uma análise do modelo linear estrutural t-Student com erros nas variáveis com um ponto de mudança desconhecido	281
5.1	O modelo t com erros independentes, com particular priori de compo- nentes independentes	281
5.1.1	O modelo que permite mudança em todos os parâmetros após o ponto de mudança k	281
5.1.2	Implementação do algoritmo de Gibbs no modelo que permite mudança em todos os parâmetros	284
5.1.3	Implementação do algoritmo de Gibbs no modelo com mu- dança em um subconjunto qualquer de parâmetros	286

5.2	O modelo t -dependente com particular priori de componentes independentes	288
5.2.1	O modelo que permite mudança em todos os parâmetros após o ponto k	288
5.2.2	Implementação do algoritmo de Gibbs no modelo com mudança em um subconjunto qualquer de parâmetros	290
5.3	Deteccção de mudança e estimação do modelo	291
5.4	Condições para a existência da posteriori no modelo t -dependente sob diferentes escolhas de prioris impróprias	291
5.4.1	Condições no caso do modelo que permite mudança em todos os parâmetros	292
5.4.2	Condições no caso do modelo com mudança apenas em σ_x^2	295
5.5	Distribuição preditiva e avaliação do ajuste	298
6	Considerações Finais	299
6.0.1	Resultados do estudo sobre ponto de mudança	299
6.0.2	Perspectivas para trabalhos futuros	303
	Apêndice	305
	Seção 2.4 Métodos Computacionais Descritos na Literatura	305
2.4.1	Classificação Geral	305
2.4.2	Métodos MCMC: idéias básicas	310
	O algoritmo de Metrópolis-Hastings	312
	O algoritmo original de Gibbs	316
2.4.3	O algoritmo de Gibbs em Blocos	318
2.4.4	O algoritmo “Collapsed Gibbs”	320
2.4.5	O algoritmo Híbrido de Metropolis-Hastings em Gibbs	324
2.4.6	Diagnósticos de Convergência	326
	Seção 2.5.1	334
	Seção 2.8	340

Seção 4.4.3	344
(a) O critério do Fator de Bayes	344
(b) O critério do Pseudo Fator de Bayes – formulação geral	347
(c) O critério “BIC” (“Bayesian Information Criterion”)	347

Referências bibliográficas	348
-----------------------------------	------------

Capítulo 1

Introdução

1.1 Caracterização geral dos problemas de ponto de mudança

Dizemos que um problema tem um ponto de mudança se dada uma seqüência de observações ordenadas pelo tempo ou espaço ou outra variável qualquer, a distribuição dessas observações muda após um determinado ponto desconhecido dessa seqüência. Podemos definir o problema da seguinte maneira:

Dada a seqüência de variáveis aleatórias (ou vetores aleatórios) z_1, z_2, \dots com distribuições dadas respectivamente por F_1, F_2, \dots onde os índices $1, 2, \dots$ são, por exemplo, “instantes de tempo”. Dizemos que o índice k , ou o ponto k , é o ponto de mudança desconhecido dessa seqüência se $F_1 = F_2 = \dots = F_k \neq F_{k+1} = F_{k+2} = \dots$

Vale salientar ainda o fato de que as observações não precisam ser tomadas necessariamente em intervalos “de tempo” igualmente espaçados, e que a definição no caso de múltiplos pontos de mudança é semelhante.

Um exemplo de aplicação é ilustrado em Worsley (1983) que estudou a incidência de um defeito de nascença denominado “Talipes” numa região da Nova Zelândia entre 1960 e 1976. No ano de 1965 o herbicida 245-T foi utilizado pela primeira vez nessa região e o interesse era saber se isso coincidia com um aumento aparente da incidência de crianças nascidas com essa doença. O modelo proposto considerou o

número de crianças nascidas com Talipes no ano i como tendo distribuição binomial $B(n_i, p_i)$, $i = 1, 2, \dots, 17$, onde n_i = número total de nascimentos no ano i e p_i = probabilidade de uma criança nascer com a doença no ano i . O autor propôs um teste para testar $H_0 : p_1 = p_2 = \dots = p_{17}$ versus $H_1 : p_1 = \dots = p_k \neq p_{k+1} = p_{k+2} = \dots = p_{17}$, onde k é o ponto de mudança desconhecido. O interesse era então saber se o ponto de mudança k coincidia com a data de introdução do pesticida, para que providências necessárias pudessem ser tomadas (como a verificação, por meio de uma análise bioquímica, de uma possível relação causal e, nesse caso, a tomada de medidas governamentais para a comercialização do produto).

Segundo Brodsky e Darkvosky (1993) os problemas de ponto de mudança tem vários aspectos e podem ser classificados por exemplo quanto às seguintes características:

- (1) Quanto ao método de obtenção dos dados. Temos dois tipos de problemas, os seqüenciais e os de tamanho de amostra fixados. Nos procedimentos seqüenciais os dados são examinados (quanto à homogeneidade) a cada nova observação. Nesse caso, a seqüência de observações $\{z_i\}_i$ é infinita (já que o tamanho da amostra é determinado por uma regra de parada que depende dos dados). Nos problemas de amostras fixas (problemas retrospectivos), o processo de aquisição de dados é completado antes da hipótese de homogeneidade ser testada. Nesse caso, temos uma seqüência finita $\{z_i\}_{i=1}^n$.
- (2) Quanto à informação à priori sobre F_i . Nesse caso, podemos ter problemas de ponto de mudança paramétricos, semi-paramétricos e não-paramétricos.
- (3) Quanto às características dos dados. Podemos ter modelos com tempo discreto ou tempo contínuo, uni ou multidimensionais, como também os modelos com observações dependentes ou independentes.

- (4) Quanto ao tipo de mudança. Temos os modelos com mudanças abruptas e os modelos com mudanças graduais. Um exemplo de mudança abrupta num modelo de regressão linear simples sem erro de medida é quando $Y_i = \alpha_1 + \beta_1 X_i$, para $i = 1, \dots, k$, e $Y_i = \alpha_2 + \beta_2 X_i$, para $i = k + 1, \dots, n$, e os X_i 's não estão necessariamente em ordem crescente no tempo. No caso, por exemplo, em que $X_1 \leq X_2 \leq \dots \leq X_n$ e além disso existe γ tal que $\alpha_1 + \beta_1 \gamma = \alpha_2 + \beta_2 \gamma$, onde $X_k \leq \gamma < X_{k+1}$, temos um modelo com mudança gradual, ou sem descontinuidade, conhecido como “regressão segmentada” ou “two-phase regression”.
- (5) Quanto ao número de mudanças. Nessa categoria podemos ter os modelos com um único ponto de mudança e os modelos com múltiplos pontos de mudança.

Para a inferência em modelos com pontos de mudança temos basicamente dois aspectos a serem considerados. O primeiro é o de determinar se existem mudanças e o segundo é o de localizar a posição desses pontos (dado que há mudança), como também estimar as outras quantidades de interesse e também fazer a análise preditiva.

1.2 O problema alvo desse estudo

Nesse trabalho o objetivo principal é fazer inferência sobre o modelo de regressão linear estrutural com erros nas variáveis e com mudança abrupta nos parâmetros após o k -ésimo “instante de tempo” (desconhecido) numa seqüência finita de vetores de observações independentes $\begin{pmatrix} Y_1 \\ X_1 \end{pmatrix}, \dots, \begin{pmatrix} Y_n \\ X_n \end{pmatrix}$ indexadas pelo “tempo”. Nesse trabalho vamos estudar também modelos que permitem mudanças em um subconjunto fixado qualquer de parâmetros e que a relação entre as variáveis observáveis, Y_i e X_i , permanece linear ao longo do “tempo” (isto é, antes e depois do ponto de mudança).

k); além disso vamos assumir que tanto a variável aleatória latente como os erros de medida são normalmente distribuídos, com parâmetros que podem mudar ao longo do “tempo”. O caso em que os erros têm uma distribuição t -Student também é analisado.

A análise inferencial, que envolve o problema de detetar a existência de mudança, localizar a(s) posição(ões) do(s) ponto(s) de mudança, estimar os demais parâmetros e fazer a análise preditiva é desenvolvida utilizando-se métodos bayesianos.

Na literatura, até o momento, muito poucos resultados têm sido publicados nessa linha; temos basicamente os artigos de Chang e Huang (1997) e Chong (1999).

Chang e Huang (1997), utilizando metodologia clássica (frequentista) analisaram o modelo que permite mudança apenas na média da covariável e somente no caso em que os erros (e a covariável) são normalmente distribuídos. Nas aplicações o modelo de Chang e Huang (1997) é muito limitado, já que permite mudança apenas na média da covariável, o que não é muito realista para a maioria dos problemas de regressão na prática. Outra limitação é que o procedimento para detectar mudança proposto pelos autores está baseado nas propriedades assintóticas da estatística da razão de verossimilhança e a eficiência desse teste para pequenas (ou médias) amostras ainda não foi estudada. Além disso, os resultados assintóticos sobre a consistência dos estimadores no caso em que há mudança estão baseados em hipóteses extremamente fortes.

Chong (1999) estudou também o modelo linear com erros nas variáveis do ponto de vista da estatística clássica e se limitou ao caso de mudança apenas no coeficiente angular da reta de regressão (mantendo os demais parâmetros sem mudança); além disso, mostrou a consistência do estimador do ponto de mudança e derivou a distribuição assintótica de um teste tipo Wald para detetar mudança.

Nossa proposta é mais geral em vários sentidos:

- (1) Podemos analisar modelos que permitem mudanças em todos os parâmetros (como também em qualquer subconjunto de parâmetros de interesse); conseqüentemente, podemos escolher o “melhor modelo”. Nas aplicações isso é vital, já que uma análise exploratória dos dados pode não “revelar” a modelagem mais adequada pelo fato de que o ponto de mudança é desconhecido.
- (2) Deixando à parte razões filosóficas, a metodologia bayesiana para tratar de problemas de pontos de mudança é a mais natural e conveniente por diversas razões práticas.
 - (a) Não se apóia em resultados assintóticos: Do ponto de vista bayesiano, adotada uma função perda, os estimadores que minimizam o risco à posteriori são ótimos. Segundo Zacks (1983), a obtenção, por exemplo, dos estimadores de máxima verossimilhança e a derivação de suas propriedades assintóticas pode ser bem complicada, mesmo nos problemas mais simples de ponto de mudança. Segundo Smith (1975) e Hinkley (1970), além do problema assintótico, os procedimentos baseados somente na verossimilhança apresentam dificuldades nos problemas de ponto de mudança, pelo fato de que a função de verossimilhança é frequentemente multimodal.
 - (b) A mesma metodologia de análise que é usada no modelo que permite mudança em um subconjunto qualquer de parâmetros é usada também no modelo que permite mudança em todos os parâmetros, ao contrário do procedimento de Chang e Huang (1997) que funciona somente para o caso de mudança na média (e que talvez possa eventualmente ser estendido somente para o caso de mudança no intercepto da reta de regressão, mantendo-se os demais parâmetros sem mudança).
 - (c) Usando metodologia bayesiana podemos investigar a existência de múltiplos pontos de mudança.

tiplos pontos de mudança de forma bem simples, usando por exemplo o método da segmentação binária descrito no capítulo 4, que consiste simplesmente em aplicar repetidas vezes o mesmo procedimento construído para detetar um único ponto de mudança. Na estatística clássica temos o problema do nível de significância e portanto nem sempre isso pode ser feito.

- (d) Do ponto de vista computacional, a análise bayesiana também é relativamente simples, principalmente utilizando as técnicas computacionais do tipo MCMC disponíveis atualmente na literatura. Além disso, é suficiente implementar basicamente um único programa computacional.

Assim, de modo geral, a estatística bayesiana para problemas de ponto de mudança é (a) mais natural, (b) tecnicamente mais simples, (c) permite incorporar ao modelo informação subjetiva disponível na área da aplicação e (d) permite-nos obter mais resultados.

1.3 Levantamento bibliográfico

Na literatura, os primeiros estudos sobre pontos de mudança apareceram na década de 50 principalmente na área de controle de qualidade e tratavam principalmente do problema de mudança na média de uma seqüência de variáveis aleatórias independentes e normalmente distribuídas. Posteriormente estendeu-se o estudo nas áreas de regressão com observações independentes, séries temporais, análise de sobrevivência e outras, de modo que a literatura sobre os problemas de ponto de mudança é atualmente muito vasta.

O modelo de regressão linear simples com mudança abrupta (não contínua) após um ponto de mudança desconhecido, no caso em que a covariável é medida *sem*

erros e o tamanho da amostra é fixado (caso não seqüencial, ou retrospectivo) foi descrito por Quandt (1958, 1960), que usou o método de máxima verossimilhança para estimar os parâmetros do modelo e derivou um teste baseado no critério da razão de verossimilhança para testar a existência de mudança nos parâmetros da equação de regressão. Nessa linha (Estatística Clássica), muitos trabalhos têm sido publicados desde então para tratar desse problema. Sob o enfoque da Estatística Bayesiana esse modelo foi estudado por Ferreira (1975) para o caso em que o número de pontos de mudança é conhecido. As posteriores dos parâmetros, seus valores esperados como também intervalos de credibilidade e testes foram derivados.

Continuando a focalizar na literatura bayesiana desses modelos de regressão, apresentamos a seguir alguns dos trabalhos relevantes nessa linha:

Broemeling e Choy (1981) desenvolveram um teste bayesiano para detectar mudança nos parâmetros da equação de regressão do modelo com no máximo um ponto de mudança. Esse teste é baseado na distribuição marginal à posteriori do ponto de mudança, onde a probabilidade à posteriori de que não haja mudança, $p(K = n)$, é comparada com a probabilidade de que haja uma mudança ($1 - p(K = n)$).

Holbert (1982) investigou o problema de estimação do modelo de regressão linear múltipla onde ocorre mudança abrupta em um único ponto, como também a estimação do modelo de regressão simples segmentado (“two-phase regression”) usando priori imprópria e obtendo a distribuição de probabilidade marginal à posteriori do ponto de mudança.

Moen et al. (1985) generalizou o trabalho de Choy e Broemeling (1980) estudando o modelo linear multivariado usando uma priori própria Normal Multivariada – Wishart e obtendo as marginais à posteriori do ponto de mudança, bem como das matrizes de regressão e da matriz de precisão.

Land e Broemeling (1983) estudaram o modelo linear geral que admite um único

ponto de mudança (mudança abrupta) num período fixado, assumindo que todos os parâmetros são desconhecidos e tomando uma particular priori própria. Eles também derivaram a distribuição preditiva à posteriori de observações futuras.

Moen e Broemeling (1985) obtiveram a densidade preditiva à posteriori para o modelo linear multivariado com um único ponto de mudança (mudança abrupta) na matriz dos parâmetros da regressão, também sob o enfoque retrospectivo, ou seja, não sequencial.

Kim (1991) apresentou um teste bayesiano para testar se existe mudança na reta de regressão usando a região de credibilidade HPD.

Carlin (1992) fez uma abordagem geral de modelos hierárquicos bayesianos com um único ponto de mudança (abrupta) incluindo uma aplicação em regressão, utilizando o amostrador de Gibbs para estimar as densidades marginais à posteriori de interesse.

Wang e Lee (1993) estudaram o modelo com mudança abrupta apenas no intercepto da reta de regressão, obtendo a distribuição marginal à posteriori tanto do ponto de mudança como da diferença dos interceptos (e também dos parâmetros que não mudam); além disso utilizaram a priori de Jeffreys comparando com a priori “uniforme” encontrada na maioria dos estudos.

Leiva (1994) estendeu o trabalho de Broemeling e Choy (1981) para o caso em que o número de pontos de mudança é desconhecido.

Lyazrhi (1997) estudou o problema de mudança abrupta no modelo de regressão linear normal com no máximo um ponto de mudança propondo um procedimento bayesiano ótimo para testar a hipótese H_0 de que não há nenhuma mudança versus a hipótese $\bigcup_{i=1}^{n-1} H_i$, onde H_i é a hipótese que pelo menos um parâmetro muda após a i -ésima observação. O procedimento é baseado na função perda quadrática quando escolhermos a hipótese H_i ao invés da verdadeira hipótese H_j .

Lyazrhi e Caussin (1997) propõem um procedimento global baseado na teoria de decisão bayesiana para detetar no máximo p pontos de mudança (onde p é desconhecido) no modelo de regressão linear Normal.

Os trabalhos citados até este ponto tratam dos problemas não sequenciais de regressão linear com erros independentes onde a(s) covariável(eis) *são medidas sem erros* e a(s) mudança(s) não são contínuas (isto é são abruptas).

No caso dos modelos com erros nas variáveis e mudança nos parâmetros temos muito poucos resultados na literatura e somente sob o enfoque clássico. Fizemos um levantamento bibliográfico pelo CIS (Current Index to Statistics) desde 1975 e apuramos os seguintes trabalhos de Teeter (1979, 1985), Gbur e Dahm (1985), Chang e Huang (1997) e Chong (1999).

Teeter (1979) estudou o modelo linear segmentado com erros nas variáveis formado por duas retas que se interceptam e mostrou que o estimador do ponto de intersecção das duas retas, baseado nos estimadores de mínimos quadrados dos coeficientes das retas, não é assintoticamente não-viesado e derivou uma aproximação assintótica para o erro quadrático médio desse estimador. Teeter (1985), estudando ainda o mesmo modelo, propõe uma técnica geral para obter uma estimativa do ponto de mudança da intersecção dos dois submodelos que se interceptam.

Gbur e Dahm (1985) estudaram também o modelo segmentado simples com erros nas variáveis, derivando o estimador do ponto de intersecção das duas retas pelo método dos momentos e obtendo sua distribuição assintótica.

Grimshaw (1992) estudou o modelo linear-platô com erros nas variáveis em duas situações; primeiramente tomando-se observações repetidas para X_i (para um mesmo Y_i) e também tomando-se observações repetidas para cada elemento do par (X_i, Y_i) . O autor analisou o estimador de mínimos quadrados dos parâmetros quanto à consistência em ambos os casos.

Os resultados de Chang e Huang (1997) e Chong (1999), descritos brevemente na seção anterior, são os mais relevantes no contexto de regressão com erros nas variáveis com ponto de mudança.

Capítulo 2

Regressão linear Normal com erros nas variáveis

Introdução

Inicialmente, sob o enfoque da estatística clássica, comparamos o modelo linear estrutural simples com erros nas variáveis (no caso em que os erros são normalmente distribuídos e aditivos) com o modelo linear funcional sem pontos de mudança. A seguir comparamos esses dois modelos na formulação bayesiana.

No restante do capítulo concentramo-nos na análise bayesiana do modelo de regressão linear estrutural simples com erros (e covariável) normalmente distribuídos para várias escolhas de distribuição à priori. Estudamos tanto o modelo com restrição de identificabilidade como o modelo sem restrições de identificabilidade (aproveitando o fato de que a metodologia bayesiana não está limitada apenas a situações de identificabilidade).

Para amostrar das posteriores implementamos quatro algoritmos do tipo “Markov Chain Monte Carlo”: o algoritmo de Gibbs, o “Grouped Gibbs”, o “Modified and Collapsed Gibbs” e o algoritmo de “Metropolis-Hastings em Gibbs”.

Nesse capítulo também derivamos quatro teoremas que estabelecem as condições necessárias e suficientes para a existência das distribuições posteriores partindo de

um conjunto específico de prioris impróprias.

Uma análise de sensibilidade com relação a diferentes valores da razão das variâncias residuais (λ) também é feita para o modelo identificável.

A qualidade do ajuste é também avaliada usando resíduos bayesianos de validação cruzada.

Finalmente descrevemos alguns métodos de seleção ou escolha de modelos encontrados na literatura aplicados ao modelo estrutural Normal.

2.1 O modelo estrutural e o modelo funcional com erros Normais na estatística clássica

O Modelo Estrutural

O modelo linear estrutural aditivo simples com erros de medida e covariável normalmente distribuídos é definido pelas equações

$$Y_i = y_i + e_i, \text{ com } y_i = \alpha + \beta x_i,$$

$$\text{e } X_i = x_i + u_i,$$

onde x_i é uma variável aleatória latente (não observável), e_i e u_i são as componentes de erro e X_i e Y_i são variáveis aleatórias observáveis para $i = 1, 2, \dots, n$. Além disso, assumimos que $(e_i, u_i, x_i)^T$ são independentes e identicamente distribuídos (*i.i.d.*) com

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \stackrel{\text{i.i.d.}}{\sim} N_3 \left(\begin{pmatrix} 0 \\ 0 \\ \mu \end{pmatrix}; \begin{pmatrix} \sigma_e^2 & 0 & 0 \\ 0 & \sigma_u^2 & 0 \\ 0 & 0 & \sigma_x^2 \end{pmatrix} \right), \text{ para } i = 1, 2, \dots, n. \quad (2.1.1)$$

O espaço paramétrico do modelo é o conjunto

$$\Theta = \{ \theta = (\alpha, \beta, \mu, \sigma_x^2, \sigma_u^2, \sigma_e^2)^T : \alpha, \beta, \mu \in \mathbb{R} \text{ e } \sigma_x^2, \sigma_u^2, \sigma_e^2 \in \mathbb{R}^+ \}.$$

Como consequência de (2.1.1), os dados “completos” $(X_i, Y_i, x_i)^T, i = 1, \dots, n$

têm a seguinte distribuição normal:

$$\begin{pmatrix} Y_i \\ X_i \\ x_i \end{pmatrix} \stackrel{\text{i.i.d.}}{\sim} N_3 \left(\begin{pmatrix} \alpha + \beta\mu \\ \mu \\ \mu \end{pmatrix}; \begin{pmatrix} \beta^2\sigma_x^2 + \sigma_e^2 & \beta\sigma_x^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma_e^2 & \sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 & \sigma_x^2 \end{pmatrix} \right), \quad (2.1.2)$$

para $i = 1, 2, \dots, n$.

Os dados observados $(Y_i, X_i)^T, i = 1, \dots, n$, têm a seguinte distribuição conjunta:

$$\begin{pmatrix} Y_i \\ X_i \end{pmatrix} \stackrel{\text{i.i.d.}}{\sim} N_2 \left(\begin{pmatrix} \alpha + \beta\mu \\ \mu \end{pmatrix}; \begin{pmatrix} \beta^2\sigma_x^2 + \sigma_e^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma_u^2 \end{pmatrix} \right), \text{ para } i = 1, 2, \dots, n, \quad (2.1.3)$$

Seja $D_{\text{obs}} = \{(Y_i, X_i)^T, i = 1, \dots, n\}$ o conjunto dos dados observados e $D = \{(Y_i, X_i, x_i)^T, i = 1, \dots, n\}$ os “dados completos”. Então a **função de verossimilhança baseada nos dados observados** é dada por

$$\begin{aligned} L(\theta|D_{\text{obs}}) &\propto |\Sigma|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2|\Sigma|} \left[(\beta^2\sigma_x^2 + \sigma_e^2) \sum_{i=1}^n (X_i - \mu)^2 \right. \right. \\ &\quad \left. \left. - 2\beta\sigma_x^2 \sum_{i=1}^n (Y_i - \alpha - \beta\mu)(X_i - \mu) + (\sigma_x^2 + \sigma_u^2) \sum_{i=1}^n (Y_i - \alpha - \beta\mu)^2 \right] \right\}, \quad (2.1.4) \end{aligned}$$

onde $|\Sigma| = \beta^2\sigma_u^2\sigma_x^2 + \sigma_e^2\sigma_x^2 + \sigma_u^2\sigma_e^2$, e a **função de verossimilhança baseada nos dados completos (D)** é dada por

$$\begin{aligned} L(\theta|D) &\propto \quad (2.1.5) \\ &(\sigma_u^2\sigma_e^2\sigma_x^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \left[\frac{\sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2}{\sigma_e^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma_u^2} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_x^2} \right] \right\}. \end{aligned}$$

É fato bastante conhecido que este modelo não é identificável (Fuller, 1987), já que a distribuição conjunta dos dados observados dada em (2.1.2) depende de seis parâmetros: $\alpha, \beta, \mu, \sigma_x^2, \sigma_u^2$ e σ_e^2 mas somente cinco podem ser estimados (já que a estatística suficiente tem dimensão 5).

Sabemos por definição que um modelo é identificável se e somente se para todo θ_1 e θ_2 pertencentes ao espaço paramétrico, com $\theta_1 \neq \theta_2$ então $f(\theta_1) \neq f(\theta_2)$. Portanto um modelo é não identificável se e somente se $\exists \theta_1 \neq \theta_2$ tal que $f(\theta_1) =$

$f(\theta_2)$. No caso do modelo estrutural com $\theta = (\alpha, \beta, \mu, \sigma_x^2, \sigma_e^2, \sigma_u^2)$, podemos exibir dois pontos $\theta_1 = (1, 1, 1, 1, 1, 1)^T$ e $\theta_2 = (1.5, 0.5, 1.0, 2.0, 1.5, 0.0)^T$ que levam à mesma distribuição.

$$\begin{pmatrix} Y_i \\ X_i \end{pmatrix} \sim N \left(\begin{pmatrix} 2 \\ 1 \end{pmatrix}; \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \right).$$

Para assegurar identificabilidade deve-se fazer alguma suposição extra sobre o modelo. Algumas das suposições mais comumente encontradas na literatura são as seguintes:

- (a) $\sigma_e^2/\sigma_u^2 = \lambda$, onde λ é conhecido;
 - (b) σ_e^2 ou σ_u^2 é conhecido;
 - (c) $\sigma_x^2/(\sigma_x^2 + \sigma_e^2)$ é conhecido.
- (2.1.6)

O modelo funcional

Na estatística clássica o modelo funcional aditivo com erros normalmente distribuídos é definido pelas equações

$$\begin{aligned} Y_i &= \alpha + \beta x_i + e_i \\ \text{e } X_i &= x_i + u_i, \end{aligned}$$

onde Y_i e X_i são variáveis aleatórias observáveis, e_i e u_i são os erros aleatórios, x_i é parâmetro desconhecido, onde

$$\begin{pmatrix} e_i \\ u_i \end{pmatrix} \stackrel{\text{i.i.d.}}{\sim} N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix} \right), \text{ para } i = 1, 2, \dots, n. \quad (2.1.7)$$

O espaço paramétrico é o conjunto

$$\Theta = \{ \theta = (\alpha, \beta, \sigma_e^2, \sigma_u^2, x_1, \dots, x_n)^T \in \mathbb{R}^2 \times \mathbb{R}^{+2} \times \mathbb{R}^n \}.$$

A função de verossimilhança nesse caso é dada por

$$L(\theta | D_{\text{obs}}) \propto (\sigma_u^2 \cdot \sigma_e^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \left[\frac{\sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2}{\sigma_e^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma_u^2} \right] \right\} \quad (2.1.8)$$

que é ilimitada. Entretanto, pode-se eliminar este problema acrescentando-se ao modelo, por exemplo, a suposição que $\sigma_e^2 = \lambda\sigma_u^2$, onde λ é conhecido.

No modelo funcional temos que x_1, x_2, \dots, x_n são constantes desconhecidas e portanto ele é apropriado quando as únicas unidades experimentais de interesse estão representadas nos dados disponíveis (isto é, os x_i 's não são uma amostra aleatória de alguma população mas a própria população) (Carrol e Stefanski, 1995).

Quando as unidades experimentais representadas nos dados constituem uma amostra aleatória de alguma população, então é apropriado tratar x_1, x_2, \dots, x_n como variáveis aleatórias i.i.d. e nesse caso é razoável utilizar um modelo estrutural (Stefansky, 2000).

2.2 O modelo estrutural e o modelo funcional na estatística bayesiana

Como na literatura bayesiana, a diferença entre os modelos funcional e estrutural não está muito clara, nossa intençãoo nessa seção é procurar esclarecer o leitor sobre esse ponto.

Inicialmente apresentamos duas formulações alternativas do modelo estrutural (a formulação 1 e a formulação 2). A formulação 2 é a que aparece mais comumente na literatura no contexto do modelos com erros nas variáveis, além disso é mais fácil comparar os modelos estrutural e funcional usando essa formulação. Entretanto a formulação 1 é a que traz maior vantagem do ponto de vista computacional.

Na subseção 2.2.1 a seguir, descrevemos o modelo estrutural aditivo Normal por meio de duas formulações equivalentes. Na subseção 2.2.2 definimos o modelo funcional aditivo Normal fazendo uma comparação com o modelo estrutural.

2.2.1 O modelo bayesiano estrutural

Formulação 1 do modelo estrutural Normal

Tomamos o modelo clássico definido pelas equações

$$\begin{aligned} Y_i &= \alpha + \beta x_i + e_i \\ \text{e } X_i &= x_i + u_i, \end{aligned} \quad (2.2.1)$$

onde

$(Y_i, X_i)^T, i = 1, \dots, n$ são os vetores de observações, e x_1, \dots, x_n são as v.a.

latentes (não observáveis),

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} | \theta \stackrel{\text{i.i.d.}}{\sim} N \left(\begin{pmatrix} 0 \\ 0 \\ \mu \end{pmatrix}; \begin{pmatrix} \sigma_e^2 & 0 & 0 \\ 0 & \sigma_u^2 & 0 \\ 0 & 0 & \sigma_x^2 \end{pmatrix} \right), \text{ para } i = 1, 2, \dots, n,$$

e acrescentamos a suposição de que o vetor de parâmetros $\theta = (\alpha, \beta, \mu, \sigma_x^2, \sigma_e^2, \sigma_u^2)^T$ tem distribuição à priori $\pi(\theta)$. Nessa formulação, por convenção não incluímos as variáveis latentes x_1, x_2, \dots, x_n no vetor de parâmetros pelo fato de que esses x_i 's, além de desconhecidos (isto é, não observáveis), são realizações de um processo amostral (diferentemente das quantidades $\alpha, \beta, \mu, \sigma_x^2, \sigma_e^2, \sigma_u^2$). Assim, o espaço paramétrico nesse caso é o conjunto

$$\Theta = \{ \theta = (\alpha, \beta, \mu, \sigma_x^2, \sigma_e^2, \sigma_u^2)^T : \theta \in \mathbb{R}^3 \times \mathbb{R}_+^3 \}.$$

Distribuição à posteriori

Chamando de D_{obs} o conjunto de dados observados, $\{(X_i, Y_i)^T, i = 1, \dots, n\}$, temos que a distribuição à posteriori de $\theta = (\alpha, \beta, \mu, \sigma_x^2, \sigma_e^2, \sigma_u^2)^T$ é dada por

$$p(\theta | D_{\text{obs}}) \propto L(\theta | D_{\text{obs}}) \pi(\theta),$$

onde $L(\theta | D_{\text{obs}})$ é a verossimilhança baseada nos dados observados dada em (2.1.4)

e $\pi(\theta)$ a distribuição à priori de θ . Portanto,

$$p(\theta | D_{\text{obs}}) \propto |\Sigma|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2|\Sigma|} \left[(\beta^2 \sigma_x^2 + \sigma_e^2) \sum_{i=1}^n (X_i - \mu)^2 \right] \right\} \quad (2.2.2)$$

$$-2\beta\sigma_x^2 \sum_{i=1}^n (Y_i - \alpha - \beta\mu)(X_i - \mu) + (\sigma_x^2 + \sigma^2) \sum_{i=1}^n (Y_i - \alpha - \beta\mu)^2 \Big] \cdot \pi(\theta),$$

onde $|\Sigma| = \beta^2\sigma_u^2\sigma_x^2 + \sigma_e^2\sigma_x^2 + \sigma_u^2\sigma_e^2$.

Distribuição à posteriori em função dos “dados completos”

É importante ressaltar que a distribuição à posteriori $p(\theta|D_{\text{obs}})$ também é uma distribuição marginal de $p(\theta, \underline{x} | D_{\text{obs}})$, ou seja, $p(\theta|D_{\text{obs}}) = \int_{\underline{x}=\mathbb{R}^n} p(\theta, \underline{x} | D_{\text{obs}}) d\underline{x}$. Então, chamando de D o conjunto de “dados completos”, $D = \{(Y_i, X_i, x_i)^T, i = 1, \dots, n\}$, temos que

$$p(\theta|D_{\text{obs}}) = \int_{\underline{x}=\mathbb{R}^n} p(\theta, \underline{x} | D_{\text{obs}}) d\underline{x} \propto \int_{\underline{x}=\mathbb{R}^n} L(\theta|D) \pi(\theta) d\underline{x},$$

onde $L(\theta|D)$ é a verossimilhança baseada nos dados completos dada em (2.1.5) e $\pi(\theta)$ a priori de θ .

Portanto a distribuição à posteriori $p(\theta|D_{\text{obs}})$ também é proporcional a

$$\int_{\underline{x}=\mathbb{R}^n} (\sigma_u^2\sigma_e^2\sigma_x^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \left[\frac{\sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2}{\sigma_e^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma_u^2} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_x^2} \right] \right\} \pi(\theta) d\underline{x}. \quad (2.2.3)$$

Observação 1: É importante ressaltar que a distribuição à posteriori $p(\theta|D_{\text{obs}})$ pode então ser explorada computacionalmente de duas maneiras: a primeira usando a fórmula (2.2.2), que não é função de x_1, x_2, \dots, x_n , e a segunda usando o fato de que $p(\theta|D_{\text{obs}})$ é distribuição marginal de $p(\underline{x}, \theta | D_{\text{obs}})$.

Observação 2: É importante observar também que na formulação 1 do modelo, se a distribuição à priori de $\theta = (\alpha, \beta, \mu, \sigma_x^2, \sigma^2)^T$ for totalmente especificada, isto é, se os hiperparâmetros forem conhecidos, temos um modelo com apenas um nível hierárquico.

Distribuição preditiva

Seja $\underline{U}_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix}$, $i = 1, \dots, n$. A distribuição preditiva à posteriori nesse modelo é dada por

$$p(\underline{U}_f | D_{\text{obs}}) = \int_{\Theta} p(\underline{U}_f, \varrho | D_{\text{obs}}) d\varrho = \int_{\Theta} p(\underline{U}_f | \varrho) p(\varrho | D_{\text{obs}}) d\varrho,$$

onde \underline{U}_f é uma observação futura de $\begin{pmatrix} Y \\ X \end{pmatrix}$, isto é, $\underline{U}_f = \begin{pmatrix} Y_f \\ X_f \end{pmatrix}$,

$$\underline{U}_f | \varrho \sim N \left[\begin{pmatrix} \alpha + \beta\mu \\ \mu \end{pmatrix}; \begin{pmatrix} \beta^2\sigma_x^2 + \sigma_e^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma_u^2 \end{pmatrix} \right],$$

e $p(\varrho | D_{\text{obs}})$ é a distribuição à posteriori de $\theta = (\alpha, \beta, \mu, \sigma_x^2, \sigma_e^2, \sigma_u^2)^T$. Portanto,

$$p(\underline{U}_f | D_{\text{obs}}) = \int_{\Theta} p(\underline{U}_f, \varrho | D_{\text{obs}}) d\varrho = \int_{\Theta} [Y_f | X_f, \varrho] [X_f | \varrho] [\varrho | D_{\text{obs}}] d\varrho,$$

onde $[\varrho | D_{\text{obs}}]$ representa a distribuição à posteriori de ϱ , além disso $[X_f | \varrho]$ e $[Y_f | X_f, \varrho]$ representam respectivamente as distribuições condicionais de $X_f | \varrho$ e $Y_f | X_f, \varrho$ onde

$$\begin{cases} X_f | \varrho \sim N(\mu, \sigma_x^2 + \sigma_u^2) & \text{e} \\ Y_f | X_f, \varrho \sim N\left(\alpha + \beta\mu + \frac{\beta\sigma_x^2}{\sigma_x^2 + \sigma_u^2}(X_f - \mu); \beta^2\sigma_x^2 + \sigma_e^2 - \frac{\beta^2(\sigma_x^2)^2}{\sigma_x^2 + \sigma_u^2}\right). \end{cases} \quad (2.2.4)$$

Geração de uma observação futura:

Para gerar uma observação \underline{U}_f , basta gerar inicialmente um ϱ da distribuição à posteriori $[\varrho | D_{\text{obs}}]$; a seguir, para esse ϱ fixado, gerar um valor X_f da distribuição $[X_f | \varrho]$ e finalmente gerar Y_f de $[Y_f | X_f, \varrho]$.

Se quisermos gerar um valor Y_f da distribuição preditiva $[Y_f | X_f = X_0, D_{\text{obs}}]$ basta gerar ϱ de $[\varrho | D_{\text{obs}}]$ e depois Y_f de $[Y_f | X_0, \varrho]$, já que $[Y_f | X_f = X_0, D_{\text{obs}}]$ é dada por $\int_{\Theta} [Y_f | X_0, \varrho] [\varrho | D_{\text{obs}}] d\varrho$.

Semelhantemente, podemos gerar uma observação futura X_f dado que $Y_f = Y_0$. A distribuição preditiva de $X_f | Y_f = Y_0, D_{\text{obs}}$ nesse caso é dada por

$$\int_{\Theta} [X_f | Y_0, \varrho] [\varrho | D_{\text{obs}}] d\varrho,$$

onde

$$X_f|Y_f = Y_0, \theta \sim N\left(\mu + \frac{\beta\sigma_x^2}{\beta^2\sigma_x^2 + \sigma_e^2}(Y_0 - \alpha - \beta\mu); \sigma_x^2 + \sigma_u^2 - \frac{\beta^2(\sigma_x^2)^2}{\beta^2\sigma_x^2 + \sigma_e^2}\right).$$

Observação: Uma alternativa para gerar uma observação futura \underline{U}_f :

A distribuição preditiva $p(\underline{U}_f|D_{\text{obs}})$ pode ser escrita também como

$$\begin{aligned} p(\underline{U}_f|D_{\text{obs}}) &= \int_{\mathbf{x}_f=\mathbb{R}} \int_{\Theta} p(U_f, \varrho, x_f|D_{\text{obs}}) d\varrho dx_f \\ &= \int_{\mathbf{x}_f=\mathbb{R}} \int_{\Theta} [\underline{U}_f|x_f, \varrho][x_f|\varrho][\varrho|D_{\text{obs}}] d\varrho dx_f, \end{aligned}$$

onde

$$\underline{U}_f|x_f, \varrho \sim N\left[\begin{pmatrix} \alpha + \beta x_f \\ x_f \end{pmatrix}, \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix}\right]$$

e portanto para gerar uma observação \underline{U}_f basta gerar ϱ da distribuição à posteriori $[\varrho|D_{\text{obs}}]$, depois x_f da distribuição $[x_f|\varrho]$, depois X_f de $[X_f|\varrho, x_f]$ e depois Y_f de $[Y_f|X_f, x_f, \varrho]$, onde $x_f|\varrho \sim N(\mu, \sigma_x^2)$, $X_f|\varrho, x_f \sim N(x_f, \sigma_u^2)$ e $Y_f|X_f, x_f, \varrho \sim N(\alpha + \beta x_f, \sigma_e^2)$.

Formulação 2 do modelo estrutural Normal

Seja o modelo definido pelas equações

$$\begin{aligned} Y_i &= \alpha + \beta x_i + e_i \\ \text{e } X_i &= x_i + u_i, \end{aligned} \tag{2.2.5}$$

onde $\begin{pmatrix} e_i \\ u_i \end{pmatrix} | \varphi \stackrel{\text{i.i.d.}}{\sim} N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix}\right)$, para $i = 1, \dots, n$.

Nessa formulação o vetor de parâmetros agora é $\theta^* = (\alpha, \beta, \sigma_e^2, \sigma_u^2, x_1, \dots, x_n)^T$, onde à priori $x_i | \alpha, \beta, \sigma_e^2, \sigma_u^2, \varphi \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma_x^2)^T$, $i = 1, \dots, n$, onde $\varphi = (\mu, \sigma_x^2)$ é o vetor de hiperparâmetros desconhecidos.

Função de verossimilhança

A função de verossimilhança, $L(\theta^*|D_{\text{obs}})$, nesse caso será igual a:

$$(\sigma_u^2 \cdot \sigma_e^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \left[\frac{\sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2}{\sigma_e^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma_u^2} \right] \right\},$$

(já que $\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | \check{\theta}^* \stackrel{\text{i.i.d.}}{\sim} N \left(\begin{pmatrix} \alpha + \beta x_i \\ x_i \end{pmatrix}; \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix} \right)$, $i = 1, \dots, n$)

Distribuição à posteriori

$$\begin{aligned} p(\check{\theta}^*, \phi | D_{\text{obs}}) &\propto L(\check{\theta}^* | D_{\text{obs}}) \pi(\theta^*, \phi) \\ &= L(\check{\theta}^* | D_{\text{obs}}) \pi(\underline{x} | \alpha, \beta, \sigma_e^2, \sigma_u^2, \phi) \pi(\alpha, \beta, \sigma_e^2, \sigma_u^2, \phi), \end{aligned} \quad (2.2.6)$$

onde $\pi(\underline{x} | \alpha, \beta, \sigma_e^2, \sigma_u^2, \phi) \propto (\sigma_x^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_x^2} \right\}$.

Portanto,

$$\begin{aligned} p(\check{\theta}^*, \phi | D_{\text{obs}}) &\propto (\sigma_u^2 \sigma_e^2 \sigma_x^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \left[\frac{\sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2}{\sigma_e^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma_u^2} \right. \right. \\ &\quad \left. \left. + \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_x^2} \right] \right\} \pi(\alpha, \beta, \sigma_e^2, \sigma_u^2, \phi). \end{aligned} \quad (2.2.7)$$

onde $\pi(\alpha, \beta, \sigma_e^2, \sigma_u^2, \phi) = \pi(\alpha, \beta, \mu, \sigma_x^2, \sigma_e^2, \sigma_u^2)$.

Notemos que se $\pi(\alpha, \beta, \sigma_e^2, \sigma_u^2, \phi)$ tiver hiperparâmetros conhecidos temos um modelo com dois níveis hierárquicos. Notemos também que a expressão em (2.2.7) coincide com a expressão dentro do sinal de integração dada em (2.2.3) (assim, a posteriori em função dos dados completos dada pela formulação 1 em (2.2.3) é a marginal da posteriori na formulação 2, quando eliminamos as variáveis x_1, x_2, \dots, x_n por integração).

Distribuição preditiva

A distribuição preditiva $p(\underline{U}_f | D_{\text{obs}})$ na formulação 2 é dada por

$$p(\underline{U}_f | D_{\text{obs}}) = \int_{\Theta^*} \int_{\Phi} p(\underline{U}_f, \check{\theta}^*, \phi, | D_{\text{obs}}) d\phi d\check{\theta}^*,$$

onde $\underline{U}_f = \begin{pmatrix} Y_f \\ X_f \end{pmatrix}$, $\Theta^* = \{\underline{\theta} = (\alpha, \beta, \sigma_e^2, \sigma_u^2, \underline{x})^T : \underline{\theta} \in \mathbb{R}^2 \times \mathbb{R}_+^2 \times \mathbb{R}^n\}$ e $\Phi = \{(\mu, \sigma_x^2)^T : \mu \in \mathbb{R}, \sigma_x^2 \in \mathbb{R}^+\}$. Portanto,

$$p(\underline{U}_f | D_{\text{obs}}) = \int_{\Theta} \int_{\underline{x} = \mathbb{R}^n} p(\underline{U}_f, \underline{\theta}, \underline{x} | D_{\text{obs}}) d\underline{\theta} d\underline{x},$$

onde $\underline{\theta} = (\alpha, \beta, \mu, \sigma_x^2, \sigma_e^2, \sigma_u^2)^T$, $\underline{x} = (x_1, \dots, x_n)^T$, $\Theta = \{\underline{\theta} \mid \alpha, \beta, \mu \in \mathbb{R} \text{ e } \sigma_x^2, \sigma_e^2, \sigma_u^2 \in \mathbb{R}^+\}$ e $\underline{x} = \{x : x_i \in \mathbb{R}\}$. Portanto,

$$\begin{aligned} p(\underline{U}_f | D_{\text{obs}}) &= \int_{\Theta} \int_{\underline{x}} [\underline{U}_f | \underline{\theta}] [\underline{\theta}, \underline{x} | D_{\text{obs}}] d\underline{\theta} d\underline{x} \\ &= \int_{\underline{x}} \int_{\Theta} [Y_f | X_f, \underline{\theta}] [X_f | \underline{\theta}] [\underline{\theta}, \underline{x} | D_{\text{obs}}] d\underline{\theta} d\underline{x}, \end{aligned}$$

onde $[\underline{\theta}, \underline{x} | D_{\text{obs}}]$ é a distribuição à posteriori dada em (2.2.7), as distribuições $[X_f | \underline{\theta}]$ e $[Y_f | X_f, \underline{\theta}]$ são dadas em (2.2.4).

Para gerar uma observação \underline{U}_f , basta gerar $\underline{\theta}$ de $[\underline{\theta}, \underline{x} | D_{\text{obs}}]$, depois X_f de $[X_f | \underline{\theta}]$ e depois Y_f de $[Y_f | X_f, \underline{\theta}]$ (da mesma maneira descrita na formulação 1).

Observação: Outra forma de escrever a posteriori preditiva é

$$\begin{aligned} p(\underline{U}_f | D_{\text{obs}}) &= \int_{\Theta^*} \int_{\Phi} \int_{\{x_f \in \mathbb{R}\}} p(\underline{U}_f, \underline{\theta}, \phi, x_f | D_{\text{obs}}) dx_f d\phi d\underline{\theta} \\ &= \int_{\underline{x}} \int_{\Theta} \int_{\{x_f \in \mathbb{R}\}} p(\underline{U}_f, \underline{\theta}, \underline{x}, x_f | D_{\text{obs}}) dx_f d\underline{\theta} d\underline{x}, \end{aligned}$$

onde $\underline{\theta} = (\alpha, \beta, \mu, \sigma_x^2, \sigma_e^2, \sigma_u^2)^T$ e $\underline{x} = (x_1, \dots, x_n)^T$ e portanto

$$p(\underline{U}_f | D_{\text{obs}}) = \int_{\underline{x}} \int_{\Theta} \int_{\{x_f \in \mathbb{R}\}} [\underline{U}_f | \underline{\theta}, x_f] [x_f | \underline{\theta}] [\underline{\theta}, \underline{x} | D_{\text{obs}}] dx_f d\underline{\theta} d\underline{x},$$

onde $x_f | \underline{\theta} \sim N(\mu, \sigma_x^2)$ e $\underline{U}_f | \underline{\theta}, x_f \sim N\left(\begin{pmatrix} \alpha + \beta x_f \\ x_f \end{pmatrix}, \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix}\right)$.

Observamos que podemos gerar \underline{U}_f da mesma maneira que na Observação da formulação 1.

Do que foi apresentado, podemos ver que as duas formulações são equivalentes em termos de inferência. Entretanto, no restante desse trabalho estaremos utilizando

a formulação 1 também por razões didáticas e computacionais. Nessa formulação é mais direto ver que a posteriori de $(\alpha, \beta, \mu, \sigma_x^2, \sigma_e^2, \sigma_u^2)$ pode ser estudada de duas formas: uma em função dos dados completos, como em (2.2.3), e outra em função somente dos dados observados como em (2.2.2), o que amplia as opções de algoritmos para amostrar da posteriori.

Comentários: O modelo estrutural bayesiano com erros nas variáveis, como sabemos, não é identificável (já que a função de verossimilhança não é identificável). Em princípio, teoricamente, isso não é problema. Entretanto nas aplicações o algoritmo de Gibbs para amostrar da posteriori de modelos não identificáveis (que será descrito nesse capítulo) pode apresentar problemas de convergência segundo Gelfand e Sahu (1999). Além disso, se a priori não é “informativa”, o problema pode piorar bastante e a recomendação dos autores nesse caso é que se escolha uma priori “adequada”, que não seja nem “muito pouco informativa”, mas também que não seja “extremamente precisa” para não limitar o aprendizado proveniente dos dados. Segundo os autores, o problema é achar esse “meio termo” o que geralmente requer um trabalho considerável de simulação.

Uma outra saída para o problema computacional decorrente da falta de identificabilidade é simplesmente acrescentar uma suposição ao modelo como, por exemplo, adotar uma das restrições de identificabilidade descritas em (2.1.6).

Com relação à existência da posteriori, ressaltamos que sempre que a priori for própria, a posteriori também será própria, entretanto podemos ter posteriores próprias para certas escolhas de prioris impróprias. Na Seção 2.5.1 estudamos as condições para a existência da posteriori do modelo estrutural sob diferentes escolhas de prioris impróprias quando incorporamos ao modelo a restrição de identificabilidade: $\sigma_e^2 = \lambda\sigma_u^2$, onde λ é conhecido e na Seção 2.5.2 fazemos um estudo semelhante para

o modelo sem restrições.

2.2.2 O modelo bayesiano funcional

Tomamos o modelo aditivo definido pelas equações

$$\begin{aligned} Y_i &= \alpha + \beta x_i + e_i \\ \text{e} \quad X_i &= x_i + u_i, \end{aligned}$$

onde Y_i e X_i são as variáveis aleatórias observáveis, x_i são quantidades fixadas desconhecidas, e_i e u_i os erros aleatórios, onde

$$\begin{pmatrix} e_i \\ u_i \end{pmatrix} \stackrel{\text{i.i.d.}}{\sim} N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix} \right), \text{ para } i = 1, 2, \dots, n,$$

e onde $\check{\theta} = (\alpha, \beta, \sigma_u^2, \sigma_e^2, \underline{x})^T$ onde $(x_1, \dots, x_n)^T$ é o vetor de parâmetros, e incorporamos a distribuição à priori $\pi(\check{\theta})$ de $\check{\theta}$.

Observação: Se $\pi(\check{\theta})$, pudesse ser totalmente especificada (isto é, se os hiperparâmetros de $\pi(\check{\theta})$ fossem conhecidos), então teríamos um modelo não hierárquico. Se, entretanto, a distribuição dos x_i 's depende de hiperparâmetros ϕ_i desconhecidos, para $i = 1, \dots, n$, então temos um modelo hierárquico.

Distribuição à posteriori

A distribuição à posteriori sob o modelo, onde ϕ_i é desconhecido, é dada por

$$p(\check{\theta}, \phi | D_{\text{obs}}) \propto L(\check{\theta} | D_{\text{obs}}) \pi(\check{\theta}, \phi),$$

onde $\phi = (\phi_1, \phi_2, \dots, \phi_n)^T$, $\phi_i = (\mu_i, \sigma_{x_i}^2)$, $i = 1, \dots, n$, $L(\check{\theta} | D_{\text{obs}})$ é a verossimilhança e $\pi(\check{\theta}, \phi)$ é a distribuição à priori de $\check{\theta}$ e ϕ , onde

$$L(\check{\theta} | D_{\text{obs}}) \propto (\sigma_u^2 \sigma_e^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \left[\frac{\sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2}{\sigma_e^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma_u^2} \right] \right\}$$

(já que $\begin{pmatrix} X_i \\ Y_i \end{pmatrix} | \theta^* \sim N \left(\begin{pmatrix} \alpha + \beta x_i \\ x_i \end{pmatrix}; \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix} \right)$, $i = 1, \dots, n$).

É interessante comparar esta posteriori com a posteriori sob o modelo estrutural dada em (2.2.6) e (2.2.7) e notar que se pudéssemos assumir que $x_i|\phi_i, \alpha, \beta, \sigma_e^2, \sigma_u^2 \stackrel{i.i.d.}{\sim} N(\mu, \sigma_x^2)$ para $i = 1, \dots, n$, onde $\phi_i = \phi = (\mu, \sigma_x^2)^T$ é desconhecido, então as posterioris coincidiriam. É fácil ver também (usando a formulação 2 do modelo estrutural) que as posterioris dos modelos estrutural e funcional podem coincidir não somente no caso Normal (erros e covariável normais) mas em qualquer outro caso, desde que $\pi(\underline{x}|\alpha, \beta, \sigma_e^2, \sigma_u^2, \phi)$ seja a mesma nos dois modelos (ou seja, $x_i|\phi_i, \alpha, \beta, \sigma_e^2, \sigma_u^2$ sejam independentes e identicamente distribuídos para $i = 1, \dots, n$).

Na estatística Bayesiana, portanto, a diferença entre os modelos estrutural e funcional está somente na escolha da distribuição de $\underline{x} = (x_1, x_2, \dots, x_n)^T$, isto é, na especificação do modelo. No caso estrutural assumimos que $x_i|\alpha, \beta, \sigma_u^2, \sigma_e^2, \phi_i$ são independentes e indenticamente distribuídos para $i = 1, \dots, n$, isto é, $\underline{x} = (x_1, \dots, x_n)^T$ constitui uma amostra aleatória (na Seção 2.2.1, tomamos $\pi(x_i|\alpha, \beta, \sigma_u^2, \sigma_e^2, \phi) \stackrel{i.i.d.}{\sim} N(\mu, \sigma_x^2)$, onde $\phi = (\mu, \sigma_x^2)$ para $i = 1, \dots, n$). No caso funcional, conceitualmente, os x_i 's são quantidades (desconhecidas) que não estão sujeitas a flutuações amostrais e por isso não nos parece razoável assumir que os x_i 's sejam i.i.d. devido à interpretação amostral que isso acarretaria. Em um modelo funcional a distribuição de $\pi(x_i|\alpha, \beta, \sigma_u^2, \sigma_e^2, \phi_i)$ pode também depender de i , para $i = 1, \dots, n$ (Carrol e Stefanski, 1995) como também pode ser imprópria, já que dificilmente teríamos uma interpretação amostral nesse caso (Florens et al., 1974).

Segundo Stephens e Dellaportas (1992), embora na estatística clássica os modelos funcional e estrutural difiram bastante quanto à análise e quanto aos procedimentos de estimação, na formulação Bayesiana a diferença é apenas uma questão de especificação do modelo ou seja uma questão de especificar a distribuição de $\underline{x}|\alpha, \beta, \sigma_e^2, \sigma_u^2, \phi$ onde $\phi = (\phi_1, \dots, \phi_n)$.

2.3 O modelo estrutural identificável com particular priori de componentes independentes

Nesta seção apresentamos o modelo linear estrutural Normal com erros nas variáveis no caso em que a distribuição à priori é própria e tem componentes independentes Normais ou Gama Invertida, assumindo a restrição de identificabilidade $\sigma_e^2 = \lambda\sigma_u^2$, onde λ é conhecido.

O modelo

No modelo estrutural descrito em (2.2.1) acrescentamos a restrição $\sigma_e^2 = \lambda\sigma_u^2$, onde $\sigma_u^2 = \sigma^2$ e λ é conhecido, obtendo o seguinte modelo identificável, definido pelas equações

$$\begin{aligned} Y_i &= \alpha + \beta x_i + e_i \\ \text{e } X_i &= x_i + u_i, \end{aligned} \quad (2.3.1)$$

onde

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \stackrel{\text{i.i.d.}}{\sim} N \left(\begin{pmatrix} 0 \\ 0 \\ \mu \end{pmatrix}, \begin{pmatrix} \lambda\sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma_x^2 \end{pmatrix} \right), \quad i = 1, 2, \dots, n,$$

onde $\theta = (\alpha, \beta, \mu, \sigma_x^2, \sigma^2)^T$ é o vetor de parâmetros.

Acrescentamos a esse modelo a suposição à priori de que $\alpha, \beta, \mu, \sigma_x^2$ e σ^2 são variáveis aleatórias independentes, com α, β, μ tendo distribuição normal e σ_x^2, σ^2 tendo distribuição Gama Invertida, como segue:

$$\begin{aligned} \pi(\alpha) &\sim N(a, \sigma_\alpha^2), \quad \pi(\beta) \sim N(b, \sigma_\beta^2), \quad \pi(\mu) \sim N(m, \sigma_\mu^2), \\ \sigma_x^2 &\sim IG(c, d) \quad \text{e} \quad \sigma^2 \sim IG(f, g) \quad (\text{isto é, } \pi(\sigma^2) \propto (\sigma^2)^{-(f+1)} \exp\left\{-\frac{g}{\sigma^2}\right\}), \end{aligned} \quad (2.3.2)$$

onde $\lambda, a, \sigma_\alpha^2, b, \sigma_\beta^2, m, \sigma_\mu^2, c, d, f$ e g são os hiperparâmetros conhecidos.

Observação: Embora vamos manter (até o final da tese) o valor de λ em todas as equações que envolvem o modelo com restrição de identificabilidade $\sigma_e^2 = \lambda\sigma_u^2$. É

importante lembrar que não existe perda de generalidade em trabalhar sempre com $\lambda = 1$ (pois sempre podemos dividir a equação $Y_i = \alpha + \beta x_i + e_i$ por $\sqrt{\lambda}$ e fazer $Y_i^* = \frac{Y_i}{\sqrt{\lambda}}$, $\alpha^* = \frac{\alpha}{\sqrt{\lambda}}$, $\beta^* = \frac{\beta}{\sqrt{\lambda}}$ e $e_i^* = \frac{e_i}{\sqrt{\lambda}}$, de tal modo que $\text{Var}(e_i^*) = \frac{\sigma_e^2}{\lambda} = \frac{\lambda \sigma_u^2}{\lambda} = \sigma_u^2$ (isto é $\sigma_{e^*}^2 = \sigma_u^2$).

Assim, trabalhando-se com as variáveis observadas Y_i^* e X_i obtemos os estimadores $\hat{\alpha}^*$ e $\hat{\beta}^*$ e finalmente $\hat{\alpha} = \hat{\alpha}^* \sqrt{\lambda}$ e $\hat{\beta} = \hat{\beta}^* \sqrt{\lambda}$.

Distribuição à posteriori

Na seção 2.2.1, usando a formulação 1, vimos que a posteriori $p(\theta|D_{\text{obs}}) = p(\alpha, \beta, \mu, \sigma_x^2, \sigma^2|D_{\text{obs}})$ pode ser escrita de duas maneiras, uma em função da verossimilhança baseada nos dados observados ($L(\theta|D_{\text{obs}})$), que é dada por

$$p(\theta|D_{\text{obs}}) \propto L(\theta|D_{\text{obs}})\pi(\theta),$$

e que chamamos de “posteriori baseada nos dados observados”. Outra maneira, em função da verossimilhança baseada nos dados completos ($L(\theta|D)$ onde $D = (D_{\text{obs}}, \underline{x})$), que é dada por

$$p(\theta|D_{\text{obs}}) = \int_{\underline{x}=\mathbb{R}^n} p(\underline{x}, \theta|D_{\text{obs}})d\underline{x} \propto \int_{\underline{x}=\mathbb{R}^n} L(\theta|D)\pi(\theta)d\underline{x},$$

e que chamamos de “posteriori baseada nos dados completos”.

Posteriori baseada nos dados observados (Dobs)

No modelo definido por (2.3.1) e (2.3.2), a posteriori é dada por

$$\begin{aligned} p(\theta|D_{\text{obs}}) \propto & |\Sigma|^{-\frac{n}{2}} (\sigma_x^2)^{-(c+1)} (\sigma^2)^{-(f+1)} \exp \left\{ -\frac{1}{2} \left[\left((\beta^2 \sigma_x^2 + \lambda \sigma^2) \sum_{i=1}^n (X_i - \mu)^2 \right. \right. \right. \\ & \left. \left. - 2\beta \sigma_x^2 \sum_{i=1}^n (Y_i - \alpha - \beta \mu)(X_i - \mu) + (\sigma_x^2 + \sigma^2) \sum_{i=1}^n (Y_i - \alpha - \beta \mu)^2 \right) / |\Sigma| \right. \\ & \left. \left. + \frac{(\alpha - a)^2}{\sigma_\alpha^2} + \frac{(\beta - b)^2}{\sigma_\beta^2} + \frac{(\mu - m)^2}{\sigma_\mu^2} + \frac{2d}{\sigma_x^2} + \frac{2g}{\sigma^2} \right] \right\} \end{aligned} \quad (2.3.3)$$

onde $|\Sigma| = \beta^2 \sigma^2 \sigma_x^2 + \lambda \sigma^2 \sigma_x^2 + \lambda (\sigma^2)^2$.

Posteriori baseada nos dados completos

$$p(\theta|D_{\text{obs}}) = \int_{\mathbf{x}} p(\theta, \mathbf{x}|D_{\text{obs}}) d\mathbf{x},$$

onde $p(\theta, \mathbf{x}|D_{\text{obs}})$ é proporcional a

$$\begin{aligned} & (\sigma_x^2)^{-(\frac{n}{2}+c+1)} (\sigma^2)^{-(n+f+1)} \exp \left\{ -\frac{1}{2} \left[\frac{\sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2}{\lambda \sigma^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma^2} \right. \right. \\ & \left. \left. + \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_x^2} + \frac{(\alpha - a)^2}{\sigma_\alpha^2} + \frac{(\beta - b)^2}{\sigma_\beta^2} + \frac{(\mu - m)^2}{\sigma_\mu^2} + \frac{2d}{\sigma_x^2} + \frac{2g}{\sigma^2} \right] \right\}. \end{aligned} \quad (2.3.4)$$

A distribuição à posteriori de $\theta = (\alpha, \beta, \mu, \sigma_x^2, \sigma^2)^T$ na forma exata é dada por

$$\begin{aligned} p(\theta|D_{\text{obs}}) &= \frac{L(\theta|D_{\text{obs}})\pi(\theta)}{\int_{\Theta} L(\theta|D_{\text{obs}})\pi(\theta)d\theta} \\ &\left(\text{ou } \frac{\int_{\mathbf{x}} L(\theta|D)\pi(\theta)d\mathbf{x}}{\int_{\Theta} \int_{\mathbf{x}} L(\theta|D)\pi(\theta)d\mathbf{x}d\theta}, \text{ onde } D = (D_{\text{obs}}, \mathbf{x}) \right). \end{aligned} \quad (2.3.5)$$

Podemos ver, com base em (2.3.3) e (2.3.4), que a posteriori portanto não tem uma forma fechada pois a constante de normalização (ou proporcionalidade) dada por $\int_{\Theta} L(\theta|D_{\text{obs}})\pi(\theta)d\theta$ (ou $\int_{\Theta} \int_{\mathbf{x}} L(\theta|D)\pi(\theta)d\mathbf{x}d\theta$) não pode ser obtida analiticamente, já que essas integrais não têm primitivas na forma explícita; apenas o núcleo da posteriori baseada nos dados observados ($L(\theta|D_{\text{obs}})\pi(\theta)$) e o núcleo de $p(\theta, \mathbf{x}|D_{\text{obs}})$ têm forma conhecida, dada por (2.3.3) e (2.3.4), respectivamente. Além disso, as distribuições marginais e condicionais à posteriori, bem como as medidas resumo da posteriori (como, por exemplo, os momentos e os quantis) obviamente não tem fórmulas fechadas pois requerem ainda mais integrais para serem calculadas (agora também em relação ao numerador de (2.3.5)).

Na literatura existem vários procedimentos numéricos computacionais para tratar desse tipo de problema e na próxima seção descrevemos alguns dos mais conhecidos na estatística, incluindo por exemplo algoritmos de simulação para amostrar de uma distribuição à posteriori.

2.4 Métodos computacionais descritos na literatura

A análise do modelo estrutural Normal, bem como de todos os modelos propostos no restante desse trabalho requer a utilização de procedimentos computacionais. Nosso objetivo então é descrever alguns dos métodos mais populares utilizados na estatística e também apresentar algumas técnicas de diagnósticos de convergência. Entretanto para não quebrar a seqüência das seções colocamos esse tópico na seção 2.4 do Apêndice. As subseções dentro do Apêndice 2.4 abordam os seguintes tópicos:

Na subseção 2.4.1 fazemos uma classificação geral dos métodos computacionais, descrevendo brevemente alguns dos algoritmos mais conhecidos.

Na subseção 2.4.2 apresentamos as idéias básicas sobre os métodos de simulação do tipo MCMC (“Markov Chain Monte Carlo”) descrevendo o algoritmo de Metropolis-Hastings e o algoritmo original de Gibbs (Geman e Geman, 1984, e Gelfand e Smith, 1990).

Nas subseções 2.4.3 e 2.4.4 descrevemos duas variações do algoritmo de Gibbs, o “Gibbs em Blocos” (ou “Grouped Gibbs”) e o “Collapsed Gibbs” (Liu, 1994). Estes métodos tem o objetivo de acelerar a convergência do Gibbs original.

Na subseção 2.4.5 descrevemos o algoritmo híbrido de Metropolis-Hastings em Gibbs (Müller, 1991), conhecido como “Localized Metropolis-Hastings”.

Finalmente, na subseção 2.4.6 do Apêndice apresentamos algumas das técnicas de diagnóstico de convergência e critérios de parada dos algoritmos encontrados na literatura.

2.5 Implementação de algoritmos para o modelo da seção 2.3

Nessa seção apresentamos inicialmente as distribuições condicionais completas à posteriori obtidas a partir do núcleo da “distribuição à posteriori baseada nos dados observados” (que é dada em (2.3.3)). Apresentamos também as condicionais completas à posteriori obtidas a partir do núcleo da distribuição conjunta $p(\underline{x}, \theta | D_{\text{obs}})$ (dada em (2.3.4)). Essas distribuições condicionais serão úteis para implementar os algoritmos MCMC dessa seção.

2.5.1 Distribuições condicionais completas à posteriori baseadas nos dados observados

Baseando-se na posteriori em (2.3.3) (sob o modelo estrutural na formulação 1), obtemos a seguir as distribuições condicionais completas univariadas de α e μ (que são Normais), a distribuição condicional completa conjunta de α e μ (que também é Normal), e as distribuições condicionais completas univariadas de β , σ_x^2 e σ^2 que são distribuições conhecidas a menos de constantes de proporcionalidade (ou seja, apenas os seus núcleos são conhecidos). Essas distribuições estão descritas a seguir.

- A distribuição de α dado $\beta, \mu, \sigma_x^2, \sigma^2$ e D_{obs} é dada por

$$\alpha | \beta, \mu, \sigma_x^2, \sigma^2, D_{\text{obs}} \sim N \left(\frac{n\sigma_\alpha^2[\sigma_x^2(\bar{Y} - \beta\bar{X}) + \sigma^2(\bar{Y} - \beta\mu)] + a|\Sigma|}{n\sigma_\alpha^2(\sigma_x^2 + \sigma^2) + |\Sigma|}; \frac{\sigma_\alpha^2|\Sigma|}{n\sigma_\alpha^2(\sigma_x^2 + \sigma^2) + |\Sigma|} \right), \quad (2.5.1)$$

onde
$$|\Sigma| = \beta^2\sigma^2\sigma_x^2 + \lambda\sigma^2\sigma_x^2 + \lambda(\sigma^2)^2. \quad (2.5.2)$$

- A distribuição de μ dado $\alpha, \beta, \sigma_x^2, \sigma^2$ e D_{obs} é a seguinte:

$$\mu | \alpha, \beta, \sigma_x^2, \sigma^2, D_{\text{obs}} \sim N \left(\frac{n\sigma_\mu^2[\lambda\sigma^2\bar{X} + \beta\sigma^2(\bar{Y} - \alpha)] + m|\Sigma|}{n\sigma_\mu^2[\sigma^2\lambda + \beta^2\sigma^2] + |\Sigma|}; \frac{\sigma_\mu^2|\Sigma|}{n\sigma_\mu^2[\sigma^2\lambda + \beta^2\sigma^2] + |\Sigma|} \right), \quad (2.5.3)$$

onde $|\Sigma|$ é dada em (2.5.2).

- A distribuição conjunta de α e μ dado $\beta, \sigma_x^2, \sigma^2$ e D_{obs} é dada por

$$\pi(\alpha, \mu | \beta \sigma_x^2, \sigma^2, D_{\text{obs}}) = \pi(\alpha | \mu, \beta, \sigma_x^2, \sigma^2, D_{\text{obs}}) \cdot \pi(\mu | \beta, \sigma_x^2, \sigma^2, D_{\text{obs}}), \quad (2.5.4)$$

onde $\pi(\alpha | \mu, \beta, \sigma_x^2, \sigma^2, D_{\text{obs}})$ é dada em (2.5.1) e

$$\mu | \beta, \sigma_x^2, \sigma^2, D_{\text{obs}} \sim N\left(\frac{B}{A}, \frac{1}{A}\right) \quad (2.5.5)$$

onde

$$A = \frac{n\sigma^2(\lambda + \beta^2)}{|\Sigma|} + \frac{1}{\sigma_\mu^2} - \frac{n^2\sigma_\alpha^2\beta^2(\sigma^2)^2}{|\Sigma|(n\sigma_\alpha^2(\sigma_x^2 + \sigma^2) + |\Sigma|)},$$

$$B = \frac{n\sigma^2(\lambda\bar{X} + \beta\bar{Y})}{|\Sigma|} + \frac{m}{\sigma_\mu^2} - \frac{n\beta\sigma^2[n\sigma_\alpha^2(\bar{Y}(\sigma_x^2 + \sigma^2) - \beta\sigma_x^2\bar{X}) + a|\Sigma|]}{|\Sigma|(n\sigma_\alpha^2(\sigma_x^2 + \sigma^2) + |\Sigma|)},$$

e $|\Sigma|$ é dada em (2.5.2). A distribuição (2.5.5) foi obtida integrando-se a posteriori em (2.3.3) em relação a α e depois obtendo-se a condicional completa em relação a μ .

- A distribuição condicional de β dado $\alpha, \mu, \sigma_x^2, \sigma^2$ e D_{obs} é

$$\pi(\beta | \alpha, \mu, \sigma_x^2, \sigma^2, D_{\text{obs}}) \propto |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_\beta^2|\Sigma|}[\beta^2.A - 2\beta.B]\right\}, \quad (2.5.6)$$

onde

$$A = \sigma_\beta^2 \left[\sigma_x^2 \sum_{i=1}^n (X_i - \mu)^2 + 2n\mu\sigma_x^2(\bar{X} - \mu) + n\mu^2(\sigma_x^2 + \sigma^2) \right] + |\Sigma|,$$

$$B = \sigma_\beta^2 \left[n\mu\sigma^2(\bar{Y} - \alpha) + \sigma_x^2 \left(\sum_{i=1}^n X_i Y_i - n\alpha\bar{X} \right) \right] + b|\Sigma|$$

e $|\Sigma|$ é dado por (2.5.2).

De (2.5.3) podemos ver que a distribuição condicional de β baseada nos dados observados não é Normal, já que A , B e $|\Sigma|$ são funções de β .

- A distribuição condicional de σ_x^2 dado $\alpha, \beta, \mu, \sigma^2, D_{\text{obs}}$ é dada por

$$\pi(\sigma_x^2 | \alpha, \beta, \mu, \sigma^2, D_{\text{obs}}) \propto |\Sigma|^{-\frac{n}{2}} (\sigma_x^2)^{-(c+1)} \exp \left\{ -\frac{1}{2} \left[\frac{\sigma_x^2 \sum_{i=1}^n (Y_i - \alpha - \beta X_i)^2}{|\Sigma|} + \frac{2d}{\sigma_x^2} \right] \right\}, \quad (2.5.7)$$

onde $|\Sigma|$ é dada por (2.5.2).

- A distribuição condicional de σ^2 dado $\alpha, \beta, \mu, \sigma_x^2, D_{\text{obs}}$ é dada por

$$\begin{aligned} & \pi(\sigma^2 | \alpha, \beta, \mu, \sigma_x^2, D_{\text{obs}}) & (2.5.8) \\ & \propto |\Sigma|^{-\frac{n}{2}} (\sigma^2)^{-(f+1)} \exp \left\{ -\frac{1}{2} \left\{ \frac{\sigma^2}{|\Sigma|} \left[\sum_{i=1}^n (Y_i - \alpha - \beta \mu)^2 + \lambda \sum_{i=1}^n (X_i - \mu)^2 \right] + \frac{2g}{\sigma^2} \right\} \right\}, \end{aligned}$$

onde $|\Sigma|$ é dada em (2.5.2).

As fórmulas (2.5.1) a (2.5.8) são desenvolvidas na seção 2.5.1 do Apêndice A.

Observação: Das distribuições condicionais completas à posteriori dadas em (2.5.6), (2.5.7) e (2.5.8) conhecemos apenas o núcleo. Nesse caso, podemos utilizar por exemplo o algoritmo híbrido de Metropolis-Hastings dentro do Gibbs, que utiliza o algoritmo Metropolis-Hastings para amostrar de cada condicional completa univariada desconhecida dada acima (na subseção 2.4.5 descrevemos os passos desse algoritmo com detalhe). Uma outra saída seria utilizar a “posteriori baseada nos dados completos” e amostrar da distribuição conjunta $p(\underline{x}, \underline{\theta} | D_{\text{obs}})$ dada em (2.3.4) por meio do algoritmo de Gibbs, que nesse caso é muito simples de ser implementado, já que todas as distribuições condicionais completas necessárias são conhecidas como vemos na subseção a seguir. Note que, uma vez obtida uma amostra de $p(\underline{x}, \underline{\theta} | D_{\text{obs}})$, podemos então descartar a parte da amostra referente a \underline{x} , já que $p(\underline{\theta} | D_{\text{obs}})$ é uma distribuição marginal de $p(\underline{x}, \underline{\theta} | D_{\text{obs}})$.

2.5.2 Distribuições condicionais completas à posteriori baseadas nos dados completos (D)

A seguir apresentamos as distribuições condicionais completas e $\alpha, \beta, \mu, \sigma_x^2, \sigma^2$ e \underline{x} obtidas a partir de $p(\underline{x}, \theta) \propto L(\theta|D)\pi(\theta)$ dada em (2.3.4).

Nesse caso as distribuições condicionais de α, β e μ são Normais e as de σ_x^2 e σ^2 são Gama Invertida, com os respectivos parâmetros dados por:

$$\alpha|\beta, \mu, \sigma_x^2, \sigma^2, D \sim N\left(\frac{n\sigma_\alpha^2(\bar{Y} - \beta\bar{x}) + \lambda a\sigma^2}{n\sigma_\alpha^2 + \lambda\sigma^2}; \frac{\lambda\sigma^2\sigma_\alpha^2}{n\sigma_\alpha^2 + \lambda\sigma^2}\right), \quad (2.5.9)$$

$$\beta|\alpha, \mu, \sigma_x^2, \sigma^2, D \sim N\left(\frac{\sigma_\beta^2(\sum_{i=1}^n x_i Y_i - n\alpha\bar{x}) + \lambda\sigma^2 b}{\sigma_\beta^2 \sum_{i=1}^n x_i^2 + \lambda\sigma^2}; \frac{\lambda\sigma^2\sigma_\beta^2}{\sigma_\beta^2 \sum_{i=1}^n x_i^2 + \lambda\sigma^2}\right), \quad (2.5.10)$$

$$\mu|\alpha, \beta, \sigma_x^2, \sigma^2, D \sim N\left(\frac{n\bar{x}\sigma_\mu^2 + \sigma_x^2 m}{n\sigma_\mu^2 + \sigma_x^2}; \frac{\sigma_x^2\sigma_\mu^2}{n\sigma_\mu^2 + \sigma_x^2}\right), \quad (2.5.11)$$

$$\sigma_x^2|\alpha, \beta, \mu, \sigma^2, D \sim IG\left(\frac{n}{2} + c; d + \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2\right), \quad (2.5.12)$$

$$\sigma^2|\alpha, \beta, \mu, \sigma_x^2, D \sim IG\left(n + f; g + \frac{1}{2} \left(\frac{\sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2}{\lambda} + \sum_{i=1}^n (X_i - x_i)^2\right)\right) \quad (2.5.13)$$

e cada uma das variáveis $x_i, i = 1, \dots, n$, tem a seguinte distribuição condicional completa:

$$x_i|\alpha, \beta, \mu, \sigma_x^2, D_{\text{obs}}, \underline{x}_{(-i)} \sim N\left(\frac{\sigma_x^2(\beta Y_i - \beta\alpha + \lambda X_i) + \lambda\sigma^2\mu}{\sigma_x^2(\beta^2 + \lambda) + \lambda\sigma^2}; \frac{\lambda\sigma^2\sigma_x^2}{\sigma_x^2(\beta^2 + \lambda) + \lambda\sigma^2}\right). \quad (2.5.14)$$

Observação: A “introdução” das variáveis latentes x_1, \dots, x_n leva às condicionais à posteriori serem Normais ou Gammas Invertidas e, portanto, mais simples de serem amostradas computacionalmente.

2.5.3 Implementação do algoritmo original de Gibbs

Para amostrar da posteriori $p(\theta|D_{\text{obs}})$ sob o modelo em (2.3.1) podemos amostrar da distribuição $p(\underline{x}, \theta|D_{\text{obs}})$ dada em (2.3.4) utilizando o algoritmo de Gibbs. Para isso

usamos as distribuições condicionais completas unidimensionais dadas por (2.5.9) até (2.5.14). Os passos do algoritmo são os seguintes:

Passo 0: Escolher arbitrariamente um ponto inicial $\vartheta^{(0)} = (\alpha^{(0)}, \beta^{(0)}, \mu^{(0)}, \sigma_x^{2(0)}, \sigma^2)^T$.

Faça $j = 0$.

Passo 1: Gerar $\underline{x}^{(j+1)}, \vartheta^{(j+1)} = (\underline{x}^{(j+1)}, \alpha^{(j+1)}, \beta^{(j+1)}, \mu^{(j+1)}, \sigma_x^{2(j+1)}, \sigma^2)^T$ como segue

- gerar $\underline{x}^{(j+1)} = (x_1^{(j+1)}, \dots, x_n^{(j+1)})^T$, onde

$$x_i^{(j+1)} \sim \pi(x_i | \alpha^{(j)}, \beta^{(j)}, \mu^{(j)}, \sigma_x^{2(j)}, \sigma^2, D_{\text{obs}}),$$

isto é,

$$x_i^{(j+1)} \sim N \left(\frac{\sigma_x^{2(j)} (\beta^{(j)} Y_i - \beta^{(j)} \alpha^{(j)} + \lambda X_i) + \lambda \sigma^{2(j)} \mu^{(j)}}{\sigma_x^{2(j)} (\beta^{2(j)} + \lambda) + \lambda \sigma^{2(j)}}; \frac{\lambda \sigma^{2(j)} \sigma_x^{2(j)}}{\sigma_x^{2(j)} (\beta^{2(j)} + \lambda) + \lambda \sigma^{2(j)}} \right),$$

$i = 1, \dots, n.$

- gerar $\alpha^{(j+1)} \sim \pi(\alpha | \underline{x}^{(j+1)}, \beta^{(j)}, \mu^{(j)}, \sigma_x^{2(j)}, \sigma^2, D_{\text{obs}})$, isto é,

$$\alpha^{(j+1)} \sim N \left(\frac{n \sigma_\alpha^2 (\bar{Y} - \beta^{(j)} \bar{x}^{(j+1)}) + \lambda \sigma^{2(j)}}{n \sigma_\alpha^2 + \lambda \sigma^{2(j)}}; \frac{\lambda \sigma^{2(j)} \sigma_\alpha^2}{n \sigma_\alpha^2 + \lambda \sigma^{2(j)}} \right),$$

- gerar $\beta^{(j+1)} \sim \pi(\beta | \underline{x}^{(j+1)}, \alpha^{(j+1)}, \mu^{(j)}, \sigma_x^{2(j)}, \sigma^2, D_{\text{obs}})$, isto é,

$$\beta^{(j+1)} \sim N \left(\frac{\sigma_\beta^2 \left(\sum_{i=1}^n x_i^{(j+1)} Y_i - n \alpha^{(j+1)} \bar{x}^{(j+1)} \right) + \lambda \sigma^{2(j)} b}{\sigma_\beta^2 \sum_{i=1}^n (x_i^{(j+1)})^2 + \lambda \sigma^{2(j)}}; \frac{\lambda \sigma^{2(j)} \sigma_\beta^2}{\sigma_\beta^2 \sum_{i=1}^n (x_i^{(j+1)})^2 + \lambda \sigma^{2(j)}} \right),$$

- gerar $\mu^{(j+1)} \sim \pi(\mu | \underline{x}^{(j+1)}, \alpha^{(j+1)}, \beta^{(j+1)}, \sigma_x^{2(j)}, \sigma^2, D_{\text{obs}})$, isto é,

$$\mu^{(j+1)} \sim N \left(\frac{n \bar{x}^{(j+1)} \sigma_\mu^2 + \sigma_x^{2(j)} m}{n \sigma_\mu^2 + \sigma_x^{2(j)}}; \frac{\sigma_x^{2(j)} \sigma_\mu^2}{n \sigma_\mu^2 + \sigma_x^{2(j)}} \right),$$

- gerar $\sigma_x^{2(j+1)} \sim \pi(\sigma_x^2 | \underline{x}^{(j+1)}, \alpha^{(j+1)}, \beta^{(j+1)}, \mu^{(j+1)}, \sigma^2, D_{\text{obs}})$, isto é,

$$\sigma_x^{2(j+1)} \sim IG \left(\frac{n}{2} + c; d + \frac{1}{2} \sum_{i=1}^n (x_i^{(j+1)} - \mu^{(j+1)})^2 \right),$$

- gerar $\sigma^{2^{(j+1)}} \sim \pi(\sigma^2 | \underline{x}^{(j+1)}, \alpha^{(j+1)}, \beta^{(j+1)}, \mu^{(j+1)}, \sigma_x^{2^{(j+1)}}, D_{\text{obs}})$, isto é,

$$\sigma^{2^{(j+1)}} \sim IG \left(n+f; g + \frac{1}{2} \left(\frac{\sum_{i=1}^n (Y_i - \alpha^{(j+1)} - \beta^{(j+1)} x_i^{(j+1)})^2}{\lambda} + \sum_{i=1}^n (X_i - x_i^{(j+1)})^2 \right) \right).$$

Passo 2: Faça $j = j + 1$ e vá para o Passo 1 (até a convergência).

2.5.4 Implementação do algoritmo de Gibbs em blocos

Na seção anterior aplicamos o amostrador de Gibbs para amostrar da posteriori $p(\theta, \underline{x} | D_{\text{obs}})$ em (2.3.4), usando as distribuições condicionais completas em (2.5.9) a (2.5.14). É importante notar entretanto que $\underline{x} = (x_1, \dots, x_n)$ aparece em cada uma das fórmulas de 2.5.9) a (2.5.14), o que significa que existe uma considerável dependência entre cada uma das componentes de θ e \underline{x} (e também entre as iterações), o que pode eventualmente prejudicar o desempenho desse algoritmo. Uma alternativa nesse caso seria usar o algoritmo de Gibbs em blocos.

Nessa seção vamos utilizar o algoritmo de Gibbs em blocos para amostrar de $p(\underline{x}, \theta | D_{\text{obs}})$ dada em (2.3.4). Para isso vamos considerar quatro grupos de componentes, $\{\alpha, \mu, \underline{x}\}$, $\{\beta\}$, $\{\sigma_x^2\}$, $\{\sigma^2\}$, onde a distribuição conjunta $\pi(\alpha, \mu, \underline{x} | \beta, \sigma_x^2, \sigma^2, D_{\text{obs}})$ é uma Normal de dimensão $n+2$ e portanto muito simples de ser amostrada já que $\pi(\alpha, \mu, \underline{x} | \beta, \sigma_x^2, \sigma^2, D_{\text{obs}}) = \pi(\underline{x} | \alpha, \mu, \beta, \sigma_x^2, \sigma^2, D_{\text{obs}}) \pi(\alpha, \mu | \beta, \sigma_x^2, \sigma^2, D_{\text{obs}})$, onde os $x_i | \alpha, \beta, \mu, \sigma_x^2, \sigma^2, D_{\text{obs}}$, para $i = 1, \dots, n$, são condicionalmente independentes com distribuição dada por (2.5.14), além disso, a distribuição conjunta de $\alpha, \mu | \beta, \sigma_x^2, \sigma^2, D_{\text{obs}}$ é dada em (2.5.4). As componentes β, σ_x^2 e σ^2 são amostradas das distribuições em (2.5.10), (2.5.12) e (2.5.13), respectivamente. Com essas informações podemos implementar o algoritmo de Gibbs em blocos e a seguir damos os passos desse algoritmo.

Algoritmo de Gibbs em Blocos

A seguir damos os passos desse algoritmo para amostrar da distribuição conjunta

$p(\vartheta, \underline{x} | D_{\text{obs}})$ dada em (2.3.4).

Passo 0: escolher $\vartheta^{(0)} = (\beta^{(0)}, \sigma_x^{2(0)}, \sigma^2)^T$ e faça $j = 0$.

Passo 1: gerar $(\mu^{(j+1)}, \alpha^{(j+1)}, \underline{x}^{(j+1)})^T$ conjuntamente da seguinte maneira:

- gerar $\mu^{(j+1)} \sim \pi(\mu | \beta^{(j)}, \sigma_x^{2(j)}, \sigma^2, D_{\text{obs}})$ que é

$$N\left(\frac{B^{(j)}}{A^{(j)}}, \frac{1}{A^{(j)}}\right), \quad \text{onde} \quad (2.5.15)$$

$$A^{(j)} = \frac{n\sigma^{2(j)}(\lambda + (\beta^{(j)})^2)}{|\Sigma^{(j)}|} + \frac{1}{\sigma_\mu^2} - \frac{n^2\sigma_\alpha^2(\beta^{(j)})^2(\sigma^{2(j)})^2}{|\Sigma^{(j)}|(n\sigma_\alpha^2(\sigma_x^{2(j)} + \sigma^{2(j)}) + |\Sigma^{(j)}|)},$$

$$B^{(j)} = \frac{n\sigma^{2(j)}(\lambda\bar{X} + \beta^{(j)}\bar{Y})}{|\Sigma^{(j)}|} + \frac{m}{\sigma_\mu^2} - \frac{n\beta^{(j)}\sigma^{2(j)}[n\sigma_\alpha^2(\bar{Y}(\sigma_x^{2(j)} + \sigma^{2(j)}) - \beta^{(j)}\sigma_x^{2(j)}\bar{X}) + a|\Sigma^{(j)}|]}{|\Sigma^{(j)}|(n\sigma_\alpha^2(\sigma_x^{2(j)} + \sigma^{2(j)}) + |\Sigma^{(j)}|)},$$

$$\text{onde } |\Sigma^{(j)}| = (\beta^{(j)})^2\sigma^{2(j)}\sigma_x^{2(j)} + \lambda\sigma^{2(j)}\sigma_x^{2(j)} + \lambda(\sigma^{2(j)})^2 \quad (2.5.16)$$

- gerar $\alpha^{(j+1)} \sim \pi(\alpha | \mu^{(j+1)}, \beta^{(j)}, \sigma_x^{2(j)}, \sigma^2, D_{\text{obs}})$ que é a seguinte normal

$$N\left(\frac{n\sigma_\alpha^2[\sigma_x^{2(j)}(\bar{Y} - \beta^{(j)}\bar{X}) + \sigma^{2(j)}(\bar{Y} - \beta^{(j)}\mu^{(j+1)})] + a|\Sigma^{(j)}|}{n\sigma_\alpha^2(\sigma_x^{2(j)} + \sigma^{2(j)}) + |\Sigma^{(j)}|}, \frac{\sigma_\alpha^2|\Sigma^{(j)}|}{n\sigma_\alpha^2(\sigma_x^{2(j)} + \sigma^{2(j)}) + |\Sigma^{(j)}|}\right), \quad (2.5.17)$$

onde $|\Sigma^{(j)}|$ é dada por (2.5.16),

- gerar as variáveis latentes $\underline{x}^{(j+1)} = (x_1^{(j+1)}, \dots, x_n^{(j+1)})^T$ do seguinte modo:

$$\text{gerar } x_i^{(j+1)} \sim \pi(x_i | \mu^{(j+1)}, \alpha^{(j+1)}, \beta^{(j)}, \sigma_x^{2(j)}, \sigma^2, D_{\text{obs}}), \text{ para}$$

$$i = 1, 2, \dots, n, \text{ onde } \pi(x_i | \mu, \alpha, \beta, \sigma_x^2, \sigma^2, D_{\text{obs}}) \text{ é dada} \quad (2.5.18)$$

por (2.5.14).

Passo 2: Gerar $(\beta^{(j+1)}, \sigma_x^{2(j+1)}, \sigma^{2(j+1)})$ como no algoritmo original de Gibbs da subseção 2.5.2:

- gerar $\beta^{(j+1)} \sim \pi(\beta | \mu^{(j+1)}, \alpha^{(j+1)}, \sigma_x^{2(j)}, \sigma^2, \underline{x}^{(j+1)}, D_{\text{obs}}),$ (2.5.19)

onde $\pi(\beta | \mu, \alpha, \sigma_x^2, \sigma^2, \underline{x}, D_{\text{obs}})$ é dada em (2.5.10),

- gerar $\sigma_x^{2(j+1)} \sim \pi(\sigma_x^2 | \mu^{(j+1)}, \alpha^{(j+1)}, \beta^{(j+1)}, \sigma^{2(j)}, \underline{x}^{(j+1)}, D_{\text{obs}})$,

$$(2.5.20)$$

onde $\pi(\sigma_x^2 | \mu, \alpha, \beta, \sigma^2, \underline{x}, D_{\text{obs}})$ é dada em (2.5.12),

- gerar $\sigma^{2(j+1)} \sim \pi(\sigma^2 | \mu^{(j+1)}, \alpha^{(j+1)}, \beta^{(j+1)}, \sigma_x^{2(j+1)}, \underline{x}^{(j+1)}, D_{\text{obs}})$,

$$(2.5.21)$$

onde $\pi(\sigma^2 | \mu, \alpha, \beta, \sigma_x^2, \underline{x}, D_{\text{obs}})$ é dada em (2.5.13).

Passo 3: Fazer $i = i + 1$ e voltar para o passo 1.

2.5.5 Implementação do algoritmo “modified collapsed Gibbs”

Nesta subseção vamos utilizar o algoritmo “modified collapsed Gibbs” para amostrar da distribuição $p(\theta, \underline{x} | D_{\text{obs}})$ dada em (2.3.4).

A idéia é num primeiro passo gerar (α, μ) , não de $\pi(\alpha, \mu | \beta, \sigma_x^2, \sigma^2, \underline{x}, D_{\text{obs}})$, mas de $\pi(\alpha, \mu | \beta, \sigma_x^2, \sigma^2, D_{\text{obs}})$, onde \underline{x} foi eliminado por integração (isto é, $\pi(\alpha, \mu | \beta, \sigma_x^2, \sigma^2, D_{\text{obs}}) = \int_{\underline{x}} \pi(\alpha, \mu, \underline{x} | \beta, \sigma_x^2, \sigma^2, D_{\text{obs}}) d\underline{x}$) e num segundo passo gerar $(\underline{x}, \beta, \sigma_x^2, \sigma^2)$ de $\pi(\underline{x}, \beta, \sigma_x^2, \sigma^2 | \alpha, \mu, D_{\text{obs}})$, que pode ser feito por meio de um mini-Gibbs utilizando as condicionais completas $\pi(\underline{x} | \alpha, \beta, \mu, \sigma_x^2, \sigma^2, D_{\text{obs}})$, $\pi(\beta | \alpha, \mu, \sigma_x^2, \sigma^2, \underline{x}, D_{\text{obs}})$, $\pi(\sigma_x^2 | \alpha, \beta, \mu, \sigma^2, \underline{x}, D_{\text{obs}})$ e $\pi(\sigma^2 | \alpha, \beta, \mu, \sigma_x^2, \underline{x}, D_{\text{obs}})$, dadas respectivamente por (2.5.14), (2.5.10), (2.5.12) e (2.5.13)

Os passos do algoritmo “modified collapsed Gibbs” são os seguintes:

Passo 0: faça $i = 0$, escolha $(\beta^{(0)}, \sigma_x^{2(0)}, \sigma^{2(0)})$ e faça $j = 1$;

Passo 1: gere $(\alpha^{(i+1)}, \mu^{(i+1)})$ de $\pi(\alpha, \mu | \beta^{(i)}, \sigma_x^{2(i)}, \sigma^{2(i)})$ do seguinte modo:

- gere $\mu^{(i+1)}$ de $\pi(\mu | \beta^{(i)}, \sigma_x^{2(i)}, \sigma^{2(i)}, D_{\text{obs}})$, onde $\pi(\mu | \beta, \sigma_x^2, \sigma^2, D_{\text{obs}})$ é dada em (2.5.5),
- gere $\alpha^{(i+1)}$ de $\pi(\alpha | \mu^{(i+1)}, \beta^{(i)}, \sigma_x^{2(i)}, \sigma^{2(i)}, D_{\text{obs}})$, onde $\pi(\alpha | \mu, \beta, \sigma_x^2, \sigma^2, D_{\text{obs}})$ é dada em (2.5.1);

Passo 2: gere $(\underline{x}_{(j)}^{(i+1)}, \beta_{(j)}^{(i+1)}, \sigma_{x_{(j)}}^{2(i+1)}, \sigma_{(j)}^{2(i+1)})$ de $\pi(\underline{x}, \beta, \sigma_x^2, \sigma_j^2 | \alpha^{(i+1)}, \mu^{(i+1)}, D_{\text{obs}})$ da seguinte maneira:

- gere $\underline{x}_{(j)}^{(i+1)}$ de $\pi(\underline{x} | \alpha^{(i+1)}, \mu^{(i+1)}, \beta_{(j)}^{(i)}, \sigma_{x_{(j)}}^{2(i)}, \sigma_{(j)}^{2(i)}, D_{\text{obs}})$ do seguinte modo:
 - gere $x_{k(j)}^{(i+1)}$ de $\pi(x_k | \alpha^{(i+1)}, \mu^{(i+1)}, \beta_{(j)}^{(i)}, \sigma_{x_{(j)}}^{2(i)}, \sigma_{(j)}^{2(i)}, D_{\text{obs}})$ para $k = 1, 2, \dots, n$, onde $\pi(x_k | \alpha, \mu, \beta, \sigma_x^2, \sigma^2, D_{\text{obs}})$ é dada em (2.5.14),
- gere $\beta_{(j)}^{(i+1)}$ de $\pi(\beta | \alpha^{(i+1)}, \mu^{(i+1)}, \underline{x}_{(j)}^{(i+1)}, \sigma_{(j)}^{2(i)}, \sigma_{(j)}^{(i)}, D_{\text{obs}})$, onde $\pi(\beta | \alpha, \mu, \underline{x}, \sigma_x^2, \sigma^2, D_{\text{obs}})$ é dada em (2.5.10),
- gere $\sigma_{x_{(j)}}^{2(i+1)}$ de $\pi(\sigma_x^2 | \alpha^{(i+1)}, \mu^{(i+1)}, \underline{x}_{(j)}^{(i+1)}, \beta_{(j)}^{(i+1)}, \sigma_{(j)}^{2(i)}, D_{\text{obs}})$, onde $\pi(\sigma_x^2 | \alpha, \mu, \beta, \sigma^2, \underline{x}, D_{\text{obs}})$ é dada em (2.5.12)
- gere $\sigma_{(j)}^{2(i+1)}$ de $\pi(\sigma^2 | \alpha^{(i+1)}, \mu^{(i+1)}, \underline{x}_{(j)}^{(i+1)}, \beta_{(j)}^{(i+1)}, \sigma_{x_{(j)}}^{2(i+1)}, D_{\text{obs}})$, onde $\pi(\sigma^2 | \alpha, \beta, \mu, \sigma_x^2, \sigma^2, \underline{x}, D_{\text{obs}})$ é dada em (2.5.13);

Passo 3: se $j < N$, faça $j = j + 1$ e volte para o passo 2 (onde N é o número de iterações do mini-Gibbs);

Passo 4: faça $i = i + 1$ e volte para o passo 1.

Observações

1. A diferença entre o “grouped” e o “modified collapsed Gibbs” para amostrar da distribuição em (2.3.4) (quando apenas o vetor \underline{x} é eliminado por integração) é que no último utilizamos um mini-Gibbs nas componentes \underline{x} , β , σ_x^2 e σ^2 .
2. Uma outra maneira de implementar o algoritmo “modified collapsed Gibbs” é no passo 1 amostrar apenas o vetor \underline{x} , e não da distribuição $\pi(\underline{x} | \theta, D_{\text{obs}})$, mas de $\pi(\underline{x} | D_{\text{obs}})$ (onde θ é eliminado integrando-se $p(\theta, \underline{x} | D_{\text{obs}})$ em relação a θ) e no passo 2 amostrar θ de $\pi(\theta | \underline{x}, D_{\text{obs}})$. No primeiro passo aplicamos

um mini-Gibbs até a convergência, utilizando para isso as distribuições condicionais completas $\pi(x_i|x_{(-i)}, D_{\text{obs}})$, para $i = 1, \dots, n$, e no segundo passo aplicamos outro mini-Gibbs até que convirja, utilizando para isso as condicionais completas $\pi(\alpha|\beta, \mu, \sigma_x^2, \sigma^2, \underline{x}, D_{\text{obs}})$, $\pi(\beta|\alpha, \mu, \sigma_x^2, \sigma^2, \underline{x}, D_{\text{obs}})$, $\pi(\mu|\alpha, \beta, \sigma_x^2, \sigma^2, \underline{x}, D_{\text{obs}})$, $\pi(\sigma_x^2|\alpha, \beta, \mu, \sigma^2, \underline{x}, D_{\text{obs}})$ e $\pi(\sigma^2|\alpha, \beta, \mu, \sigma_x^2, \underline{x}, D_{\text{obs}})$.

Para obter a distribuição $\pi(\underline{x}|D_{\text{obs}})$ precisamos integrar a posteriori conjunta $p(\underline{\theta}, \underline{x}|D_{\text{obs}})$ em relação a $\underline{\theta} = (\alpha, \beta, \mu, \sigma_x^2, \sigma^2)$, para então obter as condicionais completas $\pi(x_i|x_{(-i)}, D_{\text{obs}})$ para $i = 1, \dots, n$.

2.5.6 Implementação do algoritmo híbrido de Metropolis-Hastings em Gibbs

Para amostrar da “posteriori baseada nos dados observados”, dada em (2.3.3), utilizando o algoritmo de Metropolis-Hastings em Gibbs, devemos fazer inicialmente as seguintes observações:

1. α e μ podem ser facilmente gerados da distribuição $\pi(\alpha, \mu|\beta, \sigma_x^2, \sigma^2, D_{\text{obs}})$ que é Normal bivariada

$$\pi(\alpha, \mu|\beta, \sigma_x^2, \sigma^2, D_{\text{obs}}) = \pi(\alpha|\mu, \beta, \sigma_x^2, \sigma^2, D_{\text{obs}})\pi(\mu|\beta, \sigma_x^2, \sigma^2, D_{\text{obs}}),$$

onde $\pi(\alpha|\mu, \beta, \sigma_x^2, \sigma^2, D_{\text{obs}})$ é uma Normal dada em (2.5.1) e $\pi(\mu|\beta, \sigma_x^2, \sigma^2, D_{\text{obs}})$ é uma Normal dada por (2.5.5).

2. A distribuição condicional completa à posteriori de β dada em (2.5.6) é proporcional ao produto

$$\pi(\beta|\alpha, \mu, \sigma_x^2, \sigma^2, D_{\text{obs}}) \propto \psi_1(\beta|\alpha, \mu, \sigma_x^2, \sigma^2, D_{\text{obs}}) \exp \left\{ -\frac{1}{2\sigma_\beta^2}(\beta - b)^2 \right\},$$

onde o segundo fator é o núcleo de uma $N(b, \sigma_\beta^2)$, onde σ_β^2 é conhecido e

$$\psi_1(\beta|\alpha, \mu, \sigma_x^2, \sigma^2, D_{\text{obs}}) = |\Sigma|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma_\beta^2|\Sigma|}(\beta^2 C - 2\beta D) \right\},$$

é uma função uniformemente limitada em β , onde

$$\begin{aligned} C &= \sigma_\beta^2 \left[\sigma_x^2 \sum_{i=1}^n (X_i - \mu)^2 + 2n\mu\sigma_x^2(\bar{X} - \mu) + n\mu^2(\sigma_x^2 + \sigma^2) \right], \\ D &= \sigma_\beta^2 \left[n\mu\sigma^2(\bar{Y} - \alpha) + \sigma_x^2 \left(\sum_{i=1}^n X_i Y_i - n\alpha\bar{X} \right) \right] \\ e \quad |\Sigma| &= \beta^2 \sigma^2 \sigma_x^2 + \lambda \sigma^2 (\sigma_x^2 + \sigma^2). \end{aligned}$$

- A distribuição condicional completa à posteriori de σ_x^2 , dada em (2.5.7), é proporcional ao produto

$$\pi(\sigma_x^2 | \alpha, \beta, \mu, \sigma^2, D_{\text{obs}}) \propto \psi_2(\sigma_x^2 | \alpha, \beta, \mu, \sigma^2, D_{\text{obs}}) \cdot (\sigma_x^2)^{-(c+1)} \exp \left\{ -\frac{d}{\sigma_x^2} \right\},$$

onde o segundo fator é o núcleo de uma distribuição $IG(c, d)$ e

$$\psi_2(\sigma_x^2 | \alpha, \beta, \mu, \sigma^2, D_{\text{obs}}) = |\Sigma|^{-\frac{n}{2}} \exp \left\{ \frac{-\sigma_x^2 \sum_{i=1}^n (Y_i - \alpha - \beta X_i)^2}{2|\Sigma|} \right\},$$

onde $|\Sigma| = \sigma_x^2 (\beta^2 \sigma^2 + \lambda \sigma^2) + \lambda (\sigma^2)^2$. Além disso, ψ_2 é uma função uniformemente limitada (em σ_x^2).

- A distribuição condicional completa à posteriori de σ^2 , dada em (2.5.8) pode ser escrita como

$$\pi(\sigma^2 | \alpha, \beta, \mu, \sigma_x^2, D_{\text{obs}}) \propto \psi_3(\alpha, \beta, \mu, \sigma_x^2, \sigma^2, D_{\text{obs}}) \cdot (\sigma^2)^{-(f+1+\frac{n}{2})} \exp \left\{ -\frac{g}{\sigma^2} \right\},$$

onde o segundo fator é o núcleo de uma $IG(f + \frac{n}{2}; g)$ e

$$\psi_3(\alpha, \beta, \mu, \sigma_x^2, \sigma^2, D_{\text{obs}}) = \left(\frac{|\Sigma|}{\sigma^2} \right)^{-\frac{n}{2}} \exp \left\{ -\frac{\sigma^2}{|\Sigma|} \left[\sum_{i=1}^n (Y_i - \alpha - \beta \mu)^2 + \lambda \sum_{i=1}^n (X_i - \mu)^2 \right] \right\},$$

onde $|\Sigma| = \sigma^2 [(\beta^2 \sigma_x^2 + \lambda \sigma_x^2) + \lambda \sigma^2]$. Além disso, ψ_3 é uniformemente limitada (em σ^2).

A seguir apresentamos os passos do algoritmo híbrido de Metropolis-Hastings em Gibbs para esse problema.

Passos do Algoritmo de Metropolis-Hastings em Gibbs (para o caso particular em que $T = 1$)

Passo 0: Escolher $\theta^{(0)} = (\beta^{(0)}, \sigma_x^{2(0)}, \sigma^2)^T$ e faça $j = 0$.

Passo 1: Gerar $\theta^{(i+1)} = (\alpha^{(i+1)}, \mu^{(i+1)}, \beta^{(i+1)}, \sigma_x^{2(i+1)}, \sigma^{2(i+1)})^T$ da seguinte maneira:

- gerar $(\alpha^{(j+1)}, \mu^{(j+1)})$ conjuntamente do seguinte modo:
 - gerar $\mu^{(j+1)} \sim \pi(\mu|\beta^{(j)}, \sigma_x^{2(j)}, \sigma^{2(j)}, D_{\text{obs}})$ que é dada em (2.5.15) e (2.5.16)
 - gerar $\alpha^{(j+1)} \sim \pi(\alpha|\mu^{(j+1)}, \beta^{(j)}, \sigma_x^{2(j)}, \sigma^{2(j)}, D_{\text{obs}})$,
- onde $\pi(\alpha|\mu^{(j+1)}, \beta^{(j)}, \sigma_x^{2(j)}, \sigma^{2(j)}, D_{\text{obs}})$ é dada em (2.5.17).

- gerar $\beta^{(j+1)} \sim \pi(\beta|\alpha^{(j+1)}, \mu^{(j+1)}, \sigma_x^{2(j)}, \sigma^{2(j)}, D_{\text{obs}})$ da seguinte maneira:

- gerar o candidato $\beta^{*(j+1)}$ a partir de $N(b, \sigma_\beta^2)$ e U da $U(0, 1)$,
- fazer

$$\alpha(\beta^{*(j+1)}|\beta^{(j)}) = \min \left\{ \frac{\psi_1(\beta^{*(j+1)}|\alpha^{(j+1)}, \mu^{(j+1)}, \sigma_x^{2(j)}, \sigma^{2(j)}, D_{\text{obs}})}{\psi_1(\beta^{(j)}|\alpha^{(j+1)}, \mu^{(j+1)}, \sigma_x^{2(j)}, \sigma^{2(j)}, D_{\text{obs}})}; 1 \right\},$$

onde

$$\begin{aligned} & \psi_1(\beta^{*(j+1)}|\alpha^{(j+1)}, \mu^{(j+1)}, \sigma_x^{2(j)}, \sigma^{2(j)}, D_{\text{obs}}) \\ &= |\Sigma^*|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma_\beta^2|\Sigma^*|} ((\beta^{*(j+1)})^2 C - \beta^{*(j+1)}) D \right\} \\ & \psi_1(\beta^{(j)}|\alpha^{(j+1)}, \mu^{(j+1)}, \sigma_x^{2(j)}, \sigma^{2(j)}) \\ &= |\Sigma^{(j)}|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma_\beta^2|\Sigma^{(j)}|} ((\beta^{(j)})^2 C - \beta^{(j)}) D \right\} \end{aligned}$$

onde

$$\begin{aligned} C &= \sigma_\beta^2 \left[\sigma_x^{2(j)} \sum_{i=1}^n (X_i - \mu^{(j+1)})^2 + 2n\mu^{(j+1)}\sigma_x^{2(j)} (\bar{X} - \mu^{(j+1)}) \right. \\ & \quad \left. + n(\mu^{(j+1)})^2(\sigma_x^{2(j)} + \sigma^{2(j)}) \right], \\ D &= \sigma_\beta^2 \left[n\mu^{(j+1)}\sigma^{2(j)} (\bar{Y} - \alpha^{(j+1)}) + \sigma_x^{2(j)} \left(\sum_{i=1}^n X_i Y_i - n\alpha^{(j+1)}\bar{X} \right) \right], \\ |\Sigma^*| &= (\beta^{*(j+1)})^2 \sigma^{2(j)} \sigma_e^2 + \lambda \sigma^{2(j)} (\sigma_x^{2(j)} + \sigma^{2(j)}), \\ |\Sigma^{(j)}| &= (\beta^{(j)})^2 \sigma^{2(j)} \sigma_e^2 + \lambda \sigma^{2(j)} (\sigma_x^{2(j)} + \sigma^{2(j)}). \end{aligned}$$

- Se $U < \alpha(\beta^{*(j+1)}|\beta^{(j)})$, faça $\beta^{(j+1)} = \beta^{*(j+1)}$; caso contrário faça $\beta^{(j+1)} = \beta^{(j)}$.

■ Gerar $\sigma_x^{2^{(j+1)}} \sim \pi(\sigma_x^2|\alpha^{(j+1)}, \mu^{(j+1)}, \beta^{(j+1)}, \sigma^{2^{(j)}}, D_{\text{obs}})$ do seguinte modo:

- gerar um candidato $\sigma_x^{2^{*(j+1)}}$ a partir da $IG(c, d)$ e gerar $U \sim U(0, 1)$,
- fazer

$$\alpha(\sigma_x^{2^{*(j+1)}}|\sigma_x^{2^{(j)}}) = \min \left\{ \frac{\psi_2(\sigma_x^{2^{*(j+1)}}|\alpha^{(j+1)}, \mu^{(j+1)}, \beta^{(j+1)}, \sigma^{2^{(j)}}, D_{\text{obs}})}{\psi_2(\sigma_x^{2^{(j)}}|\alpha^{(j+1)}, \mu^{(j+1)}, \beta^{(j+1)}, \sigma^{2^{(j)}}, D_{\text{obs}})}; 1 \right\},$$

onde

$$\psi_2(\sigma_x^{2^{*(j+1)}}|\alpha^{(j+1)}, \mu^{(j+1)}, \beta^{(j+1)}, \sigma^{2^{(j)}}, D_{\text{obs}}) = |\Sigma^*|^{-\frac{n}{2}} \exp \left\{ -\frac{\sigma_x^{2^{*(j+1)}}}{2|\Sigma^*|} \cdot A \right\},$$

$$\psi_2(\sigma_x^{2^{(j)}}|\alpha^{(j+1)}, \mu^{(j+1)}, \beta^{(j+1)}, \sigma^{2^{(j)}}, D_{\text{obs}}) = |\Sigma|^{-\frac{n}{2}} \exp \left\{ -\frac{\sigma_x^{2^{(j)}}}{2|\Sigma|} \cdot A \right\},$$

onde

$$A = \sum_{i=1}^n (Y_i - \alpha^{(i+1)} - \beta^{(i+1)} X_i)^2,$$

$$|\Sigma^*| = \sigma_x^{2^{*(j+1)}} [(\beta^{(j+1)})^2 \sigma^{2^{(j)}} + \lambda \sigma^{2^{(j)}}] + \lambda (\sigma^{2^{(j)}})^2,$$

$$|\Sigma| = \sigma_x^{2^{(j)}} ((\beta^{(j+1)})^2 \sigma^{2^{(j)}} + \lambda \sigma^{2^{(j)}}) + \lambda (\sigma^{2^{(j)}})^2.$$

- Se $U < \alpha(\sigma_x^{2^{*(j+1)}}|\sigma_x^{2^{(j)}})$, faça $\sigma_x^{2^{(j+1)}} = \sigma_x^{2^{*(j+1)}}$; caso contrário faça $\sigma_x^{2^{(j+1)}} = \sigma_x^{2^{(j)}}$.

■ Gerar $\sigma^{2^{(j+1)}} \sim \pi(\sigma^2|\alpha^{(j+1)}, \mu^{(j+1)}, \beta^{(j+1)}, \sigma_x^{2^{(j+1)}}, D_{\text{obs}})$ da seguinte forma:

- gerar um candidato $\sigma^{2^{*(j+1)}}$ da $IG(f + \frac{n}{2}; g)$ e gerar U da $U(0, 10)$,
- fazer

$$\alpha(\sigma^{2^{*(j+1)}}|\sigma^{2^{(j)}}) = \min \left\{ \frac{\psi_3(\sigma^{2^{*(j+1)}}|\alpha^{(j+1)}, \mu^{(j+1)}, \beta^{(j+1)}, \sigma_x^{2^{(j+1)}}, D_{\text{obs}})}{\psi_3(\sigma^{2^{(j)}}|\alpha^{(j+1)}, \mu^{(j+1)}, \beta^{(j+1)}, \sigma_x^{2^{(j+1)}}, D_{\text{obs}})}; 1 \right\},$$

onde

$$\psi_3(\sigma^{2^{*(j+1)}}|\alpha^{(j+1)}, \beta^{(j+1)}, \mu^{(j+1)}, \sigma_x^{2^{(j+1)}}, D_{\text{obs}})$$

$$= (A^*)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{A^*} \left[\sum_{i=1}^n (Y_i - \alpha^{(i+1)} - \beta^{(i+1)} \mu^{(i+1)})^2 + \lambda \sum_{i=1}^n (X_i - \mu^{(i+1)})^2 \right] \right\}$$

$$\begin{aligned} & \psi_3(\sigma^{2^{(j)}} | \alpha^{(j+1)}, \beta^{(j+1)}, \mu^{(j+1)}, \sigma_x^{2^{(j+1)}}, D_{\text{obs}}) \\ &= (A)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{A} \left[\sum_{i=1}^n (Y_i - \alpha^{(i+1)} - \beta^{(i+1)} \mu^{(i+1)})^2 + \lambda \sum_{i=1}^n (X_i - \mu^{(i+1)})^2 \right] \right\}, \end{aligned}$$

$$\begin{aligned} \text{onde} \quad A^* &= (\beta^{(j+1)})^2 \sigma_x^{2^{(j+1)}} + \lambda \sigma_x^{2^{(j+1)}} + \lambda \sigma^{2^{*(j+1)}}, \\ A &= (\beta^{(j+1)})^2 \sigma_x^{2^{(j+1)}} + \lambda \sigma_x^{2^{(j+1)}} + \lambda \sigma^{2^{(j)}}. \end{aligned}$$

• se $U < \alpha(\sigma^{2^{*(j+1)}} | \sigma^{2^{(j)}})$ faça $\sigma^{2^{(j+1)}} = \sigma^{2^{*(j+1)}}$; caso contrário, $\sigma^{2^{(j+1)}} = \sigma^{2^{(j)}}$.

Passo 2: Faça $i = i + 1$ e volte ao passo 1.

Observação: Agrupando-se as componentes α e μ (isto é, amostrando-se conjuntamente α e μ , ao invés de individualmente) ganha-se em termos de velocidade de convergência do algoritmo.

2.6 O modelo estrutural não identificável com particular priori de componentes independentes

No modelo estrutural Normal sem restrição de identificabilidade em (2.2.1), dado pelas equações

$$\begin{aligned} & Y_i = \alpha + \beta x_i + e_i, \\ \text{e} \quad & X_i = x_i + u_i, \quad \text{onde} \\ & \begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \stackrel{\text{i.i.d.}}{\sim} N \left(\begin{pmatrix} 0 \\ 0 \\ \mu \end{pmatrix}; \begin{pmatrix} \sigma_e^2 & 0 & 0 \\ 0 & \sigma_u^2 & 0 \\ 0 & 0 & \sigma_x^2 \end{pmatrix} \right), \quad i = 1, 2, \dots, n, \end{aligned} \quad (2.6.1)$$

e onde $\theta = (\alpha, \beta, \mu, \sigma_x^2, \sigma_u^2, \sigma_e^2)^T$, acrescentamos a suposição à priori que α , β , μ , σ_x^2 , σ_e^2 e σ_u^2 são independentes, com $\alpha \sim N(a, \sigma_\alpha^2)$, $\beta \sim N(b, \sigma_\beta^2)$, $\mu \sim N(m, \sigma_\mu^2)$, $\sigma_x^2 \sim IG(c, d)$, $\sigma_u^2 \sim IG(f, g)$ e $\sigma_e^2 \sim IG(p, q)$, onde $a, b, m, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\mu^2, c, d, f, g, p$ e q são hiperparâmetros conhecidos.

Distribuição à posteriori baseada nos dados observados (D_{obs})

A posteriori baseada nos dados observados sob esse modelo é dada por

$$p(\theta|D_{\text{obs}}) \propto L(\theta|D_{\text{obs}})\pi(\theta), \text{ onde } L(\theta|D_{\text{obs}}) \text{ é dada por (2.1.4)}$$

e $\pi(\theta) = \pi(\alpha).\pi(\beta).\pi(\mu).\pi(\sigma_x^2).\pi(\sigma_e^2).\pi(\sigma_u^2)$, isto é,

$$p(\theta|D_{\text{obs}}) \propto |\Sigma|^{-\frac{n}{2}} (\sigma_x^2)^{-(c+1)} (\sigma_u^2)^{-(f+1)} (\sigma_e^2)^{-(p+1)} \exp\left\{-\frac{1}{2}\left[\left((\beta^2\sigma_x^2 + \sigma_e^2) \sum_{i=1}^n (X_i - \mu)^2 - 2\beta\sigma_x^2 \sum (Y_i - \alpha - \beta\mu)(X_i - \mu) + (\sigma_x^2 + \sigma_u^2) \sum_{i=1}^n (Y_i - \alpha - \beta\mu)^2\right) / |\Sigma| + \frac{(\alpha - a)^2}{\sigma_\alpha^2} + \frac{(\beta - b)^2}{\sigma_\beta^2} + \frac{(\mu - m)^2}{\sigma_\mu^2} + \frac{2d}{\sigma_x^2} + \frac{2g}{\sigma_u^2} + \frac{2q}{\sigma_e^2}\right]\right\}, \quad (2.6.2)$$

$$\text{onde} \quad |\Sigma| = \beta^2\sigma_u^2\sigma_x^2 + \sigma_e^2\sigma_x^2 + \sigma_e^2\sigma_u^2. \quad (2.6.3)$$

Posteriori baseada nos dados completos (D)

$$p(\theta|D_{\text{obs}}) = \int_{\mathfrak{X}} p(\mathfrak{x}, \theta|D_{\text{obs}}) d\mathfrak{x}$$

onde $p(\mathfrak{x}, \theta|D_{\text{obs}}) \propto L(\theta|D)\pi(\theta)$ e $L(\theta|D)$ é dada por (2.1.5) e $\pi(\theta) = \pi(\alpha)\pi(\beta).\pi(\mu)\pi(\sigma_x^2)\pi(\sigma_u^2)\pi(\sigma_e^2)$, ou seja,

$$p(\mathfrak{x}, \theta|D_{\text{obs}}) \propto (\sigma_x^2)^{-(\frac{n}{2}+c+1)} (\sigma_u^2)^{-(\frac{n}{2}+f+1)} (\sigma_e^2)^{-(\frac{n}{2}+p+1)} \exp\left\{-\frac{1}{2}\left[\frac{\sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2}{\sigma_e^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma_u^2} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_x^2} + \frac{(\alpha - a)^2}{\sigma_\alpha^2} + \frac{(\beta - b)^2}{\sigma_\beta^2} + \frac{(\mu - m)^2}{\sigma_\mu^2} + \frac{2d}{\sigma_x^2} + \frac{2g}{\sigma_u^2} + \frac{2q}{\sigma_e^2}\right]\right\}. \quad (2.6.4)$$

Distribuições condicionais completas “baseadas nos dados observados”

Usando a posteriori em (2.6.2) obtemos as condicionais completas univariadas de α e de μ que são Normais, a distribuição condicional completa conjunta de α e μ (que também é Normal) e as distribuições condicionais completas univariadas de

$\beta, \sigma_x^2, \sigma_u^2, \sigma_e^2$ que são conhecidas apenas por seus núcleos. Essas distribuições são apresentadas a seguir.

- $\alpha|\beta, \mu, \sigma_x^2, \sigma_u^2, \sigma_e^2, D_{\text{obs}}$

$$\sim N\left(\frac{n\sigma_\alpha^2[\sigma_x^2(\bar{Y}-\beta\bar{X})+\sigma_u^2(\bar{Y}-\beta\mu)]+a|\Sigma|}{n\sigma_\alpha^2(\sigma_x^2+\sigma_u^2)+|\Sigma|}, \frac{\sigma_\alpha^2|\Sigma|}{n\sigma_\alpha^2(\sigma_x^2+\sigma_u^2)+|\Sigma|}\right), \quad (2.6.5)$$

onde $|\Sigma|$ é dada em (2.6.3),

- $\mu|\alpha, \beta, \sigma_x^2, \sigma_u^2, \sigma_e^2, D_{\text{obs}}$

$$\sim N\left(\frac{n\sigma_\mu^2[\lambda\sigma^2\bar{X}+\beta\sigma_u^2(\bar{Y}-\alpha)]+m|\Sigma|}{n\sigma_\mu^2(\sigma_e^2+\beta^2\sigma_u^2)+|\Sigma|}, \frac{\sigma_\mu^2|\Sigma|}{n\sigma_\mu^2(\sigma_e^2+\beta^2\sigma_u^2)+|\Sigma|}\right), \quad (2.6.6)$$

onde $|\Sigma|$ é dada em (2.6.3),

- $\pi(\alpha, \mu|\beta, \sigma_x^2, \sigma_e^2, \sigma_u^2, D_{\text{obs}}) =$

$$\pi(\alpha|\beta, \mu, \sigma_x^2, \sigma_e^2, \sigma_u^2, D_{\text{obs}})\pi(\mu|\beta, \sigma_x^2, \sigma_e^2, \sigma_u^2, D_{\text{obs}}),$$

onde $\alpha|\beta, \mu, \sigma_x^2, \sigma_e^2, \sigma_u^2, D_{\text{obs}}$ tem distr. dada por (2.6.5) e

$$\mu|\beta, \sigma_x^2, \sigma_u^2, \sigma_e^2, D_{\text{obs}} \sim N\left(\frac{B}{A}, \frac{1}{A}\right), \quad \text{onde} \quad (2.6.7)$$

$$A = \frac{n(\sigma_e^2 + \beta^2\sigma_u^2)}{|\Sigma|} + \frac{1}{\sigma_\mu^2} - \frac{n^2\sigma_\alpha^2\beta^2(\sigma_u^2)^2}{|\Sigma|[n\sigma_\alpha^2(\sigma_x^2 + \sigma_u^2) + |\Sigma|]},$$

$$B = \frac{n(\bar{X}\sigma_e^2 + \bar{Y}\beta\sigma_u^2)}{|\Sigma|} + \frac{m}{\sigma_\mu^2} - \frac{n\beta\sigma_u^2[n\sigma_\alpha^2(\bar{Y}(\sigma_x^2 + \sigma_u^2) - \beta\sigma_x^2\bar{X}) + a|\Sigma|]}{|\Sigma|[n\sigma_\alpha^2(\sigma_x^2 + \sigma_u^2) + |\Sigma|]}$$

e $|\Sigma|$ é dada por (2.6.3).

- $\pi(\beta|\alpha, \mu, \sigma_x^2, \sigma_u^2, \sigma_e^2, D_{\text{obs}}) \propto |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_\beta^2|\Sigma|}[\beta^2C - 2\beta D]\right\}, \quad (2.6.8)$

onde

$$C = \sigma_\beta^2 \left[\sigma_x^2 \sum_{i=1}^n (X_i - \mu)^2 + 2n\mu\sigma_x^2(\bar{X} - \mu) + n\mu^2(\sigma_x^2 + \sigma_u^2) \right] + |\Sigma|,$$

$$D = \sigma_\beta^2 \left[n\mu\sigma^2(\bar{Y} - \alpha) + \sigma_x^2 \left(\sum_{i=1}^n X_i Y_i - n\alpha\bar{X} \right) \right] + b|\Sigma|,$$

$|\Sigma|$ é dada por (2.6.3),

- $\pi(\sigma_x^2|\alpha, \mu, \sigma_u^2, \sigma_e^2, D_{\text{obs}})$

$$\propto |\Sigma|^{-\frac{n}{2}} (\sigma_x^2)^{-(c+1)} \exp\left\{-\frac{1}{2} \left[\frac{\sigma_x^2 \sum_{i=1}^n (Y_i - \alpha - \beta X_i)^2}{|\Sigma|} + \frac{2d}{\sigma_x^2} \right]\right\}, \quad (2.6.9)$$

onde $|\Sigma|$ é dada em (2.6.3),

- $\pi(\sigma_u^2|\alpha, \mu, \beta, \sigma_x^2, \sigma_e^2, D_{\text{obs}})$
 $\propto |\Sigma|^{-\frac{n}{2}} (\sigma_u^2)^{-(f+1)} \exp \left\{ -\frac{1}{2} \left[\frac{\sigma_u^2 \sum_{i=1}^n (Y_i - \alpha - \beta\mu)^2}{|\Sigma|} + \frac{2g}{\sigma_u^2} \right] \right\},$ (2.6.10)

onde $|\Sigma|$ é dada por (2.6.3),

- $\pi(\sigma_e^2|\alpha, \beta, \mu, \sigma_x^2, \sigma_u^2, D_{\text{obs}})$
 $\propto |\Sigma|^{-\frac{n}{2}} (\sigma_e^2)^{-(p+1)} \exp \left\{ -\frac{1}{2} \left[\frac{\sigma_e^2 \sum_{i=1}^n (X_i - \mu)^2}{|\Sigma|} + \frac{2q}{\sigma_e^2} \right] \right\},$ (2.6.11)

onde $|\Sigma|$ é dada por (2.6.3).

Distribuições condicionais completas à posteriori “baseadas nos dados completos”

Usando a posteriori em (2.6.4), obtemos as seguintes distribuições condicionais completas:

- $\pi(\alpha|\beta, \mu, \sigma_x^2, \sigma_u^2, \sigma_e^2, \underline{x}, D_{\text{obs}}) \sim N \left(\frac{n\sigma_\alpha^2(\bar{Y} - \beta\bar{x}) + a\sigma_e^2}{n\sigma_\alpha^2 + \sigma_e^2}, \frac{\sigma_e^2\sigma_\alpha^2}{n\sigma_\alpha^2 + \sigma_e^2} \right),$ (2.6.12)

- $\pi(\beta|\alpha, \mu, \sigma_x^2, \sigma_u^2, \sigma_e^2, D_{\text{obs}}) \sim N \left(\frac{\sigma_\beta^2 (\sum_{i=1}^n x_i Y_i - n\alpha\bar{x}) + b\sigma_e^2}{\sigma_\beta^2 \sum_{i=1}^n x_i^2 + \sigma_e^2}, \frac{\sigma_e^2\sigma_\beta^2}{\sigma_\beta^2 \sum_{i=1}^n x_i^2 + \sigma_e^2} \right)$ (2.6.13)

- $\pi(\mu|\alpha, \beta, \sigma_x^2, \sigma_u^2, \sigma_e^2, \underline{x}, D_{\text{obs}})$ é Normal dada em (2.5.11), (2.6.14)

- $\pi(\sigma_x^2|\alpha, \beta, \mu, \sigma_u^2, \sigma_e^2, \underline{x}, D_{\text{obs}})$ é Gama Invertida dada em (2.5.12), (2.6.15)

- $\pi(\sigma_u^2|\alpha, \beta, \mu, \sigma_x^2, \sigma_e^2, \underline{x}, D_{\text{obs}}) \sim IG \left(\frac{n}{2} + f; \frac{1}{2} \sum_{i=1}^n (X_i - x_i)^2 + g \right),$ (2.6.16)

- $\pi(\sigma_e^2|\alpha, \beta, \mu, \sigma_x^2, \sigma_u^2, \underline{x}, D_{\text{obs}}) \sim IG \left(\frac{n}{2} + p; \frac{1}{2} \sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2 + q \right),$ (2.6.17)

- $\pi(x_i|\alpha, \beta, \mu, \sigma_x^2, \sigma_u^2, \sigma_e^2, x_{(-i)}, D_{\text{obs}}) = \pi(x_i|\alpha, \beta, \mu, \sigma_x^2, \sigma_u^2, \sigma_e^2, D_{\text{obs}})$
 $\sim N \left(\frac{\sigma_x^2 [\beta\sigma_u^2(Y_i - \alpha) + X_i\sigma_e^2] + \sigma_e^2\sigma_u^2\mu}{\sigma_u^2\sigma_x^2\beta^2 + \sigma_e^2\sigma_x^2 + \sigma_e^2\sigma_u^2}, \frac{\sigma_e^2\sigma_u^2\sigma_x^2}{\sigma_u^2\sigma_x^2\beta^2 + \sigma_e^2\sigma_x^2 + \sigma_e^2\sigma_u^2} \right),$ (2.6.18)

para $i = 1, \dots, n$.

Implementação de Algoritmos

Podemos utilizar as distribuições condicionais completas acima para implementar o algoritmo original de Gibbs, o Gibbs em blocos, o Metropolis-Hastings em Gibbs e o “Collapsed Gibbs” de forma totalmente semelhante à Seção 2.5.

2.7 O modelo estrutural com particular priori própria de componentes condicionalmente independentes

Nesta seção analisamos tanto o modelo Normal com a restrição de identificabilidade $\sigma_e^2 = \lambda\sigma_u^2$, onde λ é conhecido, como também o modelo Normal sem restrições de identificabilidade, usando em cada caso uma particular priori própria com componentes de locação condicionalmente independentes.

2.7.1 O modelo com restrição de identificabilidade

No modelo estrutural dado pelas equações

$$\begin{aligned} Y_i &= \alpha + \beta x_i + e_i, \\ X_i &= x_i + u_i, \quad \text{onde} \end{aligned}$$

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ \mu \end{pmatrix}; \begin{pmatrix} \lambda\sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma_x^2 \end{pmatrix} \right), \quad i = 1, \dots, n,$$

onde $\theta = (\alpha, \beta, \mu, \sigma_x^2, \sigma^2)^T$ e $\sigma_e^2 = \lambda\sigma_u^2 = \lambda\sigma^2$ com λ conhecido e acrescentamos a suposição à priori que $\pi(\theta) = \pi(\alpha|\sigma^2)\pi(\beta|\sigma^2)\pi(\mu|\sigma_x^2)\pi(\sigma_x^2)\pi(\sigma^2)$, (2.7.1)

onde $\alpha|\sigma^2 \sim N(a, \chi_{01}^2\sigma^2)$, $\beta|\sigma^2 \sim N(b, \chi_{02}^2\sigma^2)$, $\mu|\sigma_x^2 \sim N(m, \tau_0^2\sigma_x^2)$ e $\sigma_x^2 \sim IG(\nu_{01}, \lambda_{01})$, $\sigma^2 \sim IG(\nu_{02}, \lambda_{02})$, onde $a, b, m, \chi_{01}^2, \chi_{02}^2, \tau_0^2, \nu_{01}, \nu_{02}, \lambda_{01}, \lambda_{02}$ são os hiperparâmetros conhecidos.

A escolha dessa priori baseou-se primeiramente na analogia com o modelo de regressão clássico (sem erros de medida em X) com priori conjugada, além do fato

de que em muitas situações na prática é bastante razoável assumir que parâmetros de localização dependem de parâmetros de escala.

Distribuição à posteriori, baseada nos dados observados

$$p(\vartheta|D_{\text{obs}}) \propto L(\vartheta|D_{\text{obs}})\pi(\vartheta), \quad (2.7.2)$$

onde $L(\vartheta|D_{\text{obs}})$ é obtida de (2.1.4) quando fazemos $\sigma_e^2 = \lambda\sigma_u^2$ e $\sigma_u^2 = \sigma^2$.

Posteriori baseada nos dados completos (D)

$$p(\vartheta|D_{\text{obs}}) = \int_{\mathbb{X}} p(\underline{x}, \vartheta|D_{\text{obs}})d\underline{x}, \quad (2.7.3)$$

onde $p(\underline{x}, \vartheta|D_{\text{obs}}) \propto L(\vartheta|D)\pi(\vartheta)$ e $L(\vartheta|D)$ é obtida de (2.1.5) quando fazemos $\sigma_e^2 = \lambda\sigma_u^2$ e $\sigma_u^2 = \sigma^2$.

Distribuições condicionais completas à posteriori “baseadas nos dados observados”

Usando a posteriori em (2.7.2) obtemos as seguintes distribuições:

- $\alpha|\beta, \mu, \sigma_x^2, \sigma^2, D_{\text{obs}} \sim$ Normal dada em (2.5.1), substituindo σ_α^2 por $\chi_{01}^2\sigma^2$;
- $\mu|\alpha, \beta, \sigma_x^2, \sigma^2, D_{\text{obs}} \sim$ Normal dada em (2.5.3) substituindo σ_μ^2 por $\tau_0^2\sigma_x^2$;
- $\pi(\alpha, \mu|\beta, \sigma_x^2, \sigma^2, D_{\text{obs}}) = \pi(\alpha|\beta, \mu, \sigma_x^2, \sigma^2, D_{\text{obs}})\pi(\beta|\mu, \sigma_x^2, \sigma^2, D_{\text{obs}})$, onde $\mu|\beta, \sigma_x^2, \sigma^2, D_{\text{obs}}$ tem distribuição dada em (2.5.5) substituindo σ_μ^2 por $\tau_0^2\sigma_x^2$ e σ_α^2 por $\chi_{01}^2\sigma^2$;
- $\beta|\alpha, \mu, \sigma_x^2, \sigma^2, D_{\text{obs}}$ tem distribuição dada em (2.5.6) substituindo σ_β^2 por $\chi_{02}^2\sigma^2$;
- $\sigma_x^2|\alpha, \beta, \mu, \sigma^2, D_{\text{obs}}$ tem distribuição dada em (2.5.7) substituindo c por ν_{01} e d por λ_{01} ;

- $\sigma^2|\alpha, \beta, \mu, \sigma_x^2, D_{\text{obs}}$ tem distribuição dada em (2.5.8) substituindo f por ν_{02} e g por λ_{02} .

Distribuições condicionais completas à posteriori “baseadas nos dados completos”

Usando a posteriori em (2.7.3) obtemos as seguintes distribuições condicionais:

- $\alpha|\beta, \mu, \sigma_x^2, \sigma^2, \underline{x}, D_{\text{obs}} \sim \text{Normal}$ dada em (2.5.9) substituindo σ_α por $\chi_{01}^2\sigma^2$, ou seja,

$$\alpha|\beta, \mu, \sigma_x^2, \sigma^2, \underline{x}, D_{\text{obs}} \sim N\left(\frac{n\chi_{01}^2(\bar{Y} - \beta\bar{x}) + \lambda a}{n\chi_{01}^2 + \lambda}; \frac{\lambda\chi_{01}^2\sigma^2}{n\chi_{01}^2 + \lambda}\right);$$

- $\beta|\alpha, \mu, \sigma_x^2, \sigma^2, \underline{x}, D_{\text{obs}} \sim \text{Normal}$ dada em (2.5.10) substituindo σ_β^2 por $\chi_{02}^2\sigma^2$, ou seja,

$$\beta|\alpha, \mu, \sigma_x^2, \sigma^2, \underline{x}, D_{\text{obs}} \sim N\left(\frac{\chi_{02}^2[\sum_{i=1}^n x_i Y_i - n\alpha\bar{x}] + \lambda b}{\chi_{02}^2 \sum_{i=1}^n x_i^2 + \lambda}; \frac{\lambda\chi_{02}^2\sigma^2}{\chi_{02}^2 \sum_{i=1}^n x_i^2 + \lambda}\right);$$

- $\mu|\alpha, \beta, \sigma_x^2, \sigma^2, \underline{x}, D_{\text{obs}} \sim \text{Normal}$ dada em (2.5.11) substituindo σ_μ^2 por $\tau_0^2\sigma_x^2$, ou seja,

$$\mu|\alpha, \beta, \sigma_x^2, \sigma^2, \underline{x}, D_{\text{obs}} \sim N\left(\frac{n\bar{x}\tau_0^2 + m}{n\tau_0^2 + 1}; \frac{\tau_0^2\sigma_x^2}{n\tau_0^2 + 1}\right);$$

- $\sigma_x^2|\alpha, \beta, \mu, \sigma^2, \underline{x}, D_{\text{obs}} \sim \text{Gama Invertida}$ dada em (2.5.12) substituindo c por ν_{01} e d por λ_{01} ;

- $\sigma^2|\alpha, \beta, \mu, \sigma_x^2, \underline{x}, D_{\text{obs}} \sim \text{Gama Invertida}$ dada em (2.5.13) substituindo f por ν_{02} e g por λ_{02} ;

- $x_i|\alpha, \beta, \mu, \sigma_x^2, \sigma^2, x_{(-i)}, D_{\text{obs}} \sim \text{Normal}$ dada em (2.5.14).

Implementação de algoritmos

Utilizando as distribuições condicionais completas à posteriori obtidas acima, podemos implementar os algoritmos de Gibbs, “grouped Gibbs”, “collapsed Gibbs” e Metropolis em Gibbs (para amostrar da posteriori), de forma totalmente semelhante à seção 2.5.

2.7.2 O modelo sem restrições de identificabilidade

No modelo estrutural

$$\begin{aligned} Y_i &= \alpha + \beta x_i + e_i, \\ X_i &= x_i + u_i, \end{aligned}$$

onde

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ \mu \end{pmatrix}; \begin{pmatrix} \sigma_e^2 & 0 & 0 \\ 0 & \sigma_u^2 & 0 \\ 0 & 0 & \sigma_x^2 \end{pmatrix} \right), \quad i = 1, \dots, n,$$

onde $\theta = (\alpha, \beta, \mu, \sigma_x^2, \sigma_u^2, \sigma_e^2)^T$, acrescentamos a suposição à priori sobre θ que

$$\pi(\theta) = \pi(\alpha|\sigma_e^2)\pi(\beta|\sigma_e^2)\pi(\mu|\sigma_x^2)\pi(\sigma_x^2)\pi(\sigma_u^2)\pi(\sigma_e^2),$$

onde $\alpha|\sigma_e^2 \sim N(a, \chi_{01}^2\sigma_e^2)$, $\beta|\sigma_e^2 \sim N(b, \chi_{02}^2\sigma_e^2)$, $\mu|\sigma_x^2 \sim N(m, \tau_0^2\sigma_x^2)$, $\sigma_x^2 \sim IG(\nu_{01}, \lambda_{01})$, $\sigma_u^2 \sim IG(\nu_{02}, \lambda_{02})$, $\sigma_e^2 \sim IG(\nu_{03}, \lambda_{03})$.

Distribuição à posteriori

A posteriori baseada nos dados observados, então é dada por

$$p(\theta|D_{\text{obs}}) \propto L(\theta|D_{\text{obs}})\pi(\theta), \quad (2.7.4)$$

onde $L(\theta|D_{\text{obs}})$ é dado por (2.1.4) e $\pi(\theta)$ é dada acima.

A posteriori “baseada nos dados completos” (D) é dada por

$$p(\theta|D) = \int_{\mathfrak{F}} p(\mathfrak{x}, \theta|D) d\mathfrak{x}, \quad (2.7.5)$$

onde $L(\theta|D)$ é dada em (2.1.5) e $\pi(\theta)$ é dada acima.

Distribuições condicionais completas à posteriori baseadas nos dados observados

Usando a posteriori em (2.7.4), obtemos as seguintes distribuições condicionais:

- $\alpha|\beta, \mu, \sigma_x^2, \sigma_u^2, \sigma_e^2, D_{\text{obs}} \sim \text{Normal}$ dada por (2.6.5), substituindo σ_α^2 por $\chi_{01}^2 \sigma_e^2$;
- $\mu|\alpha, \beta, \sigma_x^2, \sigma_u^2, \sigma_e^2, D_{\text{obs}} \sim \text{Normal}$ dada por (2.6.6), substituindo σ_μ^2 por $\tau_0^2 \sigma_x^2$;
- $\pi(\alpha, \mu|\beta, \sigma_x^2, \sigma^2, D_{\text{obs}}) = \pi(\alpha|\beta, \mu, \sigma_x^2, \sigma^2, D_{\text{obs}})\pi(\beta|\mu, \sigma_x^2, \sigma^2, D_{\text{obs}})$, onde
 $\mu|\beta, \sigma_x^2, \sigma_u^2, \sigma_e^2, D_{\text{obs}} \sim \text{Normal}$ dada por (2.6.7), substituindo σ_μ^2 por $\tau_0^2 \sigma_x^2$ e σ_α por $\chi_{01}^2 \sigma_x^2$;
- $\beta|\alpha, \mu, \sigma_x^2, \sigma_u^2, \sigma_e^2, D_{\text{obs}}$ tem distribuição dada por (2.6.8), substituindo σ_β^2 por $\chi_{02}^2 \sigma_e^2$;
- $\sigma_x^2|\alpha, \beta, \mu, \sigma_u^2, \sigma_e^2, D_{\text{obs}}$ tem distribuição dada por (2.6.9), substituindo c por ν_{01} e d por λ_{01} ;
- $\sigma_u^2|\alpha, \beta, \mu, \sigma_x^2, \sigma_e^2, D_{\text{obs}}$ tem distribuição dada por (2.6.10), substituindo f por ν_{02} e g por λ_{02} ;
- $\sigma_e^2|\alpha, \beta, \mu, \sigma_x^2, \sigma_u^2, D_{\text{obs}}$ tem distribuição dada por (2.6.11), substituindo p por ν_{03} e q por λ_{03} .

Distribuições condicionais completas à posteriori “baseada nos dados completos”, (D)

Usando a posteriori em (2.7.5) obtemos as seguintes distribuições condicionais completas:

- $\alpha|\beta, \mu, \sigma_x^2, \sigma_u^2, \sigma_e^2, \underline{x}, D_{\text{obs}} \sim \text{Normal}$ dada em (2.6.12), substituindo σ_α^2 por $\chi_{01}^2 \sigma_e^2$, ou seja,

$$N\left(\frac{n\chi_{01}^2(\bar{Y} - \beta\bar{x}) + a}{n\chi_{01}^2 + 1}; \frac{\chi_{01}^2 \sigma_e^2}{n\chi_{01}^2 + 1}\right);$$

- $\beta|\alpha, \mu, \sigma_x^2, \sigma_u^2, \sigma_e^2, \underline{x}, D_{\text{obs}} \sim \text{Normal}$ dada em (2.6.13), substituindo σ_β^2 por $\chi_{02}^2 \sigma_e^2$, isto é,

$$N\left(\frac{\chi_{02}^2 (\sum_{i=1}^n x_i Y_i - n\alpha\bar{x}) + b}{\chi_{02}^2 \sum_{i=1}^n x_i^2 + 1}; \frac{\chi_{02}^2 \sigma_e^2}{\chi_{02}^2 \sum_{i=1}^n x_i^2 + 1}\right);$$

- $\mu|\alpha, \beta, \sigma_x^2, \sigma_e^2, \underline{x}, D_{\text{obs}} \sim \text{Normal}$ dada em (2.6.14), substituindo σ_μ^2 por $\tau_0^2 \sigma_x^2$, isto é,

$$N\left(\frac{n\bar{x}\tau_0^2 + m}{n\tau_0^2 + 1}; \frac{\tau_0^2 \sigma_x^2}{n\tau_0^2 + 1}\right);$$

- $\sigma_x^2|\alpha, \beta, \mu, \sigma_u^2, \sigma_e^2, \underline{x}, D_{\text{obs}} \sim \text{Gama Invertida}$ dada em (2.6.15), substituindo c por ν_{01} e d por λ_{01} , isto é,

$$IG\left(\frac{n}{2} + \nu_{01}; \lambda_{01} + \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2\right);$$

- $\sigma_u^2|\alpha, \beta, \mu, \sigma_x^2, \sigma_e^2, \underline{x}, D_{\text{obs}} \sim \text{Gama Invertida}$ dada em (2.6.16), substituindo f por ν_{02} e g por λ_{02} , isto é,

$$IG\left(\frac{n}{2} + \nu_{02}; \frac{1}{2} \sum_{i=1}^n (X_i - x_i)^2 + \lambda_{02}\right);$$

- $\sigma_e^2|\alpha, \beta, \mu, \sigma_x^2, \sigma_u^2, \underline{x}, D_{\text{obs}} \sim \text{Gama Invertida}$ dada em (2.6.17), substituindo p por ν_{03} e q por λ_{03} , isto é,

$$IG\left(\frac{n}{2} + \nu_{03}; \frac{1}{2} \sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2 + \lambda_{03}\right);$$

- $x_i|\alpha, \beta, \mu, \sigma_x^2, \sigma_u^2, \sigma_e^2, x_{(-i)}, D_{\text{obs}} \sim \text{Normal}$ dada por (2.6.18), para $i = 1, \dots, n$.

Implementação de algoritmos

Para amostrar da posteriori, podemos utilizar as distribuições condicionais completas acima para implementar os algoritmos de Gibbs, “grouped Gibbs”, “collapsed Gibbs” e Metropolis-Hastings de forma semelhante à Seção 2.5.

2.8 Estabelecimento das condições para a existência da distribuição à posteriori sob diferentes escolhas de priors impróprias

Nesta seção investigamos se a distribuição à posteriori sob o modelo estrutural normal é própria quando escolhemos um conjunto específico de priors impróprias, tanto no caso com restrição como no caso sem restrição de identificabilidade. O objetivo é estabelecer restrições sobre o tamanho da amostra e sobre os hiperparâmetros dessas priors para que as correspondentes distribuições à posteriori sejam próprias.

Obviamente, é importante verificar se a posteriori é própria para que se possa fazer inferência. Sabemos que o uso de priori imprópria tem sido alvo de críticas por violar o “Princípio da Coerência” na estatística bayesiana, mas sabemos também que nem sempre uma priori própria é garantia de bons resultados. Por exemplo, é muito comum (quando não temos “muita informação” à priori) escolhermos uma distribuição à priori própria com variância grande (ou seja, uma “priori própria vaga”), como por exemplo a distribuição $\pi(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)} \exp\{\frac{-\beta}{\sigma^2}\}$, com $|\beta|$ bem pequeno, em vez da distribuição imprópria $\pi(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)}$ (que é o limite da distribuição acima quando $\beta \rightarrow 0$). Sobre esse tipo de problema, Berger (2000) faz a seguinte afirmação: “... A noção comum de que é mais seguro ou mais garantido (em termos de desempenho) usar uma priori “vaga” do que uma priori imprópria (pelo fato de que teoricamente a posteriori é própria no primeiro caso) é simplesmente

errada". O autor comenta esse problema apresentando um exemplo em que o uso de uma priori vaga pode levar a resultados absurdos.

2.8.1 Condições necessárias e suficientes no caso do modelo estrutural identificável

O modelo

$$\begin{aligned} Y_i &= \alpha + \beta x_i + e_i, \\ X_i &= x_i + u_i, \end{aligned} \quad (2.8.1)$$

onde $\{(X_i, Y_i)^T, i = 1, 2, \dots, n\}$ é o conjunto dos dados observados (D_{obs}),

$\{x_i, i = 1, \dots, n\}$: dados latentes,

$\{(e_i, u_i)^T, i = 1, \dots, n\}$: erros aleatórios.

Com a suposição

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \stackrel{\text{i.i.d.}}{\sim} \left(\begin{pmatrix} 0 \\ 0 \\ \mu \end{pmatrix}; \begin{pmatrix} \sigma_e^2 & 0 & 0 \\ 0 & \sigma_u^2 & 0 \\ 0 & 0 & \sigma_x^2 \end{pmatrix} \right), \quad i = 1, \dots, n,$$

e com a restrição de identificabilidade $\sigma_e^2 = \lambda \sigma_u^2$ e $\sigma_u^2 = \sigma^2$ com λ conhecido.

$\Theta = \{\theta = (\alpha, \beta, \mu, \sigma_x^2, \sigma^2)^T : \alpha, \beta, \mu \in \mathbb{R} \text{ e } \sigma_x^2, \sigma^2 \in \mathbb{R}^+\}$, onde Θ é o espaço paramétrico.

Distribuição à priori para θ

Nessa subseção apresentamos quatro modelos de distribuição à priori imprópria não informativas e parcialmente informativas. As posteriores obtidas dessas prioris é que serão avaliadas quanto à existência.

Nesses modelos as distribuições à priori $\pi(\theta)$ são da forma

$$\pi(\theta) = \pi(\alpha)\pi(\beta)\pi(\mu)\pi(\sigma_x^2)\pi(\sigma^2),$$

onde as "marginais" de $\alpha, \beta, \mu, \sigma_x^2$ e σ^2 são dadas a seguir.

Modelo 1: $\pi(\alpha) \propto 1$; $\pi(\beta) \propto 1$; $\pi(\mu) \propto 1$; $\pi(\sigma_x^2) \propto (\sigma_x^2)^{-(c+1)} \exp\left\{-\frac{d}{\sigma_x^2}\right\}$, onde $c, d > 0$ e $\pi(\sigma^2) \propto (\sigma^2)^{-(f+1)} \exp\left\{-\frac{g}{\sigma^2}\right\}$, onde $f, g > 0$ (isto é, $\sigma_x^2 \sim IG(c, d)$ e $\sigma^2 \sim IG(f, g)$).

Modelo 2: $\pi(\alpha) \propto 1$; $\pi(\beta) \propto 1$; $\pi(\mu) \propto 1$; $\pi(\sigma_x^2) \propto (\sigma_x^2)^{-r}$, $r > 0$ e $\pi(\sigma^2) \sim IG(f, g)$ (esta priori coincide com a do modelo 1 quando $c = r - 1$ e $d = 0$).

Modelo 3: $\pi(\alpha) \propto 1$; $\pi(\beta) \propto 1$; $\pi(\mu) \propto 1$; $\pi(\sigma_x^2) \sim IG(c, d)$; $\pi(\sigma^2) \propto (\sigma^2)^{-s}$ (coincide com a priori do modelo 1 quando $f = s - 1$ e $g = 0$).

Modelo 4: $\pi(\alpha) \propto 1$; $\pi(\beta) \propto 1$; $\pi(\mu) \propto 1$; $\pi(\sigma_x^2) \propto (\sigma_x^2)^{-r}$, $r > 0$ e $\pi(\sigma^2) \propto (\sigma^2)^{-s}$, $s > 0$ (é idêntica à priori do modelo 1 quando $c = r - 1$, $f = s - 1$, $d = 0$ e $g = 0$).

A escolha dos modelos acima baseou-se em alguns argumentos que daremos a seguir e principalmente na analogia com o modelo linear sem erros nas variáveis.

Um resultado bastante conhecido demonstrado por Berger (1980) é que se uma variável aleatória z tem densidade da forma $f(z|\theta) = f(z - \theta)$ (ou seja, se θ é um parâmetro de locação), então a priori de θ deve ser uniforme no domínio de variação de θ para que ela seja invariante sob transformação de locação (isto é, sob transformação da forma $Y = z + c$, onde c é uma constante); portanto, se $-\infty < \theta < \infty$, a priori “não informativa” para θ é dado por $\pi(\theta) \propto 1$, $\theta \in \mathbb{R}$. Segundo Carlin e Louis (2000), embora a distribuição uniforme na reta seja imprópria, inferência bayesiana ainda pode ser feita se a distribuição à posteriori correspondente for própria. Berger (1985) demonstrou também que se σ^2 é um parâmetro de escala então a priori de σ^2 deve ser da forma $\pi(\sigma^2) \propto \frac{1}{\sigma^2}$ para ser invariante sob transformação de escala (e portanto a priori “não informativa” para $\sigma^2 > 0$ é dada por $\pi(\sigma^2) \propto \frac{1}{\sigma^2}$ (que nesse caso é também imprópria)).

Segundo Carlin e Louis (2000) e Berger (1980) quando o modelo tem simultaneamente um parâmetro de locação θ e outro de escala σ^2 , o procedimento mais simples comumente usado para obter priori “não informativa” é considerar independência à priori entre esses parâmetros (embora a rigor o conceito de independência está reservado apenas para as distribuições próprias) e tomar a priori $\pi(\theta, \sigma^2)$ como o produto das distribuições de θ e de σ^2 separadamente. Assim, a priori não informativa nesse caso é dada por $\pi(\theta, \sigma) \propto \frac{1}{\sigma}$ (que é imprópria quando $\theta \in \mathbb{R}$ ou $\sigma^2 \in \mathbb{R}^+$).

No modelo linear Normal sem erros nas covariáveis, da forma $Y_i = \beta_1 + \beta_2 x_{2i} + \dots + \beta_k x_{ki}$ com $e_i \sim N(0, \sigma^2)$, $i = 1, 2, \dots, n$ independentes, temos que σ^2 é um parâmetro de escala e $\underline{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$ é um “parâmetro de locação num sentido mais amplo” (já que a verossimilhança pode ser escrita como

$$L(\underline{\beta}, \sigma^2 | \underline{Y}) \propto \left(\frac{s^2}{\sigma^2}\right)^{\frac{n}{2}} \exp \left\{ -\frac{s^2}{2\sigma^2} \left[(n-k) + \frac{(\underline{\beta} - \hat{\underline{\beta}})^T \underline{X}^T \underline{X} (\underline{\beta} - \hat{\underline{\beta}})}{s^2} \right] \right\},$$

onde $s^2 = \frac{(\underline{Y} - \underline{Y})^T (\underline{Y} - \underline{Y})}{n-k}$ e $\hat{\underline{\beta}} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{Y}$ (Box e Tiao, 1973) e assim podemos ver que $\underline{\beta}$ é um parâmetro de locação da distribuição de $\hat{\underline{\beta}}$. Nesse caso, a distribuição à priori não informativa seria dada por

$$\pi(\underline{\beta}, \sigma) = p(\underline{\beta})p(\sigma^2) \propto \frac{1}{\sigma^2}.$$

Nesse modelo, a priori (não informativa) de Jeffreys, que é a raiz quadrada do determinante da matriz de informação de Fisher, é dada por

$$\pi(\underline{\beta}, \sigma^2) \propto \frac{1}{(\sigma^2)^{\frac{k}{2}-1}}, \quad \sigma^2 > 0.$$

No caso do modelo Normal com erros nas variáveis a priori de Jeffreys não pode ser obtida pois a matriz de informação de Fisher não pode ser obtida analiticamente (lembrando que para a análise das condições de existência da posteriori é necessário ter a expressão analítica da priori)

Segundo Carlin e Louis (2000), a obtenção da priori de Jeffreys no caso multi-dimensional pode ser impraticável especialmente quando a dimensão do problema é alta e um procedimento bastante usado na prática é tomar o produto das distribuições de cada componente. Segundo os autores, esse procedimento é frequentemente justificado com base no argumento de que a idéia de ignorância é consistente com a idéia de independência (embora numa distribuição imprópria a noção formal de independência não se aplica por não se tratar de uma distribuição de probabilidade).

As prioris para o modelo com erros nas variáveis, dados pelos modelos de 1 a 4 do início da seção, foram construídos levando-se em conta os argumentos acima. Para todos os quatro modelos, a priori conjunta $\pi(\theta)$ é tomada como o produto das distribuições de cada componente; além disso, fizemos analogia com os modelos lineares sem erros nas variáveis considerando α e β como “parâmetros de locação” num sentido mais amplo. Como μ é claramente um parâmetro de locação, tomamos para α , β e μ a priori não informativa uniforme na reta. Para as componentes de escala σ_x^2 e σ^2 atribuímos ou a distribuição Gama Invertida com os respectivos parâmetros, ou então as distribuições impróprias não informativas $\pi(\sigma_x^2) \propto \frac{1}{(\sigma_x^2)^r}$ para σ_x^2 e $\pi(\sigma^2) \propto \frac{1}{(\sigma^2)^s}$ para σ^2 , onde os hiperparâmetros r e s são escolhidos de tal modo que a posteriori seja própria, com isso quisemos dar uma certa flexibilidade na escolha. Em cada um dos teoremas nas próximas subseções obtemos os intervalos dos possíveis valores desses hiperparâmetros.

Achamos que vale a pena notar ainda que as distribuições impróprias $\frac{1}{(\sigma_x^2)^r}$ e $\frac{1}{(\sigma^2)^s}$ podem ser pensadas como sendo o limite, respectivamente, das distribuições $IG(c, d)$ e $IG(f, g)$ quando d e $g \rightarrow 0$ e $r = c + 1$ e $s = f + 1$.

Vale a pena também notar que o “grau de informação” das prioris dos modelos propostos no início, vai diminuindo quando passamos do modelo 1 para os modelos

2, 3 e 4. No modelo 1, as componentes σ_x^2 e σ^2 tem distribuição própria (e portanto informativas para valores convenientes dos hiperparâmetros), já no modelo 4 essas duas componentes tem distribuição não informativa.

Observação: Usamos o termo “priori não informativa” várias vezes no texto acima. A seguir apresentamos esse conceito como é conhecido na literatura.

Segundo Box e Tiao (1973), uma priori não informativa (sobre θ) não é aquela que representa ignorância total (sobre θ) mas a que representa uma situação em que o conhecimento à priori (sobre θ) é pequeno relativamente ao conhecimento (ou informação) que se espera obter dos dados; em outras palavras, $\pi(\theta)$ é uma priori não informativa se ela tem um impacto mínimo na distribuição à posteriori de θ (ou ainda, que a verossimilhança domina sobre $\pi(\theta)$). Na literatura existem vários métodos com o objetivo de construir priores com essas características. Uma possível maneira de expressar essa idéia segundo Box e Tiao (1973), é dizer que $\pi(\theta)$ não favorece nenhum valor de θ em especial. Assim, as seguintes distribuições, $\pi(\theta) = \frac{1}{n}$, $\theta \in \{\theta_1, \dots, \theta_n\}$, $\pi(\theta) = \frac{1}{b-a}$, $a < \theta < b$, e a distribuição imprópria $\pi(\theta) \propto 1$, $\theta \in \mathbb{R}$, seriam exemplos de priores não informativas (Box e Tiao expressam este estado de indiferença com prioris localmente uniformes). Essa maneira de definir uma priori não informativa é bastante questionada pelo fato de que priores uniformes não são invariantes sob reparametrizações. Uma saída para remediar o problema é tratar cada problema como um problema diferente e assim procurar a reparametrização que seja mais adequada para o particular problema que se está estudando, e somente então aí aplicar uma priori uniforme nesta escala (para mais detalhes sobre este assunto ver Box e Tiao, 1973, páginas 23 e 24).

Pode-se também argumentar por exemplo que uma “boa priori não informativa” não deveria ser afetada quando se usa diferentes parametrizações do mesmo mod-

elo. Na literatura temos a conhecida priori de Jeffreys que é invariante sob transformações (ou reparametrizações) bijetoras. Outro exemplo é o caso das famílias de distribuições com parâmetros de locação (ou de escala); nesse caso uma priori “não informativa razoável” deveria ser invariante sob transformações de locação (ou escala) nos dados. Berger (1980) demonstra que a distribuição à priori de um parâmetro de locação com essa característica deve ser do tipo $\pi(\theta) \propto 1$ e a priori de um parâmetro de escala que seja invariante sob transformação de escala nos dados deve ser do tipo $\pi(\sigma^2) \propto \frac{1}{\sigma^2}$.

Outro tipo de priori, que está relacionada bem de perto com o conceito de priori não informativa é a chamada “priori de referência” (ver Kass e Wasserman, 1996, e Bernardo, 1979).

Distribuição à posteriori

Seja $p(\theta|D_{\text{obs}})$ a distribuição a posteriori de θ e, portanto,

$$p(\theta|D_{\text{obs}}) \propto L(\theta|D_{\text{obs}})\pi(\theta);$$

onde $L(\theta|D_{\text{obs}})$ é a função de verossimilhança e $\pi(\theta)$ é a distribuição à priori de θ .

Nosso objetivo é encontrar as condições necessárias e suficientes para que a posteriori seja própria, isto é, $\int_{\Theta} L(\theta|D_{\text{obs}})\pi(\theta)d\theta < \infty$, onde

$$L(\theta|D_{\text{obs}}) \propto \frac{1}{[\beta^2\sigma_x^2\sigma^2 + \lambda\sigma_x^2\sigma^2 + \lambda(\sigma^2)^2]^{n/2}} \exp \left\{ -\frac{1}{2} \left[(\beta^2\sigma_x^2 + \lambda\sigma^2) \sum_{i=1}^n (X_i - \mu)^2 - 2\beta\sigma_x^2 \sum_{i=1}^n (Y_i - \alpha - \beta\mu)(X_i - \mu) + (\sigma_x^2 + \sigma^2) \sum_{i=1}^n (Y_i - \alpha - \beta\mu)^2 \right] \right\}.$$

O problema é que, independentemente da escolha de $\pi(\theta)$, o cálculo da integral $\int_{\Theta} L(\theta|D_{\text{obs}})\pi(\theta)d\theta$ é muito complicado. Para contornar o problema “reintroduzimos” as variáveis latentes $\underline{x} = (x_1, \dots, x_n)^T$ escrevendo a posteriori como

$$p(\theta|D_{\text{obs}}) = \int_{\underline{x} = \{x: x \in \mathbb{R}^n\}} p(\theta, \underline{x}|D_{\text{obs}})d\underline{x} \propto \int_{\underline{x}} L(\theta|D)\pi(\theta)d\underline{x},$$

onde $L(\theta|D)$ é a verossimilhança para os dados completos, dada por

$$L(\theta|D) \propto [(\sigma^2)^2 \sigma_x^2]^{-n/2} \exp \left\{ -\frac{1}{2} \left[\frac{\sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2}{\lambda \sigma^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma^2} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_x^2} \right] \right\}.$$

e portanto

$$\int_{\Theta} p(\theta|D_{\text{obs}}) d\theta \propto \int_{\Theta} \int_{\mathbf{x}} L(\theta|D) \pi(\theta) d\mathbf{x} d\theta.$$

Esta última integral é muito mais tratável do que a anterior e, portanto, ela será utilizada para demonstrar a existência da posteriori.

Mais à frente enunciaremos e demonstramos os teoremas que estabelecem as condições necessárias e suficientes para a existência das posteriores que correspondem às priores especificadas anteriormente. Entretanto, antes disso iremos demonstrar um lema cujo resultado será muito útil nas demonstrações.

Lema 2.8.1 *Se $\mathbf{x}_{p \times 1}$ é um vetor de p componentes, então*

$$(a) \int_{\{\mathbf{x} \in \mathbb{R}^p : \|\mathbf{x}\| < \varepsilon\}} \frac{1}{\|\mathbf{x}\|^\lambda} d\mathbf{x} < \infty \text{ se e somente se } \lambda < p;$$

$$(b) \int_{\{\mathbf{x} \in \mathbb{R}^p : \|\mathbf{x}\| > \varepsilon\}} \frac{1}{\|\mathbf{x}\|^\lambda} d\mathbf{x} < \infty \Leftrightarrow \lambda > p$$

, onde $\|\mathbf{x}\|$ é a norma Euclideana de \mathbf{x} .

Demonstração do caso (a): Se $p = 1$, a demonstração é trivial. Se $p \geq 2$,

podemos usar a seguinte mudança de coordenadas: $\mathbf{x} = r \mathbf{d}$, onde

$$\mathbf{d} = \begin{pmatrix} \text{sen}(\theta_1) \text{sen}(\theta_2) \text{sen}(\theta_3) \cdots \text{sen}(\theta_{p-3}) \text{sen}(\theta_{p-2}) \text{sen}(\theta_{p-1}) \\ \text{sen}(\theta_1) \text{sen}(\theta_2) \text{sen}(\theta_3) \cdots \text{sen}(\theta_{p-3}) \text{sen}(\theta_{p-2}) \cos(\theta_{p-1}) \\ \text{sen}(\theta_1) \text{sen}(\theta_2) \text{sen}(\theta_3) \cdots \text{sen}(\theta_{p-3}) \cos(\theta_{p-2}) \\ \text{sen}(\theta_1) \text{sen}(\theta_2) \text{sen}(\theta_3) \cdots \cos(\theta_{p-3}) \\ \vdots \\ \text{sen}(\theta_1) \text{sen}(\theta_2) \cos(\theta_3) \\ \text{sen}(\theta_1) \cos(\theta_2) \\ \cos(\theta_1) \end{pmatrix} \quad \text{e} \quad r \geq 0.$$

Esta transformação é uma generalização da chamada “transformação em “coordenadas esféricas” (quando $p = 2$ temos a transformação em coordenadas polares e quando $p = 3$ a transformação em coordenadas esféricas). É fácil ver que $\|\underline{d}\| = 1$ e o valor absoluto do jacobiano da transformação é

$|J| = r^{p-1} |\text{sen}^{p-2}\theta_1 \text{sen}^{p-3}\theta_2 \cdots \text{sen}^1\theta_{p-2}|$. Portanto,

$$\begin{aligned} & \int_{\|\underline{x}\| < \varepsilon} \frac{1}{\|\underline{x}\|^\lambda} d\underline{x} \\ &= \int_0^\varepsilon \int_{\Theta = \{\underline{\theta} = (\theta_1, \dots, \theta_{p-1}) : 0 < \theta_i < 2\pi\}} \frac{1}{r^\lambda} r^{p-1} |\text{sen}^{p-2}\theta_1 \text{sen}^{p-3}\theta_2 \cdots \text{sen}^1\theta_{p-2}| d\underline{\theta} dr \\ &= \int_0^\varepsilon \int_{\Theta = \{\underline{\theta} = (\theta_1, \dots, \theta_{p-1}) : 0 < \theta_i < 2\pi\}} \frac{1}{r^{\lambda-p+1}} |\text{sen}^{p-2}\theta_1 \cdots \text{sen}^1\theta_{p-2}| d\underline{\theta} dr \\ &= \int_{\Theta = \{\underline{\theta} = (\theta_1, \dots, \theta_{p-1}) : 0 < \theta_i < 2\pi\}} |\text{sen}^{p-2}\theta_1 \cdots \text{sen}^1\theta_{p-2}| d\underline{\theta} \cdot \int_0^\varepsilon \frac{1}{r^{\lambda-p+1}} dr, \end{aligned}$$

que é finita se e somente se $\lambda - p + 1 < 1$, ou seja, $\lambda < p$. (Note que a integral em Θ é finita já que a função no integrando é limitada e o domínio de integração Θ é um conjunto limitado.)

A demonstração do Lema 2.8.1(b) é semelhante.

Teorema 2.8.2 *A condição necessária e suficiente para a existência da distribuição à posteriori do modelo Normal em (2.8.1) com priori dada por $\pi(\underline{\theta}) \propto \pi(\sigma_x^2)\pi(\sigma^2)$, onde $\sigma_x^2 \sim IG(c, d)$ e $\sigma^2 \sim IG(f, g)$, é que $n \geq 3$ (onde n é o tamanho da amostra).*

Demonstração: Primeiramente devemos calcular a integral

$$\int_{\Theta} p(\underline{\theta} | D_{\text{obs}}) d\underline{\theta} = \int_{\underline{x}} \int_{\Theta} p(\underline{\theta}, \underline{x} | D_{\text{obs}}) d\underline{x} d\underline{\theta},$$

onde $p(\underline{\theta}, \underline{x} | D_{\text{obs}}) \propto L(\underline{\theta} | D)\pi(\underline{\theta})$ e depois verificar sob que condições ela é finita.

A distribuição conjunta à posteriori de $\underline{\theta}$ e \underline{x} , $p(\underline{\theta}, \underline{x} | D_{\text{obs}})$, sob o modelo do Teorema 2.8.2 é

$$\begin{aligned} p(\underline{\theta}, \underline{x} | D_{\text{obs}}) \propto & (\sigma_x^2)^{-\left(\frac{n}{2} + c + 1\right)} (\sigma^2)^{-(n+f+1)} \exp \left\{ -\frac{1}{2} \left[\frac{2d}{\sigma_x^2} + \frac{2g}{\sigma^2} \right. \right. \\ & \left. \left. + \frac{\sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2}{\lambda \sigma^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma^2} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_x^2} \right] \right\}. \end{aligned}$$

Integrando inicialmente esta função em relação a α , temos que (ver seção 2.8 do Apêndice A)

$$\int_{-\infty}^{\infty} p(\theta, \underline{x} | D_{\text{obs}}) d\alpha \propto (\sigma_x^2)^{-(\frac{n}{2}+c+\frac{1}{2})} (\sigma^2)^{-(n+f+1)} \exp \left\{ -\frac{1}{2} \left[\frac{2d}{\sigma_x^2} + \frac{2g}{\sigma^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma^2} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_x^2} + \frac{\sum_{i=1}^n (Y_i - \beta x_i)^2 - n(\bar{Y} - \beta \bar{x})^2}{\lambda \sigma^2} \right] \right\}. \quad (2.8.2)$$

A seguir, integramos em relação μ (ver Apêndice A, seção 2.8), obtendo

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\theta, \underline{x} | D_{\text{obs}}) d\alpha d\mu \propto (\sigma_x^2)^{-(\frac{n}{2}+c+\frac{1}{2})} (\sigma^2)^{-(n+f+\frac{1}{2})} \exp \left\{ -\frac{1}{2} \left[\frac{2d}{\sigma_x^2} + \frac{2g}{\sigma^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma^2} + \frac{\sum_{i=1}^n (Y_i - \beta x_i)^2 - n(\bar{Y} - \beta \bar{x})^2}{\lambda \sigma^2} + \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_x^2} \right] \right\}. \quad (2.8.3)$$

A seguir, integramos em relação a β (ver seção 2.8 do Apêndice A),

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\theta, \underline{x} | D_{\text{obs}}) d\alpha d\mu d\beta \propto (\sigma_x^2)^{-(\frac{n}{2}+c+\frac{1}{2})} (\sigma^2)^{-(n+f)} (S_{xx})^{-\frac{1}{2}} \cdot \exp \left\{ -\frac{1}{2} \left[\frac{2d}{\sigma_x^2} + \frac{2g}{\sigma^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma^2} + \frac{S_{xx}}{\sigma_x^2} + \frac{S_{YY} S_{xx} - S_{xY}^2}{\lambda \sigma^2 S_{xx}} \right] \right\}, \quad (2.8.4)$$

onde $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$, $S_{YY} = \sum_{i=1}^n (Y_i - \bar{Y})^2$, $S_{xY} = \sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})$.

Agora, integrando com relação a σ_x^2 (ver seção 2.8 do Apêndice A),

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_0^{\infty} p(\theta, \underline{x} | D_{\text{obs}}) d\alpha d\mu d\beta d\sigma_x^2 \propto \frac{(\sigma^2)^{-(n+f)}}{S_{xx}^{\frac{1}{2}} (2d + S_{xx})^{\frac{n+2c-1}{2}}} \cdot \exp \left\{ -\frac{1}{2} \left[\frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma^2} + \frac{(S_{YY} S_{xx} - S_{xY}^2)}{\lambda \sigma^2 S_{xx}} + \frac{2g}{\sigma^2} \right] \right\}. \quad (2.8.5)$$

Finalmente, integrando com respeito a σ^2 (ver seção 2.8 do Apêndice A), obtemos

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{\infty} p(\theta, \underline{x} | D_{\text{obs}}) d\alpha d\mu d\beta d\sigma_x^2 d\sigma^2 \quad (2.8.6)$$

$$\propto \frac{1}{S_{xx}^{\frac{1}{2}} (2d + S_{xx})^{\frac{n+2c-1}{2}} \left[\sum_{i=1}^n (X_i - x_i)^2 + \frac{(S_{YY} S_{xx} - S_{xY}^2)}{\lambda S_{xx}} + 2g \right]^{n+f-1}}.$$

Portanto, a integral

$$\int_{\Theta} p(\varrho|D_{\text{obs}})d\varrho = \int_{\mathfrak{X}} \int_{\Theta} p(\varrho, \mathfrak{x}|D_{\text{obs}})d\varrho d\mathfrak{x}$$

é proporcional a

$$\int_{\mathfrak{X}} \frac{1}{(S_{xx})^{\frac{1}{2}}(2d + S_{xx})^{\frac{n+2c-1}{2}} \left(\sum_{i=1}^n (X_i - \bar{x})^2 + \frac{S_{YY}S_{xx} - S_{xY}^2}{\lambda S_{xx}} + 2g \right)^{n+f-1}} d\mathfrak{x}, \quad (2.8.7)$$

onde $\mathfrak{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$, $S_{YY} = \sum_{i=1}^n (Y_i - \bar{Y})^2$ e $S_{xY} = \sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})$.

Notemos que para $n = 1$, $\mathfrak{x} = x_1$, $S_{x_1x_1} = 0$ e $S_{x_1Y_1} = 0$, $\forall x_1 \in \mathbb{R}$ e, portanto, o integrando não está definido, qualquer que seja $x_1 \in \mathbb{R}$.

O cálculo dessa integral fica mais simples se, por exemplo, no lugar da soma de quadrados $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$ tivermos uma soma de quadrados do tipo $\sum_i z_i^2$. Isso pode ser conseguido por meio de uma mudança de variáveis do tipo $z = H\mathfrak{x}$, onde

$$z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \text{ e } H_{n \times n} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & \dots & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & 0 & 0 & \dots & 0 \\ \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & -\frac{3}{\sqrt{12}} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \dots & -\frac{(n-1)}{\sqrt{n(n-1)}} \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \dots & \frac{1}{\sqrt{n}} \end{bmatrix}, \quad (2.8.8)$$

onde a matriz H é conhecida como a matriz de Helmert (Searle, 1971), que é ortogonal e está definida para $n \geq 2$. A transformação $\mathfrak{x} \mapsto H\mathfrak{x}$ é uma rotação de \mathfrak{x} (no caso bidimensional, por exemplo, $H = \begin{pmatrix} \cos(\theta) & -\text{sen}(\theta) \\ \text{sen}(\theta) & \cos(\theta) \end{pmatrix}$ onde $\theta = \frac{\pi}{4}$).

No novo espaço (rotacionado), os dados observados $\tilde{X}_{n \times 1}$ e $\tilde{Y}_{n \times 1}$ são representados, respectivamente, por $\tilde{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = H\tilde{X}$ e $\tilde{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = H\tilde{Y}$.

Observação: No nosso contexto usaremos a notação \underline{b} para representar tanto o ponto $(b_1, \dots, b_n)^\top$ (no espaço n -dimensional) como ao vetor \vec{b} (onde \vec{b} pode ser representado geometricamente pelo segmento orientado que passa pela origem e tem extremidade no ponto $(b_1, \dots, b_n)^\top$). Da mesma forma, \underline{z} e $\underline{\xi}$ podem representar tanto pontos como vetores mas a notação \vec{z} , \vec{c} é reservada apenas para vetores.

No novo espaço as quantidades S_{xx} , S_{YY} , S_{xY} e $\sum_{i=1}^n (X_i - x_i)^2$, que aparecem na integral (2.8.7), são iguais a $\sum_{i=1}^{n-1} z_i^2$, $\sum_{i=1}^{n-1} c_i^2$, $\sum_{i=1}^{n-1} z_i c_i$ e $\sum_{i=1}^{n-1} (b_i - z_i)^2 + (\sqrt{n}\bar{X} - z_n)$, respectivamente. A justificativa segue mais abaixo.

Justificativa

$S_{xx} = \sum_{i=1}^n x_i^2 = \frac{(\sum_{i=1}^n x_i)^2}{n} = \underline{x}^\top \underline{x} - \frac{1}{n}(\underline{x}^\top \underline{1})^2$, onde $\underline{x} = H^\top \underline{z}$ (já que $\underline{z} = H \underline{x}$ e H é ortogonal). Assim,

$$S_{xx} = \underline{z}^\top H H^\top \underline{z} - \frac{1}{n}(\underline{z}^\top H \underline{1})^2 = \underline{z}^\top \underline{z} - \frac{1}{n}(\underline{z}^\top H \underline{1})^2,$$

onde $H \underline{1} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \sqrt{n} \end{pmatrix}$ e, portanto,

$$S_{xx} = \sum_{i=1}^n z_i^2 - \frac{1}{n}(\sqrt{n}z_n)^2 = \sum_{i=1}^{n-1} z_i^2.$$

Semelhantemente,

$$S_{YY} = \sum_{i=1}^n Y_i^2 - \frac{(\sum_{i=1}^n Y_i)^2}{n} = \underline{Y}^\top \underline{Y} - \frac{1}{n}(\underline{Y}^\top \underline{1})^2,$$

onde $\underline{Y} = H^\top \underline{\xi}$ (pois $\underline{\xi} = H \underline{Y}$). Portanto,

$$S_{YY} = \underline{\xi}^\top \underline{\xi} - \frac{1}{n}(\underline{\xi}^\top H \underline{1})^2 = \sum_{i=1}^n c_i^2 - \frac{1}{n}(\sqrt{n}c_n)^2 = \sum_{i=1}^{n-1} c_i^2.$$

Também, semelhantemente,

$$\begin{aligned} S_{xY} &= \sum_{i=1}^n x_i Y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n Y_i}{n} = \underline{x}^\top \underline{Y} - \frac{1}{n} (\underline{x}^\top \underline{1})(\underline{Y}^\top \underline{1}) \\ &= \underline{z}^\top H H^\top \underline{\epsilon} - \frac{1}{n} (\underline{z}^\top H \underline{1})(\underline{\epsilon}^\top H \underline{1}) = \underline{z}^\top \underline{\epsilon} - \frac{(\sqrt{n} z_n)(\sqrt{n} c_n)}{n} = \sum_{i=1}^{n-1} z_i c_i. \end{aligned}$$

Finalmente,

$$\begin{aligned} \sum_{i=1}^n (X_i - x_i)^2 &= (\underline{X} - \underline{x})^\top (\underline{X} - \underline{x}) = (H^\top \underline{b} - H^\top \underline{z})^\top (H^\top \underline{b} - H^\top \underline{z}) \\ &= (\underline{b} - \underline{z})^\top H H^\top (\underline{b} - \underline{z}) = (\underline{b} - \underline{z})^\top (\underline{b} - \underline{z}) = \sum_{i=1}^n (b_i - z_i)^2 \\ &= \sum_{i=1}^{n-1} (b_i - z_i)^2 + (b_n - z_n)^2, \end{aligned}$$

onde

$$\underline{b}_n = \left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}} \right) \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \sqrt{n} \bar{X}.$$

Portanto,

$$\sum_{i=1}^n (X_i - x_i)^2 = \sum_{i=1}^{n-1} (b_i - z_i)^2 + (\sqrt{n} \bar{X} - z_n)^2.$$

Agora vamos chamar de \underline{z}_x (ou \vec{z}_x) o vetor formado pelas $n - 1$ primeiras componentes de $\underline{z}_{n \times 1} = H \underline{x}$ (isto é, $\underline{z}_x = A \underline{x}$, onde A é a matriz formada pelas $n - 1$ primeiras linhas da matriz H). Analogamente, vamos chamar de \underline{b}_X (ou \vec{b}_X) e de $\underline{\epsilon}_Y$ (ou \vec{c}_Y) os vetores formados pelas primeiras $n - 1$ componentes dos vetores \underline{b} e $\underline{\epsilon}$, respectivamente. Então,

$$S_{xx} = \|\underline{z}_x\|^2, \quad S_{YY} = \|\underline{\epsilon}_Y\|^2 \quad \text{e} \quad \sum_i^n (X_i - x_i)^2 = (\underline{b}_X - \underline{z}_x)^\top (\underline{b}_X - \underline{z}_x) + (\sqrt{n} \bar{X} - z_n)^2$$

Finalmente,

$$S_{xY}^2 = \left(\sum_{i=1}^{n-1} z_i c_i \right)^2 = (\underline{z}_x^\top \underline{\epsilon}_Y)^2 = |\vec{z}_x \cdot \vec{c}_Y|^2,$$

onde $\vec{z}_x \cdot \vec{c}_Y$ é o produto escalar (ou produto interno) dos vetores \vec{z}_x e \vec{c}_Y , ou seja, $\vec{z}_x \cdot \vec{c}_Y = \|\vec{z}_x\| \cdot \|\vec{c}_Y\| \cos \theta$, onde θ é o ângulo entre \vec{z}_x e \vec{c}_Y . Portanto,

$$\frac{S_{xx} S_{YY} - S_{xY}^2}{\lambda S_{xx}} = \frac{\|\underline{z}_x\|^2 \|\underline{\epsilon}_Y\|^2 - (\underline{z}_x^\top \underline{\epsilon}_Y)^2}{\lambda \|\underline{z}_x\|^2} = \frac{\|\underline{z}_x\|^2 \|\underline{\epsilon}_Y\|^2 - \|\underline{z}_x\|^2 \|\underline{\epsilon}_Y\|^2 \cos^2 \theta}{\lambda \|\underline{z}_x\|^2}$$

$$= \frac{\|\xi_Y\|^2(\sin^2\theta)}{\lambda}$$

quando $z_x \neq 0$.

Portanto, a quantidade $\frac{\|z_x\|^2\|\xi_Y\|^2 - |z_x c_y|^2}{\lambda\|z_x\|^2}$ está bem definida para $z_x \neq 0$ e não depende de $\|z_x\|$ mas sim do ângulo θ entre \vec{z}_x e \vec{c}_Y . Além disso, ela assume valores maiores ou iguais a zero, atingindo o zero se e somente se $\theta = 0$ (isto é, quando os vetores \vec{z}_x e \vec{c}_Y forem linearmente dependentes).

Finalmente, o Jacobiano da transformação é 1 e, portanto, a integral em (2.8.7) é igual a

$$\int_{Z=\{z \in \mathbb{R}^n\}} \frac{1}{\|z_x\|(\|z_x\|^2 + 2d)^{\frac{n+2c-1}{2}} \left[\|z_x - \underline{b}_X\|^2 + \frac{\|\xi_Y\|^2 \sin^2\theta}{\lambda} + 2g + (\sqrt{n}\bar{X} - z_n)^2 \right]^{n+f-1}} dz. \quad (2.8.9)$$

Para obter a integral dessa função, fazemos primeiramente a integração em relação à variável z_n (com o objetivo de eliminar z_n), isto é, calculamos a seguinte integral:

$$\int_{-\infty}^{\infty} \frac{1}{\|z_x\|(\|z_x\|^2 + 2d)^{\frac{n+2c-1}{2}} \left[\|z_x - \underline{b}_X\|^2 + \frac{\|\xi_Y\|^2 \sin^2\theta}{\lambda} + (\sqrt{n}\bar{X} - z_n)^2 + 2g \right]^{n+f-1}} dz_n. \quad (2.8.10)$$

Fazendo

$$k_1 = \|z_x\|(\|z_x\|^2 + 2d)^{\frac{n+2c-1}{2}} \quad \text{e} \quad k_2 = \|z_x - \underline{b}_X\|^2 + \frac{\|\xi_Y\|^2 \sin^2\theta}{\lambda} + 2g,$$

temos que $k_2 > 0$ (mesmo quando $g = 0$), pois as duas primeiras parcelas nunca se anulam simultaneamente, como veremos no Teorema 2.8.4.

Assim, a integral em (2.8.10) torna-se igual a

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{1}{k_1 \left[k_2 + (\sqrt{n}\bar{X} - z_n)^2 \right]^{n+f-1}} dz_n &= \frac{1}{k_1 k_2^{n+f-1}} \int_{-\infty}^{\infty} \frac{1}{\left[1 + \frac{(\sqrt{n}\bar{X} - z_n)^2}{k_2} \right]^{n+f-1}} dz_n \\ &= \frac{1}{k_1 k_2^{n+f-1}} \int_{-\infty}^{\infty} \frac{1}{\left[1 + \left(\frac{\sqrt{n}\bar{X} - z_n}{\sqrt{k_2}} \right)^2 \right]^{n+f-1}} dz_n. \end{aligned} \quad (2.8.11)$$

Com a mudança de variável $u = \frac{\sqrt{nX} - z_n}{\sqrt{k_2}}$, a última expressão em (2.8.11) torna-se igual a $\frac{1}{k_1 k_2^{n+f-\frac{3}{2}}} \int_{-\infty}^{\infty} \frac{1}{[1+u^2]^{n+f-1}} du$, onde $\int_{-\infty}^{\infty} \frac{1}{(1+u^2)^{n+f-1}} du$ é finita se e somente se $2(n+f-1) > 1$, isto é,

$$n+f > \frac{3}{2}. \quad (2.8.12)$$

Portanto, a integral em (2.8.10) é proporcional a $\frac{1}{k_1 k_2^{n+f-\frac{3}{2}}}$, ou seja, ela é proporcional à função

$$f(\underline{z}_x) = \frac{1}{\|\underline{z}_x\| [\|\underline{z}_x\|^2 + 2d]^{\frac{n+2c-1}{2}} \left[\|\underline{z}_x - \underline{b}_X\|^2 + \frac{\|\underline{c}_Y\|^2 \text{sen}^2 \theta}{\lambda} + 2g \right]^{n+f-\frac{3}{2}}}, \quad (2.8.13)$$

onde d e g são positivos. Portanto,

$$\int_{\Theta} p(\theta | D_{\text{obs}}) d\theta \propto \int_{Z_x = \{\underline{z}_x \in \mathbb{R}^{n-1}\}} f(\underline{z}_x) dz_1 dz_2 \cdots dz_{n-1} = \int_{Z_x} f(\underline{z}_x) d\underline{z}_x.$$

Note que a função $f(\underline{z}_x)$ em (2.8.13) é contínua e possui uma singularidade no ponto $\underline{z}_x = \underline{0}$. Esta singularidade pode ser tratada particionando-se a região de integração Z_x em dois conjuntos R_1 e R_2 , onde

$$R_1 = \{\underline{z}_x \in \mathbb{R}^{n-1} : \|\underline{z}_x\| < \varepsilon\} \quad \text{e} \quad R_2 = \{\underline{z}_x \in \mathbb{R}^{n-1} : \|\underline{z}_x\| \geq \varepsilon\},$$

com ε suficientemente pequeno.

Avaliação da integral de $f(\underline{z}_x)$ na região R_1

É fácil ver que em R_1 a integral de $f(\underline{z}_x)$ é finita se e somente se $n > 2$. Uma forma de ver isso é notar primeiramente que nessa região o termo $[\|\underline{z}_x\|^2 + 2d]$, no denominador de $f(\underline{z}_x)$, é “dominado” pela constante positiva $2d$, isto é, satisfaz a desigualdade $2d \leq \|\underline{z}_x\|^2 + 2d \leq 2d + \varepsilon^2$ e portanto para ε pequeno, $\|\underline{z}_x\|^2 + 2d \approx 2d$ (isto é, para $\varepsilon \rightarrow 0$ (ou $\|\underline{z}_x\| \rightarrow 0$), temos que $\|\underline{z}_x\|^2 + 2d \rightarrow 2d$). Da mesma forma, o termo $\left[\|\underline{z}_x - \underline{b}_X\|^2 + \frac{\|\underline{c}_Y\|^2 \text{sen}^2 \theta}{\lambda} + 2g \right]$ é “dominado” por uma constante positiva,

pois

$$\begin{aligned} [\|b_X\| - \varepsilon]^2 &\leq [\|z_x\| - \|b_X\|]^2 \leq \|z_x - b_X\|^2 \\ &\leq (\|z_x\| + \|b_X\|)^2 \leq (\varepsilon + \|b_X\|)^2 \end{aligned}$$

e portanto

$$\lim_{\varepsilon \rightarrow 0} \left[\|z_x - b_X\|^2 + \frac{\|\xi_Y\|^2 \text{sen}^2 \theta}{\lambda} + 2g \right] = \|b_X\|^2 + \frac{\|\xi_Y\|^2 \text{sen}^2 \theta}{\lambda} + 2g,$$

onde $\|b_X\| \geq 0$, $\frac{\|\xi_Y\|^2 \text{sen}^2 \theta}{\lambda} \geq 0$ e g é uma constante positiva.

Portanto, a integral $\int_{R_1} f(z_x) dz_x$ é finita se e somente se $\int_{R_1} \frac{1}{\|z_x\|} dz_x < \infty$.

Mas, pelo Lema 2.8.1(a), a integral $\int_{R_1} \frac{1}{\|z_x\|} dz_x$ é finita se e somente se $1 < n-1$, isto é, $n > 2$.

Avaliação da integral de $f(z_x)$ na região R_2

Para analisar a integral de $f(z_x)$ na região R_2 vamos particionar essa região em dois subconjuntos:

$$R_{2a} = \{ z_x \in \mathbb{R}^{n-1} : \varepsilon \leq \|z_x\| < k \} \quad \text{e} \quad R_{2b} = \{ z_x \in \mathbb{R}^{n-1} : \|z_x\| \geq k \},$$

onde k é escolhido suficientemente grande.

Na região R_{2a} a integral $\int_{R_{2a}} f(z_x) dz_x$ é finita pois $f(z_x)$ é uma função limitada e o conjunto R_{2a} também é limitado. Na região R_{2b} , a função $f(z_x)$ é limitada, porém R_{2b} não é um conjunto limitado, entretanto a quantidade $\|z_x\|^2$ domina os termos $[\|z_x\|^2 + 2d]$ e $[\|z_x - b_X\|^2 + \frac{\|\xi_Y\|^2 \text{sen}^2 \theta}{\lambda} + 2g]$ no denominador de $f(z_x)$ e portanto a integral

$$\int_{R_{2b}} f(z_x) dz_x < \infty \Leftrightarrow \int_{R_{2b}} \frac{1}{\|z_x\|^{1 + \frac{2(n+2c-1)}{2} + 2(n+f-\frac{3}{2})}} dz_x < \infty.$$

Portanto, $\int_{R_{2b}} f(z_x) dz_x$ converge se e somente se $1 + \frac{2(n+2c-1)}{2} + 2(n+f-\frac{3}{2}) > n-1$, isto é, $(n+f+c > 1)$, o que é garantido pelo Lema 2.8.1(b).

É fácil ver que na região R_{2b} o termo $[\|z_x\|^2 + 2d]$ é dominado por $\|z_x\|^2$, pois nessa região vale a desigualdade

$$k^2 \leq \|z_x\|^2 \leq [\|z_x\|^2 + 2d]$$

e, portanto, para k suficientemente grande, $[\|z_x\|^2 + 2d] \approx \|z_x\|^2$, isto é, para $k \rightarrow \infty$, $\Rightarrow \|z_x\| \rightarrow \infty$ e portanto $\lim_{k \rightarrow \infty} (\|z_x\|^2 + 2d) = \lim_{k \rightarrow \infty} \|z_x\|^2$.

Analogamente, é simples verificar que na região R_{2b} o termo $[\|z_x - \hat{b}_X\|^2 + \frac{\|\zeta_Y\|^2 \text{sen}^2 \theta}{\lambda} + 2g]$ é dominado por $\|z_x\|^2$.

Finalmente, analisando-se simultaneamente todas as regiões R_1 , R_{2a} e R_{2b} simultaneamente (levando em conta também a desigualdade em (2.8.12)), concluímos que $\int_{\Theta} p(\theta|D_{\text{obs}})d\theta$ é finita se e somente se $n > 2$ e $n + f + c > 1$, isto é, a posteriori é própria em Θ se e somente se $n \geq 3$. ■

Teorema 2.8.3 *A condição necessária e suficiente para a existência da distribuição à posteriori do modelo Normal em (2.8.1) com priori dada por $\pi(\theta) \propto \pi(\sigma_x^2)\pi(\sigma^2)$, onde $\pi(\sigma_x^2) \propto (\sigma_x^2)^{-r}$, $r > 0$, e $\sigma^2 \sim IG(f, g)$, é que $0 < r < \frac{1}{2}$ e $n \geq 2$.*

Demonstração: Para demonstrar esse teorema podemos aproveitar os cálculos obtidos na demonstração do Teorema 2.8.2, já que a distribuição a priori especificada nesse teorema coincide com a priori do Teorema 2.8.2 quando substituímos os hiperparâmetros c e d , respectivamente, por $(r - 1)$ e zero.

Portanto, a integral em Θ da distribuição à posteriori

$$\begin{aligned} \int_{\Theta} p(\theta|D_{\text{obs}})d\theta &= \int_{\mathfrak{X}} \int_{\Theta} p(x, \theta|D_{\text{obs}})d\theta d\mathfrak{X} \propto \int_{\mathfrak{X}} \int_{\Theta} L(\theta|D)\pi(\theta)d\theta d\mathfrak{X} \\ &\propto \int_{Z_x = \{z_x \in \mathbb{R}^{n-1}\}} g(z_x)dz_x, \end{aligned}$$

onde

$$g(z_x) = \frac{1}{\|z_x\|^{1 + \frac{2(n+2r-3)}{2}} \left[\|z_x - \hat{b}_X\|^2 + \frac{\|\zeta_Y\|^2 \text{sen}^2 \theta}{\lambda} + 2g \right]^{n+f-\frac{3}{2}}}$$

$$= \frac{1}{\|z_x\|^{n+2r-2} \left[\|z_x - b_X\|^2 + \frac{\|c_Y\|^2 \sin^2 \theta}{\lambda} + 2g \right]^{n+f-\frac{3}{2}}}. \quad (2.8.14)$$

Note que a função $g(z_x)$ foi obtida a partir de $f(z_x)$ dada em (2.8.13) substituindo-se c por $r - 1$ e d por zero e assumindo que a desigualdade em (2.8.12) é válida.

Para demonstrar esse teorema particionamos a região de integração Z_x da mesma forma que no Teorema 2.8.2, e o objetivo é encontrar as condições para que a integral $\int_{Z_x} g(z_x) dz_x$ seja finita.

Avaliação da integral na região R_1

Na região R_1 a integral $\int_{R_1} g(z_x) dz_x$ é finita se e somente se $\int_{R_1} \frac{1}{\|z_x\|^{n+2r-2}} dz_x$ for finita (pois o termo $\left[\|z_x - b_X\|^2 + \frac{\|c_Y\|^2 \sin^2 \theta}{\lambda} + 2g \right]$ que aparece no denominador, é dominado por uma constante positiva (como vimos na demonstração do Teorema 2.8.2). Portanto, a integral em R_1 é finita se e somente se $0 < r < \frac{1}{2}$ (pois pelo Lema 2.8.1 a integral $\int_{R_1} \frac{1}{\|z_x\|^{n+2r-2}} dz_x$ converge se e somente se $n+2r-2 < n-1$).

Avaliação da integral na região R_2

Na região R_{2a} não existe problema pois $g(z_x)$ é limitada e R_{2a} é limitado. Em R_{2b} a integral $\int_{R_{2b}} g(z_x) dz_x$ converge se e somente se $\int_{R_{2b}} \frac{1}{\|z_x\|^{n+2r-2+2(n+f-\frac{3}{2})}} dz_x$ convergir (pois $\|z_x\|^2$ domina o termo $\left[\|z_x - b_X\|^2 + \frac{\|c_Y\|^2 \sin^2 \theta}{\lambda} + 2g \right]$ que aparece no denominador como já discutimos na demonstração do Teorema 2.8.2. Portanto, pelo Lema 2.8.1(b), a integral em R_{2b} é finita se e somente se $n+2r-2+2(n+f-\frac{3}{2}) > n-1$, isto é, $n+f+r > 2$.

Observação: Se $n = 1$, a integral

$$\int_{\Theta} p(\theta | D_{\text{obs}}) d\theta = \int_{\mathbb{R}} \frac{(S_{x_1 x_1})^{f-r+\frac{1}{2}}}{\left[S_{x_1 x_1} (X_1 - x_1)^2 + \frac{(S_{x_1 x_1} S_{Y_1 Y_1} - S_{x_1 Y_1}^2)}{\lambda} + S_{x_1 x_1} 2g \right]^f} dx_1$$

(que é igual a (2.8.7) quando substituimos c por $(r - 1)$, d por zero, n por 1, S_{xx} por $S_{x_1x_1}$, S_{xY} por $S_{x_1Y_1}$ e S_{YY} por $S_{Y_1Y_1}$) está bem definida e é finita se e somente se $f = 0$ e $0 < r < \frac{1}{2}$ (pois $S_{x_1x_1}$ e $S_{x_1Y_1}$ são nulos para todo $x_1 \in \mathbb{R}$). Entretanto, por hipótese, f é positivo e portanto n deve ser necessariamente maior ou igual a 2.

Analisando todos esses resultados simultaneamente, concluímos que a condição necessária e suficiente para a existência da posteriori é que $n \geq 2$ e $0 < r < \frac{1}{2}$. ■

Teorema 2.8.4 *Suponhamos que os vetores de dados $\underline{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ e $\underline{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$ são linearmente independentes (isto é, pelo menos um ponto (X_i, Y_i) não é colinear com os demais) e que as componentes de \underline{X} , bem como as de \underline{Y} , não são todas iguais. Então se a distribuição à priori é dada por $\pi(\theta) = \pi(\sigma_x^2)\pi(\sigma^2)$, onde $\sigma_x^2 \sim IG(c, d)$ e $\pi(\sigma^2) \propto (\sigma^2)^{-s}$, $c, d, s > 0$, a condição necessária e suficiente para que a posteriori do modelo Normal em (2.8.1) seja própria é que $n \geq 3$.*

Demonstração: A demonstração desse teorema é muito semelhante à do Teorema 2.8.5. ■

Teorema 2.8.5 *Suponha que \underline{X} e \underline{Y} são linearmente independentes e as componentes, tanto de \underline{X} como de \underline{Y} , não são todas iguais. Então, se a distribuição à priori é dada por $\pi(\theta) = \pi(\sigma_x^2)\pi(\sigma^2)$, onde $\pi(\sigma_x^2) \propto (\sigma_x^2)^{-r}$ e $\pi(\sigma^2) \propto (\sigma^2)^{-s}$, com $r, s > 0$; a condição necessária e suficiente para que a posteriori do modelo em (2.8.1) seja própria é que $n + r + s > 3$ e $0 < r < \frac{1}{2}$.*

Demonstração: Seja $p(\theta|D_{\text{obs}})$ a distribuição a posteriori. A integral em Θ de $p(\theta|D_{\text{obs}})$ é dada por

$$\begin{aligned} \int_{\Theta} p(\theta|D_{\text{obs}})d\theta &= \int_{\underline{x}} \int_{\Theta} p(\underline{x}, \theta|D_{\text{obs}})d\theta d\underline{x} \propto \int_{\underline{x}} \int_{\Theta} L(\theta|D)\pi(\theta)d\theta d\underline{x} \\ &\propto \int_{Z=\mathbb{R}^n} \frac{1}{\|\underline{z}_x\|^{n+2r-2} \left[\|\underline{z}_x - \underline{b}_X\|^2 + \frac{\|\underline{z}_Y\|^2 \text{sen}^2 \theta}{\lambda} + (\sqrt{n}\bar{X} - z_n)^2 \right]^{n+s-2}} d\underline{z}. \end{aligned}$$

Essa integral foi obtida substituindo-se na integral em (2.8.9) as constantes c por $(r - 1)$, d por zero, f por $(s - 1)$ e g por zero.

Para estudar essa integral, primeiramente fazemos a integração em relação à z_n , ou seja, calculamos

$$\frac{1}{\|z_x\|^{n+2r-2}} \int_{-\infty}^{\infty} \frac{1}{\left[\|z_x - \tilde{b}_X\|^2 + \frac{\|\xi_Y\|^2 \text{sen}^2 \theta}{\lambda} + (\sqrt{n} \bar{X} - z_n)^2 \right]^{n+s-2}} dz_n.$$

Fazendo agora $k = \|z_x - \tilde{b}_X\|^2 + \frac{\|\xi_Y\|^2 \text{sen}^2 \theta}{\lambda}$, a integral em z_n acima é igual a

$$\int_{-\infty}^{\infty} \frac{1}{(k + (\sqrt{n} \bar{X} - z_n)^2)^{n+s-2}} dz_n.$$

Mas como k é positivo, já que os termos $\|z_x - \tilde{b}_X\|^2$ e $\frac{\|\xi_Y\|^2 \text{sen}^2 \theta}{\lambda}$ não se anulam simultaneamente (como justificaremos mais adiante), então

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{1}{[k + (\sqrt{n} \bar{X} - z_n)^2]^{n+s-2}} dz_n &= \frac{1}{k^{n+s-2}} \int_{-\infty}^{\infty} \frac{1}{\left[1 + \left(\frac{\sqrt{n} \bar{X} - z_n}{\sqrt{k}} \right)^2 \right]^{n+s-2}} dz_n \\ &= \frac{1}{k^{n+s-\frac{5}{2}}} \int_{-\infty}^{\infty} \frac{1}{(1 + u^2)^{n+s-2}} du, \end{aligned}$$

onde $\int_{-\infty}^{\infty} \frac{1}{(1+u^2)^{n+s-2}} du < \infty$ se e somente se $2(n+s-2) > 1$, isto é, $n+s > \frac{5}{2}$ (que é igual à desigualdade em (2.8.12) quando substituimos f por $(s-1)$).

Assumindo que $n+s > \frac{5}{2}$ então $\int_{-\infty}^{\infty} \frac{1}{[k + (\sqrt{n} \bar{X} - z_n)^2]^{n+s-2}} dz_n$ é proporcional a $\frac{1}{k^{n+s-\frac{5}{2}}}$ e portanto

$$\begin{aligned} \int_{\Theta} p(\theta | D_{\text{obs}}) d\theta &\propto \int_{Z_x = \mathbb{R}^{n-1}} \frac{1}{\|z_x\|^{n+2r-2} \left[\|z_x - \tilde{b}_X\|^2 + \frac{\|\xi_Y\|^2 \text{sen}^2 \theta}{\lambda} \right]^{n+s-\frac{5}{2}}} dz_x \\ &= \int_{Z_x} h(z_x) dz_x, \quad \text{onde} \\ h(z_x) &= \frac{1}{\|z_x\|^{n+2r-2} \left[\|z_x - \tilde{b}_X\|^2 + \frac{\|\xi_Y\|^2 \text{sen}^2 \theta}{\lambda} \right]^{n+s-\frac{5}{2}}} \end{aligned} \quad (2.8.15)$$

Note que $h(z_x)$ é contínua e possui uma única singularidade na origem (e outras duas singularidades removíveis). Para avaliar quais as condições em que a integral

$\int_{\mathbb{R}^{n-1}} h(z_x) dz_x$ converge, particionamos a região de integração em vários subconjuntos. A Figura 2.0 ilustra essa partição para o caso $n = 3$.

A seguir definimos as partições da região de integração:

$R_1 = \{z_x \in \mathbb{R}^{n-1} : \|z_x\| < \varepsilon_1\}$, onde ε_1 é suficientemente pequeno, por exemplo, $\varepsilon_1 = \frac{\|b_x\|}{10}$ (R_1 é a bola de centro na origem e raio ε_1 e isola a singularidade na origem);

$R_3 = \{z_x \in \mathbb{R}^{n-1} : \|z_x\| \geq k\}$;

$R_2 = \{z_x \in \mathbb{R}^{n-1} : \varepsilon_1 \leq \|z_x\| \leq k\}$, onde k é suficientemente grande, por exemplo, $k = 2\|b_x\|$. Esta região é particionada ainda em três sub-regiões, R_{2a} , R_{2b} e R_{2c} , que definimos a seguir:

$R_{2a} = R_2 \cap M$, onde M é a região compreendida entre as retas s e t da Figura 2.0.

As retas s e t passam pela origem e formam um ângulo θ_ε com \vec{c}_Y , onde θ_ε é escolhido suficientemente pequeno (por exemplo, $\theta_\varepsilon = \varphi/10$, onde φ é o ângulo entre \vec{c}_Y e \vec{b}_X). Portanto, a região M pode ser escrita como

$M = \{z_x = (z_1, \dots, z_{n-1})^T \in \mathbb{R}^{n-1} \mid \vec{z}_x \text{ e } \vec{c}_Y \text{ formam um ângulo } \theta \text{ menor do que } \theta_\varepsilon = \varphi/10\}$.

A região R_{2b} é a bola de centro em b_x e raio ε_2 ,

$$R_{2b} = \{z_x \in \mathbb{R}^{n-1} : \|z_x - b_x\| < \varepsilon_2\},$$

onde, ε_2 é fixado arbitrariamente, tão pequeno quanto se queira, por exemplo, $\varepsilon_2 = \min \left\{ \frac{\|b_x\| - \varepsilon_1}{10}, \frac{d(b_x, r) - d(P, r)}{10} \right\}$, onde

$d(b_x, r)$ é a distância perpendicular do ponto b_x à reta r , ou seja, $d(b_x, r) = \frac{\|\vec{b}_X \times \vec{c}_Y\|}{\|\vec{c}_Y\|}$ onde $\vec{b}_X \times \vec{c}_Y = \|\vec{b}_X\| \sin \varphi$ (isto é, $\vec{b}_X \times \vec{c}_Y$ é o produto vetorial de \vec{b}_X e \vec{c}_Y);

P é o ponto de interseção de duas retas: a reta s e a reta perpendicular a r que passa pelo ponto \underline{b}_X ;

$$d(P, r) = \text{distância perpendicular do ponto } P \text{ à reta } r, d(P, r) = \frac{\|\vec{OP} \times \vec{c}_Y\|}{\|\vec{c}_Y\|}.$$

Finalmente, a região R_{2c} é a região complementar, isto é, $R_{2c} = R_2 \cap (R_{2a} \cup R_{2b})$.

É importante observar que os vetores \underline{b}_X e \underline{c}_Y são distintos do vetor nulo, pois pela hipótese do Teorema 2.8.5, os vetores \underline{X} e \underline{Y} não têm todas as componentes iguais (lembrando que $\underline{b}_X = A\underline{X}$ e $\underline{c}_Y = A\underline{Y}$, onde A é formada pelas primeiras $n - 1$ linhas da matriz de Helmert definida em (2.8.8)).

O ângulo φ entre \vec{b}_X e \vec{c}_Y é diferente de zero pois por hipótese \vec{X} e \vec{Y} são linearmente independentes.

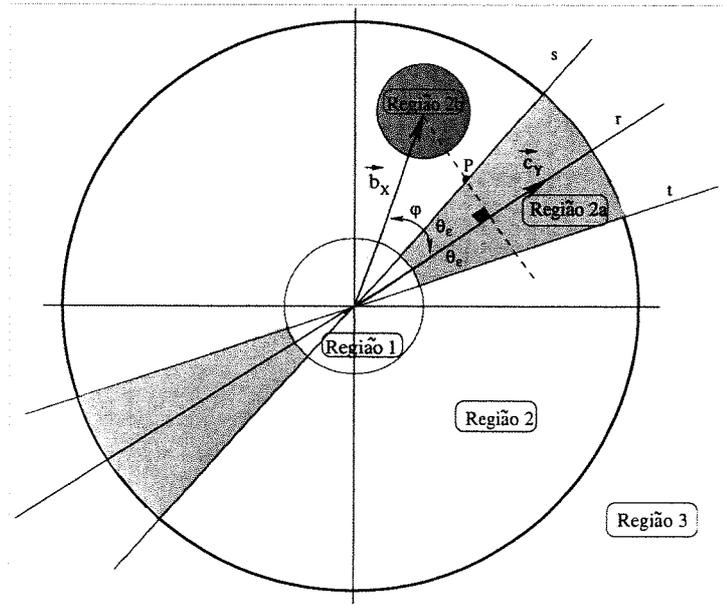


Figura 2.0: Região de integração da distribuição à posteriori no Teorema 2.8.4 quando $n = 3$. Os vetores observados \vec{b}_X e \vec{c}_Y são representados pelos segmentos orientados que passam pela origem e tem extremidades nos pontos $\underline{b}_X = (b_1, \dots, b_{n-1})^\top$ e $\underline{c}_Y = (c_1, \dots, c_{n-1})^\top$ respectivamente. O ângulo formado por \vec{b}_X e \vec{c}_Y é φ . A reta r passa pela origem e tem a direção de \vec{c}_Y . Cada uma das retas s e t forma um ângulo θ_ε com a reta r .

AVALIAÇÃO DA INTEGRAL DE $h(z_x)$ NA REGIÃO R_1 (onde $h(z_x)$ é dada em (2.8.15))

Na região R_1 o termo $\left[\|z_x - \underline{b}_X\|^2 + \frac{\|\xi_Y\|^2 \text{sen}^2 \theta}{\lambda}\right]$ no denominador de $h(z_x)$ é dominado por $\left[\|\underline{b}_X\|^2 + \frac{\|\xi_Y\|^2 \text{sen}^2 \theta}{\lambda}\right]$, que é positivo. De fato,

$$(\|\underline{b}_X\| - \varepsilon_1)^2 \leq \|z_x\| - \|\underline{b}_X\|^2 \leq \|z_x - \underline{b}_X\|^2 \leq (\|z_x\| + \|\underline{b}_X\|)^2 \leq (\|\underline{b}_X\| + \varepsilon_1)^2$$

e, portanto, para ε_1 suficientemente pequeno, $\|\underline{b}_X\|^2 \cong \|z_x - \underline{b}_X\|^2$. Note que embora $\frac{\|\xi_Y\|^2 \text{sen}^2 \theta}{\lambda}$ possa ser nulo, $\|\underline{b}_X\|^2$ é diferente de zero pela hipótese do Teorema 2.8.4 e, portanto, $\|\underline{b}_X\|^2 + \frac{\|\xi_Y\|^2 \text{sen}^2 \theta}{\lambda} > 0$.

Portanto, $\int_{R_1} h(z_x) d z_x$ é finita se e somente se $\int_{R_1} \frac{1}{\|z_x\|^{n+2r-2}} d z_x$ convergir. Pelo Lema 2.8.1 esta integral converge se e somente se $n+2r-2 < n-1$, isto é, $0 < r < \frac{1}{2}$.

AVALIAÇÃO DA INTEGRAL NA REGIÃO R_{2a}

Nessa região a função $h(z_x)$ é limitada e portanto a integral $\int_{R_{2a}} h(z_x) d z_x$ é finita já que o domínio de integração R_{2a} é um conjunto limitado. É fácil ver que $h(z_x)$ é limitada pois nessa região vale a desigualdade

$$0 \leq \frac{\|\xi_Y\|^2 \text{sen}^2 \theta}{\lambda} < \frac{\|\xi_Y\|^2 \text{sen}^2 \theta_\varepsilon}{\lambda}$$

(pelo fato que o ângulo θ entre \vec{z}_x e \vec{c}_Y é menor que θ_ε) e, além disso, quando $z_x \in R_{2a}$ então $z_x \in R_2$ e $z_x \notin R_{2b}$ e, portanto, as seguintes desigualdades também são válidas:

$$\varepsilon_1 \leq \|z_x\| < k \quad \text{e} \quad \varepsilon_2^2 \leq \|z_x - \underline{b}_X\|^2 \leq (\|z_x\| + \|\underline{b}_X\|)^2 \leq (k + \|\underline{b}_X\|)^2.$$

Isso implica que

$$0 < \frac{1}{k^{n+2r-2} \left[(k + \|\underline{b}_X\|)^2 + \frac{\|\xi_Y\|^2 \text{sen}^2 \theta_\varepsilon}{\lambda} \right]^{n+s-\frac{5}{2}}} \leq h(z_x) \leq \frac{1}{\varepsilon_1^{n+2r-2} [\varepsilon_2]^{2(n+s-\frac{5}{2})}} < \infty.$$

AVALIAÇÃO DA INTEGRAL EM R_{2b}

Nessa região a integral $\int_{R_{2b}} h(z_x) dz_x$ também é finita pois $h(z_x)$ é limitada e R_{2b} é um conjunto limitado. Aqui também é simples verificar que $h(z_x)$ é limitada pois quando $z_x \in R_{2b}$, então $z_x \in R_2$ e $z_x \notin R_{2a}$ e, portanto, valem as seguintes desigualdades:

$$\varepsilon_1 \leq \|z_x\| < k \quad \text{e} \quad \frac{\|\xi_Y\|^2 \text{sen}^2 \theta_\varepsilon}{\lambda} \leq \frac{\|\xi_Y\|^2 \text{sen}^2 \theta}{\lambda} \leq \frac{\|\xi_Y\|^2}{\lambda}.$$

Além disso, vale que $0 \leq \|z_x - \underline{b}_X\| < \varepsilon_2$ e portanto

$$\frac{1}{k^{n+2r-2} \left[\varepsilon_2^2 + \frac{\|\xi_Y\|^2}{\lambda} \right]^{n+s-\frac{5}{2}}} \leq h(z_x) \leq \frac{1}{\varepsilon_1^{n+2r-2} \left[\frac{\|\xi_Y\|^2 \text{sen}^2 \theta_\varepsilon}{\lambda} \right]^{n+s-\frac{5}{2}}}.$$

AVALIAÇÃO DA INTEGRAL EM R_{2c}

Nesta região também a integral $\int_{R_{2c}} h(z_x) dz_x$ é finita pois a função $h(z_x)$ é limitada e R_{2c} é limitada. De fato, se $z_x \in R_{2c}$, então $z_x \in R_2 \cap R_{2a}^c \cap R_{2b}^c$ e valem as seguintes desigualdades:

$$\varepsilon_1 \leq \|z_x\| < k, \quad \frac{\|\xi_Y\|^2 \text{sen}^2 \theta_\varepsilon}{\lambda} \leq \frac{\|\xi_Y\|^2 \text{sen}^2 \theta}{\lambda} \leq \frac{\|\xi_Y\|^2}{\lambda}$$

e $\varepsilon_2 \leq \|z_x - \underline{b}_X\| \leq (k + \|\underline{b}_X\|)$ e portanto

$$\frac{1}{k^{n+2r-2} \left[(k + \|\underline{b}_X\|)^2 + \frac{\|\xi_Y\|^2 \text{sen}^2 \theta}{\lambda} \right]^{n+s-\frac{5}{2}}} \leq h(z_x) \leq \frac{1}{\varepsilon_1^{n+2r-2} \left[\varepsilon_2^2 + \frac{\|\xi_Y\|^2 \text{sen}^2 \theta_\varepsilon}{\lambda} \right]^{n+s-\frac{5}{2}}}.$$

AVALIAÇÃO DA INTEGRAL EM R_3

Em R_3 a função $h(z_x)$ é limitada, mas a região de integração não é um conjunto limitado, entretanto, o termo $\left[\|z_x - \underline{b}_X\|^2 + \frac{\|\xi_Y\|^2 \text{sen}^2 \theta}{\lambda} \right]$ do denominador de $h(z_x)$ é dominado por $\|z_x\|^2$. (De fato, para k suficientemente grande, por exemplo $k = 2\|\underline{b}_X\|$, vale que $\|\underline{b}_X\| < \frac{\|z_x\|}{2}$ e portanto valem as desigualdades

$$\frac{k^2}{4} \leq \left(\frac{\|z_x\|}{2} \right)^2 \leq (\|z_x\| - \|\underline{b}_X\|)^2 \leq \left[\|z_x - \underline{b}_X\|^2 + \frac{\|\xi_Y\|^2 \text{sen}^2 \theta}{\lambda} \right];$$

então, para k suficientemente grande, $\frac{\|z_x\|^2}{2} \equiv \left[\|z_x - b_X\|^2 + \frac{\|C_Y\|^2 \sin^2 \theta}{\lambda} \right]$, isto é, para $k \rightarrow \infty \Rightarrow \|z_x\| \rightarrow \infty \Rightarrow \lim_{k \rightarrow \infty} \frac{\|z_x\|^2}{2} = \lim_{k \rightarrow \infty} \left[\|z_x - b_X\|^2 + \frac{\|C_Y\|^2 \sin^2 \theta}{\lambda} \right]$. Portanto, a integral

$$\int_{R_3} \frac{1}{\|z_x\|^{n+2r-2} \left[\|z_x - b_X\|^2 + \frac{\|C_Y\|^2 \sin^2 \theta}{\lambda} \right]^{n+s-\frac{5}{2}}} d z_x$$

é finita se e somente se $\int_{R_3} \frac{1}{\|z_x\|^{n+2r-2+2(n+s-\frac{5}{2})}} d z_x$ é finita. Mas isso ocorre se e somente se $n + 2r - 2 + 2(n + s - \frac{5}{2}) > n - 1$, isto é, $n + r + s > 3$.

Conclusão da demonstração:

Analisando-se simultaneamente todas essas informações, concluímos que a condição necessária e suficiente para que a posteriori no Teorema 2.8.5 seja própria é que $0 < r < \frac{1}{2}$ e $n + r + s > 3$. ■

2.8.2 Condições necessárias e suficientes no caso do modelo não identificável

É fato conhecido que mesmo um modelo não identificável com priori imprópria pode ter posteriori própria (Gelfand e Sahu, 1999, e Erickson, 1989). Nessa seção discutimos a existência da posteriori do modelo estrutural normal sem restrições de identificabilidade sob diferentes escolhas de prioris impróprias.

Na seção 2.6 definimos o modelo

$$\begin{aligned} Y_i &= \alpha + \beta x_i + e_i \\ \text{e} \quad X_i &= x_i + e_i, \end{aligned} \tag{2.8.16}$$

onde

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \stackrel{\text{i.i.d.}}{\sim} N \left(\begin{pmatrix} 0 \\ 0 \\ \mu \end{pmatrix}; \begin{pmatrix} \sigma_e^2 & 0 & 0 \\ 0 & \sigma_u^2 & 0 \\ 0 & 0 & \sigma_x^2 \end{pmatrix} \right), \quad i = 1, \dots, n,$$

e $\Theta = \{ \theta = (\alpha, \beta, \mu, \sigma_x^2, \sigma_u^2, \sigma_e^2)^T \mid \alpha, \beta, \mu \in \mathbb{R} \text{ e } \sigma_x^2, \sigma_u^2, \sigma_e^2 \in \mathbb{R}^+ \}$.

A distribuição à posteriori, baseada nos dados completos (D) é dada por

$$p(\theta | D_{\text{obs}}) = \int_{\mathfrak{X} = \mathbb{R}^n} p(\theta, \mathfrak{X} | D_{\text{obs}}) d \mathfrak{X} = \int_{\mathfrak{X}} L(\theta | D) \pi(\theta) d \mathfrak{X},$$

onde $L(\varrho|D)$ é dada por (2.1.5) e $\pi(\vartheta)$ é a distribuição à priori de ϑ .

As prioris consideradas nessa seção serão da forma $\pi(\vartheta) = \prod_{i=1}^6 \pi(\vartheta_i)$, onde $\pi(\alpha)$, $\pi(\beta)$ e $\pi(\mu)$ são proporcionais a 1, $\sigma_e^2 \sim IG(p, q)$ e as componentes σ_x^2 e σ_u^2 podendo ter distribuição própria ou imprópria de acordo com os teoremas abaixo.

A seguir apresentamos teoremas que garantem posteriori própria para quatro diferentes distribuições à priori.

Teoremas:

Teorema 2.8.6 *No modelo Normal com erros nas variáveis, dado em (2.8.16), e com priori $\pi(\vartheta) \propto \pi(\sigma_x^2)\pi(\sigma^2)\pi(\sigma_e^2)$, onde $\sigma_x^2 \sim IG(c, d)$, $\sigma_u^2 \sim IG(f, g)$ e $\sigma_e^2 \sim IG(p, q)$, a condição necessária e suficiente para a existência da posteriori correspondente é que $n \geq 3$.*

Demonstração: A integral da posteriori no espaço paramétrico Θ é

$$\begin{aligned} \int_{\Theta} p(\varrho|D_{\text{obs}})d\varrho &= \int_{\Theta} \int_{\mathfrak{X}=\mathbb{R}^n} p(\mathfrak{x}, \varrho|D_{\text{obs}})d\mathfrak{x}d\varrho \\ &= \int_{\mathfrak{X}} \int_{\Theta} p(\mathfrak{x}, \varrho|D_{\text{obs}})d\varrho d\mathfrak{x} \propto \int_{\mathfrak{X}} \int_{\Theta} L(\varrho|D)\pi(\varrho)d\varrho d\mathfrak{x} \\ &\propto \int_{\mathfrak{X}} \frac{1}{S_{xx}^{\frac{1}{2}} [2d + S_{xx}]^{\frac{n+2c-1}{2}} [2g + \sum_{i=1}^n (X_i - x_i)^2]^{\frac{n}{2}+f} \left[2q + \frac{S_{xx}S_{YY} - S_{xY}^2}{S_{xx}}\right]^{\frac{n}{2}+p-1}} d\mathfrak{x}, \end{aligned} \quad (2.8.17)$$

onde $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$, $S_{YY} = \sum_{i=1}^n (Y_i - \bar{Y})^2$, $S_{xY} = \sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})$.

Usando a transformação que a cada (\mathfrak{x}, X, Y) associa $(z, \underline{b}, \underline{c}) = (H \mathfrak{x}, H X, H Y)$, onde H é definida em (2.8.8), a integral acima torna-se igual a

$$\int_{Z=\mathbb{R}^n} \frac{1}{\left\{ \|z_x\| [2d + \|z_x\|^2]^{\frac{n+2c-1}{2}} [2g + \|z_x - \underline{b}_X\|^2 + (\sqrt{n}\bar{X} - z_n)^2]^{\frac{n}{2}+f} \cdot [2q + \|\underline{c}_Y\|^2 \text{sen}^2 \theta]^{\frac{n}{2}+p-1} \right\}} dz,$$

onde z_x , \underline{b}_X e \underline{c}_Y são os vetores formados pelas primeiras $n-1$ componentes dos vetores z , \underline{b} e \underline{c} , respectivamente, e $z = (z_1, z_2, \dots, z_n) = (z_x, z_n)$.

Agora, eliminando z_n por integração e assumindo que a desigualdade $(n+2f > 1)$ é válida, temos que a integral acima é proporcional a

$$\int_{Z_x=\mathbb{R}^{n-1}} 1/\left\{\|z_x\|[2d + \|z_x\|^2]^{\frac{n+2c-1}{2}} [2g + \|z_x - \underline{b}_X\|^2]^{\frac{n+2f-1}{2}} \cdot [2q + \|\xi_Y\|^2 \text{sen}^2 \theta]^{\frac{n}{2}+p-1}\right\} d z_x. \quad (2.8.18)$$

Para analisar essa integral, particionamos Z_x da mesma forma que no Teorema 2.8.2, e assim temos que

Na região $R_1 = \{z_x \in Z_x \mid \|z_x\| < \varepsilon\}$, ela converge se e somente se a integral $\int_{R_1} \frac{1}{\|z_x\|} d z_x < \infty$, isto é, $n \geq 3$. Na região complementar, a integral converge se e somente se $1 + \frac{2(n+2c-1)}{2} + \frac{2(n+2f-1)}{2} > n - 1$, isto é, sempre.

Analisando simultaneamente todas as informações, concluímos que a posteriori é própria se e somente se $n \geq 3$. ■

Teorema 2.8.7 *No modelo Normal em (2.8.16) e com priori $\pi(\theta) \propto \pi(\sigma_x^2)\pi(\sigma_u^2) \cdot \pi(\sigma_e^2)$, onde $\pi(\sigma_x^2) \propto (\sigma_x^2)^{-r}$, $r > 0$, $\sigma_u^2 \sim IG(f, g)$ e $\sigma_e^2 \sim IG(p, q)$, a condição necessária e suficiente para a existência da posteriori correspondente é que $0 < r < \frac{1}{2}$ e $n + 2(r + f) > 2$.*

Demonstração: Substituindo-se d por zero e c por $r - 1$ em (2.8.18), temos que

$$\int_{\Theta} p(\theta | D_{\text{obs}}) d\theta \propto \int_{Z_x=\mathbb{R}^{n-1}} \frac{1}{\|z_x\|^{n+2r-2} [2g + \|z_x - \underline{b}_X\|^2]^{\frac{n+2f-1}{2}} [2q + \|\xi_Y\|^2 \text{sen}^2 \theta]^{\frac{n}{2}+p-1}} d z_x. \quad (2.8.19)$$

Para analisar essa integral particionamos Z_x da mesma forma que no Teorema 2.8.3 e assim temos que:

Na região $R_1 = \{z_x \in Z_x \mid \|z_x\| < \varepsilon\}$, ela converge se e somente se $n + 2r - 2 < n - 1$, isto é, $0 < r < \frac{1}{2}$.

Na região complementar, ela converge se e somente se $n + 2r - 2 + 2 \left(\frac{n+2f-1}{2}\right) > n - 1$, isto é, $n + 2(r + f) > 2$.

Analisando simultaneamente essas informações, concluímos que a posteriori é própria se e somente se $0 < r < \frac{1}{2}$ e $n + 2(r + f) > 2$. ■

Teorema 2.8.8 *No modelo Normal com erros nas variáveis dado em (2.8.16) e com priori $\pi(\theta) \propto \pi(\sigma_x^2)\pi(\sigma_u^2)\pi(\sigma_e^2)$, onde $\sigma_x^2 \sim IG(c, d)$, $\pi(\sigma_u^2) \propto (\sigma_u^2)^{-s}$ e $\sigma_e^2 \sim IG(p, q)$, onde \underline{X} e \underline{Y} são linearmente independentes e as componentes de \underline{X} , bem como as de \underline{Y} , não são todas iguais, então a condição necessária e suficiente para que a posteriori seja própria é que $n \geq 3$ e $0 < s < 1$.*

Demonstração: A demonstração é bastante parecida com a do Teorema 2.8.9 a seguir. Por este motivo a omitimos. ■

Teorema 2.8.9 *No modelo Normal com erros nas variáveis dado em (2.8.16) e com priori para θ dada por $\pi(\theta) \propto \pi(\sigma_x^2)\pi(\sigma_u^2)\pi(\sigma_e^2)$, onde $\pi(\sigma_x^2) \propto (\sigma_x^2)^{-r}$, $\pi(\sigma_u^2) \propto (\sigma_u^2)^{-s}$, $\pi(\sigma_e^2) \sim IG(p, q)$, onde \underline{X} e \underline{Y} são linearmente independentes e as componentes de \underline{X} , bem como as de \underline{Y} , não são todas iguais, então a condição necessária e suficiente para que a posteriori seja própria é que $n + 2(r + s) > 4$, $0 < r < \frac{1}{2}$, $0 < s < 1$.*

Demonstração: Substituindo d e g por zero, c por $r - 1$ e f por $s - 1$ em (2.8.17), temos que

$$\begin{aligned} \int_{\Theta} p(\theta|D_{\text{obs}})d\theta &\propto \int_{\mathfrak{X}} \frac{1}{S_{xx}^{\frac{n+2r-2}{2}} [\sum_{i=1}^n (X_i - x_i)^2]^{\frac{n+2s-2}{2}} \left[2q + \frac{S_{xx}S_{YY} - S_{xY}^2}{S_{xx}}\right]^{\frac{n+2p-2}{2}}} d\mathfrak{x} \\ &= \int_Z \frac{1}{\|\underline{z}_x\|^{n+2r-2} [\|\underline{z}_x - \underline{b}_X\|^2 + (\sqrt{n}\bar{X} - z_n)^2]^{\frac{n+2s-2}{2}} [2q + \|\xi_Y\|^2 \text{sen}^2\theta]^{\frac{n+2p-2}{2}}} d\mathfrak{z}. \end{aligned} \quad (2.8.20)$$

Integrando a função dentro do sinal de integração em (2.8.20), com relação a z_n , obtemos a função

$$\frac{1}{\|\underline{z}_x\|^{n+2r-2} [2q + \|\xi_Y\|^2 \text{sen}^2\theta]^{\frac{n+2p-2}{2}} (\|\underline{z}_x - \underline{b}_X\|^2)^{\frac{n+2s-2}{2}}} \int_{-\infty}^{\infty} \frac{1}{\left(1 + \left(\frac{\sqrt{n}\bar{X} - z_n}{\|\underline{z}_x - \underline{b}_X\|}\right)^2\right)^{\frac{n+2s-2}{2}}} dz_n.$$

Agora, fazendo $u = \frac{\sqrt{n}\bar{X}-z_n}{\|\tilde{z}_x - \tilde{b}_X\|}$, a integral

$$\int_{-\infty}^{\infty} \frac{1}{\left(1 + \left(\frac{\sqrt{n}\bar{X}-z_n}{\|\tilde{z}_x - \tilde{b}_X\|}\right)^2\right)^{\frac{n+2s-2}{2}}} dz_n = \|\tilde{z}_x - \tilde{b}_X\| \int_{-\infty}^{\infty} \frac{1}{(1+u^2)^{\frac{n+2s-2}{2}}} du,$$

onde $\int_{-\infty}^{\infty} \frac{1}{(1+u^2)^{\frac{n+2s-2}{2}}} du < \infty$ se e somente se $2\left(\frac{n+2s-2}{2}\right) > 1$, isto é, $n+2s > 3$ (notemos que a função $\frac{1}{[1+u^2]^{\frac{n+2s-2}{2}}}$ só não está definida num único ponto u quando $z_x = \tilde{b}_X$ e, portanto, não traz problemas para a integral de Riemann).

Portanto, a integral em (2.8.20) é proporcional a

$$\int_{Z_x} \frac{1}{\|\tilde{z}_x\|^{n+2r-2} [\|\tilde{z}_x - \tilde{b}_X\|]^{n+2s-3} [2q + \|\zeta_Y\|^2 \text{sen } 2\theta]^{\frac{n+2p-2}{2}}} d\tilde{z}_x.$$

Para analisar essa integral, particionamos a região de integração Z_x nas regiões:

$$R_1 = \{z_x \in \mathbb{R}^{n-1} : \|z_x\| < \varepsilon_1\} \text{ para } \varepsilon_1 \text{ suficientemente pequeno (por exemplo, } \varepsilon_1 = \frac{\|\tilde{b}_X\|}{10}\text{),}$$

$$R_2 = \{z_x \in \mathbb{R}^{n-1} : \varepsilon_1 \leq \|z_x\| < k\}, \text{ onde } k \text{ é suficientemente grande (por exemplo, } k = 2\|\tilde{b}_X\|\text{),}$$

$$R_{2a} = \{z_x \in R_2 : \|z_x - \tilde{b}_X\| < \varepsilon_2\}, \text{ onde } \varepsilon_2 \text{ é suficientemente pequeno (por exemplo, } \varepsilon_2 = \frac{\|\tilde{b}_X\| - \varepsilon_1}{10}\text{),}$$

$$R_3 = \{z_x \in \mathbb{R}^{n-1} : \|z_x\| \geq k\}.$$

Em R_1 , a integral é finita se e somente se $n+2r-2 < n-1$, isto é, $0 < r < \frac{1}{2}$.

Em R_3 , ela é finita se e somente se $n+2r-2+n+2s-3 > n-1$, isto é, $n+2(r+s) > 4$.

Em R_{2a} , fazendo-se a transformação $y = z_x - \tilde{b}_X$, a integral acima é proporcional a

$$\int_{\|y\| < \varepsilon_2} \frac{1}{\|y + \tilde{b}_X\|^{n+2r-2} [y]^{n+2s-3}} dy,$$

que é finita se e somente se $n+2s-3 < n-1$, isto é, $0 < s < 1$.

Em $R_2 \cap R_{2a}^c$, a integral é finita pois a função dentro do sinal de integração é limitada e o domínio de integração ao $R_2 \cap R_{2a}^c$ é um conjunto limitado.

Analisando simultaneamente todas essas informações, concluímos que a posteriori é própria se e somente se $0 < r < \frac{1}{2}$, $n + 2(r + s) > 4$ e $0 < s < 1$. ■

Observação 1: As condições para a existência da posteriori nos modelos com e sem restrição de identificabilidade são bem distintas (principalmente entre os teoremas 2.8.4 e 2.8.8 e entre os teoremas 2.8.5 e 2.8.9)

Observação 2: Os modelos dessa seção apresentam dois problemas distintos: envolvem uma verossimilhança não identificável e incorporam prioris impróprias. Embora tenhamos conseguido provar a existência das posterioris a partir de certas prioris impróprias, ainda há o problema da falta de identificabilidade que pode acarretar em falha na convergência do algoritmo de Gibbs especialmente porque a priori é pouco informativa. (Na seção 2.12, entretanto, exibimos um conjunto de dados onde aparentemente não ocorre problema de convergência do algoritmo de Gibbs, onde este foi baseado no modelo não identificável com priori praticamente não informativa).

2.9 Estimação do modelo

Para estimar os parâmetros do modelo, podemos usar vários estimadores que correspondem a diferentes funções de perda. É fato bastante conhecido que tomando-se a função perda quadrática, o estimador de Bayes de $\tau(\theta)$, que é o estimador que minimiza a perda esperada com relação à distribuição à posteriori (isto é, o estimador que minimiza o risco à posteriori) é a média da distribuição de $\tau(\theta)$. No caso da função perda ser igual ao desvio absoluto, o estimador de Bayes é a mediana da posteriori e no caso da função perda 0-1 é a moda da posteriori.

Nos exemplos de aplicação vamos apresentar não somente as estimativas pontuais

da média, mediana, quantis e variância da posteriori de cada um dos parâmetros, como também os intervalos de credibilidade de comprimento mínimo para cada parâmetro (que coincidem com os intervalos “Highest Posterior Density”, ou HPD, quando as distribuições forem unimodais). Além das estimativas pontuais e por intervalo, vamos apresentar também as estimativas das densidades marginais à posteriori de cada um dos parâmetros do modelo.

Partindo do pressuposto que os $(T - T_0)$ vetores de observações gerados pelo algoritmo MCMC (onde $T_0 =$ “burn in period”) podem ser considerados como uma amostra da distribuição à posteriori que desejamos analisar, então as medidas acima podem ser usadas para fazer inferência sobre essa posteriori.

Se tivermos gerado m cadeias paralelas, podemos utilizar essa informação na estimação. Na tabela abaixo apresentamos o diagrama das observações provenientes de m cadeias paralelas geradas por um algoritmo MCMC.

cadeia 1	cadeia 2	...	cadeia m
$\vartheta_1^{(1)}$	$\vartheta_2^{(1)}$...	$\vartheta_m^{(1)}$
\vdots			\vdots
$\vartheta_1^{(T_0)}$	$\vartheta_2^{(T_0)}$...	$\vartheta_m^{(T_0)}$
\vdots			\vdots
$\vartheta_1^{(T)}$	$\vartheta_2^{(T)}$...	$\vartheta_m^{(T)}$

No caso, por exemplo, do modelo estrutural com restrição de identificabilidade $\sigma_e^2 = \lambda\sigma_u^2 = \lambda\sigma^2$, $\vartheta_i^{(j)}$ é igual a $(\alpha_i^{(j)}, \beta_i^{(j)}, \mu_i^{(j)}, \sigma_{x(i)}^{(j)}, \sigma_{(i)}^{2(j)})$. Assim, por exemplo, uma estimativa simples da média e variância da posteriori de α , é dada por

$$\hat{E}(\alpha|D_{\text{obs}}) = \frac{1}{m(T - T_0)} \sum_{i=1}^m \sum_{j=T_0+1}^T \alpha_i^{(j)},$$

$$\widehat{\text{Var}}(\alpha|D_{\text{obs}}) = \frac{1}{m(T - T_0) - 1} \sum_{i=1}^m \sum_{j=T_0+1}^T (\alpha_i^{(j)} - \hat{E}(\alpha|D_{\text{obs}}))^2$$

e uma estimativa simples da densidade marginal à posteriori de α é obtida fazendo-se um histograma com os valores $\alpha_i^{(j)}$, $i = 1, \dots, m$, $j = T_0, \dots, T$.

Observação 1: No modelo estrutural com restrição de identificabilidade $\sigma_e^2 = \lambda\sigma^2$, a densidade condicional completa à posteriori de α , $\pi(\alpha|\beta, \mu, \sigma_x^2, \sigma^2, \mathfrak{z}, D_{\text{obs}})$ é conhecida (e tem esperança e variância finita) dada em (2.5.9) por

$$\alpha|\beta, \mu, \sigma_x, \sigma^2, \mathfrak{z}, D_{\text{obs}} \sim N(A, B), \text{ onde}$$

$$A = \frac{n\sigma_\alpha^2(\bar{Y} - \beta\bar{x}) + \lambda a\sigma_\alpha^2}{n + \sigma_\alpha^2 + \lambda\sigma^2}, \text{ onde } \bar{x} = \frac{1}{n} \sum_{t=1}^n x_t,$$

e

$$B = \frac{\lambda\sigma_\alpha^2\sigma_\alpha^2}{n\sigma_\alpha^2 + \lambda\sigma^2}.$$

Então a média da distribuição marginal de α , $E(\alpha|D_{\text{obs}})$, pode ser estimada pelo método de Monte Carlo por

$$\frac{1}{m(T-T_0)} \sum_{i=1}^m \sum_{j=T_0+1}^T E(\alpha|\beta_i^{(j)}, \mu_i^{(j)}, \sigma_{x(i)}^{2(j)}, \sigma_i^{2(j)}, \mathfrak{z}_i^{(j)}, D_{\text{obs}}) = \frac{1}{m(T-T_0)} \sum_{i=1}^m \sum_{j=T_0+1}^T A_i^{(j)}.$$

Analogamente, a variância, $\text{Var}(\alpha|D_{\text{obs}})$, pode ser estimada por

$$\frac{1}{m(T-T_0)} \sum_{i=1}^m \sum_{j=T_0+1}^T B_i^{(j)} + \frac{1}{m(T-T_0)} \sum_{i=1}^m \sum_{j=T_0+1}^T (A_i^{(j)})^2 - \left(\frac{\sum_{i=1}^m \sum_{j=T_0+1}^T A_i^{(j)}}{m(T-T_0)} \right)^2,$$

onde

$$A_i^{(j)} = \frac{n\sigma_\alpha^2(\bar{Y} - \beta_i^{(j)}\bar{\mathfrak{z}}_i^{(j)}) + \lambda a\sigma_i^{2(j)}}{n\sigma_\alpha^2 + \lambda\sigma_i^{2(j)}}, \text{ onde } \bar{\mathfrak{z}}_i^{(j)} = \frac{1}{n} \sum_{t=1}^n x_{t(i)}^{(j)},$$

e

$$B_i^{(j)} = \frac{\lambda\sigma_i^{2(j)}\sigma_\alpha^2}{n\sigma_\alpha^2 + \lambda\sigma_i^{2(j)}}.$$

Esses estimadores são conhecidos como “estimadores Rao-Blackwelizados”.

Usando o mesmo raciocínio, estima-se a própria densidade marginal de α por

$$\hat{p}(\alpha|D_{\text{obs}}) = \frac{1}{m(T-T_0)} \sum_{i=1}^m \sum_{j=T_0+1}^T \pi(\alpha|\beta_i^{(j)}, \mu_i^{(j)}, \sigma_{x(i)}^{2(j)}, \sigma_i^{2(j)}, \mathfrak{z}_i^{(j)}, D_{\text{obs}}),$$

onde

$$\pi(\alpha|\beta_i^{(j)}, \mu_i^{(j)}, \sigma_{x(i)}^{2(j)}, \sigma_i^{2(j)}, \mathfrak{z}_i^{(j)}) = \frac{1}{\sqrt{2\pi B_i^{(j)}}} \exp \left\{ -\frac{1}{2B_i^{(j)}} (\alpha - A_i^{(j)})^2 \right\}.$$

Observação 2: As distribuições marginais à posteriori dos parâmetros β , μ , σ_x^2 e σ^2 no modelo estrutural dado em (2.3.1), bem como suas médias e variâncias podem ser estimadas de forma análoga à descrita na Observação 1, utilizando-se respectivamente as condicionais completas dadas em (2.5.9) a (2.5.13).

Observação 3: Para estimar a variância, por exemplo, de $\hat{E}(\alpha|D_{\text{obs}})$, utilizando uma única e longa cadeia de comprimento $T - T_0$, podemos formar n grupos de $k = \frac{(T-T_0)}{n}$ observações e calcular $\bar{\alpha}_1, \bar{\alpha}_2, \dots, \bar{\alpha}_n$, onde $\bar{\alpha}_i$ é a média amostral do i -ésimo grupo (k deve ser suficientemente grande de tal modo que a autocorrelação amostral obtida com base nos $\bar{\alpha}_i$ é bem pequena, digamos menor que 0,1). Nesse caso, um estimador de $\text{Var}(\hat{E}(\alpha|D_{\text{obs}}))$ é dado por

$$\frac{1}{n(n-1)} \sum_{i=1}^n (\bar{\alpha}_i - \bar{\alpha})^2, \text{ onde } \bar{\alpha} = \frac{1}{n} \sum_{i=1}^n \bar{\alpha}_i.$$

Outro estimador (proposto por Kass et al., 1998) é dado por

$$\widehat{\text{Var}}(\hat{\alpha}) = \frac{S_\alpha^2}{ESS},$$

onde $S_\alpha^2 = \sum_{i=T_0+1}^T \frac{(\alpha_i - \hat{\alpha})}{(T - T_0) - 1}$ e $ESS = \frac{N}{1 + 2 \sum_{k=1}^{\infty} \hat{\rho}_k(\alpha)}$, onde $\rho_k(\alpha)$ é a autocorrelação no lag k para o parâmetro α (que pode ser estimada usando as autocorrelações amostrais a partir dos dados gerados da cadeia. A série $\sum_{k=1}^{\infty} \hat{\rho}_k(\alpha)$ é aproximada desprezando-se todos os termos a partir do $k_0 + 1$ -ésimo termo, quando $|\rho_{k_0}(\alpha)| < 0,1$).

As variâncias das médias dos demais parâmetros do modelo podem ser estimadas da mesma forma.

2.10 Avaliação do ajuste e análise de sensibilidade

Na subseção 2.10.1 analisamos a sensibilidade na estimação da média dos parâmetros do modelo estrutural a diferentes valores do hiperparâmetro λ , onde $\lambda = \sigma_e^2 / \sigma_u^2$, e

na seção 2.10.2 avaliamos a qualidade do ajuste.

2.10.1 Sensibilidade do modelo a diferentes valores da razão das variâncias residuais

No modelo estrutural Normal com restrição de identificabilidade $\sigma_e^2 = \lambda\sigma_u^2$, a constante λ é tratada como hiperparâmetro conhecido, e é importante fazer uma avaliação da sensibilidade na estimação dos parâmetros quando diferentes valores de λ são fixados. Nesse trabalho limitamo-nos apenas a uma análise exploratória, construindo gráficos das médias amostrais de cada parâmetro em função de λ mantendo todos os demais parâmetros fixados (isto é, variando apenas o valor de λ , onde $\lambda > 0$) e usando amostras da posteriori obtidas por exemplo por algum método MCMC. Na seção 2.12 esse procedimento é ilustrado por meio de um exemplo com dados simulados. Um estudo mais completo sobre esse tópico deve ser conduzido futuramente.

2.10.2 Avaliação da qualidade do ajuste

Após ajustarmos o modelo aos dados devemos verificar se esse ajuste é adequado, isto é, se o modelo é compatível com os dados. A literatura bayesiana sobre o assunto é bem vasta e muitas das técnicas de diagnóstico utilizam a distribuição preditiva à posteriori de alguma forma, por exemplo, comparando os dados observados com os dados simulados pela priori preditiva (já que se o modelo está bem ajustado esses dois conjuntos de dados devem se assemelhar); um exemplo é o conjunto de técnicas denominado “posterior predictive checks” (Gelman et al., 1997, cap. 6). Outra possibilidade também é estudar os resíduos baseados na diferença entre os valores observados e os valores previstos a partir da posteriori preditiva. Para uma breve apresentação das técnicas de diagnóstico mais conhecidas, ver Carlin e Louis (2000), Chen et al. (2000), Gelman et al. (1997). Outras referências muito importantes são

Spiegelhalter et al. (2002) e o relatório técnico de Dey e Maiti (2002).

Nesse capítulo vamos nos restringir somente ao cálculo da densidade condicional preditiva de validação cruzada e dos resíduos de validação cruzada para o modelo estrutural Normal.

Observação: Convém lembrar entretanto que além da avaliação da qualidade do ajuste ou adequação do modelo (“internal checking”) onde são utilizadas medidas de diagnóstico para verificar se o modelo é compatível com os dados, isto é, se o modelo se ajusta adequadamente aos dados, é importante também avaliar o efeito nas inferências quando diferentes modelos alternativos são utilizados para o mesmo problema (“external checking”) como por exemplo modelos com diferentes prioris, diferentes hiperparâmetros, diferentes verossimilhanças, diferentes estruturas hierárquicas, etc., já que esses modelos podem estar igualmente bem ajustados mas nos levarem a diferentes conclusões.

(a) Densidade preditiva de validação cruzada

Quando o tamanho da amostra for relativamente pequeno, que não possamos nos dar ao luxo de reservar uma parte dos dados (digamos 20 a 30%) apenas para a validação do modelo (ou seja, usar somente 70 a 80% dos dados para o ajuste do modelo), então uma solução é utilizar a “densidade condicional preditiva de validação” cruzada, $\pi(Y_i|Y_{(-i)})$, $i = 1, \dots, n$, onde $\underline{Y} = (Y_1, \dots, Y_n)$ é a amostra dos dados e $Y_{(-i)}$ é a amostra resultante após a exclusão da observação Y_i .

Essa densidade é conhecida como “CPO_i” (“Conditional Predictive Ordinate”), e dá a “verossimilhança” de cada ponto Y_i condicionada ao restante das observações $Y_{(-i)}$,

$$\begin{aligned} \text{CPO}_i &= \pi(Y_i|Y_{(-i)}) = \int_{\Theta} \pi(Y_i, \theta|Y_{(-i)})d\theta \\ &= \int_{\Theta} \pi(Y_i|\theta, Y_{(-i)})\pi(\theta|Y_{(-i)})d\theta, \end{aligned}$$

onde θ é o vetor de parâmetros do modelo considerado.

Essa medida foi proposta inicialmente por Geisser (1980) e estudada posteriormente por Gelfand et al. (1992). O gráfico dos valores da CPO_i versus i , por exemplo, pode ser usado como diagnóstico para detectar “outliers”, já que um valor “alto” de CPO_i indica que a observação Y_i está bem ajustada, e um valor “baixo”, o contrário.

No modelo estrutural, a amostra das observações é $D_{\text{obs}} = (\underline{U}_1, \dots, \underline{U}_n)$, onde $\underline{U}_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix}$ é a i -ésima observação e $\underline{U}_{(-i)} = D_{\text{obs}(-i)}$ é a amostra após a extração de \underline{U}_i .

A seguir apresentamos três fórmulas alternativas de estimar a CPO_i para o modelo estrutural Normal, usando o fato que

$$\begin{aligned} CPO_i &= \pi(\underline{U}_i | D_{\text{obs}(-i)}) = \int_{\Theta} \pi(\underline{U}_i, \theta | D_{\text{obs}(-i)}) d\theta \\ &= \int_{\mathcal{X} = \mathbb{R}^n} \int_{\Theta} \pi(\underline{U}_i, \theta, \underline{x} | D_{\text{obs}(-i)}) d\theta d\underline{x} \\ &= \int_{\mathcal{X}_i = \mathbb{R}} \int_{\Theta} \pi(\underline{U}_i, \theta, x_i | D_{\text{obs}(-i)}) d\theta dx_i. \end{aligned}$$

Cálculo da CPO_i

As fórmulas a seguir referem-se ao modelo sem restrições de identificabilidade. No caso do modelo com a restrição $\sigma_e^2 = \lambda\sigma_u^2 = \lambda\sigma^2$, basta substituir nessas fórmulas essa restrição.

(1) Fórmula 1

$$\begin{aligned} CPO_i = \pi(\underline{U}_i | D_{\text{obs}(-i)}) &= \int_{\Theta} \pi(\underline{U}_i | \theta) \pi(\theta | D_{\text{obs}(-i)}) d\theta \\ &= \left[\int_{\Theta} \frac{1}{\pi(\underline{U}_i | \theta)} \pi(\theta | D_{\text{obs}}) d\theta \right]^{-1}, i = 1, \dots, n, (2.10.1) \end{aligned}$$

onde $\underline{U}_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix}$ e $\pi(\theta | D_{\text{obs}})$ é a posteriori de θ .

$$\underline{U}_i | \theta \stackrel{\text{iid}}{\sim} N \left(\begin{pmatrix} \alpha + \beta\mu \\ \mu \end{pmatrix}; \begin{pmatrix} \beta^2\sigma_x^2 + \sigma_e^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma_u^2 \end{pmatrix} \right),$$

isto é,

$$\begin{aligned} \pi(\underline{U}_i|\vartheta) = & (2\pi|\Sigma|)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2|\Sigma|}\left[(\beta^2\sigma_x^2 + \sigma_e^2)(X_i - \mu)^2 \right. \right. \\ & \left. \left. - 2\beta\sigma_x^2(Y_i - \alpha - \beta\mu)(X_i - \mu) + (\sigma_x^2 + \sigma_u^2)(Y_i - \alpha - \beta\mu)^2\right]\right\}, \end{aligned} \quad (2.10.2)$$

onde $|\Sigma| = \beta^2\sigma_u^2\sigma_x^2 + \sigma_e^2\sigma_x^2 + \sigma_e^2\sigma_u^2$.

$$\begin{aligned} \text{Demonstração: } \pi(\underline{U}_i|D_{\text{obs}(-i)}) &= \int_{\Theta} \pi(\underline{U}_i, \vartheta|D_{\text{obs}(-i)})d\vartheta \\ &= \int_{\Theta} \pi(\underline{U}_i|\vartheta, D_{\text{obs}(-i)})\pi(\vartheta|D_{\text{obs}(-i)})d\vartheta = \int_{\Theta} \pi(\underline{U}_i|\vartheta)\pi(\vartheta|D_{\text{obs}(-i)})d\vartheta. \end{aligned}$$

A última igualdade segue pelo fato de que dado ϑ , então \underline{U}_i e $D_{\text{obs}(-i)}$ são independentes. Para completar a demonstração, usamos a seguinte igualdade:

$$\pi(\vartheta|D_{\text{obs}(-i)}) = \frac{\pi(\vartheta|D_{\text{obs}})}{\pi(\underline{U}_i|\vartheta)} \cdot \left[\int_{\Theta} \frac{\pi(\vartheta|D_{\text{obs}})}{\pi(\underline{U}_i|\vartheta)} d\vartheta \right]^{-1} \quad (2.10.3)$$

Portanto,

$$\begin{aligned} \text{CPO}_i &= \left[\int_{\Theta} \frac{\pi(\vartheta|D_{\text{obs}})}{\pi(\underline{U}_i|\vartheta)} d\vartheta \right]^{-1} \int_{\Theta} \pi(\vartheta|D_{\text{obs}})d\vartheta \\ &= \left[\int_{\Theta} \frac{1}{\pi(\underline{U}_i|\vartheta)} \pi(\vartheta|D_{\text{obs}})d\vartheta \right]^{-1}. \end{aligned}$$

A igualdade em (2.10.3) sai da seguinte relação:

$$\pi(\vartheta|D_{\text{obs}}) = \frac{\pi(\vartheta, \underline{U}_i, \underline{U}_{(-i)})}{\pi(D_{\text{obs}})} \propto \pi(\underline{U}_i|\vartheta, \underline{U}_{(-i)})\pi(\vartheta|\underline{U}_{(-i)}),$$

onde $\pi(\underline{U}_i|\vartheta, \underline{U}_{(-i)}) = \pi(\underline{U}_i|\vartheta)$ e portanto

$$\pi(\vartheta|D_{\text{obs}(-i)}) \propto \frac{\pi(\vartheta|D_{\text{obs}})}{\pi(\underline{U}_i|\vartheta)}.$$

Assim,

$$\pi(\vartheta|D_{\text{obs}(-i)}) = \frac{\pi(\vartheta|D_{\text{obs}})}{\pi(\underline{U}_i|\vartheta)} \Big/ \int_{\Theta} \frac{\pi(\vartheta|D_{\text{obs}})}{\pi(\underline{U}_i|\vartheta)} d\vartheta.$$

Observação: É importante ressaltar que o truque usado acima permite que o cálculo da CPO_i para todo $i = 1, \dots, n$ seja feito usando a mesma posteriori $\pi(\vartheta|D_{\text{obs}})$.

Estimativa de CPO_i

Dada uma amostra $(\vartheta^{(1)}, \dots, \vartheta^{(G)})$ da posteriori $p(\vartheta|D_{\text{obs}})$ do modelo estrutural Normal, obtida por exemplo por algum dos métodos MCMC descritos no Capítulo 2, então uma estimativa de Monte Carlo de CPO_i é dada por

$$\widehat{\text{CPO}}_i = \left[\frac{1}{G} \sum_{g=1}^G \frac{1}{\pi(\underline{U}_i|\vartheta^{(g)})} \right]^{-1}, \quad (2.10.4)$$

onde $\pi(\underline{U}_i|\vartheta)$ é dada em (2.10.2).

Observação: Computacionalmente esse método é muito eficiente pois a amostragem é feita de uma única posteriori. Caso contrário deveríamos obter para cada i uma amostra $(\vartheta_{(i)}^{(1)}, \dots, \vartheta_{(i)}^{(G)})$, da posteriori correspondente $p(\vartheta|D_{\text{obs}(-i)})$ e uma estimativa de Monte Carlo de CPO_i seria

$$\widehat{\text{CPO}}_i = \frac{1}{G} \sum_{g=1}^G \pi(\underline{U}_i|\vartheta_{(i)}^{(g)}), \quad i = 1, \dots, n.$$

O problema agora é que isso traria um alto “custo computacional”, especialmente se n é muito grande (já que amostramos G elementos de cada uma das n posteriores distintas $p(\vartheta|D_{\text{obs}(-1)}), \dots, p(\vartheta|D_{\text{obs}(-n)})$ para obter $\widehat{\text{CPO}}_1, \dots, \widehat{\text{CPO}}_n$).

(2) Fórmula 2

$$\text{CPO}_i = \pi(\underline{U}_i|D_{\text{obs}}) = \left[\int_{\mathbf{x}=\mathbb{R}^n} \int_{\Theta} \frac{1}{\pi(\underline{U}_i|x_i, \vartheta)} \pi(\vartheta, \underline{x}|D_{\text{obs}}) d\vartheta d\underline{x} \right]^{-1}, \quad i = 1, \dots, n, \quad (2.10.5)$$

onde $\underline{x} = (x_1, \dots, x_n)^T$, $\pi(\vartheta, \underline{x}|D_{\text{obs}})$ é a distribuição conjunta de \underline{x} e ϑ à posteriori e

$$\underline{U}_i|x_i, \vartheta \stackrel{\text{iid}}{\sim} N \left(\begin{pmatrix} \alpha + \beta x_i \\ x_i \end{pmatrix}; \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix} \right),$$

isto é,

$$\pi(\underline{U}_i|x_i, \vartheta) = (2\pi\sigma_e^2\sigma_u^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[\frac{(Y_i - \alpha - \beta x_i)^2}{\sigma_e^2} + \frac{(X_i - x_i)^2}{\sigma_u^2} \right] \right\} \quad (2.10.6)$$

Demonstração:

$$\begin{aligned}
\text{CPO}_i &= \pi(\underline{U}|D_{\text{obs}(-i)}) = \int_{\underline{x}} \int_{\Theta} \pi(\underline{U}_i, \underline{\theta}, \underline{x}|D_{\text{obs}(-i)}) d\underline{\theta} d\underline{x} \\
&= \int_{\underline{x}} \int_{\Theta} \pi(\underline{U}_i|\underline{\theta}, \underline{x}, D_{\text{obs}(-i)}) \pi(\underline{\theta}, \underline{x}|D_{\text{obs}(-i)}) d\underline{\theta} d\underline{x} \\
&= \int_{\underline{x}} \int_{\Theta} \pi(\underline{U}_i|x_i, \underline{\theta}) \pi(\underline{\theta}, \underline{x}|D_{\text{obs}(-i)}) d\underline{\theta} d\underline{x},
\end{aligned}$$

onde

$$\pi(\underline{\theta}, \underline{x}|D_{\text{obs}(-i)}) = \frac{\pi(\underline{\theta}, \underline{x}|D_{\text{obs}})}{\pi(\underline{U}_i|x_i, \underline{\theta})} \bigg/ \int_{\underline{x}} \int_{\Theta} \frac{\pi(\underline{\theta}, \underline{x}|D_{\text{obs}})}{\pi(\underline{U}_i|x_i, \underline{\theta})} d\underline{\theta} d\underline{x} \quad (2.10.7)$$

e portanto

$$\begin{aligned}
\text{CPO}_i &= \left[\int_{\underline{x}} \int_{\Theta} \frac{\pi(\underline{\theta}, \underline{x}|D_{\text{obs}})}{\pi(\underline{U}_i|x_i, \underline{\theta})} d\underline{\theta} d\underline{x} \right]^{-1} \int_{\underline{x}} \int_{\Theta} \pi(\underline{\theta}, \underline{x}|D_{\text{obs}}) d\underline{\theta} d\underline{x} \\
&= \left[\int_{\underline{x}} \int_{\Theta} \frac{\pi(\underline{\theta}, \underline{x}|D_{\text{obs}})}{\pi(\underline{U}_i|x_i, \underline{\theta})} d\underline{\theta} d\underline{x} \right]^{-1}.
\end{aligned}$$

A igualdade em (2.10.7) é obtida a partir da relação

$$\pi(\underline{\theta}, \underline{x}|D_{\text{obs}}) \propto \pi(\underline{\theta}, \underline{x}, U_i, D_{\text{obs}(-i)}) = \pi(\underline{U}_i|x_i, \underline{\theta}) \pi(\underline{x}, \underline{\theta}|D_{\text{obs}(-i)})$$

e portanto

$$\begin{aligned}
\pi(\underline{\theta}, \underline{x}|D_{\text{obs}(-i)}) &\propto \frac{\pi(\underline{\theta}, \underline{x}|D_{\text{obs}})}{\pi(\underline{U}_i|x_i, \underline{\theta})} \\
&= \frac{\pi(\underline{\theta}, \underline{x}|D_{\text{obs}})}{\pi(\underline{U}_i|x_i, \underline{\theta})} \bigg/ \int_{\underline{x}} \int_{\Theta} \frac{\pi(\underline{\theta}, \underline{x}|D_{\text{obs}})}{\pi(\underline{U}_i|x_i, \underline{\theta})} d\underline{\theta} d\underline{x}.
\end{aligned}$$

Estimativa de CPO_i

Dada uma amostra $(\underline{\theta}^{(1)}, \underline{x}^{(1)}, \dots, \underline{\theta}^{(G)}, \underline{x}^{(G)})$ da distribuição $p(\underline{\theta}, \underline{x}|D_{\text{obs}})$, obtida por exemplo pelo algoritmo de Gibbs (desde que ele convirja), então uma estimativa de Monte Carlo da CPO_i é dada por

$$\widehat{\text{CPO}}_i = \left[\frac{1}{G} \sum_{g=1}^G \frac{1}{\pi(\underline{U}_i|x_i^{(g)}, \underline{\theta}^{(g)})} \right]^{-1}, \quad (2.10.8)$$

onde $\pi(\underline{U}_i|x_i, \underline{\theta})$ é dada em (2.10.6) e onde $x_i^{(g)}$ é a i -ésima componente do vetor gerado $\underline{x}^{(g)}$.

(3) Fórmula 3

$$\begin{aligned} \text{CPO}_i &= \pi(\underline{U}_i|D_{\text{obs}(-i)}) \\ &= \left[\int_{\underline{x}_i=\mathbb{R}} \int_{\Theta} \frac{1}{\pi(\underline{U}_i|x_i, \underline{\theta})} \pi(x_i|\underline{\theta}) \pi(\underline{\theta}|D_{\text{obs}}) d\underline{\theta} dx_i \right]^{-1}, \end{aligned} \quad (2.10.9)$$

$i = 1, \dots, n$, onde $x_i|\underline{\theta} \sim N(\mu, \sigma_x^2)$, $\pi(\underline{\theta}|D_{\text{obs}})$ é a posteriori de $\underline{\theta}$ e $\pi(\underline{U}_i|x_i, \underline{\theta})$ é dada em (2.10.6).

Demonstração

$$\begin{aligned} \pi(\underline{U}_i|D_{\text{obs}(-i)}) &= \int_{\underline{x}_i} \int_{\Theta} \pi(\underline{U}_i, x_i, \underline{\theta}|D_{\text{obs}(-i)}) d\underline{\theta} dx_i \\ &= \int_{\underline{x}_i} \int_{\Theta} \pi(\underline{U}_i|x_i, \underline{\theta}, D_{\text{obs}(-i)}) \pi(x_i, \underline{\theta}|D_{\text{obs}(-i)}) d\underline{\theta} dx_i, \end{aligned}$$

onde $\pi(\underline{U}_i|x_i, \underline{\theta}, D_{\text{obs}(-i)}) = \pi(\underline{U}_i|x_i, \underline{\theta})$ e

$$\pi(x_i, \underline{\theta}|D_{\text{obs}(-i)}) = \frac{\pi(x_i, \underline{\theta}|D_{\text{obs}})}{\pi(\underline{U}_i|x_i, \underline{\theta})} \bigg/ \int_{\underline{x}_i} \int_{\Theta} \frac{\pi(\underline{\theta}, \underline{x}|D_{\text{obs}})}{\pi(\underline{U}_i|x_i, \underline{\theta})} d\underline{\theta} dx_i \quad (2.10.10)$$

e portanto

$$\begin{aligned} \text{CPO}_i &= \int_{\underline{x}_i} \int_{\Theta} \pi(\underline{U}_i|x_i, \underline{\theta}) \frac{\pi(x_i, \underline{\theta}|D_{\text{obs}})}{\pi(\underline{U}_i|x_i, \underline{\theta})} \left[\int_{\underline{x}_i} \int_{\Theta} \frac{\pi(x_i, \underline{\theta}|D_{\text{obs}})}{\pi(\underline{U}_i|x_i, \underline{\theta})} d\underline{\theta} dx_i \right]^{-1} d\underline{\theta} dx_i \\ &= \left[\int_{\underline{x}_i} \int_{\Theta} \frac{\pi(x_i, \underline{\theta}|D_{\text{obs}})}{\pi(\underline{U}_i|x_i, \underline{\theta})} d\underline{\theta} dx_i \right]^{-1} \\ &= \left[\int_{\underline{x}_i} \int_{\Theta} \frac{1}{\pi(\underline{U}_i|x_i, \underline{\theta})} \pi(x_i|\underline{\theta}) \pi(\underline{\theta}|D_{\text{obs}}) d\underline{\theta} dx_i \right]^{-1}. \end{aligned}$$

Estimativa de CPO_i

Gerada uma amostra $(\underline{\theta}^{(1)}, \dots, \underline{\theta}^{(g)})$ da posteriori de $\pi(\underline{\theta}|D_{\text{obs}})$ e gerado $x_i^{(g)}$ de $\pi(x_i|\underline{\theta}^{(g)})$, onde $x_i|\underline{\theta}^{(g)} \sim N(\mu^{(g)}, \sigma_x^{2(g)})$, então uma estimativa de Monte Carlo de

CPO_{*i*} é dada por

$$\widehat{\text{CPO}}_i = \left(\frac{1}{G} \sum_{g=1}^G \frac{1}{\pi(\underline{U}_i | x_i^{(g)}, \underline{\varrho}^{(g)})} \right)^{-1}. \quad (2.10.11)$$

Notemos que, embora a fórmula (2.10.11) coincida com a fórmula (2.10.8), a diferença está basicamente no processo de geração de $x_i^{(g)}$; além de termos, em princípio, a opção de usar também outros algoritmos (além do Gibbs) para amostrar $\underline{\varrho}^{(g)}$ de $\pi(\underline{\varrho} | D_{\text{obs}})$.

Observação: Podemos plotar $\widehat{\text{CPO}}_i$ versus i ou versus X_i (ou mesmo versus Y_i) para analisar o ajuste do modelo.

(b) Resíduos de validação cruzada

Se $\underline{Y} = (Y_1, \dots, Y_n)$ é a amostra de observações e $\underline{Y}_{(-i)}$ é a amostra após extrair Y_i , então o resíduo estimado de “validação cruzada” é a diferença entre o valor observado Y_i e o valor ajustado ou previsto pelo modelo condicional aos dados $\underline{Y}_{(-i)}$. Gelfand et al. (1992) definiram este resíduo como $r_i = Y_i - E(Y_i | Y_{(-i)})$, $i = 1, \dots, n$, onde $E(Y_i | Y_{(-i)})$ é a média com relação à distribuição preditiva de validação cruzada $\pi(Y_i | Y_{(-i)})$. O resíduo padronizado então é definido por

$$r'_i = \frac{Y_i - E(Y_i | Y_{(-i)})}{\sqrt{\text{Var}(Y_i | Y_{(-i)})}}, \quad i = 1, \dots, n.$$

Nessa seção, o objetivo é obter os resíduos de validação cruzada padronizados do modelo estrutural Normal.

No modelo estrutural Normal os resíduos bidimensionais são

$$\begin{pmatrix} e_i \\ u_i \end{pmatrix} \stackrel{\text{iid}}{\sim} N_2(\underline{0}, V) \quad \text{onde } V = \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix}, \quad i = 1, \dots, n.$$

Estes resíduos podem ser escritos como

$$\begin{pmatrix} e_i \\ u_i \end{pmatrix} = \begin{pmatrix} Y_i \\ X_i \end{pmatrix} - \begin{pmatrix} \alpha + \beta x_i \\ x_i \end{pmatrix} = \begin{pmatrix} Y_i \\ X_i \end{pmatrix} - E \left[\begin{pmatrix} Y_i \\ X_i \end{pmatrix} \middle| \underline{\varrho}, x_i \right]$$

e portanto, os resíduos padronizados são

$$V^{-\frac{1}{2}} \begin{pmatrix} e_i \\ u_i \end{pmatrix} = \text{Var}^{-\frac{1}{2}} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} \middle| \varrho, x_i \right) \left[\begin{pmatrix} Y_i \\ X_i \end{pmatrix} - E \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} \middle| \varrho, x_i \right) \right] \stackrel{\text{iid}}{\sim} N_2(\underline{0}, I_{2 \times 2}),$$

$i = 1, \dots, n$.

Fazendo $\underline{U}_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix}$, os resíduos de validação cruzada padronizados são definidos por

$$\underline{d}_i = \text{Var}^{-\frac{1}{2}}(\underline{U}_i | D_{\text{obs}(-i)}) [\underline{U}_i - E(\underline{U}_i | D_{\text{obs}(-i)})], \quad i = 1, \dots, n,$$

onde $D_{\text{obs}(-i)} = \underline{U}_{(-i)} = (\underline{U}_1, \dots, \underline{U}_{i-1}, \underline{U}_{i+1}, \dots, \underline{U}_n)$ e onde $E(\underline{U}_i | D_{\text{obs}(-i)})$ e $\text{Var}(\underline{U}_i | D_{\text{obs}(-i)})$ são calculados a seguir.

• Cálculo de $E(\underline{U}_i | D_{\text{obs}(-i)})$

$$\begin{aligned} E(\underline{U}_i | D_{\text{obs}(-i)}) &= E[E(\underline{U}_i | \varrho, \underline{x}) | D_{\text{obs}(-i)}] \\ &= \int_{\underline{x}} \int_{\Theta} E(\underline{U}_i | \varrho, \underline{x}) \pi(\varrho, \underline{x} | D_{\text{obs}(-i)}) d\varrho d\underline{x}, \end{aligned}$$

onde

$$\begin{aligned} \pi(\varrho, \underline{x} | D_{\text{obs}(-i)}) &= \frac{\pi(\varrho, \underline{x} | D_{\text{obs}})}{\pi(\underline{U}_i | \varrho, x_i)} \bigg/ \int_{\underline{x}} \int_{\Theta} \frac{\pi(\varrho, \underline{x} | D_{\text{obs}})}{\pi(\underline{U}_i | \varrho, x_i)} d\varrho d\underline{x} \\ &= \text{CPO}_i \frac{\pi(\varrho, \underline{x} | D_{\text{obs}})}{\pi(\underline{U}_i | \varrho, x)}, \end{aligned}$$

onde

$$\text{CPO}_i = \left(\int_{\underline{x}} \int_{\Theta} \frac{1}{\pi(\underline{U}_i | \varrho, x_i)} \pi(\varrho, \underline{x} | D_{\text{obs}}) d\varrho d\underline{x} \right)^{-1}$$

(como em (2.10.7)). Portanto,

$$E(\underline{U}_i | D_{\text{obs}(-i)}) = \text{CPO}_i \int_{\underline{x}} \int_{\Theta} \frac{E(\underline{U}_i | \varrho, \underline{x})}{\pi(\underline{U}_i | \varrho, x_i)} \pi(\varrho, \underline{x} | D_{\text{obs}}) d\varrho d\underline{x}, \quad (2.10.12)$$

onde

$$E(\underline{U}_i | \varrho, \underline{x}) = E(\underline{U}_i | \varrho, x_i) = \begin{pmatrix} \alpha + 3x_i \\ x_i \end{pmatrix},$$

$\pi(\underline{U}_i|\underline{\theta}, x_i)$ é dada em (2.10.6) e $\pi(\underline{\theta}, \underline{x}|D_{\text{obs}})$ é a distribuição conjunta à posteriori de $\underline{\theta}$ e \underline{x} .

• **Estimação de $E(\underline{U}_i|D_{\text{obs}(-i)})$ por Monte Carlo**

Seja $(\underline{\theta}^{(1)}, \underline{x}^{(1)}, \dots, \underline{\theta}^{(G)}, \underline{x}^{(G)})$ uma amostra da distribuição à posteriori $p(\underline{\theta}, \underline{x}|D_{\text{obs}})$, obtida por exemplo pelo algoritmo de Gibbs.

Uma estimativa de Monte Carlo de $E(\underline{U}_i|D_{\text{obs}(-i)})$ é dada por

$$\hat{E}(\underline{U}_i|D_{\text{obs}(-i)}) = \widehat{\text{CPO}}_i \frac{1}{G} \sum_{g=1}^G \frac{1}{\pi(\underline{U}_i|\underline{\theta}^{(g)}, x_i^{(g)})} \begin{pmatrix} \alpha^{(g)} + \beta^{(g)} x_i^{(g)} \\ x_i^{(g)} \end{pmatrix} \quad (2.10.13)$$

onde

$$\widehat{\text{CPO}}_i = \left(\frac{1}{G} \sum_{g=1}^G \frac{1}{\pi(\underline{U}_i|\underline{\theta}^{(g)}, x_i^{(g)})} \right)^{-1}$$

e $\pi(\underline{U}_i|\underline{\theta}, x_i)$ é dada em (2.10.6).

• **Cálculo da $\text{Var}(\underline{U}_i|D_{\text{obs}(-i)})$**

$$\begin{aligned} \text{Var}(\underline{U}_i|D_{\text{obs}(-i)}) &= E[\text{Var}(\underline{U}_i|\underline{\theta}, \underline{x})|D_{\text{obs}(-i)}] + \text{Var}[E(\underline{U}_i|\underline{\theta}, \underline{x})|D_{\text{obs}(-i)}] \\ &= E_{\underline{\theta}, \underline{x}|D_{\text{obs}(-i)}}[\text{Var}(\underline{U}_i|\underline{\theta}, \underline{x})] + E_{\underline{\theta}, \underline{x}|D_{\text{obs}(-i)}}([E(\underline{U}_i|\underline{\theta}, \underline{x})][E(\underline{U}_i|\underline{\theta}, \underline{x})]^\top) \\ &\quad - (E_{\underline{\theta}, \underline{x}|D_{\text{obs}(-i)}}[E(\underline{U}_i|\underline{\theta}, \underline{x})])(E_{\underline{\theta}, \underline{x}|D_{\text{obs}(-i)}}[E(\underline{U}_i|\underline{\theta}, \underline{x})])^\top \\ &= \int_{\underline{x}} \int_{\Theta} \text{Var}(\underline{U}_i|\underline{\theta}, \underline{x}) \pi(\underline{\theta}, \underline{x}|D_{\text{obs}(-i)}) d\underline{\theta} d\underline{x} \\ &\quad + \int_{\underline{x}} \int_{\Theta} [E(\underline{U}_i|\underline{\theta}, \underline{x})][E(\underline{U}_i|\underline{\theta}, \underline{x})]^\top \pi(\underline{\theta}, \underline{x}|D_{\text{obs}(-i)}) d\underline{\theta} d\underline{x} \\ &\quad - \left[\int_{\underline{x}} \int_{\Theta} E(\underline{U}_i|\underline{\theta}, \underline{x}) \pi(\underline{\theta}, \underline{x}|D_{\text{obs}(-i)}) d\underline{\theta} d\underline{x} \right] \\ &\quad \cdot \left[\int_{\underline{x}} \int_{\Theta} E(\underline{U}_i|\underline{\theta}, \underline{x}) \pi(\underline{\theta}, \underline{x}|D_{\text{obs}(-i)}) d\underline{\theta} d\underline{x} \right]^\top. \end{aligned}$$

Mas

$$\pi(\underline{\theta}, \underline{x}|D_{\text{obs}(-i)}) = \text{CPO}_i \frac{\pi(\underline{\theta}, \underline{x}|D_{\text{obs}})}{\pi(\underline{U}_i|\underline{\theta}, x_i)},$$

onde

$$\text{CPO}_i = \left(\int_{\underline{x}} \int_{\Theta} \frac{1}{\pi(\underline{U}_i|\underline{\theta}, x_i)} \pi(\underline{\theta}, \underline{x}|D_{\text{obs}}) d\underline{\theta} d\underline{x} \right)^{-1}.$$

Portanto,

$$\begin{aligned}
\text{Var}(\underline{U}_i | D_{\text{obs}(-i)}) &= \text{CPO}_i \left[\int_{\underline{x}} \int_{\Theta} \frac{\text{Var}(\underline{U}_i | \underline{\theta}, \underline{x})}{\pi(\underline{U}_i | \underline{\theta}, x_i)} \pi(\underline{\theta}, \underline{x} | D_{\text{obs}}) d\underline{\theta} d\underline{x} \right. \\
&+ \left. \int_{\underline{x}} \int_{\Theta} \frac{[E(\underline{U}_i | \underline{\theta}, \underline{x})][E(\underline{U}_i | \underline{\theta}, \underline{x})]^\top}{\pi(\underline{U}_i | \underline{\theta}, x_i)} \pi(\underline{\theta}, \underline{x} | D_{\text{obs}}) d\underline{\theta} d\underline{x} \right] \\
&- \left[\text{CPO}_i \int_{\underline{x}} \int_{\Theta} \frac{E(\underline{U}_i | \underline{\theta}, \underline{x})}{\pi(\underline{U}_i | \underline{\theta}, x_i)} \pi(\underline{\theta}, \underline{x} | D_{\text{obs}}) d\underline{\theta} d\underline{x} \right] \\
&\cdot \left[\text{CPO}_i \int_{\underline{x}} \int_{\Theta} \frac{E(\underline{U}_i | \underline{\theta}, \underline{x})}{\pi(\underline{U}_i | \underline{\theta}, x_i)} \pi(\underline{\theta}, \underline{x} | D_{\text{obs}}) d\underline{\theta} d\underline{x} \right]^\top, \tag{2.10.14}
\end{aligned}$$

onde

$$\begin{aligned}
\text{Var}(\underline{U}_i | \underline{\theta}, \underline{x}) &= \text{Var}(\underline{U}_i | \underline{\theta}, x_i) = \begin{bmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_u^2 \end{bmatrix}, \\
E(\underline{U}_i | \underline{\theta}, \underline{x}) &= \begin{pmatrix} \alpha + \beta x_i \\ x_i \end{pmatrix}
\end{aligned}$$

e $\pi(\underline{\theta}, \underline{x} | D_{\text{obs}})$ é a posteriori conjunta de $\underline{\theta}$ e \underline{x} .

• **Estimação de $\text{Var}(\underline{U}_i | D_{\text{obs}(-i)})$**

Dada uma amostra $(\underline{\theta}^{(1)}, \underline{x}^{(1)}, \dots, \underline{\theta}^{(G)}, \underline{x}^{(G)})$ da posteriori conjunta $p(\underline{x}, \underline{\theta} | D_{\text{obs}})$, gerada por exemplo pelo algoritmo de Gibbs, então uma estimativa de Monte Carlo dessa matriz de covariâncias é dada por

$$\begin{aligned}
\widehat{\text{Var}}(\underline{U}_i | D_{\text{obs}(-i)}) &= \widehat{\text{CPO}}_i \frac{1}{G} \sum_{g=1}^G \frac{1}{\pi(\underline{U}_i | \underline{\theta}^{(g)}, x_i^{(g)})} \left\{ \text{Var}(\underline{U}_i | \underline{\theta}^{(g)}, x_i^{(g)}) \right. \\
&+ \left. [E(\underline{U}_i | \underline{\theta}^{(g)}, x_i^{(g)})][E(\underline{U}_i | \underline{\theta}^{(g)}, x_i^{(g)})]^\top \right\} \\
&- \widehat{\text{CPO}}_i^2 \left(\frac{1}{G} \sum_{g=1}^G \frac{E(\underline{U}_i | \underline{\theta}^{(g)}, x_i^{(g)})}{\pi(\underline{U}_i | \underline{\theta}^{(g)}, x_i^{(g)})} \right) \left(\frac{1}{G} \sum_{g=1}^G \frac{E(\underline{U}_i | \underline{\theta}^{(g)}, x_i^{(g)})}{\pi(\underline{U}_i | \underline{\theta}^{(g)}, x_i^{(g)})} \right)^\top \tag{2.10.15}
\end{aligned}$$

onde

$$\widehat{\text{CPO}}_i = \left(\frac{1}{G} \sum_{g=1}^G \frac{1}{\pi(\underline{U}_i | \underline{\theta}^{(g)}, x_i^{(g)})} \right)^{-1}$$

e $\pi(\underline{U}_i | \underline{\theta}^{(g)}, x_i^{(g)})$ é dada em (2.10.6).

Observação 1: Uma maneira alternativa de estimar $E(\underline{U}_i|D_{\text{obs}(-i)})$ e

$\text{Var}(\underline{U}_i|D_{\text{obs}(-i)})$ é a seguinte:

- (1) obter uma amostra $(\vartheta^{(1)}, \dots, \vartheta^{(G)})$ da posteriori $p(\vartheta|D_{\text{obs}})$ (usando qualquer dos métodos MCMC deste capítulo),
- (2) gerar $x_i^{(g)}$ de $\pi(x_i|\vartheta^{(g)})$ (onde $x_i|\vartheta \sim N(\mu, \sigma_x^2)$),
- (3) usar a expressão em (2.10.13) para estimar $E(\underline{U}_i|D_{\text{obs}(-i)})$ e a expressão em (2.10.15) para estimar $\text{Var}(\underline{U}_i|D_{\text{obs}(-i)})$.

Esse estimador é útil quando não conseguimos amostrar de $p(\underline{x}, \vartheta|D_{\text{obs}})$ usando, por exemplo, o algoritmo de Gibbs, mas conseguimos amostrar da posteriori $p(\vartheta|D_{\text{obs}})$ por meio do algoritmo de Metropolis-Hastings em Gibbs por exemplo.

Observação 2: Na análise de resíduos do modelo linear estrutural Normal, nosso objetivo é checar a suposição de que

$$\begin{pmatrix} \frac{e_i}{\sqrt{\sigma_e^2}} \\ \frac{u_i}{\sqrt{\sigma_u^2}} \end{pmatrix} \stackrel{\text{iid}}{\sim} N_2(\underline{0}, I_{2 \times 2}), \quad i = 1, \dots, n,$$

(isto é, $\frac{e_i}{\sqrt{\sigma_e^2}} \stackrel{\text{iid}}{\sim} N(0, 1)$, $\frac{u_i}{\sqrt{\sigma_u^2}} \stackrel{\text{iid}}{\sim} N(0, 1)$, onde e_i e u_i são independentes para $i = 1, \dots, n$), como também a linearidade do modelo. Para essa análise, vamos usar os vetores de resíduos de validação cruzada padronizados

$$\hat{d}_i = \begin{pmatrix} \hat{d}_{i1} \\ \hat{d}_{i2} \end{pmatrix} = \widehat{\text{Var}}^{-\frac{1}{2}}(\underline{U}_i|D_{\text{obs}(-i)})(\underline{U}_i - \hat{E}(\underline{U}_i|D_{\text{obs}(-i)})),$$

lembrando que os \hat{d}_i 's não são mais exatamente normalmente distribuídos e independentes, mesmo quando os resíduos $\begin{pmatrix} e_i \\ u_i \end{pmatrix}$ são Normais e independentes.

Podemos analisar esses resíduos \hat{d}_i , $i = 1, \dots, n$, descritivamente por meio de gráficos, por exemplo, plotando \hat{d}_{i1} versus Y_i (ou versus o valor ajustado $\hat{E}(Y_i|Y_{(-i)})$) e \hat{d}_{i2} versus X_i (ou $\hat{E}(X_i|X_{(-i)})$), para $i = 1, \dots, n$. Esses gráficos nos ajudam a

checar possíveis falhas na homogeneidade das variâncias e na linearidade (bem como a existência de “outliers”).

Podemos plotar também \hat{d}_{i1} versus i (e \hat{d}_{i2} versus i), para $i = 1, \dots, n$, para checar uma possível falha na independência entre os \hat{d}_i 's (bem como a existência de “outliers”).

Para checar a normalidade da primeira componente do resíduo, podemos plotar os resíduos ordenados $\hat{d}_{(1),1} \leq \hat{d}_{(2),1} \leq \dots \leq \hat{d}_{(n),1}$ versus $\Phi^{-1}(F_1), \Phi^{-1}(F_2), \dots, \Phi^{-1}(F_n)$, respectivamente, onde F_i é a proporção de observações menores ou iguais a $\hat{d}_{(i),1}$ (e portanto $\hat{d}_{(i),1}$ é o valor que deixa uma área igual a F_i à sua esquerda na distribuição empírica e portanto se não houver empates $F_i = i/n$) e $\Phi^{-1}(F_i)$ é o valor do resíduo que deixa uma área igual a F_i à sua esquerda na distribuição $N(0, 1)$. Portanto, se os resíduos tiverem de fato uma distribuição aproximadamente Normal, espera-se que os pontos $(\hat{d}_{(i),1}, \Phi^{-1}(F_i))$, para $i = 1, \dots, n$, estejam mais ou menos próximos da reta $y = x$. O mesmo procedimento pode ser usado para a segunda componente dos resíduos.

Para checar a normalidade bivariada podemos usar o dispositivo gráfico descrito em Johnson e Wichern (1998), páginas 195 e 196. O gráfico é construído da seguinte maneira:

- (1) Calcular as distâncias $d_i^* = (\hat{d}_i - \bar{\hat{d}})^T S^{-1} (\hat{d}_i - \bar{\hat{d}})$, $i = 1, \dots, n$, onde $\bar{\hat{d}} = \frac{1}{n} \sum_{i=1}^n \hat{d}_i$ e S é a matriz de covariâncias amostrais obtida da amostra $(\hat{d}_1, \dots, \hat{d}_n)$;
- (2) Ordenar os valores de d_i^* : $d_{(1)}^* \leq d_{(2)}^* \leq \dots \leq d_{(n)}^*$;
- (3) Plotar os pontos $(d_{(i)}^*, q_{c,2}(F_i))$, $i = 1, \dots, n$, onde $F_i = \frac{i-1/2}{n} \cong$ proporção de observações $\leq d_{(i)}^*$ (assim $d_{(i)}^*$ é o valor cuja área à sua esquerda é igual a F_i , na distribuição empírica) e $q_{c,2}(F_i)$ é o valor cuja área à sua esquerda, numa

χ_2^2 , é igual à F_i .

Se de fato os \hat{d}_i são aproximadamente normais, então espera-se que os pontos $(d_{(i)}^*, q_{c,2}(F_i))$, $i = 1, \dots, n$, estejam próximos da reta $y = x$.

Observação: A idéia desse procedimento se baseia no fato de que se um vetor aleatório $\underline{X}_{p \times 1}$ tem distribuição $N_p(\underline{\mu}, \underline{\Sigma})$, com $|\underline{\Sigma}| > 0$, então

$$(\underline{X} - \underline{\mu})^T \underline{\Sigma}^{-1} (\underline{X} - \underline{\mu}) \sim \chi_p^2$$

2.11 Algumas medidas de comparação de modelos

Nessa seção o objetivo é apresentar alguns critérios de escolha de modelos aplicados aos modelos estruturais Normais com erros nas variáveis.

(1) O critério da medida L

A medida L é um critério que serve tanto para a comparação de modelos, como para a avaliação do ajuste e tem a vantagem de não requerer priori própria em geral (ver Ibrahim et al., 2000, e Ibrahim et al., 2001).

A seguir, descrevemos a formulação geral da medida L aplicada a um modelo qualquer, apresentando primeiramente o caso em que as variáveis observadas são unidimensionais e em seguida o caso multidimensional.

- Definição da medida L no caso unidimensional

Seja $\underline{Y} = (Y_1, \dots, Y_n)^T$ uma amostra aleatória de $\underline{Y}|\theta \sim f(\underline{Y}|\theta)$.

Seja $\underline{Y}^* = (Y_1^*, \dots, Y_n^*)^T = (Y_{f_1}, \dots, Y_{f_2})^T$, n observações futuras com a mesma distribuição dos dados observados \underline{Y} .

Então a medida L é definida como

$$L_2(\underline{Y}, \delta) = \sum_{i=1}^n \text{Var}(Y_i^* | \underline{Y}) + \delta \sum_{i=1}^n [E(Y_i^* | \underline{Y}) - Y_i]^2,$$

onde $E(Y_i^* | \underline{Y})$ e $\text{Var}(Y_i^* | \underline{Y})$ são, respectivamente, a esperança e a variância com relação à distribuição posteriori preditiva $\pi(Y_i^* | \underline{Y})$ e $0 < \delta < 1$, onde δ pode ser interpretado como um peso do viés quadrático e Ibrahim et al. (1998) estudaram o problema de encontrar um valor ótimo para δ e baseados nesse estudo concluíram que $\nu = \frac{1}{2}$ é um valor razoável.

Interpretação: Um “bom” modelo deveria ter uma pequena $\text{Var}(Y_i^* | \underline{Y})$ e um viés $|Y_i - E(Y_i^* | \underline{Y})|$ também pequeno, isto é, um valor pequeno de $L_2(\underline{Y}, \delta)$ e portanto entre dois modelos competitivos, o “melhor” será o que tiver a menor medida L .

- Definição da medida L para variáveis multidimensionais

Seja $\underline{Y} = (Y_1, Y_2, \dots, Y_n)$ uma amostra aleatória de $\underline{Y} | \theta \sim f(\underline{Y} | \theta)$ e seja $\underline{Y}^* = (Y_1^*, Y_2^*, \dots, Y_n^*) = (Y_{f_1}, Y_{f_2}, \dots, Y_{f_n})$, o vetor formado de n vetores de observações futuras com a mesma distribuição dos dados observados.

$$L_2(\underline{Y}, \delta) = \sum_{i=1}^n \text{Var}(Y_i^* | \underline{Y}) + \delta \sum_{i=1}^n [E(Y_i^* | \underline{Y}) - Y_i][E(Y_i^* | \underline{Y}) - Y_i]^T,$$

onde $0 \leq \delta \leq 1$, $\text{Var}(Y_i^* | \underline{Y})$ é a matriz de covariâncias e $E(Y_i^* | \underline{Y})$ é o vetor esperado com relação à distribuição posteriori preditiva.

Observação: $L_2(\underline{Y}, \delta)$ é uma matriz e portanto uma saída é considerar medidas resumo como, por exemplo, o determinante dessa matriz ($\det L = |L_2(\underline{Y}, \delta)|$), ou o traço dela ($\text{tr}(L_2(\underline{Y} | \delta))$).

Calculo da Medida L no Modelo Estrutural Normal Simples:

Nesse modelo os dados \underline{Y} são agora denotados por $D_{\text{obs}} = (\underline{U}_1, \dots, \underline{U}_n)$, onde $\underline{U}_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix}$, e os vetores de observações futuras, por $\underline{U}_i^* = \begin{pmatrix} Y_{j_i} \\ X_{j_i} \end{pmatrix}$, $i = 1, 2, \dots, n$, e $\underline{\theta} = (\alpha, \beta, \mu, \sigma_x^2, \sigma_u^2, \sigma_e^2)^T$ (ou então $\underline{\theta} = (\alpha, \beta, \mu, \sigma_x^2, \sigma^2)^T$ no modelo com restrição de identificabilidade $\sigma_e^2 = \lambda\sigma_u^2 = \lambda\sigma^2$).

Para calcular $E(\underline{U}_i^*|D_{\text{obs}})$ e $\text{Var}(\underline{U}_i^*|D_{\text{obs}})$, usamos

$$\begin{aligned} E(\underline{U}_i^*|D_{\text{obs}}) &= E[E(\underline{U}_i^*|\underline{\theta})|D_{\text{obs}}] = E_{\underline{\theta}|D_{\text{obs}}}(E(\underline{U}_i^*|\underline{\theta})), \\ \text{Var}(\underline{U}_i^*|D_{\text{obs}}) &= E[\text{Var}(\underline{U}_i^*|\underline{\theta})|D_{\text{obs}}] + \text{Var}[E(\underline{U}_i^*|\underline{\theta})|D_{\text{obs}}] \\ &= E_{\underline{\theta}|D_{\text{obs}}}[\text{Var}(\underline{U}_i^*|\underline{\theta})] + E_{\underline{\theta}|D_{\text{obs}}}[(E(\underline{U}_i^*|\underline{\theta}))(E(\underline{U}_i^*|\underline{\theta}))^T] \\ &\quad - [E_{\underline{\theta}|D_{\text{obs}}}(E(\underline{U}_i^*|\underline{\theta}))][E_{\underline{\theta}|D_{\text{obs}}}(E(\underline{U}_i^*|\underline{\theta}))]^T, \end{aligned}$$

onde

$$E(\underline{U}_i^*|\underline{\theta}) = \begin{pmatrix} \alpha + \beta\mu \\ \mu \end{pmatrix} \text{ e } \text{Var}(\underline{U}_i^*|\underline{\theta}) = \begin{bmatrix} \beta^2\sigma_x^2 + \sigma_e^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma_u^2 \end{bmatrix}.$$

Portanto, a medida L pode ser facilmente calculada usando o método simples de Monte Carlo da seguinte maneira:

Seja $(\underline{\theta}^{(1)}, \dots, \underline{\theta}^{(G)})$ uma amostra da posteriori $\pi(\underline{\theta}|D_{\text{obs}})$ (gerada por exemplo pelo algoritmo de Gibbs da seção 4.2.2),

$$\begin{aligned} \widehat{L}_2(D_{\text{obs}}, \delta) &= n \left\{ \frac{1}{G} \sum_{g=1}^G [\text{Var}(\underline{U}_i^*|\underline{\theta}^{(g)}) + (E(\underline{U}_i^*|\underline{\theta}^{(g)}))(E(\underline{U}_i^*|\underline{\theta}^{(g)}))^T] \right. \\ &\quad \left. - \left(\frac{1}{G} \sum_{g=1}^G E(\underline{U}_i^*|\underline{\theta}^{(g)}) \right) \left(\frac{1}{G} \sum_{g=1}^G E(\underline{U}_i^*|\underline{\theta}^{(g)}) \right)^T \right\} \\ &\quad + \delta \sum_{i=1}^n \left[\underline{U}_i - \frac{1}{G} \sum_{g=1}^G E(\underline{U}_i^*|\underline{\theta}^{(g)}) \right] \left[\underline{U}_i - \frac{1}{G} \sum_{g=1}^G E(\underline{U}_i^*|\underline{\theta}^{(g)}) \right]^T \end{aligned} \quad (2.11.1)$$

onde $\delta \in [0, 1]$.

No caso de regressão simples, $L_2(D_{\text{obs}}, \delta)$ é uma matriz 2×2 e uma medida resumo pode ser o determinante $\det(\widehat{L}_2(D_{\text{obs}}, \delta))$ ou o traço $(\widehat{L}_2(D_{\text{obs}}, \delta))$.

(2) O critério da Informação Bayesiano (BIC)

Essa medida é descrita na seção 4.4.3 do Apêndice para o caso geral.

No modelo estrutural Normal M_i , temos que

$$\text{BIC}_{M_i} = -2 \log(\pi_{M_i}(\underline{U}_1, \dots, \underline{U}_n | \hat{\theta}_{M_i})),$$

onde $\underline{U}_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix}$ e $\pi_{M_i}(\underline{U}_1, \dots, \underline{U}_n | \theta)$ é a densidade das observações $(\underline{U}_1, \dots, \underline{U}_n)$ do modelo M_i , $\hat{\theta}_{M_i}$ é o estimador de máxima verossimilhança de θ no modelo M_i , p_i é o número de parâmetros do modelo M_i e n o tamanho da amostra (ver Fuller, 1980).

Num modelo normal com restrição de identificabilidade, essa medida é facilmente calculada pois nesse caso os estimadores de máxima verossimilhança são conhecidos explicitamente.

No modelo com restrição $\sigma_e^2 = \lambda \sigma_u^2$, a densidade $\pi_{M_i}(\underline{U}_i | \theta)$ é dada pela expressão em (2.1.4) multiplicada pela constante de proporcionalidade $(2\pi)^{-\frac{1}{2}}$.

(3) Medidas informais alternativas para seleção de modelos

- (a) O produto $\prod_{i=1}^n \text{CPO}_i$.

Para cada um dos modelos competitivos calcula-se o produto

$$\prod_{i=1}^n \text{CPO}_i \quad (\text{ou o } \log(\prod_{i=1}^n \text{CPO}_i)) \quad (2.11.2)$$

e escolhe-se o modelo que corresponder ao maior valor dessa medida. No caso do modelo estrutural Normal, podemos estimar CPO_i , por exemplo, por meio das fórmulas (2.10.4), (2.10.8) e (2.10.11).

Quando o número de modelos competitivos é baixo (digamos, até 3), podemos também plotar os valores individuais de $\widehat{\text{CPO}}_i$ versus i para cada modelo no mesmo sistema de coordenadas. No caso de exatamente dois modelos, o quociente dos valores de $\prod_{i=1}^n \widehat{\text{CPO}}_i$ dá o estimador do Pseudo Fator de Bayes (ver o Apêndice B2).

(b) A soma das normas (euclidianas) dos resíduos padronizados

$$\sum_{i=1}^n \left| \frac{U_i - E(U_i|D_{\text{obs}})}{\sqrt{\text{Var}(U_i|D_{\text{obs}})}} \right| \quad (2.11.3)$$

onde

$$\hat{E}(U_i|D_{\text{obs}}) = \frac{1}{G} \sum_{g=1}^G \begin{pmatrix} \alpha^{(g)} + \beta^{(g)} \mu^{(g)} \\ \mu^{(g)} \end{pmatrix}$$

e

$$\begin{aligned} \widehat{\text{Var}}(U_i|D_{\text{obs}}) = \frac{1}{G} \left\{ \sum_{g=1}^G \begin{pmatrix} (\beta^{(g)})^2 \sigma_x^{2(g)} + \sigma_e^{2(g)} & \beta^{(g)} \sigma_x^{2(g)} \\ \beta^{(g)} \sigma_x^{2(g)} & \sigma_x^{2(g)} + \sigma_u^{2(g)} \end{pmatrix} \right. \\ \left. + \begin{pmatrix} \alpha^{(g)} + \beta^{(g)} \mu^{(g)} \\ \mu^{(g)} \end{pmatrix} \begin{pmatrix} \alpha^{(g)} + \beta^{(g)} \mu^{(g)} \\ \mu^{(g)} \end{pmatrix}^{\top} \right\} - (\hat{E}(U_i|D_{\text{obs}}))(\hat{E}(U_i|D_{\text{obs}}))^{\top}. \end{aligned}$$

Calcula-se para cada modelo o valor dessa soma e escolhe-se o modelo que corresponder ao menor valor dessa medida.

Outra alternativa é calcular a soma dos valores absolutos dos resíduos padronizados de validação cruzada

$$\sum_{i=1}^n \left| \frac{U_i - E(U_i|D_{\text{obs}(-i)})}{\sqrt{\text{Var}(U_i|D_{\text{obs}(-i)})}} \right|,$$

onde a $E(U_i|D_{\text{obs}(-i)})$ pode ser estimada, por exemplo, usando a fórmula em (2.10.13) e a $\text{Var}(U_i|D_{\text{obs}(-i)})$, a fórmula em (2.10.15).

2.12 EXEMPLOS

Nessa seção, baseando-se em uma amostra de dados simulados, são analisados os modelos com e sem restrição de identificabilidade usando prioris próprias e impróprias; além disso são comparados tres algoritmos MCMC. Para uma melhor compreensão do texto é recomendável a leitura da seção 2.4 do Apêndice.

2.12.1 Análise do modelo estrutural normal com restrição de identificabilidade

1. DADOS

Nesta seção é utilizada uma amostra de tamanho n de dados simulados a partir do modelo:

$$\begin{cases} Y_i = \alpha + \beta x_i + e_i, & i = 1, \dots, n, \\ X_i = x_i + u_i, \end{cases} \quad (2.12.4)$$

onde

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \stackrel{\text{iid}}{\sim} N \left(\begin{pmatrix} 0 \\ 0 \\ \mu \end{pmatrix}; \begin{pmatrix} \lambda\sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma_x^2 \end{pmatrix} \right), \quad i = 1, \dots, n,$$

A Tabela 2.1 a seguir contém uma amostra de tamanho 30 de dados gerados a partir desse modelo utilizando-se um programa na linguagem OX (“Object-Oriented Matrix Programming Language”)

Tabela 2.1: Amostra de tamanho $n=30$ de dados simulados a partir do modelo (2.12.4) para $\alpha = 2.0$, $\beta = 1.0$, $\mu = 2.0$, $\sigma_x^2 = 0.6$, $\sigma^2 = 0.1$ e $\lambda = 1.0$.

i	Y_i	X_i	i	Y_i	X_i	i	Y_i	X_i
1	3.6651	0.98589	11	3.2228	0.72437	21	4.8247	2.4804
2	5.0580	2.4616	12	2.4887	0.75859	22	4.8126	2.6593
3	3.8566	2.2892	13	3.2458	1.8808	23	3.5777	1.7877
4	3.6561	1.1471	14	5.5650	3.5664	24	4.3372	1.6879
5	4.9743	2.7713	15	4.1371	2.3534	25	3.6680	1.9640
6	4.3220	2.2590	16	4.9251	2.5498	26	3.3724	1.3345
7	3.7144	1.8731	17	3.8136	1.8851	27	4.3191	2.2530
8	3.9894	2.3108	18	4.5970	2.5381	28	3.5498	1.0989
9	4.5248	3.0514	19	4.5387	2.6756	29	2.7067	1.2432
10	2.8640	0.90543	20	4.6773	2.9065	30	2.5472	0.84126

2. MODELO PROPOSTO

Consideramos o modelo estrutural dado em (2.12.4 ou 2.3.1), com distribuição à priori $\pi(\alpha, \beta, \mu, \sigma_x^2, \sigma^2) = \pi(\alpha)\pi(\beta)\pi(\mu)\pi(\sigma_x^2)\pi(\sigma^2)$, onde $\alpha \sim N(0; 1000)$, $\beta \sim N(0; 1000)$, $\mu \sim N(0; 1000)$, $\sigma_x^2 \sim IG(2.001; 1.0)$, $\sigma^2 \sim IG(2.001; 1.0)$ e $\lambda = 1.0$.

3. AJUSTE DO MODELO

Nesse item vamos ajustar o modelo estrutural normal proposto acima aos dados da tabela 2.1, procedendo como se não conhecêssemos os verdadeiros valores dos parâmetros que foram usados para gerar esses dados.

Inicialmente utilizamos o algoritmo de Gibbs descrito na subseção 2.5.3 com o objetivo de amostrar da distribuição posteriori do modelo. São geradas $M = 4$ seqüências paralelas de Gibbs de comprimento $T = 10.000$ partindo de pontos iniciais gerados pelo procedimento descrito a seguir.

(a) procedimento para gerar os dados iniciais

(i) *Localização de possíveis modas:*

– Tomamos primeiramente uma grade de 20^5 pontos igualmente espaçados na região do espaço paramétrico (aproximadamente em torno do estimador de máxima verossimilhança de θ , $\hat{\theta} = (1.0465, 1.9185, 1.9748, 0.4941, 0.0659)$), dada pelo produto cartesiano $[-47.5, 47.5] \times [-47.5, 47.5] \times [-47.5, 47.5] \times [.01, 47.51] \times [.01, 47.51]$ e avaliamos o valor da função densidade à posteriori $p(\alpha, \beta, \mu, \sigma_x^2, \sigma^2 | D_{\text{obs}})$ em cada ponto dessa grade.

– Selecionamos os pontos que produzem os 10000 maiores valores da densidade à posteriori e utilizamos o algoritmo de otimização de Quase Newton denominado BFGS (Broyden, Fletcher, Goldfarb, Shanno) partindo de cada um desses pontos para localizar possíveis máximos locais (modas). (Para isso

foi usado o comando MAXBFGS da linguagem OX).

– Repetimos os dois passos anteriores considerando agora uma grade mais fina de 20^5 pontos igualmente espaçados na região dada pelo produto cartesiano: $[-4.75, 4.75] \times [-4.75, 4.75] \times [-4.75, 4.75] \times [0.01, 4.76] \times [0.01, 4.76]$. Desse procedimento foi encontrado apenas um único ponto de máximo:

$$\hat{\theta} = (1.8927, 1.0595, 1.9748, 0.43977, 0.11100).$$

(ii) *Geração dos valores iniciais:*

– Geramos 10000 vetores, $\vartheta_1, \dots, \vartheta_{10000}$, da distribuição $t_4(\hat{\theta}, \hat{V})$, onde $\hat{\theta}$ é a moda da posteriori dada acima e $\hat{V} = - \left(\frac{\partial^2 \log p(\vartheta | D_{\text{obs}})}{\partial \vartheta \partial \vartheta^T} \right)^{-1} \Big|_{\vartheta = \hat{\theta}}$, utilizando a decomposição de Choleski de \hat{V} , ($\hat{V} = L^T L$) (ver seção 2.4 do Apêndice).

– Reamostramos dez vetores: $\vartheta_1, \dots, \vartheta_{10}$ desses 10000 (usando o procedimento de Reamostragem Ponderada descrito na seção 2.4 do Apêndice), e pelo fato desses pontos estarem muito próximos resolvemos considerar apenas quatro deles:

$$\vartheta_1^{(0)} = (1.36552, 1.39858, 1.62917, 0.19871, 0.12144)^T,$$

$$\vartheta_2^{(0)} = (1.23000, 1.40901, 1.88743, 0.61149, 0.08740)^T,$$

$$\vartheta_3^{(0)} = (1.77440, 1.12877, 2.26311, 0.50376, 0.03471)^T,$$

$$\vartheta_4^{(0)} = (2.42116, 0.68870, 2.30602, 0.35828, 0.13245)^T.$$

(b) Diagnóstico de convergência das seqüências de Gibbs

A seguir são apresentados alguns dispositivos gráficos (figuras 2.1 a 2.4) como também algumas medidas (tabela 2.2) para o diagnóstico de possíveis falhas na convergência do algoritmo de Gibbs (ver seção 2.4.6 do apêndice sobre diagnósticos de convergência).

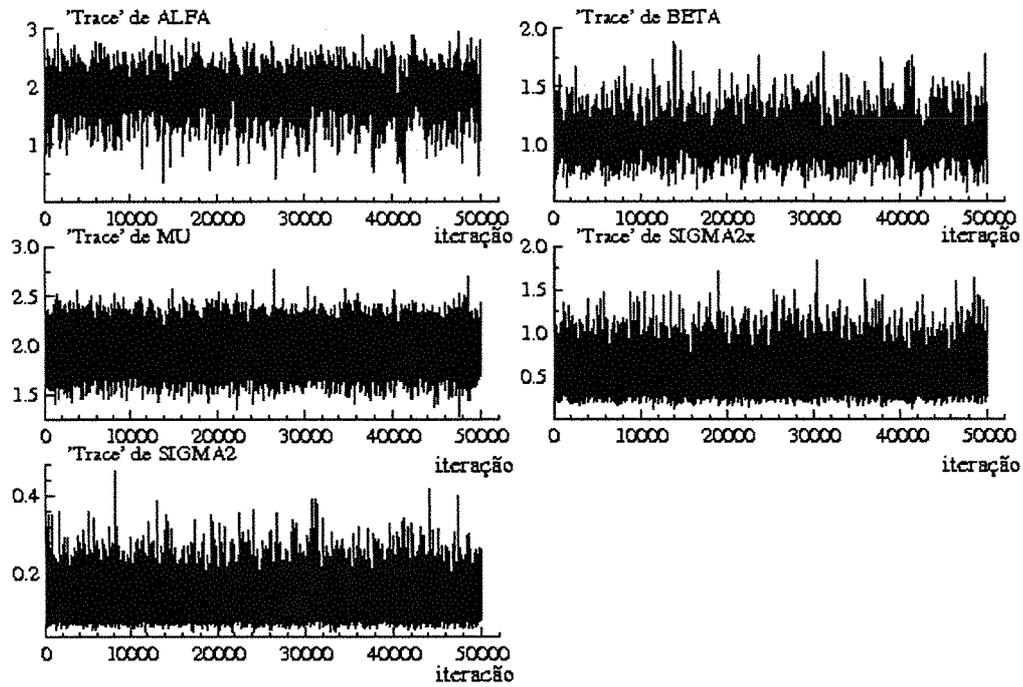


Figura 2.1: Gráficos das séries temporais (“Trace”) de α , β , μ , σ_x^2 e σ^2 , baseados em uma única sequência de Gibbs de comprimento $T = 50.000$ partindo de $\theta_1^{(0)}$.

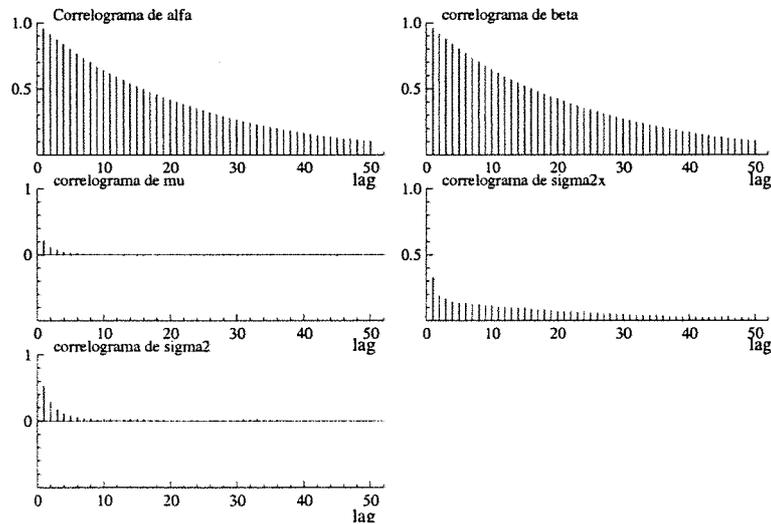


Figura 2.2: Gráficos das funções de auto-correlação de α , β , μ , σ_x^2 e σ^2 , baseados em uma única sequência de Gibbs de comprimento $T = 50.000$ partindo de $\theta_1^{(0)}$.

Brooks & Gelman Multivariate Shrink Factors

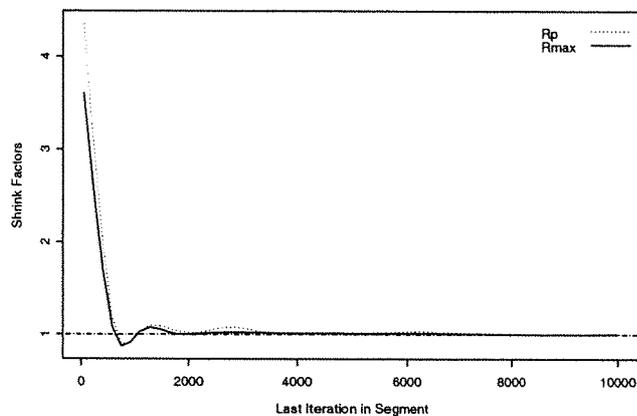


Figura 2.3: Gráfico da medida “Multivariate Potential Scale Reduction” (MPSRF) de Brooks e Gelman (1998) baseado em $M = 4$ seqüências de Gibbs de comprimento $T = 10.000$ (Rmax é o maior valor entre as medidas univariadas de $\alpha, \beta, \mu, \sigma_x^2, \sigma^2$).

Gelman & Rubin Shrink Factors

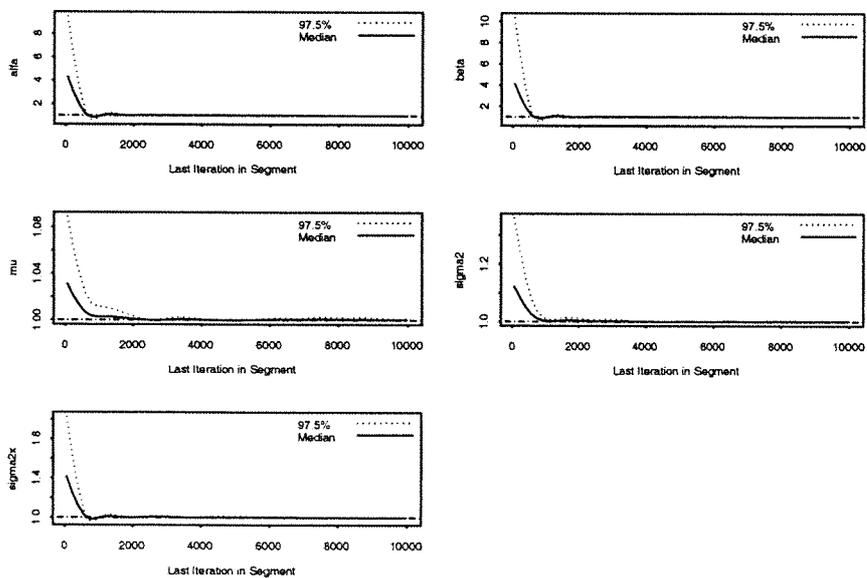


Figura 2.4: Gráfico da medida de Gelman e Rubin corrigida “Corrected potential scale reduction factors” (CSRF) para cada parâmetro $\alpha, \beta, \mu, \sigma_x^2$ e σ^2 , baseada em $M = 4$ seqüências de Gibbs de comprimento $T = 10.000$.

Tabela 2.2: Medidas de Diagnóstico de convergência de Brooks, Gelman e Rubin

	Medida de Gelman&Rubin(PSRF)	Medida Corrigida(CSRF)	Quantil 0.975
α	1.002184	1.003412	1.008271
β	1.001965	1.002994	1.007383
μ	1.000334	1.000341	1.001258
σ^2	1.000398	1.000408	1.001460
σ^2	1.000705	1.000897	1.002601

A medida multivariada (MPSRF) é igual a 1.0027.

Observação: Os dispositivos apresentados até aqui não nos sugerem que exista problema na convergência do algoritmo de Gibbs.

(c) Estimação

A tabela 2.3 apresenta as estimativas de várias medidas resumo da distribuição à posteriori. Essas estimativas foram obtidas utilizando-se a segunda metade de cada uma das $M = 4$ seqüências de Gibbs de comprimento $T = 10.000$ descritas anteriormente partindo-se dos pontos iniciais $\theta_1^{(0)}$ a $\theta_4^{(0)}$ apresentados anteriormente (lembrando que a amostra de dados utilizada é a da tabela 2.1 e a distribuição à priori é dada no item 2.).

Antes de partir para a análise dos resultados é importante salientar que neste modelo a distribuição à priori é praticamente não informativa e portanto, se de fato o procedimento de estimação é adequado, espera-se que as estimativas à posteriori dos parâmetros estejam próximas dos valores dos parâmetros que foram usados na geração dos dados. Além disso essas estimativas deveriam também estar próximas das estimativas de máximaverossimilhança.

Na tabela 2.3 observa-se que as estimativas do valor esperado da distr. à posteriori dos parâmetros, obtidas pela média aritmética e pela média Rao-Blackelizada, estão muito próximas, além disso, tanto a média como a mediana amostral estimam cada um dos parâmetros do modelo com “pequeno” viés(estimado) e somente o parâmetro σ^2 parece estar um pouco superestimado. Além disso, mais importante ainda, os valores verdadeiros dos parâmetros (aqueles usados para gerar a amostra

de dados) pertencem aos intervalos de credibilidade apresentados. Notemos também que as estimativas de máximaveross. estão bem próximas das estimativas de Bayes.

A Figura 2.5 apresenta os gráficos das estimativas “Rao-Blackwelizadas” das densidades posterioris marginais dos parâmetros, obtidas de acordo com a seção 2.9, usando as seqüências de Gibbs descritas acima.

Tabela 2.3: Estimativas do valor esperado, quantis, desvio padrão, máximo, mínimo e intervalos de credibilidade de comprimento mínimo das distribuições posterioris marginais de $\alpha, \beta, \mu, \sigma_x^2, \sigma^2$.

MEDIDAS RESUMO	α	β	μ	σ_x^2	σ^2
Média aritmética	1.8652	1.0743	1.9718	0.5045	0.1340
Média “R-Blackweliz.”	1.8653	1.0743	1.9732	0.5043	0.1337
Quantis amostrais					
0.025	1.2142	0.8116	1.6819	0.2663	0.0808
.05	1.3461	0.8518	1.7316	0.2924	0.0868
.1	1.4766	0.8973	1.7875	0.3275	0.0941
.2	1.6241	0.9537	1.8513	0.3731	0.1041
.25	1.6783	0.9746	1.8757	0.3906	0.1082
.3	1.7287	0.9937	1.8977	0.4084	0.1123
.4	1.8102	1.0296	1.9361	0.4431	0.1200
.5	1.8846	1.0641	1.9717	0.4793	0.1278
.6	1.9576	1.1011	2.0078	0.5171	0.1368
.7	2.0315	1.1407	2.0472	0.5607	0.1471
.75	2.0713	1.1636	2.0686	0.5886	0.1537
.8	2.1160	1.1896	2.0924	0.6199	0.1609
.9	2.2335	1.2610	2.1554	0.7151	0.1823
.95	2.3282	1.3301	2.2127	0.8064	0.2015
.975	2.4072	1.3966	2.2586	0.8896	0.2204
Interv.Credib. HPD					
$p = 0.95$ Lim. inf.	1.2828	0.7982	1.6783	0.2370	0.0736
Lim sup.	2.4642	1.3795	2.2543	0.8315	0.2060
$p = 0.90$ Lim. inf.	1.3914	0.8384	1.7300	0.2535	0.0806
Lim sup.	2.3658	1.3126	2.2105	0.7388	0.1897
desvio padrão (dp)	0.3024	0.1459	0.1462	0.1612	0.0360
d.p. Rao-Blackweliz.	0.3032	0.1461	0.1456	0.1602	0.0361
“Naive SE” = $\frac{dp}{\sqrt{M(T-T_0)}}$	0.002138	0.001031	0.001033	0.00114	0.000255
“Batch SE”	0.010241	0.005016	0.001376	0.002418	0.000498
Mínimo	0.4118	0.6082	1.4117	0.1513	0.0537
Máximo	2.8863	1.7473	2.7745	1.7092	0.4632
Estimativa de máxima verossimilhança	1.9185	1.0465	1.9748	0.4941	0.0659

NOTA: Os valores dos parâmetros usados na geração dos dados são: $\alpha = 2.0, \beta = 1.0, \mu = 2.0, \sigma_x^2 = 0.6$ e $\sigma^2 = 1.0$.

Os intervalos de credibilidade nesse exemplo coincidem com os intervalos HPD.

“Batch SE” é o desvio padrão amostral das médias de grupos consecutivos de tamanho 50, dividido pela raiz quadrada do número de grupos (calculado por meio do BOA (“Bayesian Output Analysis”)).

M=número de cadeias paralelas.

T=número total de iterações, T_0 =número de iterações descartadas.

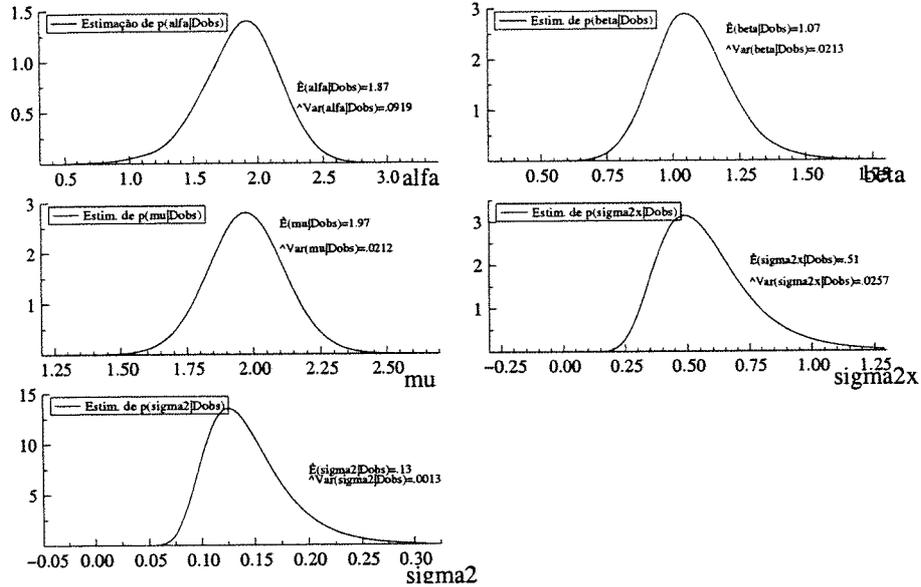


Figura 2.5: Gráficos das estimativas “Rao-Blackwelizadas” das densidades marginais à posteriori $p(\alpha|D_{obs}), \dots, p(\sigma^2|D_{obs})$, incluindo também suas médias e variâncias

(d) Avaliação do ajuste

A Figura 2.6 apresenta os gráficos dos valores da densidade condicional preditiva de validação cruzada, $[CPO_i = \pi\left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{obs}\right)]$, primeiramente em função de i , depois em função de X_i e depois em função de Y_i , para $i = 1, \dots, 30$, onde CPO_i é estimada por meio da fórmula (2.10.8). As estimativas foram baseadas na segunda metade de uma única seqüência de comprimento $T=10.000$ do algoritmo de Gibbs que parte do ponto inicial $\theta_1^{(0)}$ dado anteriormente.

Na seção 2.10.2(b), os resíduos bayesianos bidimensionais de validação cruzada podiam ser estimados por

$$\hat{d}_i = \begin{pmatrix} d_{i1} \\ d_{i2} \end{pmatrix} = \widehat{Var}^{-\frac{1}{2}} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{obs(-i)} \right) \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} - \hat{E} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{obs(-i)} \right) \right),$$

$i = 1, \dots, n$, onde $\hat{E} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{obs(-i)} \right)$ é dada em (2.10.13) e $\widehat{Var}^{-\frac{1}{2}} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{obs(-i)} \right)$ é dada em (2.10.15), para $i = 1, \dots, n$.

A Figura 2.7 apresenta os gráficos de cada uma das componentes d_{1i} e d_{2i} do vetor de resíduos.

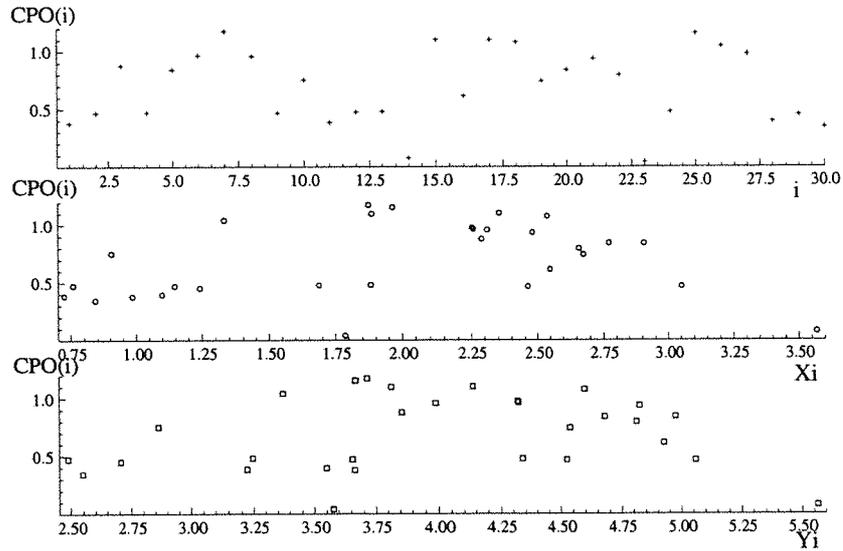


Figura 2.6: Gráficos dos valores da densidade condicional preditiva, $(CPO(i))$, versus i (posição superior), versus X_i (posição do meio) e versus Y_i (posição inferior).

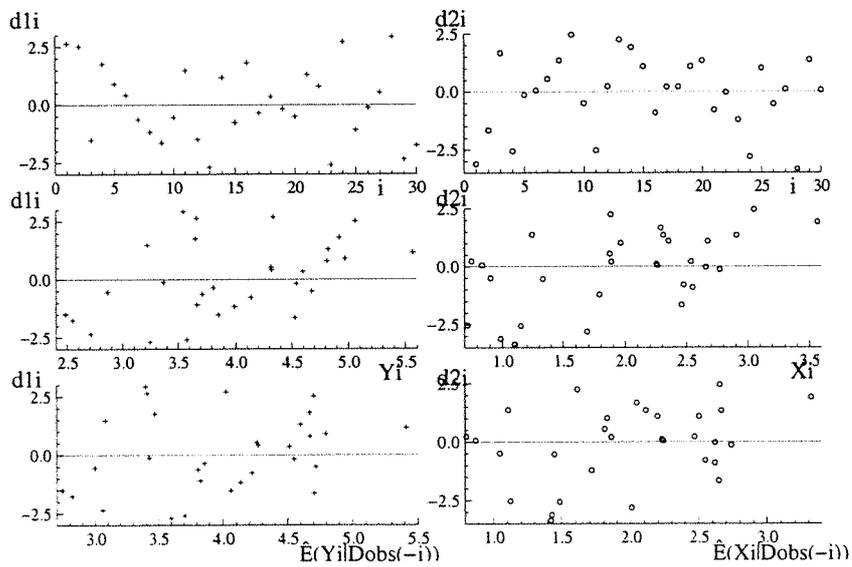


Figura 2.7: Do lado esquerdo os gráficos da componente d_{1i} versus i , versus Y_i e versus $E(Y_i | D_{\text{obs}}(-i))$. Do lado direito (d_{2i}) versus i , versus X_i e versus $E(X_i | D_{\text{obs}}(-i))$.

Os gráficos acima não sugerem que haja grandes problemas de falta de ajustamento. Entretanto na figura 2.7 existem 2 pontos nos gráficos de ordenada d_2 , cujos valores são um pouquinho menores que -3 e que aparentemente poderiam ser vistos como pontos aberrantes, portanto uma análise de resíduos mais aprofundada deve ser conduzida futuramente para examinar melhor o ajuste.

4. SENSIBILIDADE À λ – quociente das variâncias residuais

Examinamos descritivamente a variação nas estimativas dos valores esperados de α , β , μ , σ_x^2 e σ^2 à posteriori quando diferentes valores do hiperparâmetro λ são considerados (mantendo-se os demais hiperparâmetros do modelo em (2.12.2) inalterados).

Ns figuras 2.8 e 2.9, são apresentados gráficos com as estimativas “Rao-Blackwelizadas” dos valores esperados dos parâmetros da distribuição à posteriori baseadas nos dados da Tabela 2.1. Para cada λ fixado, foi gerada uma seqüência de Gibbs de comprimento $T = 10.000$ (com vetor inicial igual ao estimador de máxima verossimilhança correspondente). Para calcular as estimativas considerou-se apenas a segunda metade de cada seqüência.

Na figura 2.10, comparamos os gráficos (das médias Rao-Blackelizadas versus λ) baseados em dois diferentes tamanhos de amostra ($n = 50$, $n = 500$). Essas amostras foram geradas à partir do mesmo modelo em (2.12.1) usando a mesma semente (ranseed=17 da linguagem OX). Para cada λ fixado, foi gerada uma seqüência de Gibbs partindo-se da estimativa de máxima verossimilhança correspondente. Para o cálculo das médias foram usadas apenas a segunda metade de cada seqüência de $T=10000$ iterações .

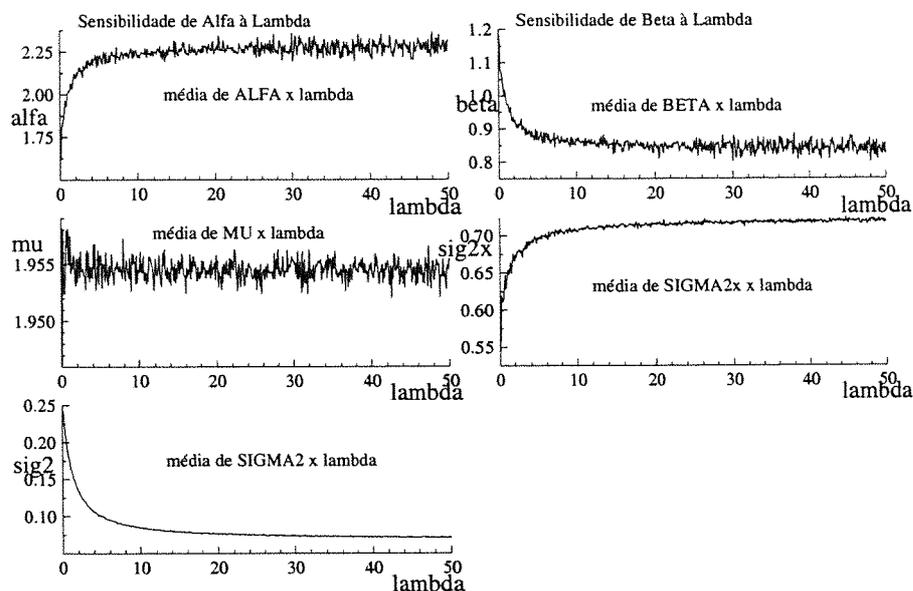


Figura 2.8: Gráficos das médias “R.-Blackwelizadas” de α , β , μ , σ_x^2 e σ^2 versus λ baseadas na amostra de tamanho $n=30$ da tabela 2.1. A grade de pontos em λ é formada por 500 pontos igualmente espaçados no intervalo $[0,50]$. (Os valores verdadeiros dos parâmetros são: $\alpha = 2.0$, $\beta = 1.0$, $\mu = 1.0$, $\sigma_x^2 = 0.6$ e $\sigma^2 = 0.1$ e do hiperparâmetro λ é 1.0.)

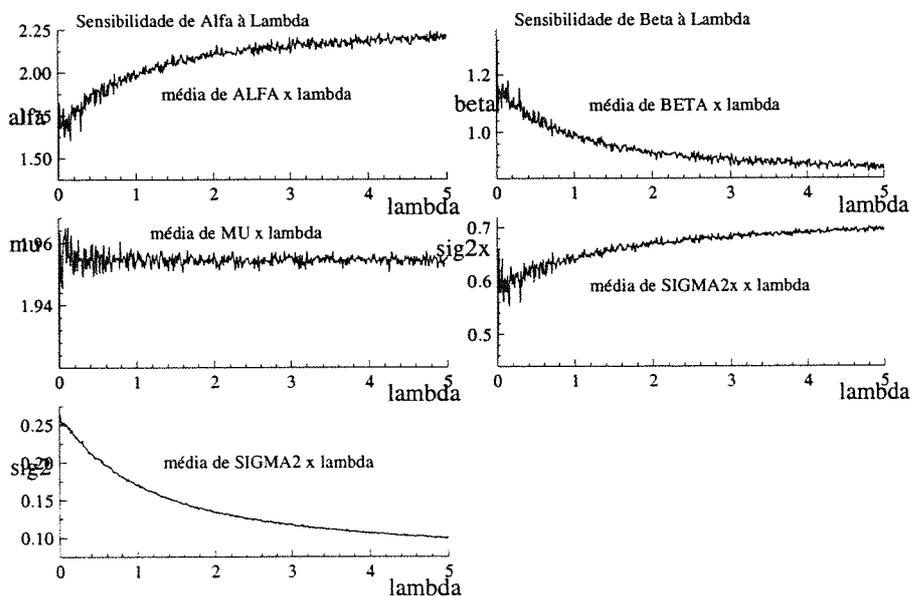


Figura 2.9: Idem à figura anterior, porém com grade de pontos em λ formada por 500 pontos igualmente espaçados no intervalo $[0,5]$.

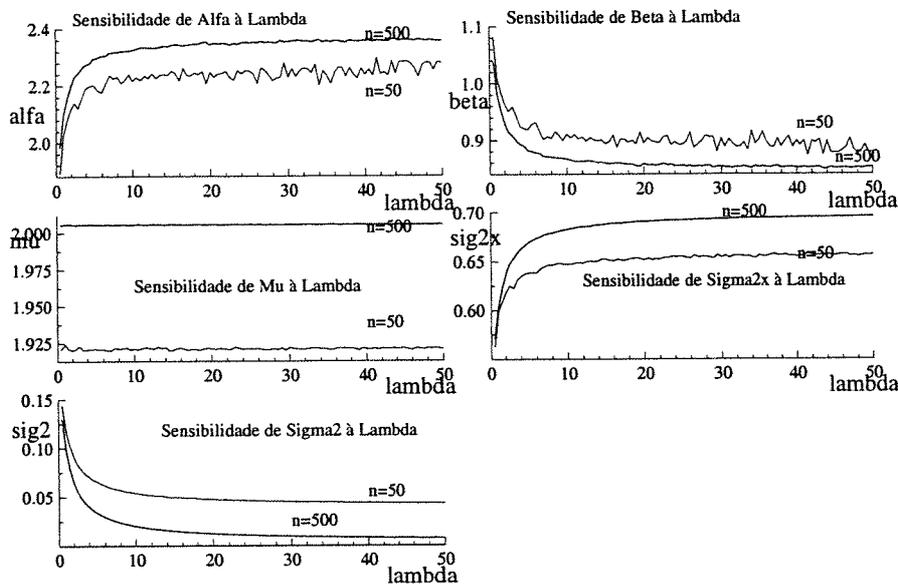


Figura 2.10: Gráfico das médias “Rao-Blackwelizadas” de α , β , μ , σ_x^2 e σ^2 versus λ , das posteriores baseadas nas amostras geradas de tamanho $n = 50$ e $n = 500$. A grade em λ é formada de 100 pontos igualmente espaçados no intervalo $[0,50]$ (os valores verdadeiros dos parâmetros são $\alpha = 2.0$, $\beta = 1.0$, $\mu = 2.0$, $\sigma_x^2 = 0.6$ e $\sigma^2 = 0.1$ e de λ é 1.0).

Observações:

Das figuras 2.8 e 2.9 observamos que:

- (1) A média amostral do parâmetro μ parece não ser influenciada pelo valor de λ .
- (2) Quando λ é subestimado (isto é, quando $0 < \lambda < 1$), aparentemente, as médias amostrais de β e σ_x^2 superestimam os valores verdadeiros de β e σ_x^2 mas a média amostral de α subestima o verdadeiro valor de α , entretanto a estimativa do viés de estimação não ultrapassa a 15% dos valores verdadeiros dos parâmetros.
- (3) Quando λ é superestimado ($\lambda > 1$), aparentemente ocorre o oposto (α é superestimado e β e σ_x^2 são subestimados) e o viés de estimação também não ultrapassa 15%(ou 20%) dos valores verdadeiros dos parâmetros. Isso ocorre mesmo quando λ é bem grande (da ordem de 50 vezes maior que o seu verdadeiro valor). Aparentemente as estimativas tendem a se estabilizar à medida que λ cresce.

(4) O valor da estimativa do viés na estimação de σ^2 é em geral bem grande. Além disso, quando o tamanho da amostra (n) não é muito grande ($n \approx 30$), o valor de σ^2 é muito superestimado quando λ está próximo do verdadeiro valor ($0 < \lambda < 2$), indicando que procedimento de estimação de σ^2 seria razoável apenas para amostras bem grandes .

Da figura 2.10 sugere que as estimativas dos parâmetros aparentemente tendem a se estabilizar independentemente do tamanho (n) da amostra, (mas o valor do viés de estimação depende de n).

5. COMPARAÇÃO ENTRE OS ALGORITMOS

Para comparar os algoritmos de Gibbs, Grouped Gibbs e M.Collapsed Gibbs, apresentamos os gráficos das funções de autocorrelação amostral e das séries temporais de cada um dos parâmetros, como também as matrizes de correlação cruzada dos parâmetros do modelo dessa seção para uma particular amostra de dados (tabela 2.1), utilizando cada um dos algoritmos acima (partindo sempre do mesmo ponto inicial e tomando-se $T=50.000$ iterações).

Foram obtidas também, a partir de cada algoritmo, as estimativas Rao-Blackwellizadas das densidades marginais à posteriori e as estimativas das medidas resumo da distr. à posteriori (com base em 4 seqüências de comprimento $T=10.000$). Entretanto serão apresentados apenas os gráficos da função de autocorrelação amostral (fig.2.11) e a tabela (2.4) com as matrizes de correlação cruzada sob os 3 algoritmos.

Da figura 2.11 a 2.13 podemos observar, com base nas autocorrelações amostrais, que aparentemente não há vantagem em usar o Grouped Gibbs em relação ao Gibbs simples, mas há um certo ganho (embora não muito grande) em utilizar o algoritmo M.Collapsed Gibbs sobre os demais para amostrar da distr. à posteriori do modelo.

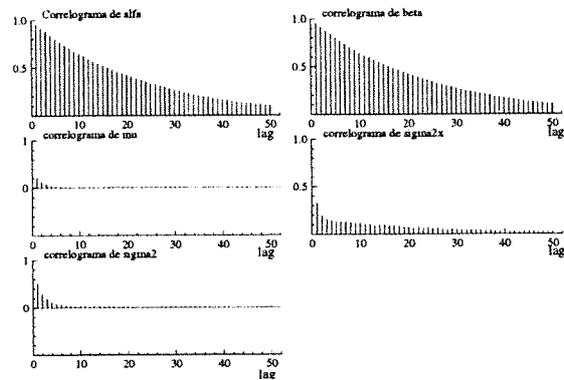


Figura 2.11: Função de autocorrelação amostral da seqüência do algoritmo de Gibbs.

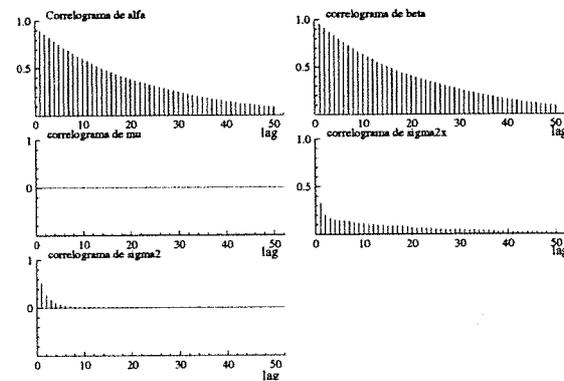


Figura 2.12: Função de autocorrelação amostral da seqüência do algoritmo “Grouped Gibbs”.

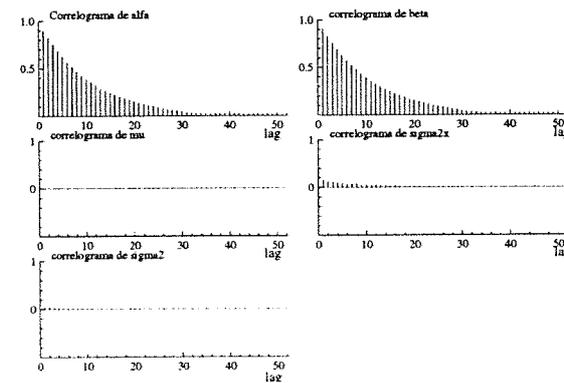


Figura 2.13: Função de autocorrelação amostral da seqüência do algoritmo “Modified and Collapsed Gibbs”, com um mini Gibbs de comprimento $N = 10$.

Tabela 2.4: Matrizes de correlação de $(\alpha, \beta, \mu, \sigma_x^2, \sigma^2)$ obtidas apartir dos algoritmos de Gibbs, Grouped Gibbs e M.Collapsed Gibbs

		α	β	μ	σ_x^2	σ^2
Gibbs	α	1.000	-0.9509	-0.0967	0.3688	0.0441
	β		1.000	-0.0054	-0.388	0.0459
	μ			1.000	-0.0007	-0.0039
	σ_x^2				1.000	-0.1136
	σ^2					1.0000
Group.Gibbs	α	1.000	-0.9484	-0.0960	0.3746	-0.0273
	β		1.000	-0.0108	-0.3961	0.0328
	μ			1.000	0.0135	0.0057
	σ_x^2				1.000	-0.1108
	σ^2					1.0000
M.Collap.Gibbs	α	1.000	-0.9494	-0.0959	0.3678	-0.0340
	β		1.000	-0.0040	-0.3853	0.0332
	μ			1.000	0.0212	0.0042
	σ_x^2				1.000	-0.1107
	σ^2					1.0000

Da tabela 2.4 vemos, com base nos valores das correlações cruzadas, que aparentemente não há vantagem no uso do algoritmo Grouped Gibbs sobre o Gibbs simples, já que metade das correlações diminuíram e metade aumentaram. Entretanto o M.Collapsed Gibbs gerou uma seqüência cujas correlações cruzadas são em geral um pouquinho menores que as correlações obtidas pelo Gibbs simples.

Observações:

- Os gráficos das séries temporais obtidos pelos tres algoritmos, para esse particular exemplo, ficaram muito parecidos, não dando para distinguir nenhuma diferença marcante entre eles.

As estimativas das medidas resumo, obtidas pelos tres algoritmos ficaram muito próximas diferindo nos piores casos apenas pela segunda casa decimal. Os gráficos das estimativas Rao-Blackwelizadas das densidades marginais à posteriori ficaram praticamente indistinguíveis.

- Com base apenas no particular exemplo estudado acima, aparentemente não há grande vantagem em usar modificações do algoritmo de Gibbs.

2.12.2 Análise do modelo estrutural normal sem restrição de identificabilidade

1. DADOS

A amostra de tamanho $n=30$ de dados a ser usada nessa seção foi gerada a partir do modelo

$$\begin{cases} Y_i = \alpha + \beta x_i + e_i, & i = 1, \dots, n, \\ X_i = x_i + u_i, \end{cases} \quad (2.12.5)$$

onde

$$\begin{pmatrix} x_i \\ u_i \\ e_i \end{pmatrix} \stackrel{\text{iid}}{\sim} N \left(\begin{pmatrix} \mu \\ 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma_e^2 \end{pmatrix} \right), \quad i = 1, \dots, n,$$

para $\alpha = 2.0$, $\beta = 1.0$, $\mu = 2.0$, $\sigma_x^2 = 0.6$, $\sigma^2 = 0.1$, $\sigma_e^2 = 0.1$.

Nesse caso, o conjunto de dados é o mesmo que o da Tabela 2.1.

2. MODELO PROPOSTO

Consideramos o modelo em (2.12.5) (ou (2.6.1), com distribuição à priori dada por $\pi(\alpha, \beta, \mu, \sigma_x^2, \sigma^2, \sigma_e^2)$ com componentes independentes, onde α , β e μ têm distribuição $N(0; 1000)$ e σ_x^2 , σ^2 e σ_e^2 têm distribuição $IG(2.001; 1.0)$.

3. AJUSTE DO MODELO

(a) **Pontos Iniciais:** Geramos $M = 4$ pontos iniciais usando procedimento semelhante ao da seção 2.12.1. Os pontos gerados foram:

$$\theta_1^{(0)} = (1.08372, 1.30058, 2.29445, 0.23498, 0.14854, 0.17795);$$

$$\theta_2^{(0)} = (2.12682, 1.03819, 1.93648, 0.33617, 0.11543, 0.10954);$$

$$\theta_3^{(0)} = (2.36979, 0.83386, 1.94213, 0.43797, 0.15633, 0.12669);$$

$$\theta_4^{(0)} = (2.83504, 0.65290, 1.85471, 0.57342; 0.17470, 0.18693);$$

Partindo desses pontos geramos $M = 4$ seqüências de Gibbs de comprimento $T = 10.000$ com o objetivo de amostrar da distribuição à posteriori. (Foi gerada também uma única seqüência de comprimento $T = 50000$ para os gráficos de diagnóstico).

(b) Diagnósticos de convergência das seqüências de Gibbs

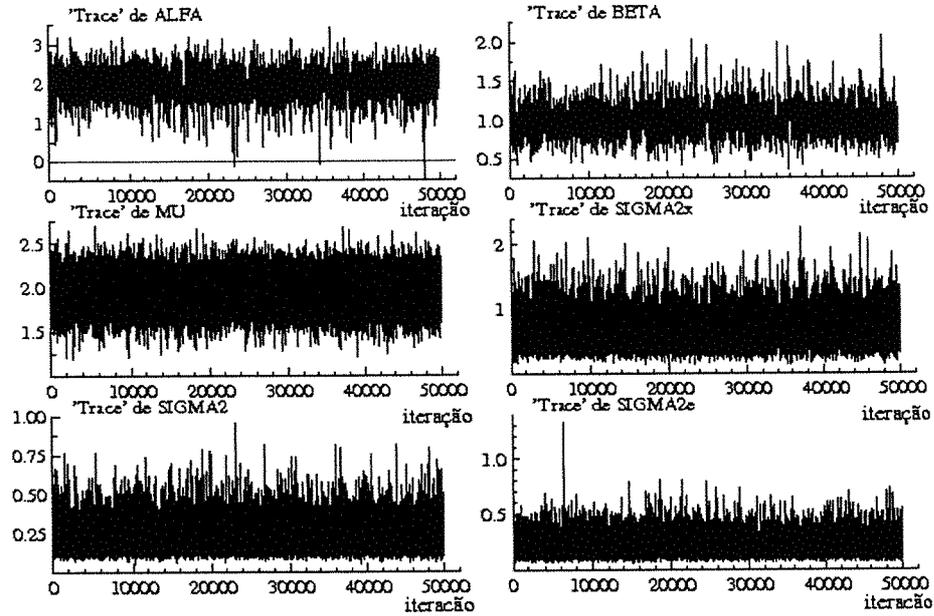


Figura 2.14: Gráficos das séries temporais (“trace”) de α , β , μ , σ_x^2 , σ^2 e σ_e^2 , baseados em uma única seqüência de Gibbs de comprimento $T = 50000$ partindo de $\theta_1^{(0)}$.

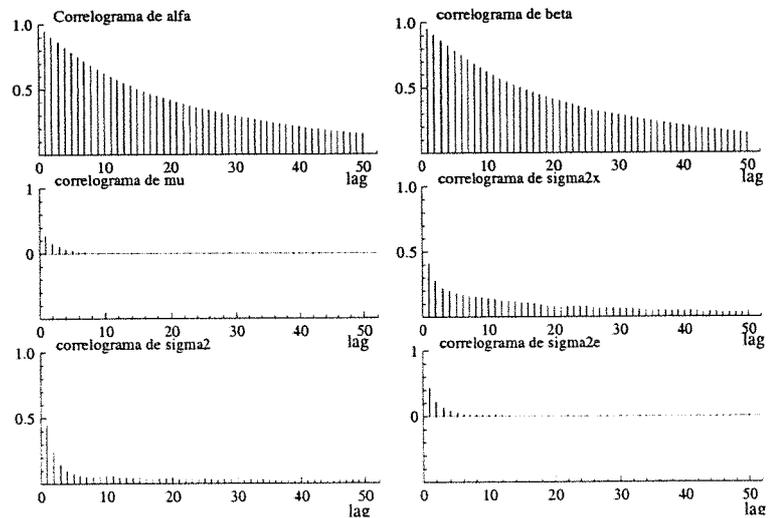


Figura 2.15: Gráficos das funções de auto-correlação de α , β , μ , σ_x^2 , σ^2 e σ_e^2 , baseados em uma única seqüência de Gibbs de comprimento $T = 50000$ partindo de $\theta_1^{(0)}$.

Brooks & Gelman Multivariate Shrink Factors

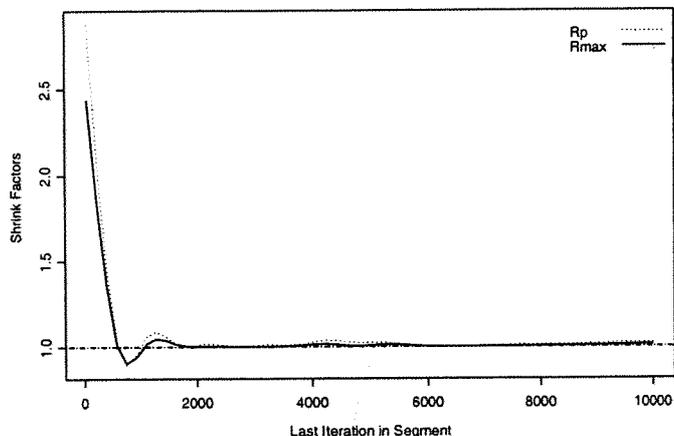


Figura 2.16: Gráfico da medida “Multivariate Potential Scale Reduction” (MPSRF) de Brooks e Gelman (1998) baseado em $M = 4$ seqüências de Gibbs de comprimento $T = 10000$ (R_{max} é o maior valor entre as medidas univariadas de $\alpha, \beta, \mu, \sigma_x^2, \sigma^2, \sigma_e^2$).

Gelman & Rubin Shrink Factors

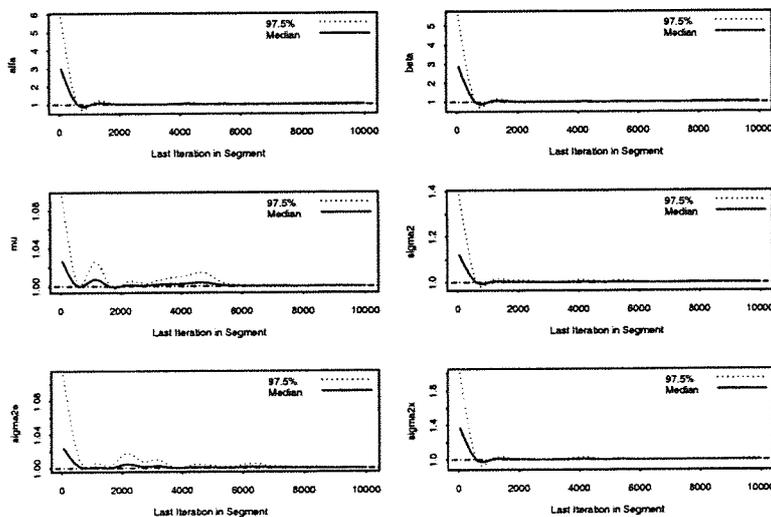


Figura 2.17: Gráfico da medida de Gelman e Rubin corrigida “Corrected potential scale reduction factors” (CSRFB) para cada parâmetro $\alpha, \beta, \mu, \sigma_x^2, \sigma^2, \sigma_e^2$, baseada em $M = 4$ seqüências de Gibbs de comprimento $T = 10.000$.

	Medida de Gelman&Rubin(PSRF)	Medida Corrigida(CSRF)	Quantil 0.975
α	1.008714	1.015572	1.035024
β	1.008708	1.015500	1.034929
μ	0.9999645	1.000034	1.000171
σ^2	1.000683	1.002484	1.006979
σ_x^2	0.9999692	1.000794	1.002451
σ_e^2	1.002024	1.000011	1.000157

A medida multivariada (MPSRF) é igual a 1.014622.

Obs. As figuras 2.14 a 2.17 e a tabela 2.5 não sugerem que exista problemas na convergência do algoritmo de Gibbs .

(c) Estimação

Na figura 2.18 apresentamos os gráficos das densidades posterioris marginais estimadas, calculadas de acordo com a seção 2.9, e na tabela 2.6 as estimativas à posteriori de várias medidas resumo. Essas estimativas baseiam-se em $M = 4$ seqüências de Gibbs de comprimento $T = 10.000$, descartando-se as primeiras $T_0 = 5.000$ iterações.

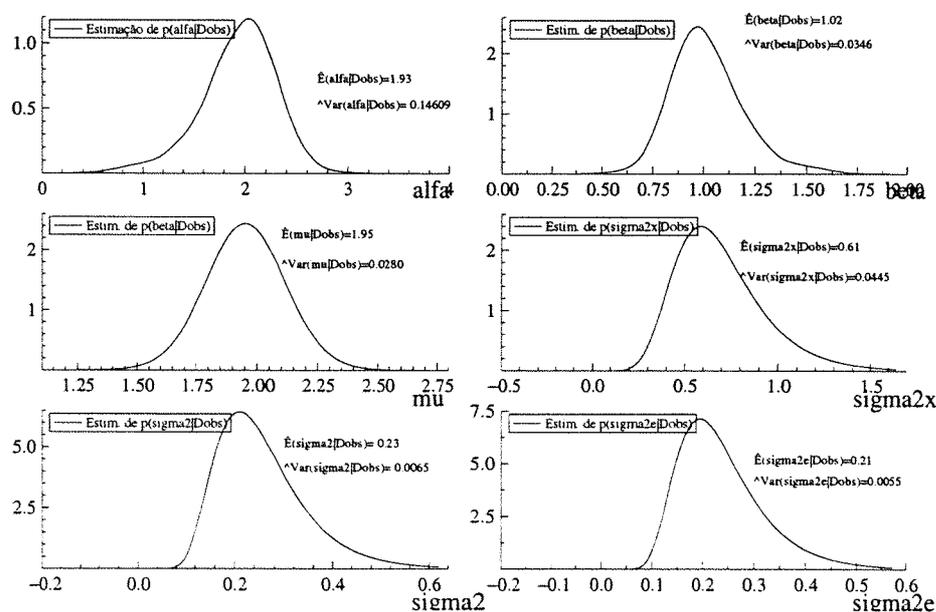


Figura 2.18: Gráficos das estimativas “Rao-Blackwelizadas” das densidades marginais à posteriori $p(\alpha|D_{Obs}), \dots, p(\sigma_e^2|D_{Obs})$.

Tabela 2.6: Estimativas do valor esperado, quantis, desvio padrão, máximo, mínimo e intervalos de credibilidade de comprimento mínimo das distribuições marginais à posteriori de $\alpha, \beta, \mu, \sigma_x^2, \sigma^2, \sigma_\epsilon^2$, usando os dados da tabela 2.1

MEDIDAS RESUMO	α	β	μ	σ_x^2	σ^2	σ_ϵ^2
Média aritmética	1.9341	1.0188	1.9527	0.6074	0.2263	0.2084
Média "R-Blackweliz."	1.9343	1.0184	1.9524	0.6089	0.2258	0.2087
Quantis amostrais						
0.025	1.0234	0.7050	1.617	0.2917	0.1122	0.1032
.05	1.2399	0.7525	1.6775	0.3270	0.1232	0.1144
.1	1.4387	0.8059	1.7408	0.3724	0.1388	0.1289
.2	1.6524	0.8701	1.8135	0.4334	0.1598	0.1478
.25	1.7208	0.8939	1.842	0.4584	0.1691	0.1559
.3	1.7791	0.9165	1.8658	0.4825	0.1776	0.1641
.4	1.8801	0.9580	1.9108	0.5284	0.1944	0.1791
.5	1.9701	0.9999	1.9522	0.5749	0.2123	0.1949
.6	2.0556	1.0432	1.9945	0.6254	0.2313	0.2126
.7	2.1434	1.0934	2.0399	0.6844	0.2539	0.2336
.75	2.1885	1.0934	2.064	0.7181	0.2674	0.2461
.8	2.2393	1.1558	2.0925	0.7609	0.2830	0.2608
.9	2.3781	1.254	2.1646	0.8805	0.3325	0.3039
.95	2.4876	1.3499	2.229	0.9984	0.3776	0.3467
.975	2.5872	1.4631	2.2851	1.1204	0.4217	0.3887
Interv. Credib. HPD						
$p = 0.95$ Lim. inf.	1.1256	0.6678	1.6253	0.2487	0.0993	0.0881
Lim sup.	2.6501	1.4027	2.2916	1.0250	0.3901	0.3554
$p = 0.90$ Lim. inf.	1.313	0.7240	1.6811	0.2828	0.1090	0.1001
Lim sup.	2.5362	1.3086	2.2305	0.9145	0.3471	0.3162
desvio padrão (dp)	0.3825	0.1862	0.1684	0.2116	0.0806	0.0744
d.p. Rao-Blackweliz.	0.3822	0.1861	0.1672	0.2110	0.0807	0.0807
"Naive SE" = $\frac{dp}{\sqrt{M(T-T_0)}}$	0.002705	0.000002	0.001191	0.001496	0.000570	0.000526
"Batch SE"	0.014119	0.006874	0.001734	0.001292	0.000990	0.004280
Mínimo	0.3045	0.3671	1.2070	0.1451	0.0507	0.0552
Máximo	3.4224	1.9439	2.6967	2.2617	0.8209	1.3220

NOTA: Os intervalos de credibilidade nesse exemplo coincidem com os intervalos HPD.
 "Batch SE" é o desvio padrão amostral das médias de grupos consecutivos de tamanho 50, dividido pela raiz quadrada do número de grupos.
 Os valores dos parâmetros usados na geração dos dados são: $\alpha = 2.0, \beta = 1.0, \mu = 2.0, \sigma_x^2 = 0.6, \sigma^2 = 0.1$ e $\sigma_\epsilon^2 = 0.1$
 M=número de cadeias paralelas.
 T=número total de iterações, T_0 =número de iterações descartadas.

Na tabela 2.6 vemos que as estimativas do valor esperado da distribuição à posteriori dos parâmetros, obtidas pela média aritmética e pela média Rao-Blackwelizada estão muito próximas, além disso, tanto a média como a mediana superestimam os valores verdadeiros dos parâmetros σ^2 e σ_ϵ^2 (valores usados na geração dos dados), mas o valor do viés estimado de cada um dos demais parâmetros é pequeno. Mais importante ainda é que todas as estimativas dos intervalos de credibilidade de 95% contém os valores verdadeiros dos parâmetros.

(d) Avaliação do ajuste

A Figura 2.19 a seguir apresenta os gráficos dos valores da densidade condicional preditiva de validação cruzada, $[(CPO_i = \pi \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{\text{obs}} \right))]$, primeiramente em função de i , depois em função de X_i e também em função de Y_i , para $i = 1, \dots, 30$, onde CPO_i é estimada por meio da fórmula (2.10.8) adaptada para o caso do modelo sem restrições de identificabilidade. As estimativas basearam-se na segunda metade de uma seqüência de comprimento $T=10.000$ do algoritmo de Gibbs cujo ponto inicial é $\theta_1^{(0)}$.

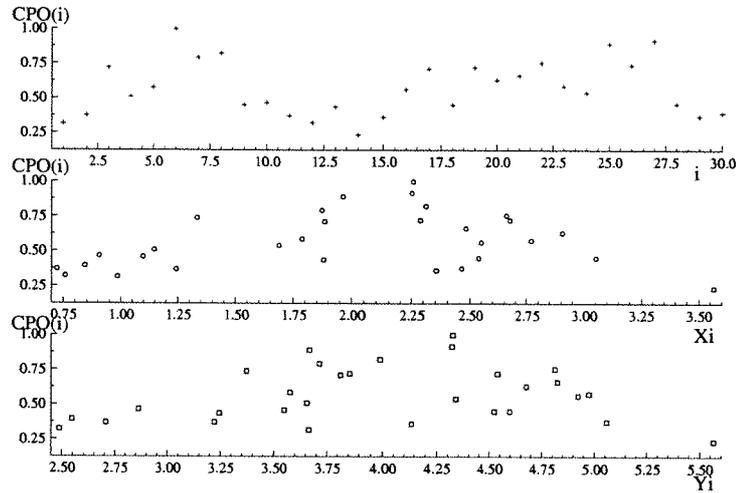


Figura 2.19: Gráficos dos valores da densidade condicional preditiva, $(CPO(i))$, versus i , versus X_i e versus Y_i , baseados na segunda metade de uma única seqüência de Gibbs de comprimento $T = 10.000$.

A figura 2.20 a seguir apresenta os gráficos das componentes d_{1i} e d_{2i} do vetor de resíduos bayesianos bidimensionais de validação cruzada

$$d_i = \begin{pmatrix} d_{1i} \\ d_{2i} \end{pmatrix} = \widehat{\text{Var}}^{-\frac{1}{2}} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{\text{obs}(-i)} \right) \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} - \hat{E} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{\text{obs}(-i)} \right) \right),$$

$i = 1, \dots, n$, dado na seção 2.10.2, onde as fórmulas de $\hat{E} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{\text{obs}(-i)} \right)$ e de $\widehat{\text{Var}}^{-\frac{1}{2}} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{\text{obs}(-i)} \right)$ foram adaptadas para o modelo sem restrição de identificabilidade a partir respectivamente das formulas (2.10.13) e (2.10.15).

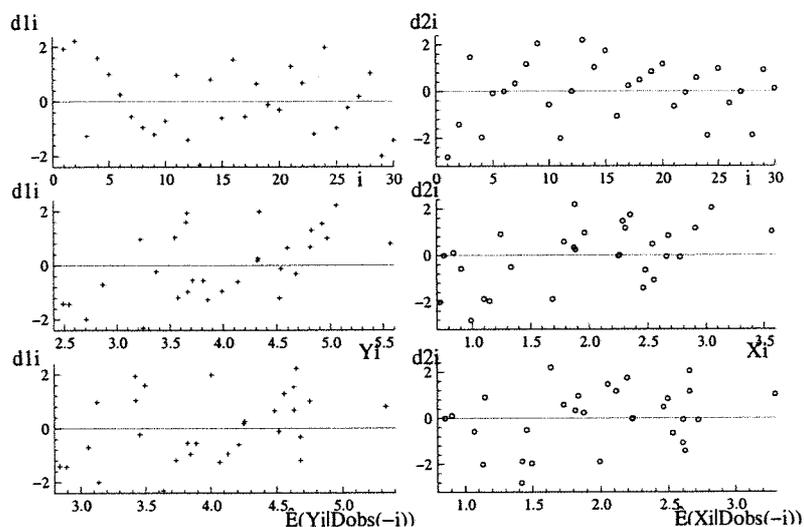


Figura 2.20: Os gráficos do lado esquerdo referem-se à primeira componente do vetor de resíduos (d_{1i}), e os do lado direito à segunda componente (d_{2i}). Eles estão baseados na mesma sequência de Gibbs utilizada na figura 2.19.

As figuras 2.19 e 2.20 não sugerem que haja problemas de falta de ajustamento, além disso, aparentemente a qualidade do ajuste do modelo sem restrição é melhor do que a do modelo com restrição de identificabilidade (ver fig. 2.6 e 2.7).

4. COMPARAÇÃO de MODELOS

A figura 2.21 compara os valores das CPO(i)'s dos modelos com e sem restrição de identificabilidade. A fórmula em (2.10.8) foi usada para o cálculo desses valores (no caso sem restr. de identif. essa fórmula foi empregada com as devidas adaptações). Apenas a segunda metade de uma sequência de Gibbs de comprimento $T=10000$ foi utilizada nos cálculos.

A tabela 2.7 a apresenta, os valores estimados das seguintes medidas de comparação de modelos: a medida L e a soma dos resíduos padronizados em valor absoluto (ver seção 2.11) para os modelos com e sem restrições de identificabilidade.

Na figura 2.21 vemos que os valores das CPO(i)'s tendem a ser maiores no modelo com restr. de identificabilidade e na tabela 2.7 as estimativas da medida L (baseadas no determinante) são consideravelmente menores no modelo com restr. de identif.,

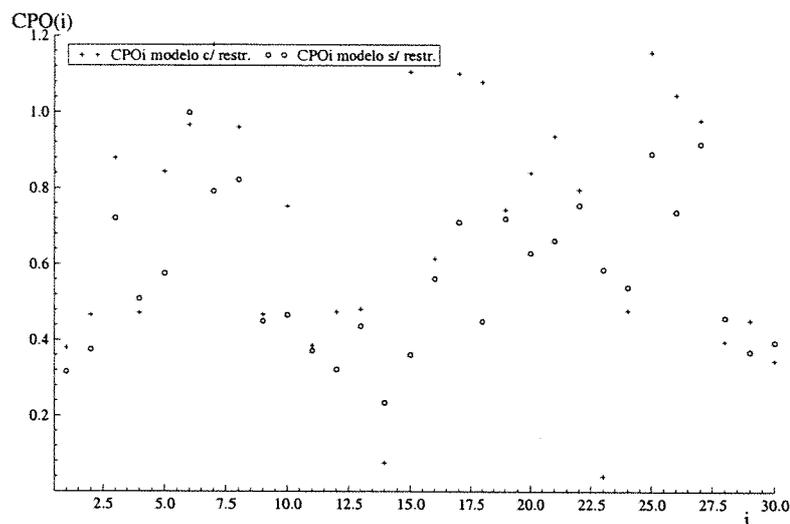


Figura 2.21: Gráfico dos valores da densidade condicional preditiva ($CPO(i)$) versus i , sob o modelo com restrição de identif. (+) e sob o modelo sem restrição de identif. (o)

Tabela 2.7: Estimativas da medida L e da soma das normas dos resíduos, nos modelos com e sem restrição de identificabilidade.

		Modelo com Restr.	Modelo sem Restr.	
medida L	det(L_2)	$\delta = 0.0$	161.43	221.19
		$\delta = 0.1$	183.74	248.74
		$\delta = 0.5$	286.02	372.0
		$\delta = 0.9$	409.18	516.14
		$\delta = 1.0$	443.24	555.44
	traço(L_2)	$\delta = 0.0$	41.73	43.39
		$\delta = 0.1$	45.23	46.89
		$\delta = 0.5$	59.24	60.89
		$\delta = 0.9$	73.24	74.90
		$\delta = 1.0$	76.74	78.40
Soma das normas dos resíduos		57.65	46.56	

o que sugere que este modelo deve ser o escolhido. Entretanto se considerarmos apenas a soma dos módulos dos resíduos padronizados como critério, o modelo sem restrição deveria ser o escolhido (É interessante notar que as estimativas dos valores da medida L calculadas pelo traço, ficaram bem próximas nos dois modelos).

Assim, se os dados amostrais são os da tabela 2.1 e o valor de λ não é conhecido, o modelo sem restr. de identif. (com distr. à priori vaga descrito nessa seção) não seria uma má escolha.

2.12.3 O modelo estrutural com restrição de identificabilidade e priori imprópria

1. DADOS

O conjunto de dados é o mesmo da Tabela 2.1.

2. MODELO PROPOSTO

Consideramos o modelo estrutural dado em (2.3.1) com distribuição à priori imprópria $\pi(\alpha, \beta, \mu, \sigma_x^2, \sigma^2)$ com $\alpha, \beta, \mu, \sigma_x^2$ e σ^2 “independentes”, onde α, β e μ são proporcionais a 1, $\pi(\sigma_x^2) \propto (\sigma_x^2)^{-r}$ e $\pi(\sigma^2) \propto (\sigma^2)^{-s}$, onde $r = .25$ e $s = 1.0$.

3. AJUSTE DO MODELO

(a) Pontos iniciais

Geramos 4 pontos iniciais usando procedimento semelhante ao da seção 2.12.1: $\theta_1^{(0)} = (2.4012, 0.8822, 2.0033, 0.1813, 0.1507)$; $\theta_2^{(0)} = (1.7943, 1.1077, 1.9321, 0.6855, 0.2603)$; $\theta_3^{(0)} = (1.1561, 1.3948, 2.1476, 0.2809, 0.0832)$ e $\theta_4^{(0)} = (1.6981, 1.1382, 1.7934, 0.3590, 0.0731)$.

A partir desses pontos geramos $M = 4$ seqüências de Gibbs de comprimento $T = 10000$ e uma seqüência mais longa de comprimento $T=50000$ partindo de $\theta_1^{(0)}$.

(b) Diagnósticos de convergência das seqüências de Gibbs

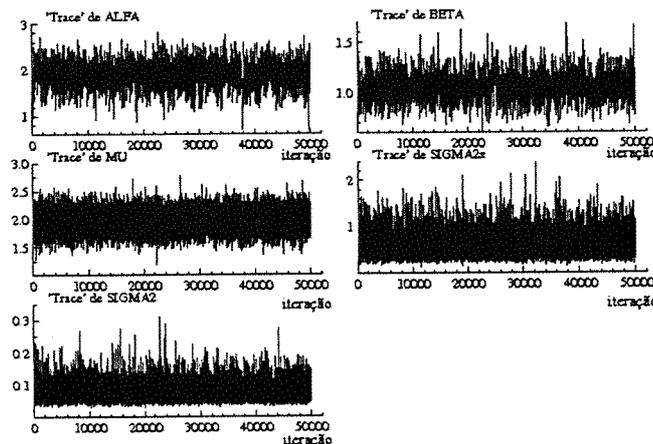


Figura 2.22: Gráficos das séries temporais de $\alpha, \beta, \mu, \sigma_x^2$ e σ^2 , baseados na seq de Gibbs de compr. $T = 50.000$.

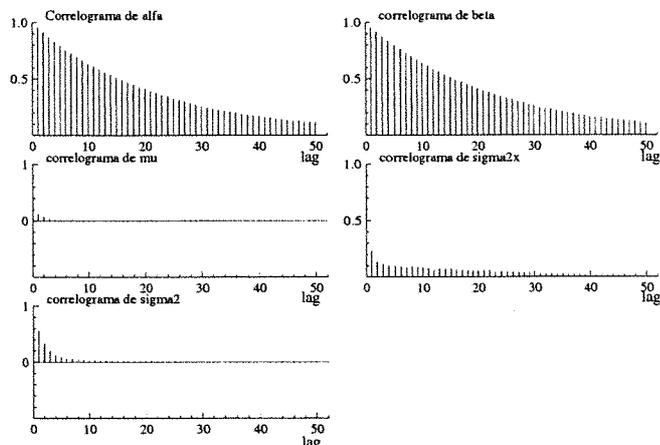


Figura 2.23: Gráficos das funções de auto-correlação de α , β , μ , σ_x^2 e σ^2 , baseados em uma única seqüência de Gibbs de comprimento $T = 50.000$.

Brooks & Gelman Multivariate Shrink Factors

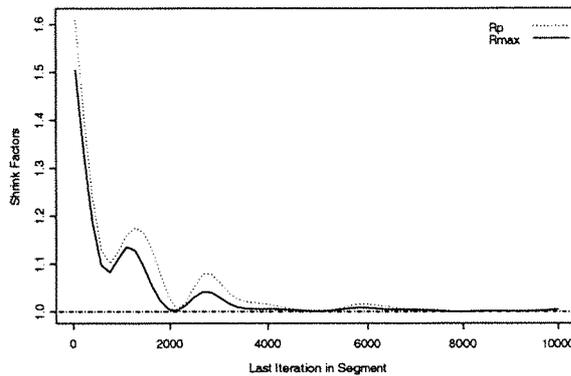


Figura 2.24: "Multivariate Potential Scale Reduction"(MPSRF) de Brooks e Gelman (1998) baseado em 4 seq. de Gibbs de compr. $T = 10000$ (R_{max} é o maior valor entre as medidas univariadas de $\alpha, \beta, \mu, \sigma_x^2, \sigma^2$).

Tabela 2.8: Medidas de Diagnóstico de convergência de Brooks, Gelman e Rubin

	Medida de Gelman&Rubin(PSRF)	Medida Corrigida(CSRF)	Quantil 0.975
α	1.004851	1.006815	1.017351
β	1.004639	1.006362	1.016433
μ	0.9999868	1.000022	1.000206
σ^2	1.000193	1.000693	1.001954
σ_x^2	1.000496	1.000298	1.000919

A medida multivariada (MPSRF) é igual a 1.0086177.

Gelman & Rubin Shrink Factors

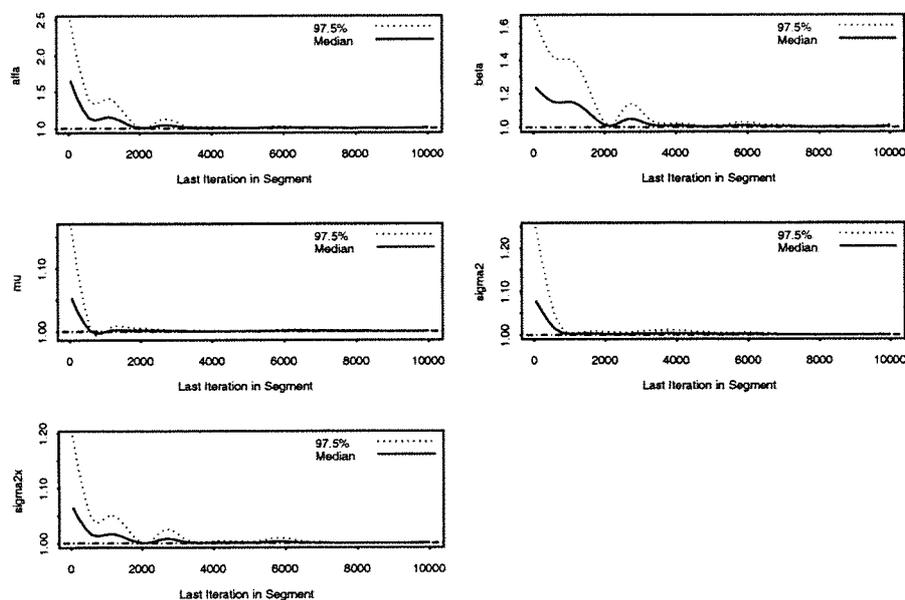


Figura 2.25: Gráfico da medida de Gelman e Rubin corrigida “Corrected potential scale reduction factors” (CSRPF) para cada parâmetro $\alpha, \beta, \mu, \sigma_x^2, \sigma^2$, baseada em $M = 4$ seqüências de Gibbs de comprimento $T = 10000$.

(c) Estimação

A tabela 2.9 apresenta as estimativas de várias medidas resumo da distribuição à posteriori e a figura 2.26 apresenta os gráficos das densidades posterioris marginais estimadas. Essas estimativas foram obtidas utilizando-se a segunda metade de cada uma das $M = 4$ seqüências de Gibbs de comprimento $T = 10.000$.

Antes de analisar os resultados, é importante lembrar que pelo fato da distribuição à priori ser imprópria (e portanto não informativa), as estimativas à posteriori dos parâmetros deveriam estar próximas dos valores dos parâmetros usados na geração dos dados se de fato o procedimento de estimação é adequado.

Na tabela 2.9 observa-se que a média aritmética e a média Rao-Blackelizada estão muito próximas, além disso, tanto a média como a mediana amostral estimam

cada um dos parâmetros do modelo com “pequeno” viés(estimado) e somente o parâmetro σ^2 parece estar um pouquinho subestimado). Mais importante ainda é que todos os intervalos estimados de credibilidade apresentados contém os valores verdadeiros dos parâmetros.

Tabela 2.9: Estimativas do valor esperado, quantis, desvio padrão, máximo,mínimo e intervalos de credibilidade de comprimento mínimo das distribuições marginais à posteriori de $\alpha, \beta, \mu, \sigma_x^2, \sigma^2$.

MEDIDAS RESUMO	α	β	μ	σ_x^2	σ^2
Média aritmética	1.9050	1.0542	1.9731	0.5792	0.0766
Média “R-Blackweliz.”	1.9047	1.0543	1.9735	0.5785	0.0765
Quantis amostrais					
0.025	1.3910	0.8500	1.6816	0.3004	0.0447
.05	1.4853	0.8827	1.7296	0.3312	0.0481
.1	1.5925	0.9198	1.7868	0.3714	0.0525
.5	1.7144	0.9611	1.8523	0.4221	0.0583
.25	1.7577	0.9767	1.8764	0.4450	0.0607
.3	1.7979	0.9905	1.8980	0.4659	0.0632
.4	1.8671	1.0169	1.9373	0.5056	0.0679
.5	1.9259	1.0434	1.9719	0.5475	0.0728
.6	1.9821	1.0717	2.0081	0.5909	0.0781
.7	2.0372	1.1062	2.0496	0.6439	0.0843
.75	2.0677	1.1250	2.0712	0.6756	0.0843
.8	2.0987	1.1463	2.0960	0.7147	0.0924
.9	2.1853	1.2049	2.1594	0.8286	0.1057
.95	2.2607	1.2579	2.2152	0.9422	0.1180
.975	2.3217	1.3008	2.2653	1.0453	0.1302
Interv.Credib. HPD					
$p = 0.95$ Lim. inf.	1.4267	0.8428	1.6763	0.2532	0.0395
Lim sup.	2.3504	1.2903	2.2594	0.9582	0.1198
$p = 0.90$ Lim. inf.	1.5215	0.8742	1.7351	0.2965	0.0432
Lim sup.	2.2877	1.2467	2.2195	0.8696	0.1088
desvio padrão (dp)	0.2366	0.11145	0.1480	0.1836	0.0223
d.p. Rao-Blackweliz.	0.2367	0.1143	0.1477	0.1924	0.0222
“Naive SE” = $\frac{dp}{\sqrt{M(T-T_0)}}$	0.001673	0.000810	0.001047	0.001362	0.000157
“Batch SE”	0.009851	0.004790	0.001356	0.003144	0.000331
Mínimo	0.7077	0.6796	1.3625	0.1459	0.0304
Máximo	2.6747	1.6771	2.7642	2.1438	0.2745

NOTA: mesma nota da tabela 2.3.

Comparando a tabela 2.9 com a tabela 2.3 notamos que as estimativas de Bayes dos parâmetros do modelo com priori imprópria tem menor viés estimado do que as estimativas de Bayes no modelo com priori própria vaga (da seção 2.12.1) (Isso já era esperado pelo fato de que no modelo com priori imprópria toda a informação vem somente dos dados). Observou-se também que os intervalos de credibilidade envolvendo o modelo com priori imprópria tendem a ter menor comprimento.

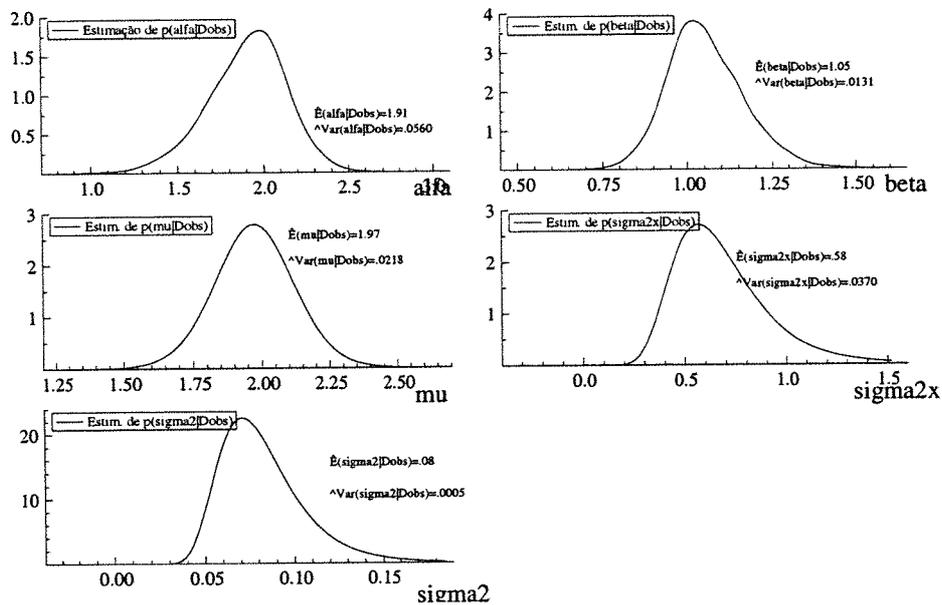


Figura 2.26: Gráficos das estimativas “Rao-Blackwelizadas” das densidades marginais à posteriori $p(\alpha|D_{obs}), \dots, p(\sigma^2|D_{obs})$.

(d) Avaliação do ajuste

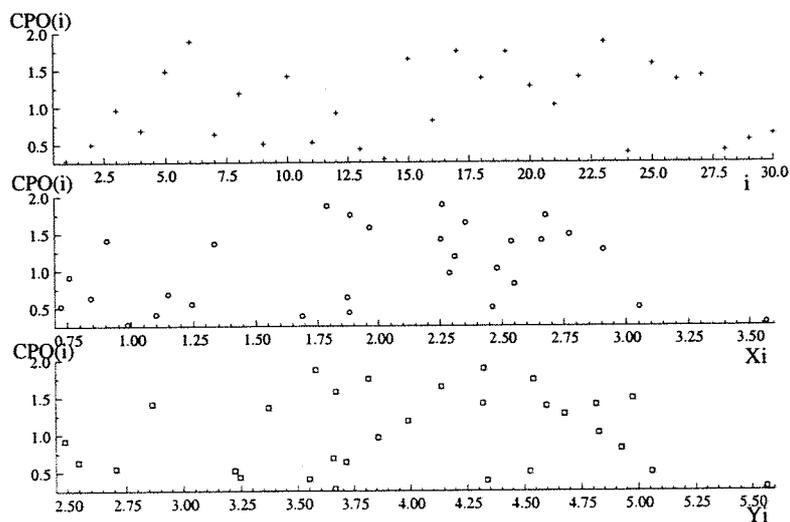


Figura 2.27: Gráficos dos valores da densidade condicional preditiva, $(CPO(i))$, versus i , versus X_i e versus Y_i , baseados na segunda metade de uma única seqüência de Gibbs de comprimento $T = 10.000$

Na figura 2.27 temos os gráficos da densidade condicional preditiva de validação cruzada, $[CPO_i = \pi \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{\text{obs}} \right)]$, em função de i , depois em função de X_i e também em função de Y_i , para $i = 1, \dots, 30$, onde CPO_i é estimada por meio da fórmula (2.10.8).

A Figura 2.28 apresenta os gráficos dos valores estimados de cada uma das componentes do vetor de resíduos bayesianos de validação cruzada (d_{1i} e d_{2i}) da seção 2.10.2(b). Esses gráficos estão baseados na mesma sequência de Gibbs utilizada na figura 2.27.

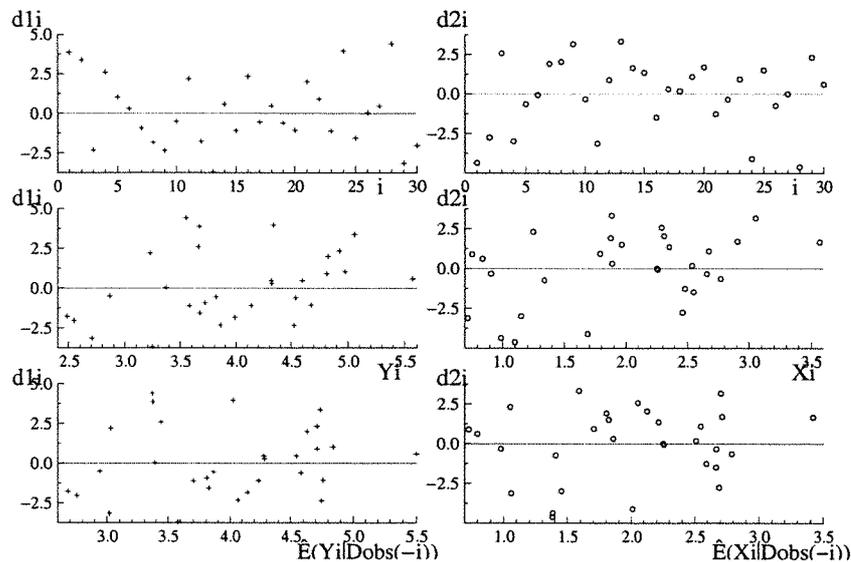


Figura 2.28: Do lado esquerdo estão os valores da componente (d_{1i}) dos resíduos bayesianos de validação cruzada em função de i , de Y_i e de $\hat{E}(Y_i | D_{\text{obs}(-i)})$, e do lado direito os valores de (d_{2i}) em função de i , de X_i e de $\hat{E}(X_i | D_{\text{obs}(-i)})$.

Na figura 2.28 os gráficos à esquerda possuem 4 ou 5 possíveis pontos aberrantes (cujas ordenadas são maiores do que 3 em valor absoluto) o que sugere que uma análise de resíduos mais detalhada deva ser realizada para examinar melhor o ajuste.

4. SENSIBILIDADE À λ – a razão das variâncias residuais

Nesse item analisamos descritivamente a variação nas estimativas dos valores es-

perados de α , β , μ , σ_x^2 e σ^2 à posteriori quando diferentes valores do hiperparâmetro λ são fixados (mantendo-se inalterados os demais hiperparâmetros do modelo).

Inicialmente nas figuras 2.29 e 2.30 exibimos gráficos com as médias “Rao-Blackwellizadas” dos parâmetros. (Para cada λ foi gerada uma seqüência de Gibbs de comprimento $T = 10.000$, cujo ponto inicial é o estimador de máxima verossimilhança correspondente, descartando-se a primeira metade de cada seqüência).

A seguir na figura 2.31 comparamos os gráficos baseados em dois diferentes tamanhos de amostra ($n = 50$, $n = 500$) geradas à partir do mesmo modelo e usando a mesma semente (ranseed=17 da linguagem OX). Para cada λ fixado, a seqüência de Gibbs foi gerada partindo-se da estimativa de máxima verossim. correspondente.

Levando-se em conta que a distribuição à priori é imprópria (e que portanto, se o procedimento é adequado, as estimativas à posteriori devem estar próximas dos valores dos parâmetros usados na geração dos dados) então das figuras 2.29 e 2.30 concluímos que:

- (1) A média amostral de μ parece não ser influenciada pelo valor de λ .
- (2) Quando λ é subestimado (isto é, quando $0 < \lambda < 1$), aparentemente, as médias amostrais de β , σ_x^2 e σ^2 superestimam respectivamente os valores verdadeiros de β , σ_x^2 e σ^2 , mas a média amostral de α subestima o verdadeiro valor de α .
- (3) Quando λ é superestimado ($\lambda > 1$), aparentemente α e σ_x^2 é superestimado e β e σ^2 são subestimados.. Isso ocorre mesmo quando λ é bem grande (da ordem de 50 vezes maior que o seu verdadeiro valor).

Da figura 2.31, temos também que as estimativas dos parâmetros aparentemente tendem a se estabilizar à medida que λ cresce independentemente do tamanho da amostra (n), sendo que o valor do viés estimado em geral é diferente para cada n .

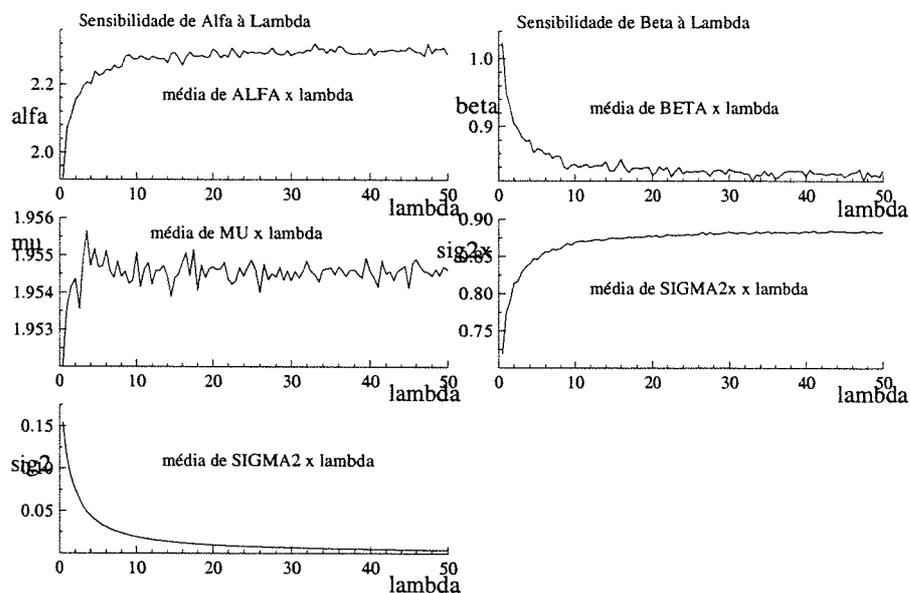


Figura 2.29: Gráficos das médias “R.-Blackwelizadas” de α , β , μ , σ_x^2 e σ^2 versus λ da distr. posteriori baseada nos dados da Tabela 2.1. A grade de pontos em λ é formada por 500 pontos igualmente espaçados no intervalo $[0,50]$.

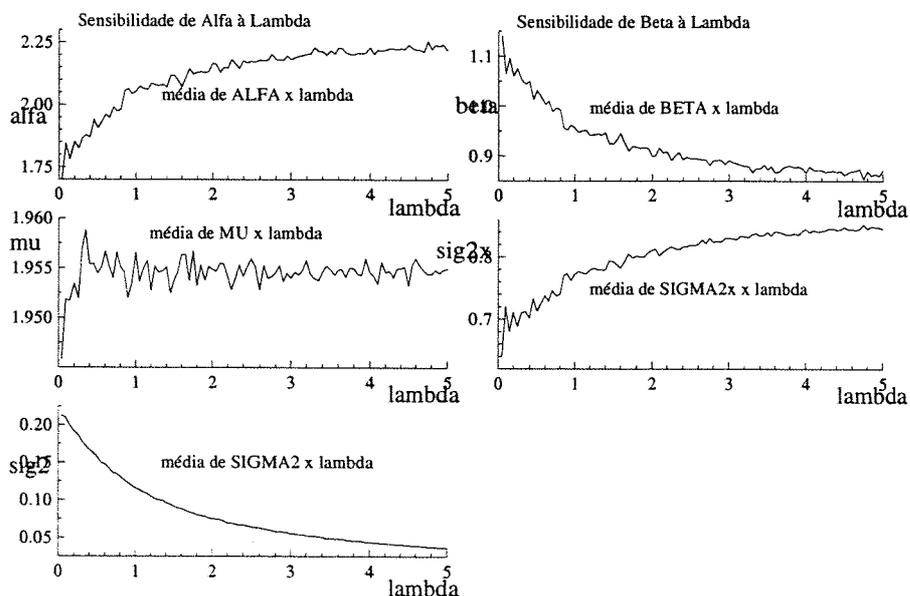


Figura 2.30: Idem à figura anterior, porém com grade em λ formada por 500 pontos igualmente espaçados no intervalo $[0,5]$.

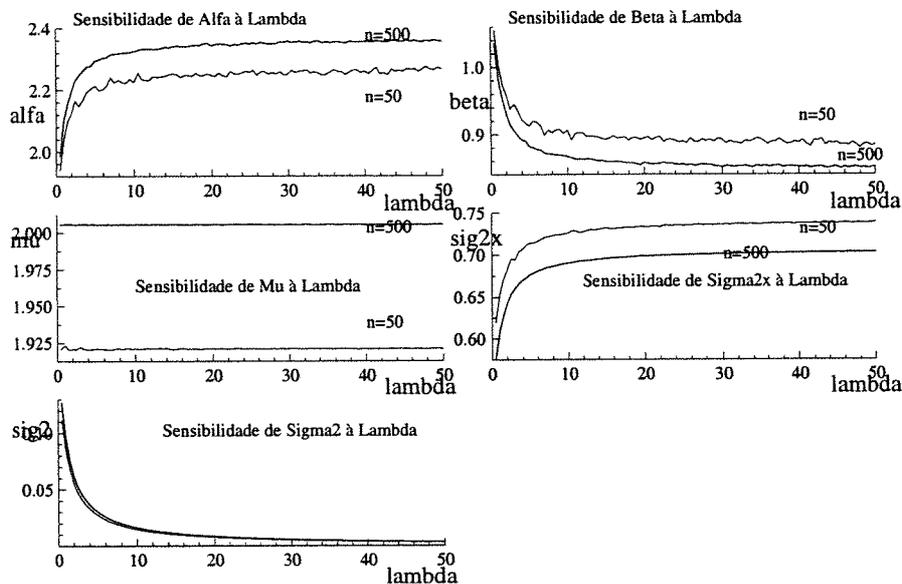


Figura 2.31: Gráfico das médias “Rao-Blackwelizadas” de α , β , μ , σ_x^2 e σ^2 versus λ , das posteriores baseadas nas amostras geradas de tamanho $n = 50$ e $n = 500$. A grade em λ é formada de 100 pontos igualmente espaçados no intervalo $[0,50]$ (os valores verdadeiros dos parâmetros são $\alpha = 2.0$, $\beta = 1.0$, $\mu = 2.0$, $\sigma_x^2 = 0.6$ e $\sigma^2 = 0.1$ e de λ é 1.0).

5. COMPARAÇÃO de MODELOS

Na figura 2.32 comparamos os valores estimados das $CPO(i)$'s no modelo com restrição de identificabilidade com priori própria e com priori imprópria (por meio da fórmula (2.10.8) e usando a segunda metade das seqüências de Gibbs de comprimento $T=10000$).

A tabela 2.10 apresenta os valores estimados das seguintes medidas de comparação de modelos: a medida L e a soma dos resíduos padronizados em valor absoluto (ver seção 2.11) para o modelo com priori própria e para o modelo com priori imprópria com restrição de identif. Os cálculos basearam-se sempre na segunda metade das seqüências de Gibbs de comprimento $T=10000$.

Na figura 2.32 vemos que os valores das $CPO(i)$'s tendem a ser maiores no modelo com priori imprópria e na tabela 2.10 as estimativas da medida L , baseadas no determinante, são menores no modelo com priori imprópria, indicando que o

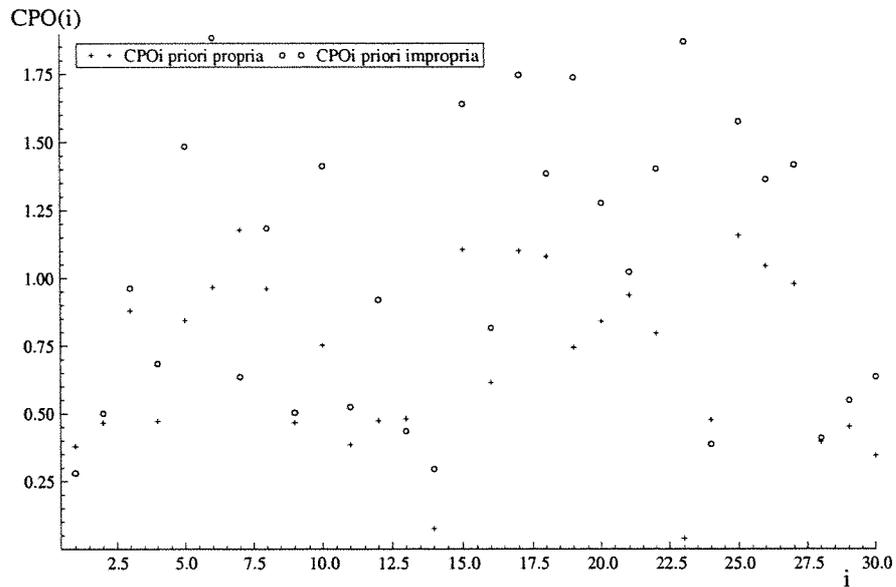


Figura 2.32: Valores da densidade condicional preditiva ($CPO(i)$) versus i , no modelo com priori própria (+) e no modelo com priori imprópria (o).

Tabela 2.10: As estimativas da medida L e da soma das normas dos resíduos, nos modelos com priori própria e com priori imprópria.

		priori própria.	priori imprópria.	
medida L	det(L_2)	$\delta = 0.0$	161.43	98.94
		$\delta = 0.1$	183.74	115.67
		$\delta = 0.5$	286.02	195.67
		$\delta = 0.9$	409.18	296.49
		$\delta = 1.0$	443.24	324.96
	tr(L_2)	$\delta = 0.0$	41.73	42.40
		$\delta = 0.1$	45.23	45.9
		$\delta = 0.5$	59.24	59.9
		$\delta = 0.9$	73.24	73.91
		$\delta = 1.0$	76.74	77.41
Soma das normas dos resíduos		57.65	76.24	

modelo com priori imprópria deve ser o escolhido. Entretanto se considerarmos a soma dos módulos dos resíduos como critério, o modelo com priori própria deveria ser o escolhido (É interessante notar também que os valores da medida L calculados pelo traço, embora numericamente menores no modelo com priori própria, estão muito próximos nos dois modelos).

2.12.4 O modelo estrutural sem restrição de identificabilidade e priori imprópria

1. DADOS

O conjunto de dados é o mesmo da Tabela 2.1

2. MODELO PROPOSTO

Consideramos o modelo estrutural dado em (2.6.1) com distribuição à priori $\pi(\alpha, \beta, \mu, \sigma_x^2, \sigma^2, \sigma_e^2)$ com componentes “independentes”, onde α , β e μ são proporcionais a 1, $\pi(\sigma_x^2) \propto (\sigma_x^2)^{-r}$, $\pi(\sigma^2) \propto (\sigma^2)^{-s}$, com $r = .25$ e $s = 0.5$, e $\sigma_e^2 \sim IG(2.001; 1.0)$.

3. AJUSTE DO MODELO

(a) Pontos iniciais

Usamos os mesmos pontos iniciais da seção 2.12.2, e partindo desses pontos geramos $M = 4$ seqüências de Gibbs de comprimento $T = 10.000$.

(b) Diagnósticos de convergência das seqüências de Gibbs

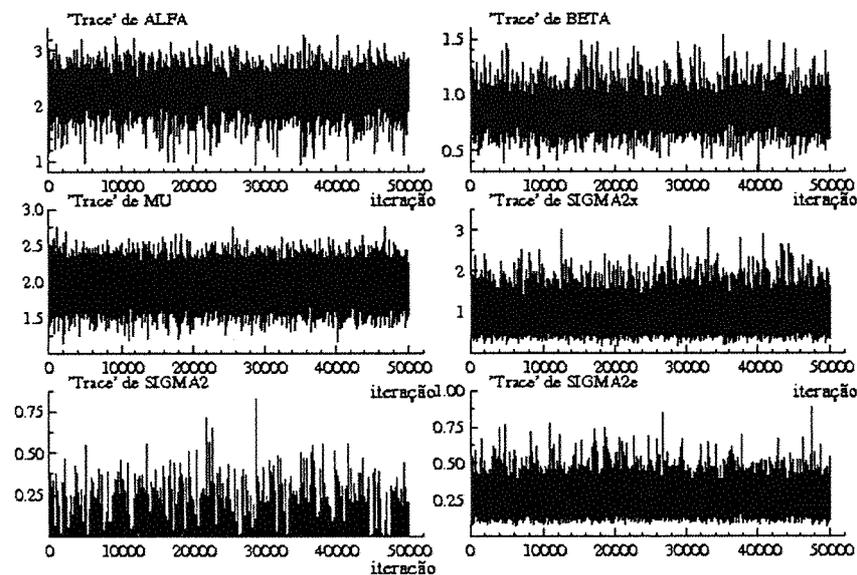


Figura 2.33: Gráficos das séries temporais (“trace”) de α , β , μ , σ_x^2 , σ^2 e σ_e^2 , baseados em uma única seqüência de Gibbs de comprimento $T = 50.000$.

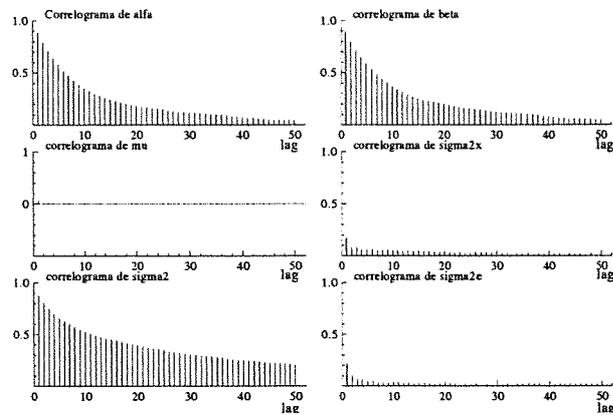


Figura 2.34: Gráficos das funções de auto-correlação de α , β , μ , σ_x^2 , σ^2 e σ_e^2 , baseados em uma única seqüência de Gibbs de comprimento $T = 50.000$.

Brooks & Gelman Multivariate Shrink Factors

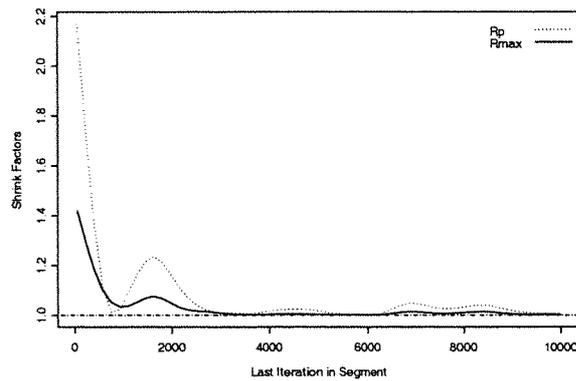


Figura 2.35: “Multivariate Potential Scale Reduction” (MPSRF) de Brooks e Gelman (1998) baseado em $M = 4$ seq. de Gibbs de comprimento $T = 10.000$ (R_{max} é o maior valor entre as medidas univariadas de α , β , μ , σ_x^2 , σ^2 , σ_e^2).

Tabela 2.11: Medidas de Diagnóstico de convergência de Brooks, Gelman e Rubin

	Medida de Gelman&Rubin(PSRF)	Medida Corrigida(CSRF)	Quantil 0.975
α	1.001738	1.002581	1.006482
β	1.001903	1.002636	1.006882
μ	1.000015	1.000045	1.000289
σ_x^2	1.000338	1.000524	1.001450
σ^2	1.003152	1.003986	1.010873
σ_e^2	1.00017	1.000260	1.000831

A medida multivariada (MPSRF) é igual a 1.0053917.

Gelman & Rubin Shrink Factors

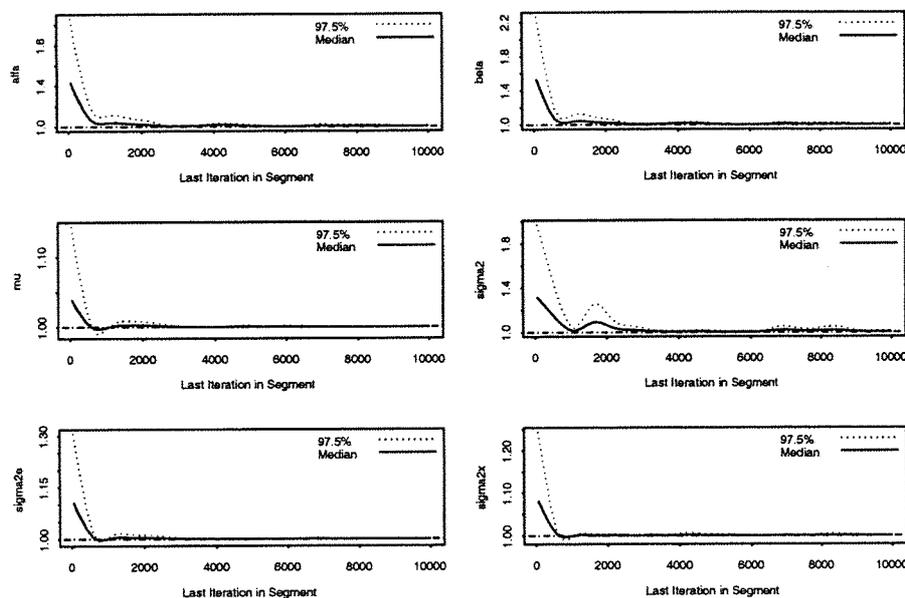


Figura 2.36: Gráfico da medida de Gelman e Rubin corrigida “Corrected potential scale reduction factors” (CSRf) para cada parâmetro $\alpha, \beta, \mu, \sigma_x^2, \sigma_e^2, \sigma^2$, baseada em $M = 4$ seqüências de Gibbs de comprimento $T = 10.000$.

(c) Estimação

Na tabela 2.12 apresentamos as estimativas de várias medidas resumo da distribuição à posteriori e na figura 2.37, os gráficos das densidades posteriores marginais estimadas. Essas estimativas baseiam-se em $M = 4$ seqüências de Gibbs de comprimento $T = 10.000$, descartando-se a primeira metade de cada uma.

Observa-se na tabela 2.12 que a média aritmética e a média Rao-Blackelizada estão muito próximas. A média e a mediana também estão bastante próximas (exceto no caso do parâmetro σ^2), além disso, elas superestimam o verdadeiro valor de σ_e^2 e subestimam o valor de σ^2 . Nesse modelo, entretanto, os intervalos estimados de credibilidade de 95% e 90% para o parâmetro σ^2 não contém o verdadeiro valor de σ^2 .

Comparando a tabela 2.12 com a tabela 2.6 nota-se que embora as estimativas de Bayes no modelo com priori imprópria têm maior viés do que as estimativas de Bayes no modelo com priori própria vaga (da seção 2.12.2), os intervalos de credibilidade de α , β , μ e σ_x^2 , no modelo com priori imprópria tem menor comprimento. O problema é que σ^2 aparentemente não foi bem estimado (como já foi comentado).

Tabela 2.12: Estimativas do valor esperado, quantis, desvio padrão, máximo, mínimo e intervalos de credibilidade de comprimento mínimo das distribuições marginais à posteriori de $\alpha, \beta, \mu, \sigma_x^2, \sigma^2, \sigma_e^2$.

MEDIDAS RESUMO	α	β	μ	σ_x^2	σ^2	σ_e^2
Média aritmética	2.2181	0.8722	1.9529	0.8375	0.0477	0.2322
Média "R-Blackweliz."	2.2186	0.8717	1.9544	0.8390	0.0476	0.2315
Quantis amostrais						
0.025	1.6050	0.6410	1.6061	0.4355	0.0000	0.1264
.05	1.7499	0.6789	1.6681	0.4883	0.0003	0.1387
.1	1.8886	0.7211	1.7340	0.5497	0.0009	0.1544
.2	2.0198	0.7713	1.8105	0.6251	0.0032	0.1751
.25	2.0666	0.7887	1.8416	0.6547	0.0050	0.1840
.3	2.1069	0.8044	1.8654	0.6844	0.0072	0.1917
.4	2.1745	0.8349	1.9119	0.7396	0.0132	0.2070
.5	2.2361	0.8632	1.9533	0.7956	0.0225	0.2222
.6	2.2953	0.8924	1.9961	0.8585	0.0355	0.2387
.7	2.3586	0.9251	2.0406	0.9317	0.0546	0.2573
.75	2.3923	0.9432	2.0672	0.9757	0.0674	0.2694
.8	2.4312	0.9648	2.0949	1.0257	0.0822	0.2821
.9	2.5348	1.0315	2.1704	1.1776	0.1305	0.3212
.95	2.6215	1.0954	2.2368	1.3232	0.1785	0.3597
.975	2.7048	1.1668	2.2940	1.4683	0.2224	0.3953
Interv. Credib. HPD						
$p = 0.95$ Lim. inf.	1.6487	0.6188	1.6122	0.3929	0.0000	0.1128
Lim sup.	2.7330	1.1334	2.2982	1.3853	0.1785	0.3692
$p = 0.90$ Lim. inf.	1.8043	0.6624	1.6701	0.4290	0.0000	0.1208
Lim sup.	2.6711	1.0736	2.2376	1.2331	0.1305	0.3319
desvio padrão (dp)	0.2721	0.1304	0.1731	0.2641	0.0629	0.0695
d.p. Rao-Blackweliz.	0.2721	0.1303	0.1718	0.2659	0.0623	0.0682
"Naive SE" = $\frac{dp}{\sqrt{M(T-T_0)}}$	0.001924	0.00092220	0.001224	0.001867	0.000444	0.000491
"Batch SE"	0.008153	0.003976	0.001393	0.003726	0.002183	0.000822
Mínimo	0.9371	0.4045	1.2028	0.1684	0.0000	0.0735
Máximo	3.2785	1.5371	2.7565	3.0913	0.8244	1.8424

NOTA: Mesma nota da tabela 2.6.

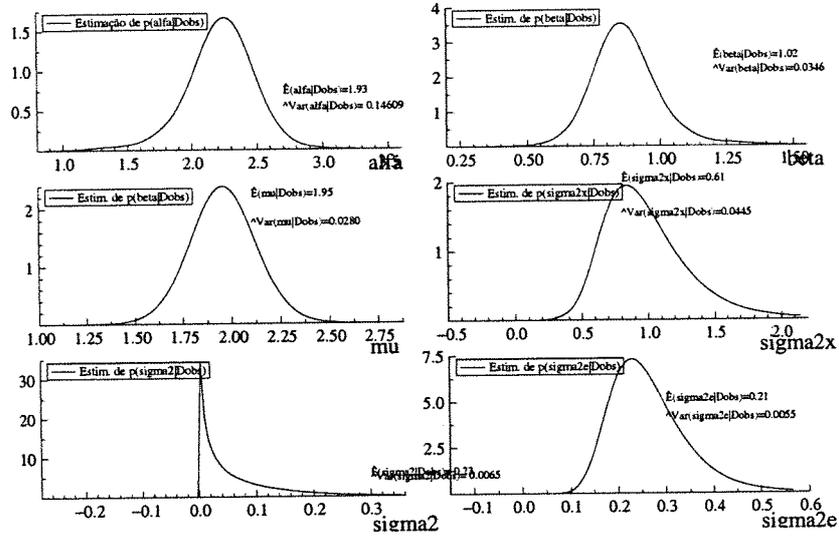


Figura 2.37: Gráficos das estimativas “Rao-Blackwelizadas” das densidades marginais à posteriori $p(\alpha|D_{obs}), \dots, p(\sigma_e^2|D_{obs})$.

(d) Avaliação do ajuste

A Figura 2.38 a seguir apresenta os gráficos dos valores da densidade condicional preditiva de validação cruzada ($CPO_i = \pi\left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{obs}\right)$) primeiramente em função de i , depois em função de X_i e também em função de Y_i , para $i = 1, \dots, 30$, onde CPO_i é estimada por meio da fórmula (2.10.8) adaptada para o modelo dessa subseção. As estimativas basearam-se na segunda metade de uma seqüência de comprimento $T=10.000$ do algoritmo de Gibbs cujo ponto inicial é $\theta_1^{(0)}$.

A figura 2.39 a seguir apresenta os gráficos das componentes d_{1_i} e d_{2_i} do vetor de resíduos bayesianos bidimensionais de validação cruzada

$$d_i = \begin{pmatrix} d_{1_i} \\ d_{2_i} \end{pmatrix} = \widehat{Var}^{-\frac{1}{2}} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{obs(-i)} \right) \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} - \hat{E} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{obs(-i)} \right) \right),$$

$i = 1, \dots, n$, dado na seção 2.10.2, onde as fórmulas de $\hat{E} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{obs(-i)} \right)$ e de $\widehat{Var}^{-\frac{1}{2}} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{obs(-i)} \right)$ foram adaptadas para o modelo dessa subseções a partir respectivamente das formulas (2.10.13) e (2.10.15).

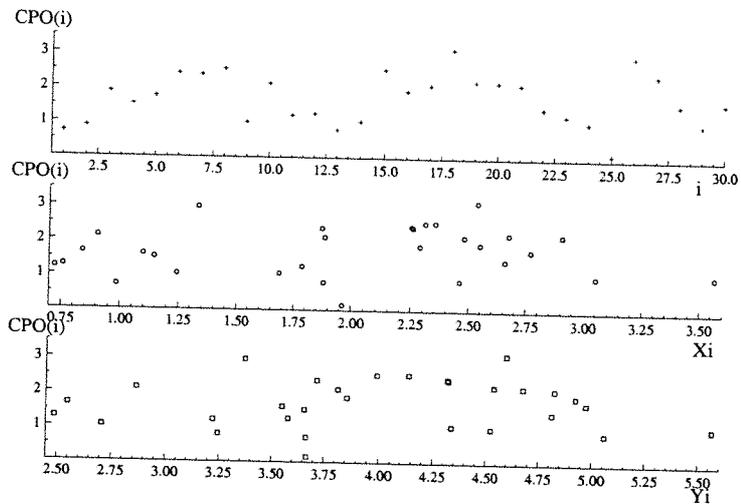


Figura 2.38: Gráficos dos valores da densidade condicional preditiva, $(CPO(i))$, versus i , versus X_i e versus Y_i .

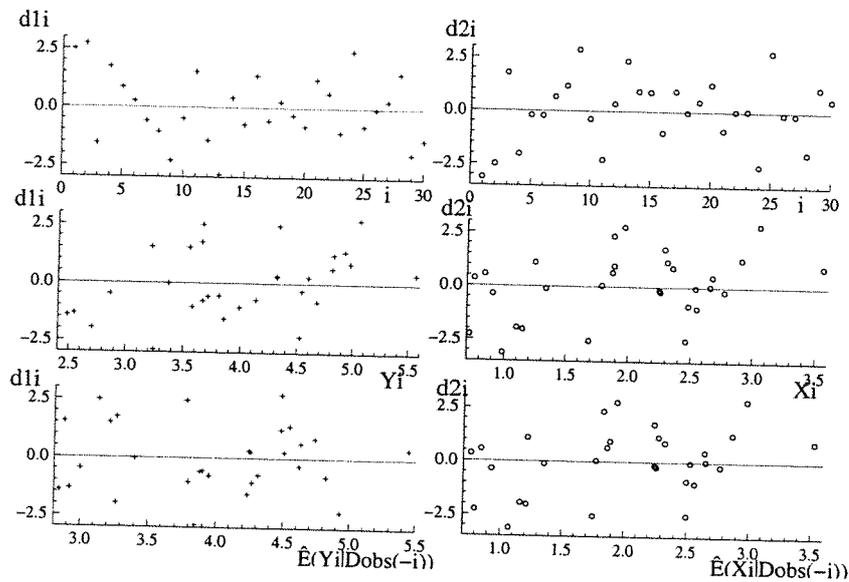


Figura 2.39: Os gráficos do lado esquerdo referem-se à primeira componente do vetor de resíduos (d_{1i}) , e os do lado direito à segunda componente (d_{2i}) . Eles estão baseados na mesma sequência de Gibbs utilizada na figura 2.38.

Obs: As figuras 2.38 e 2.39 não sugerem que haja problemas de falta de ajustamento.

4. COMPARAÇÃO de MODELOS

A figura 2.40 compara a densidade condicional preditiva nos modelos com priori própria e com priori imprópria sem restrição de identif. (usando a fórmula (2.10.8))

com as devidas adaptações). Para os cálculos, apenas a segunda metade de uma seqüência de Gibbs de comprimento $T=10000$ foi utilizada.

A tabela 2.13 apresenta, os valores estimados da medida L e da soma dos resíduos padronizados em valor absoluto nos modelos com priori própria e com priori imprópria sem restr.de identif.

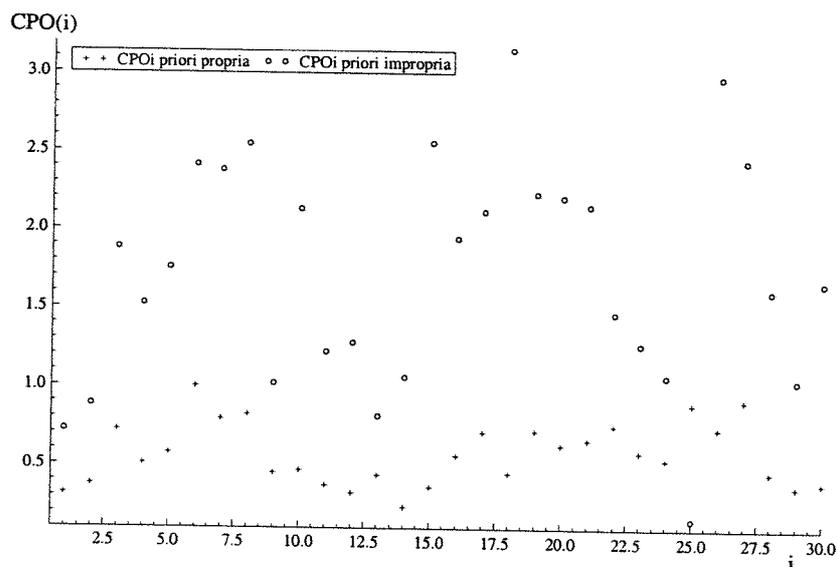


Figura 2.40: Densidade condicional preditiva ($CPO(i)$) versus i , no modelo com priori própria (+) e no modelo com priori imprópria (o).

Tabela 2.13: A medida L , o produto $\prod_{i=0}^n CPO_i$ e a soma das normas dos resíduos, nos modelos c/ priori própria e c/ priori imprópria.

		priori própria.	priori imprópria.
medida L	$\det(L_2)$		
	$\delta = 0.0$	221.19	137.66
	$\delta = 0.1$	248.74	157.58
	$\delta = 0.5$	372.0	250.30
	$\delta = 0.9$	516.14	363.91
	$\delta = 1.0$	555.44	395.68
	$\text{tr}(L_2)$		
	$\delta = 0.0$	43.39	44.12
	$\delta = 0.1$	46.89	47.62
	$\delta = 0.5$	60.89	61.63
$\delta = 0.9$	74.90	75.63	
$\delta = 1.0$	78.40	79.13	
Soma das normas dos resíduos		46.56	52.25

Na figura 2.40 vemos que os valores das $CPO(i)$'s tendem a ser maiores no modelo com priori imprópria e na tabela 2.14 as estimativas da medida L , baseadas no determinante, são consideravelmente menores no modelo com priori imprópria. Isso sugere que o modelo com priori imprópria deva ser o escolhido. Entretanto se considerarmos como critério, apenas a soma dos módulos dos resíduos padronizados então o modelo com priori própria deveria ser o escolhido. Observou-se também que os valores da medida L calculados pelo traço ficaram bem próximos nos dois modelos.

Assim, se o valor de λ (quociente das variâncias residuais) não é conhecido e os dados amostrais são os da tabela 2.1, o modelo (sem restr. de identif.) com priori imprópria não seria uma má escolha.

2.12.5 Resumo dos resultados:

Na seção 2.12 foram ajustados 4 modelos bayesianos estruturais normais utilizando o particular conjunto de dados da tabela 2.1. Os modelos são:

- 1) modelo com restr. de identif. e priori própria vaga da subseção(2.12.1) ,
- (2) modelo sem restr. de identif. com priori própria vaga (2.12.2),
- (3) modelo com restr. de identif. e priori imprópria (2.12.3)
- (4) modelo sem restr. com priori imprópria (2.12.4).

1 – comparação dos algoritmos:

Foram utilizados 3 diferentes algoritmos para amostrar da posteriori do modelo da subseção 2.12.1: O algoritmo de Gibbs simples, o algoritmo “Grouped Gibbs” e o algoritmo “M. Collapsed Gibbs”. Aparentemente o “Grouped Gibbs” não traz vantagens sobre o primeiro algoritmo e existe apenas um pequeno ganho (em termos da função de autocorrelação e da matriz de correlações cruzadas) em usar o “M.Collap. Gibbs” sobre os demais , não justificando fortemente o seu uso (nesse

particular exemplo).

2 – convergência

Os dispositivos de diagnóstico de convergência apresentados nessa seção não nos sugeriram que haja problemas na convergência do algoritmo de Gibbs em nenhum dos modelos considerados.

3 – comparação dos modelos:

– Comparando os modelos (1) e (3):

Levando em conta que no modelo (3) os intervalos de credibilidade tem menor comprimento, as estimativas dos parâmetros tem um menor viés de estimação, as estimativas da medida L (calculadas pelo determinante) são menores e os valores das CPO(i)'s tendem a ser maiores do que no modelo (1), então o modelo (3) é o escolhido.

– Comparando os modelos (2) e (4):

Do ponto de vista de estimação os dois tem vantagens e desvantagens (o modelo (4) tem intervalos de credibilidade de menor comprimento, porém maior viés de estimação e além disso o verdadeiro valor do parâmetro σ^2 não pertence ao intervalo de credibilidade estimado correspondente). Entretanto no modelo (4) as estimativas da medida L (calculadas pelo determinante) são menores do que no modelo (2) e os valores das CPO(i)'s tendem a ser maiores do que no modelo (2). Por estes motivos uma possível escolha seria o modelo (4).

Do exposto acima vemos que o uso de prioris impróprias (das seções 2.12.3 e 2.12.4) traz benefícios tanto no caso do modelo com restrição como no caso sem restrição de identificabilidade.

– Comparando os modelos (1) e (2) e os modelos (3) e (4):

Usando a medida L e os valores das CPO(i)'s vemos que o modelo (1) é escolhido sobre o modelo (2); além disso o modelo (3) é preferido ao (4).

4 – qualidade do ajuste

Nos modelos sem restrição de identificabilidade (com priori própria ou imprópria), os gráficos das CPO(i)'s e das componentes dos vetores de resíduos, não nos sugeriram que haja algum problema de falta de ajuste. Entretanto nos modelos com restr. de identif. (especialmente o modelo com priori imprópria), ocorrem alguns pontos aberrantes, o que sugere que uma análise mais completa deva ser realizada.

5 – Sensibilidade das estimativas dos parâmetros ao valor de λ

Quando a amostra de dados é a da tabela 2.1 e o modelo escolhido é o modelo (1), o valor de λ fornecido pelo usuário não afeta muito as estimativas de α , β , μ e σ_x^2 em termos do viés (já que o valor máximo observado do viés estimado, foi da ordem de 15 a 20% do verdadeiro valor do parâmetro correspondente). Entretanto a escolha do λ pode afetar bastante as estimativas de σ^2 (também em termos do viés de estimação) e nesse caso observou-se que mesmo que λ seja escolhido corretamente, ainda assim esse parâmetro fica bastante superestimado, principalmente se o tamanho da amostra não for bem grande.

Observações gerais:

- As médias aritméticas e as médias R-Blackwelizadas ficaram sempre muito próximas independentemente do modelo estudado.
- A medida L baseada no traço aparentemente não serviu para diferenciar os modelos.
- O critério da medida L (calculada pelo determinante) e o critério da comparação dos CPO(i)'s concordaram quanto à escolha dos modelos, entretanto discordaram do critério da soma dos resíduos padronizados na escolha dos modelos estudados nessa seção.

Capítulo 3

O modelo linear estrutural identificável t -Student com erros nas variáveis

Introdução

É fato bastante conhecido que os modelos que usam a distribuição Normal não são robustos à presença de valores discrepantes, no sentido de que as inferências sob tais modelos podem mudar consideravelmente na presença de uma ou mais observações discrepantes.

Uma maneira de reduzir a influência desses valores discrepantes é trocar o modelo Normal por um modelo com caudas mais pesadas como por exemplo o modelo t -Student.

A distribuição t -Student com ν graus de liberdade, $t_\nu(\mu, \Sigma)$, tem caudas mais longas que a Normal ($N(\mu, \Sigma)$) e além disso, quando o número de graus de liberdade tende a infinito ($\nu \rightarrow \infty$), ela se aproxima da Normal e quando $\nu = 1$ ela coincide com a distribuição de Cauchy. Assim, na literatura, uma alternativa robusta à Normal tem sido a distribuição t , fixando-se um valor pequeno para ν . Gelman et al. (1997) sugerem utilizar $\nu = 4$ para assegurar a existência dos momentos até terceira ordem e recomendam também que a distribuição Normal não deve ser usada

quando os dados são “muito dispersos” (isto é, quando 10% dos pontos estiverem há uma distância superior a $1,5d$ da mediana, onde d é a distância entre o primeiro e o terceiro quartil, e o tamanho da amostra é grande). Além disso, segundo esses autores, nos modelos de regressão, a distribuição t com um número pequeno de graus de liberdade fixado, trata os “outliers” como observações de alta variância e produz resultados semelhantes aos obtidos na regressão ponderada (em que os pesos são os inversos das variâncias).

A distribuição t pode ser empregada não somente para acomodar eventuais pontos discrepantes (ou dados com alta dispersão), mas também situações em que os hiperparâmetros das prioris têm valores muito extremos ou então quando temos um modelo com várias hierarquias.

Nesse capítulo estudamos os modelos lineares estruturais t -independente e t -dependente com restrição de identificabilidade no caso em que o número de graus de liberdade é conhecido, considerando uma particular distribuição à priori própria com componentes independentes.

Nesse capítulo também são demonstrados quatro teoremas que estabelecem condições suficientes para a existência da distribuição posteriori no modelo t -dependente para um conjunto particular de prioris impróprias.

Uma avaliação da qualidade do ajuste e uma análise de sensibilidade do número de graus de liberdade do modelo t (quando este é assumido conhecido) também serão feitas nesse capítulo. Finalmente, descreveremos algumas técnicas de comparação de modelos.

Antes porém de apresentar os modelos de regressão estrutural t -dependente e t -independente, apresentamos a seguir duas das formas de definir a distribuição t multivariada descritas na literatura. Um estudo bastante abrangente sobre a distribuição t , situando-a também como membro da família de distribuições elípticas,

é encontrado em Arellano-Valle (1994).

Definição 1: Dizemos que um vetor aleatório $X_{n \times 1}$ tem distribuição t n -variada e não-singular com ν graus de liberdade ($\nu > 0$), parâmetro de locação $\mu_{n \times 1}$ e parâmetro de escala $\Sigma_{n \times n}$ (Σ positiva definida), isto é, $X \sim t_n(\mu, \Sigma, \nu)$, se a função densidade de X é dada por

$$f_X(x|\mu, \Sigma, \nu) = k(n, \nu) |\Sigma|^{-\frac{1}{2}} \left[\nu + (X - \mu)^\top \Sigma^{-1} (X - \mu) \right]^{-\frac{(n+\nu)}{2}},$$

$x \in \mathbb{R}^n$, $\mu \in \mathbb{R}^n$, $\Sigma > 0$ (isto é, Σ positiva definida), onde $k(n, \nu) = \frac{\Gamma(\frac{n+\nu}{2}) \nu^{\nu/2}}{\Gamma(\frac{\nu}{2}) \pi^{n/2}}$.

Observação 1: Quando $\nu = 1$ a distribuição acima se reduz à distribuição de Cauchy n -variada (não central).

Observação 2: Quando $\nu \rightarrow \infty$ essa distribuição se aproxima de uma Normal $N(\mu, \Sigma)$.

Observação 3: Essa definição assume explicitamente a existência de uma função densidade e portanto não permite incorporar uma versão singular da distribuição t multivariada.

Definição 2: Dizemos que um vetor aleatório $X_{n \times 1}$ segue uma distribuição t n -variada com ν graus de liberdade ($\nu > 0$), parâmetro de locação μ e parâmetro de escala Σ , isto é, $X \sim t_n(\mu, \Sigma, \nu)$, se $X = \mu + w^{\frac{1}{2}} Z$, onde $Z \sim N_n(0, \Sigma)$ e $w \sim \frac{\nu}{\chi^2_\nu}$ são independentes, isto é, $X|w \sim N(\mu, w\Sigma)$ e $w \sim IG(\frac{\nu}{2}, \frac{\nu}{2})$.

Observação 1: A distribuição incondicional de X nesse caso é

$$f_X(x) = \int_0^\infty f(x|w) f(w) dw,$$

que é uma mistura de escala da distribuição normal.

Observação 2: No caso em que Σ é positiva definida, a Definição 1 e a Definição 2 são equivalentes.

Nesse capítulo apresentaremos o modelo de regressão estrutural t -independente e o modelo t -dependente, tanto no caso em que ν é conhecido como também no caso em que ν é desconhecido.

3.1 O modelo estrutural t -independente com o número de graus de liberdade conhecido

Nessa seção apresentamos o modelo de regressão linear simples t -independente com erros nas variáveis, considerando duas formulações equivalentes. Além disso, discutimos a implementação de algoritmos MCMC para amostrar da posteriori desses modelos.

3.1.1 Definição do modelo

Formulação 1

Considere as equações

$$\begin{cases} Y_i = \alpha + \beta x_i + e_i, \\ X_i = x_i + u_i, \end{cases}$$

onde Y_i e X_i são as variáveis aleatórias observáveis, x_i a variável aleatória não observável e u_i o erro na medição de x_i , com a suposição

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \Big| \theta \stackrel{\text{iid}}{\sim} t_3(\mu', \Sigma', \nu), \quad i = 1, \dots, n,$$

onde

$$\mu' = \begin{pmatrix} 0 \\ 0 \\ \mu \end{pmatrix} \quad \text{e} \quad \Sigma' = \begin{pmatrix} \sigma_e^2 & 0 & 0 \\ 0 & \sigma_u^2 & 0 \\ 0 & 0 & \sigma_x^2 \end{pmatrix},$$

onde $\sigma_e^2 = \lambda \sigma_u^2 = \lambda \sigma^2$ com λ conhecido, $\nu > 0$ e $|\Sigma'| \neq 0$.

O vetor de parâmetros é $\theta = (\alpha, \beta, \mu, \sigma_x^2, \sigma^2)^T$ (3.1.1)

e o espaço paramétrico é $\Theta = \{\theta : \alpha, \beta, \mu \in \mathbb{R} \text{ e } \sigma_x^2 \text{ e } \sigma^2 \in \mathbb{R}_+\}$.

Observação: Quando $\nu \geq 3$, a distribuição acima tem matriz de covariâncias dada por $Var\left(\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} | \theta\right) = \Sigma' \frac{\nu}{\nu-2}$, e vetor de médias μ' . Quando $\nu = 1$ caímos na distribuição de Cauchy não central com parâmetro de locação igual a μ' e parâmetro de escala igual a Σ' .

Tomando-se $Z_i = \begin{pmatrix} Y_i \\ X_i \\ x_i \end{pmatrix}$, $i = 1, \dots, n$, então a suposição em (3.1.1) implica que

$$Z_i | \theta \stackrel{\text{iid}}{\sim} t_3(\mu, \Sigma, \nu), \quad i = 1, \dots, n,$$

onde

$$\mu = \begin{pmatrix} \alpha + \beta\mu \\ \mu \\ \mu \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \beta^2\sigma_x^2 + \sigma_e^2 & \beta\sigma_x^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma_u^2 & \sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 & \sigma_x^2 \end{pmatrix}, \quad (3.1.2)$$

onde $\sigma_e^2 = \lambda\sigma_u^2 = \lambda\sigma^2$, $\nu > 0$ e Σ é positiva definida.

Notemos que se a suposição em (3.1.2) vale, então

$$Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} \sim t_{3n}(\mathbf{1}_n \otimes \mu; I_{n \times n} \otimes \Sigma, \nu)$$

(mas a recíproca não é verdadeira).

A distribuição à priori de θ nesse modelo será denotada por $\pi(\theta)$.

Observações

1. Arellano-Valle (1994) define o modelo estrutural t -independente no contexto da estatística clássica por meio da distribuição das variáveis observadas U_i

definidas por

$$\underline{U}_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix}, \text{ onde } \underline{U}_i | \underline{\theta} \stackrel{\text{iid}}{\sim} t_2(\underline{\mu}^*, \underline{\Sigma}^*, \nu), \quad i = 1, \dots, n,$$

onde

$$\underline{\mu}^* = \begin{pmatrix} \alpha + \beta\mu \\ \mu \end{pmatrix}, \quad \underline{\Sigma}^* = \begin{pmatrix} \beta^2\sigma_x^2 + \sigma_e^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma_u^2 \end{pmatrix}. \quad (3.1.3)$$

2. O modelo t -independente é não identificável e uma maneira de torná-lo identificável é incorporar por exemplo a restrição $\sigma_e^2 = \lambda\sigma_u^2 = \lambda\sigma^2$ (Arellano-Valle, 1994).
3. Mesmo no caso em que o número de graus de liberdade (ν) é desconhecido, o modelo se torna identificável quando adotamos por exemplo a restrição acima.
4. Em estatística bayesiana, como já discutimos no Capítulo 2, não há necessidade de que os modelos sejam identificáveis quando escolhermos convenientemente a distribuição à priori. Entretanto, nesse capítulo vamos estudar apenas o modelo t identificável.

A função de verossimilhança baseada nos dados completos

A função de verossimilhança baseada nos dados completos, $D = \underline{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$, é dada por

$$L(\underline{\theta} | D) \propto \prod_{i=1}^n |\underline{\Sigma}|^{-\frac{1}{2}} \left[\nu + (\underline{Z}_i - \underline{\mu}) \underline{\Sigma}^{-1} (\underline{Z}_i - \underline{\mu}) \right]^{-\left(\frac{\nu+1}{2}\right)}, \quad (3.1.4)$$

onde $\underline{\mu}$ e $\underline{\Sigma}$ são definidos em (3.1.2).

A função de verossimilhança baseada nos dados observados

A função de verossimilhança baseada nos dados observados $D_{\text{obs}} = \begin{pmatrix} U_1 \\ \vdots \\ U_n \end{pmatrix}$ é

dada por

$$L(\theta|D_{\text{obs}}) \propto \prod_{i=1}^n |\Sigma^*|^{-\frac{1}{2}} \left[\nu + (U_i - \mu^*) \Sigma^{*-1} (U_i - \mu^*) \right]^{-\left(\frac{2+\nu}{2}\right)}, \quad (3.1.5)$$

onde μ^* e Σ^* são definidos em (3.1.3).

A distribuição à posteriori

A distribuição à posteriori $p(\theta|D_{\text{obs}})$ é dada por

$$p(\theta|D_{\text{obs}}) = \int_{\mathfrak{X}} p(\theta, \mathfrak{x}|D_{\text{obs}}) d\mathfrak{x} \propto \int_{\mathfrak{X}} L(\theta|D)\pi(\theta) d\mathfrak{x}, \quad (3.1.6)$$

onde $L(\theta|D)$ é dada por (3.1.4) e $\pi(\theta)$ é a distribuição à priori de θ .

Outra forma de escrever a posteriori é

$$p(\theta|D_{\text{obs}}) \propto L(\theta|D_{\text{obs}})\pi(\theta), \quad (3.1.7)$$

onde $L(\theta|D_{\text{obs}})$ é dada por (3.1.5) e $\pi(\theta)$ é a distribuição à priori de θ .

Formulação 2 – o modelo de mistura de escala da distribuição Normal

Uma especificação equivalente do modelo acima é dada a seguir.

Seja a equação de regressão

$$\begin{cases} Y_i = \alpha + \beta x_i + e_i, & i = 1, \dots, n, \\ X_i = x_i + u_i, \end{cases} \quad \text{e seja } \underline{Z}_i = \begin{pmatrix} Y_i \\ X_i \\ x_i \end{pmatrix}.$$

O modelo assume que

$$\underline{Z}_i | \theta, w_i \stackrel{\text{indep}}{\sim} N_3(\underline{\mu}, \Sigma w_i), \quad i = 1, 2, \dots, n, \quad \text{onde } \underline{\mu} \text{ e } \Sigma \text{ estão definidos em (3.1.2).} \quad (3.1.8)$$

A distribuição à priori adotada para θ e w é da forma

$$\pi(\theta, w) = \pi(\theta)\pi(w),$$

onde $\underline{w} = (w_1, \dots, w_n)^T$ e $w_i \stackrel{\text{iid}}{\sim} IG(\frac{\nu}{2}, \frac{\nu}{2})$, $i = 1, \dots, n$.

O espaço paramétrico é $\Theta^* = \{(\underline{\theta}, \underline{w})^T : \underline{\theta} \in \Theta \text{ e } \underline{w} \in \mathbb{R}_+^n\}$.

A função de verossimilhança baseada nos dados completos

A função de verossimilhança baseada nos dados completos, $D = \underline{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$,
é dada por

$$L(\underline{\theta}, \underline{w} | D) \propto \prod_{i=1}^n |\underline{\Sigma} w_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\underline{Z}_i - \underline{\mu})^T (\underline{\Sigma} w_i)^{-1} (\underline{Z}_i - \underline{\mu}) \right\},$$

ou seja, ela é proporcional a

$$\frac{1}{\left(\prod_{i=1}^n w_i^{3/2} \right) [(\sigma^2)^2 \sigma_x^2]^{n/2}} \exp \left\{ -\frac{1}{2} \left[\sum_{i=1}^n \frac{(Y_i - \alpha - \beta x_i)^2}{\lambda \sigma^2 w_i} + \sum_{i=1}^n \frac{(X_i - x_i)^2}{\sigma^2 w_i} + \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma_x^2 w_i} \right] \right\}. \quad (3.1.9)$$

A função de verossimilhança baseada nos dados observados

A função de verossimilhança baseada nos dados observados $D_{\text{obs}} = \begin{pmatrix} U_1 \\ \vdots \\ U_n \end{pmatrix}$,
onde $U_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix}$ é

$$L(\underline{\theta}, \underline{w} | D_{\text{obs}}) \propto \prod_{i=1}^n |\underline{\Sigma}^* w_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\underline{U}_i - \underline{\mu}^*)^T (\underline{\Sigma}^* w_i)^{-1} (\underline{U}_i - \underline{\mu}^*) \right\}, \quad (3.1.10)$$

onde $\underline{\mu}^*$ e $\underline{\Sigma}^*$ são definidos em (3.1.3).

Distribuição à posteriori

A posteriori $p(\underline{\theta} | D_{\text{obs}})$ é dada por

$$p(\underline{\theta} | D_{\text{obs}}) = \int_{\underline{x} = \mathbb{R}^n} \int_{\underline{w} = \mathbb{R}_+^n} p(\underline{\theta}, \underline{w}, \underline{x} | D_{\text{obs}}) d\underline{w} d\underline{x}, \quad (3.1.11)$$

onde $p(\theta, \underline{w}, \underline{x}|D_{\text{obs}}) \propto L(\theta, \underline{w}|D)\pi(\theta)\pi(\underline{w})$ e onde $L(\theta, \underline{w}|D)$ é dada em (3.1.9).

A posteriori é dada também por

$$p(\theta|D_{\text{obs}}) \propto \int_{\underline{w}=\mathbb{R}_+^n} p(\theta, \underline{w}|D_{\text{obs}})d\underline{w}, \quad (3.1.12)$$

onde $p(\theta, \underline{w}|D_{\text{obs}}) \propto L(\theta, \underline{w}|D_{\text{obs}})\pi(\theta, \underline{w})$ e onde $L(\theta, \underline{w}|D_{\text{obs}})$ é dada em (3.1.10).

Observação 1: A formulação 1 e a formulação 2 do modelo t -independente são equivalentes, já que a posteriori de θ , dada em (3.1.11), coincide com a posteriori de θ dada em (3.1.6). A justificativa dessa afirmação é dada a seguir.

Demonstração:
$$\begin{aligned} p(\theta|D_{\text{obs}}) &= \int_{\underline{x}=\mathbb{R}^n} \int_{\underline{w}=\mathbb{R}_+^n} p(\theta, \underline{w}, \underline{x}|D_{\text{obs}})d\underline{w}d\underline{x} \\ &\propto \int_{\underline{x}} \int_{\underline{w}=\mathbb{R}_+^n} L(\theta, \underline{w}|D)\pi(\theta)\pi(\underline{w})d\underline{w}d\underline{x} \\ &= \int_{\underline{x}} \int_{\underline{w}} \prod_{i=1}^n |\underline{\Sigma} w_i|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}[(\underline{Z}_i - \underline{\mu})^\top (\underline{\Sigma} w_i)^{-1}(\underline{Z}_i - \underline{\mu})]\right\} \pi(\theta) \prod_{i=1}^n w_i^{-(\frac{\nu}{2}+1)} \\ &\quad \cdot \exp\left\{-\frac{\nu/2}{w_i}\right\} d\underline{w}d\underline{x} \\ &= |\underline{\Sigma}|^{-\frac{n}{2}} \int_{\underline{x}} \int_{\underline{w}} \prod_{i=1}^n w_i^{-(\frac{3+\nu}{2}+1)} \exp\left\{\frac{-\frac{1}{2}[(\underline{Z}_i - \underline{\mu})^\top \underline{\Sigma}^{-1}(\underline{Z}_i - \underline{\mu}) + \nu]}{w_i}\right\} \pi(\theta) \\ &\quad d\underline{w}d\underline{x} \\ &= |\underline{\Sigma}|^{-\frac{n}{2}} \pi(\theta) \int_{\underline{x}} \int_{\underline{w}} \prod_{i=1}^n w_i^{-(\frac{3+\nu}{2}+1)} \exp\left\{\frac{-\frac{1}{2}[(\underline{Z}_i - \underline{\mu})^\top \underline{\Sigma}^{-1}(\underline{Z}_i - \underline{\mu}) + \nu]}{w_i}\right\} \\ &\quad d\underline{w}d\underline{x} \\ &= |\underline{\Sigma}|^{-\frac{n}{2}} \pi(\theta) \int_{\underline{x}} \prod_{i=1}^n \int_{w_i=\mathbb{R}_+} w_i^{-(\frac{3+\nu}{2}+1)} \exp\left\{\frac{-\frac{1}{2}[(\underline{Z}_i - \underline{\mu})^\top \underline{\Sigma}^{-1}(\underline{Z}_i - \underline{\mu}) + \nu]}{w_i}\right\} \\ &\quad dw_i d\underline{x}, \end{aligned} \quad (3.1.13)$$

onde a função dentro da integral em w_i é o núcleo de uma Gama Invertida

$$IG\left(\frac{3+\nu}{2}, \frac{1}{2}[(\underline{Z}_i - \underline{\mu})^\top \underline{\Sigma}^{-1}(\underline{Z}_i - \underline{\mu}) + \nu]\right).$$

Portanto, (3.1.13) é igual a

$$|\underline{\Sigma}|^{-\frac{n}{2}} \pi(\theta) \int_{\underline{x}} \prod_{i=1}^n \frac{\Gamma\left(\frac{3+\nu}{2}\right)}{\left\{\frac{1}{2}[(\underline{Z}_i - \underline{\mu})^\top \underline{\Sigma}^{-1}(\underline{Z}_i - \underline{\mu}) + \nu]\right\}^{\frac{3+\nu}{2}}} d\underline{x}$$

e portanto

$$p(\boldsymbol{\theta} | D_{\text{obs}}) \propto \int_{\boldsymbol{x}} \prod_{i=1}^n |\boldsymbol{\Sigma}|^{-\frac{1}{2}} [\nu + (\boldsymbol{Z}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{Z}_i - \boldsymbol{\mu})]^{-\frac{3+\nu}{2}} \pi(\boldsymbol{\theta}) d\boldsymbol{x},$$

que coincide com (3.1.6).

3.1.2 Implementação de algoritmos MCMC no modelo com particular priori própria de componentes independentes

Tomando-se a distribuição à priori

$$\pi(\boldsymbol{\theta}, \boldsymbol{w}) = \pi(\boldsymbol{\theta})\pi(\boldsymbol{w}) = \pi(\alpha)\pi(\beta)\pi(\mu)\pi(\sigma_x^2)\pi(\sigma^2) \prod_{i=1}^n \pi(w_i), \quad (3.1.14)$$

onde $\alpha \sim N(a, \sigma_\alpha^2)$, $\beta \sim N(b, \sigma_\beta^2)$, $\mu \sim N(m, \sigma_\mu^2)$, $\sigma_x^2 \sim IG(c, d)$, $\sigma^2 \sim IG(f, g)$ e $w_i \sim IG(\frac{\nu}{2}, \frac{\nu}{2})$, $i = 1, \dots, n$, e considerando a formulação 2 do modelo t -independente descrita na subseção anterior, podemos implementar o algoritmo de Gibbs, de forma semelhante à do modelo Normal do Capítulo 2, utilizando as distribuições condicionais completas exibidas abaixo, que são baseadas na distribuição à posteriori $p(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{w} | D_{\text{obs}})$ envolvida em (3.1.11).

É importante lembrar que todos os algoritmos MCMC descritos na seção 2.4 do Capítulo 2 para o modelo Normal podem ser implementados para esse modelo de forma semelhante quando usamos a formulação 2.

É conveniente lembrar também que no caso do modelo t sem restrição de identificabilidade a implementação dos algoritmos seria semelhante ao caso do modelo normal sem restrições definido na seção 2.6 (a mesma observação vale no caso do modelo t com priori $\pi(\boldsymbol{\theta}, \boldsymbol{w}) = \pi(\boldsymbol{w})\pi(\boldsymbol{\theta})$, onde $\pi(\boldsymbol{\theta})$ tem componentes condicionalmente independentes como em (2.7.1).

Nessa seção, entretanto, vamos nos restringir apenas à implementação do algoritmo original de Gibbs para o modelo t -independente com restrição de identificabilidade e com priori dada em (3.1.4), para não nos tornarmos muito repetitivos.

Distribuições condicionais completas baseadas na distribuição

$p(\theta, w, \underline{x}|D_{\text{obs}})$ em (3.1.11)

$$\begin{aligned} & x_i|x_{(-i)}, \alpha, \beta, \mu, \sigma^2, w, D_{\text{obs}} \\ & \sim N\left(\frac{\sigma_x^2[\lambda X_i + \beta(Y_i - \alpha)] + \lambda\sigma_\mu^2}{\sigma_x^2[\lambda + \beta^2] + \lambda\sigma^2}; \frac{\lambda\sigma^2\sigma_x^2(w_i)}{\sigma_x^2[\lambda + \beta^2] + \lambda\sigma^2}\right), \end{aligned} \quad (3.1.15)$$

$$i = 1, \dots, n,$$

$$\begin{aligned} & \alpha|\underline{x}, \beta, \mu, \sigma_x^2, \sigma^2, w, D_{\text{obs}} \\ & \sim N\left(\frac{\sigma_\alpha^2 \sum_{i=1}^n \frac{(Y_i - \beta x_i)}{w_i} + \lambda\sigma^2 a}{\sigma_\alpha^2 \sum_{i=1}^n \frac{1}{w_i} + \lambda\sigma^2}; \frac{\lambda\sigma^2\sigma_\alpha^2}{\sigma_\alpha^2 \sum_{i=1}^n \frac{1}{w_i} + \lambda\sigma^2}\right), \end{aligned} \quad (3.1.16)$$

$$\begin{aligned} & \beta|\underline{x}, \alpha, \mu, \sigma_x^2, \sigma^2, w, D_{\text{obs}} \\ & \sim N\left(\frac{\sigma_\beta^2 \sum_{i=1}^n \frac{(Y_i - \alpha)x_i}{w_i} + \lambda b\sigma^2}{\sigma_\beta^2 \sum_{i=1}^n \frac{x_i^2}{w_i} + \lambda\sigma^2}; \frac{\lambda\sigma^2\sigma_\beta^2}{\sigma_\beta^2 \sum_{i=1}^n \frac{x_i^2}{w_i} + \lambda\sigma^2}\right), \end{aligned} \quad (3.1.17)$$

$$\begin{aligned} & \mu|\underline{x}, \alpha, \beta, \sigma_x^2, \sigma^2, w, D_{\text{obs}} \\ & \sim N\left(\frac{\sigma_\mu^2 \sum_{i=1}^n \frac{x_i}{w_i} + \sigma_x^2 m}{\sigma_\mu^2 \sum_{i=1}^n \frac{1}{w_i} + \sigma_x^2}; \frac{\sigma_\mu^2 \sigma_x^2}{\sigma_\mu^2 \sum_{i=1}^n \frac{1}{w_i} + \sigma_x^2}\right), \end{aligned} \quad (3.1.18)$$

$$\begin{aligned} & \sigma_x^2|\underline{x}, \alpha, \beta, \mu, \sigma^2, w, D_{\text{obs}} \\ & \sim IG\left(\frac{n}{2} + c; \frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{w_i} + d\right), \end{aligned} \quad (3.1.19)$$

$$\begin{aligned} & \sigma^2|\underline{x}, \alpha, \beta, \mu, \sigma_x^2, w, D_{\text{obs}} \\ & \sim IG\left(k + f; \frac{1}{2} \left(\sum_{i=1}^n \frac{(Y_i - \alpha - \beta x_i)^2}{\lambda w_i} + \sum_{i=1}^n \frac{(X_i - x_i)^2}{w_i} \right) + g\right), \end{aligned} \quad (3.1.20)$$

$$\begin{aligned} & w_i|w_{(-i)}, \underline{x}, \alpha, \beta, \mu, \sigma_x^2, \sigma^2, D_{\text{obs}} \\ & \sim IG\left(\frac{3 + \nu}{2}; \frac{1}{2} \left[\frac{(Y_i - \alpha - \beta x_i)^2 + \lambda(X_i - x_i)^2}{\lambda\sigma^2} + \frac{(x_i - \mu)^2}{\sigma_x^2} + \nu \right]\right), \end{aligned} \quad (3.1.21)$$

$$i = 1, \dots, n,$$

Observação 1: Fazendo $w_i = 1$ nas fórmulas (3.1.15) a (3.1.21) obtemos as fórmulas (2.5.9) a (2.5.14) do Capítulo 2.

Observação 2: No modelo dessa seção, o número de graus de liberdade, ν , é suposto conhecido. Entretanto, nas aplicações em geral não temos tal conhecimento. Uma maneira de tratar o problema é considerar o número de graus de liberdade ν como uma variável aleatória atribuindo uma priori para ela como por exemplo na seção 3.2 a seguir. Uma outra maneira é analisar o problema descritivamente tratando ν como um hiperparâmetro e calculando por exemplo a medida L para vários valores de ν fixados e escolhendo o ν que corresponde ao menor valor dessa medida.

3.1.3 Distribuição preditiva

Seja $U_f = \begin{pmatrix} Y_f \\ X_f \end{pmatrix}$ uma observação futura. A distribuição preditiva à posteriori no modelo t -independente definido em (3.1.8) é dada por

$$\begin{aligned} \pi(U_f|D_{\text{obs}}) &= \int_{W_f} \int_{\Theta} \pi(U_f, \theta, w_f|D_{\text{obs}}) d\theta dw_f \\ &= \int_{W_f} \int_{\Theta} \pi(U_f|\theta, w_f) \pi(w_f) \pi(\theta|D_{\text{obs}}) d\theta dw_f, \\ &\quad \text{onde } \theta = (\alpha, \beta, \mu, \sigma_x^2, \sigma^2) \\ &= \int_{W_f} \int_{\Theta} \pi(Y_f|X_f, \theta, w_f) \pi(X_f|\theta, w_f) \pi(w_f) \pi(\theta|D_{\text{obs}}) d\theta dw_f, \end{aligned} \quad (3.1.22)$$

onde $\pi(\theta|D_{\text{obs}})$ é a distribuição à posteriori,

$$\begin{aligned} w_f &\sim IG\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \\ X_f|\theta, w_f &\sim N\left(\mu, (\sigma_x^2 + \sigma^2)w_f\right) \end{aligned}$$

$$\text{e } Y_f|X_f, \theta, w_f \sim N\left(\alpha + \beta\mu + \frac{\beta\sigma_x^2}{\sigma_x^2 + \sigma^2}(X_f - \mu); \left(\beta^2\sigma_x^2 + \lambda\sigma^2 - \frac{\beta^2(\sigma_x^2)^2}{\sigma_x^2 + \sigma^2}\right)w_f\right).$$

Geração de uma observação futura

Para gerar uma observação U_f , basta seguir os seguintes passos:

- (1) gerar $\varrho^{(g)}$ de $\pi(\varrho|D_{\text{obs}})$ (ou de $\pi(\varrho, w|D_{\text{obs}})$ ou de $\pi(\varrho, w, x|D_{\text{obs}})$);
- (2) gerar $w_f^{(g)}$ de $\pi(w_f)$;
- (3) gerar X_f de $\pi(X_f|\varrho^{(g)}, w_f^{(g)})$ e
- (4) gerar Y_f de $\pi(Y_f|X_f^{(g)}, \varrho^{(g)}, w_f^{(g)})$ (e portanto, dado que $X_f = X_0$, podemos gerar Y_f de $\pi(Y_f|X_f = X_0, \varrho^{(g)}, w_f^{(g)})$).

Semelhantemente, dado que $Y_f = Y_0$, podemos gerar X_f de $\pi(X_f|Y_f = Y_0, \varrho^{(g)}, w_f^{(g)})$, onde

$$X_f|Y_f, \varrho, w_f \sim N\left(\mu + \frac{\beta\sigma_x^2}{\beta^2\sigma_x^2 + \lambda\sigma^2}(Y_f - \alpha - \beta\mu); \left[\sigma_x^2 + \sigma^2 - \frac{(\beta\sigma_x^2)^2}{\beta^2\sigma_x^2 + \lambda\sigma^2}\right] w_f\right).$$

Observação: A distribuição preditiva $\pi(\underline{U}_f|D_{\text{obs}})$ também pode ser escrita como

$$\begin{aligned} \pi(\underline{U}_f|D_{\text{obs}}) &= \int_{\mathbf{x}_f} \int_{\Theta} \int_{W_f} \pi(\underline{U}_f, \varrho, x_f, w_f|D_{\text{obs}}) dw_f d\varrho dx_f \\ &= \int_{\mathbf{x}_f} \int_{\Theta} \int_{W_f} \pi(\underline{U}_f|x_f, w_f, \varrho) \pi(x_f|\varrho, w_f) \pi(w_f) \pi(\varrho|D_{\text{obs}}) dw_f d\varrho dx_f, \end{aligned} \quad (3.1.23)$$

onde $w_f \sim IG(\frac{\nu}{2}, \frac{\nu}{2})$,

$$\begin{aligned} x_f|\varrho, w_f &\sim N(\mu; \sigma_x^2 w_f) \\ \underline{U}_f|x_f, w_f, \varrho &\sim N\left[\begin{pmatrix} \alpha + \beta x_f \\ x_f \end{pmatrix}; \begin{pmatrix} \lambda\sigma & 0 \\ 0 & \sigma^2 \end{pmatrix} w_f\right], \end{aligned}$$

isto é, Y_f e X_f são condicionalmente independentes dado ϱ, w_f, x_f .

3.1.4 Avaliação do ajuste

Podemos fazer uma avaliação da qualidade do ajuste do modelo t -independente de forma semelhante à apresentada na seção 2.10. Para isso, basta obter as fórmulas de cálculo da densidade condicional preditiva de validação cruzada (CPO_i) e dos resíduos de validação cruzada para esse modelo.

(1) Densidade preditiva de validação cruzada

A seguir apresentamos algumas fórmulas alternativas para estimar $CPO_i = \pi(\underline{U}_i | D_{\text{obs}(-i)})$, onde $\underline{U}_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix}$ é a i -ésima observação e $D_{\text{obs}(-i)} = \underline{U}_{(-i)}$, $i = 1, \dots, n$.

Fórmula 1

$$\begin{aligned} CPO_i &= \int_{\Theta} \pi(\underline{U}_i, \vartheta | D_{\text{obs}(-i)}) d\vartheta = \int_{\Theta} \pi(\underline{U}_i | \vartheta) \pi(\vartheta | D_{\text{obs}(-i)}) d\vartheta \\ &= \left[\int_{\Theta} \frac{1}{\pi(\underline{U}_i | \vartheta)} \pi(\vartheta | D_{\text{obs}}) d\vartheta \right]^{-1} \end{aligned} \quad (3.1.24)$$

onde $\vartheta = (\alpha, \beta, \mu, \sigma_x^2, \sigma^2)$, onde $\underline{U}_i | \vartheta \sim t_2(\underline{\mu}, \underline{\Sigma}, \nu)$, com

$$\underline{\mu} = \begin{pmatrix} \alpha + \beta\mu \\ \mu \end{pmatrix} \quad \text{e} \quad \underline{\Sigma} = \begin{pmatrix} \beta^2\sigma_x^2 + \lambda\sigma^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma^2 \end{pmatrix},$$

isto é,

$$\pi(\underline{U}_i | \vartheta) = k(n, \nu) |\Sigma|^{-\frac{1}{2}} [\nu + (\underline{U}_i - \underline{\mu})^T \Sigma^{-1} (\underline{U}_i - \underline{\mu})]^{-\frac{(2+\nu)}{2}},$$

onde $k(n, \nu) = \frac{\Gamma(\frac{2+\nu}{2}) \nu^{\nu/2}}{\Gamma(\frac{\nu}{2}) \pi}$ e ν é o número de graus de liberdade (conhecido).

Estimação

Uma estimativa de Monte Carlo de CPO_i é dada por

$$\widehat{CPO}_i = \left(\frac{1}{G} \sum_{g=1}^G \frac{1}{\pi(\underline{U}_i | \vartheta^{(g)})} \right)^{-1}, \quad (3.1.25)$$

onde $(\vartheta^{(1)}, \dots, \vartheta^{(G)})$ é uma amostra da posteriori $p(\vartheta | D_{\text{obs}})$, gerada por exemplo de $p(\vartheta, \underline{x}, \underline{w} | D_{\text{obs}})$ pelo algoritmo de Gibbs da seção 3.1.2, ou de $p(\vartheta, \underline{w} | D_{\text{obs}})$ pelo algoritmo de M-H em Gibbs (desde que estes algoritmos converjam).

Fórmula 2

$$CPO_i = \int_W \int_{\Theta} \pi(\underline{U}_i, \vartheta, \underline{w} | D_{\text{obs}(-i)}) d\vartheta d\underline{w}$$

$$\begin{aligned}
&= \int_{\Theta} \int_{\mathcal{W}} \pi(U_i | \varrho, \underline{w}) \pi(\varrho, \underline{w} | D_{\text{obs}(-1)}) d\varrho d\underline{w} \\
&= \left(\int_{\Theta} \int_{\mathcal{W}} \frac{1}{\pi(U_i | \varrho, w_i)} \pi(\varrho, \underline{w} | D_{\text{obs}}) d\varrho d\underline{w} \right)^{-1}, \quad (3.1.26)
\end{aligned}$$

onde $\underline{w} = (w_1, \dots, w_n)^T$, e onde

$$U_i | \varrho, w_i \sim N \left[\begin{pmatrix} \alpha + \beta \mu \\ \mu \end{pmatrix}; \begin{pmatrix} \beta \sigma_x^2 + \lambda \sigma^2 & \beta \sigma_x^2 \\ \beta \sigma_x^2 & \sigma_x^2 + \sigma^2 \end{pmatrix} w_i \right],$$

Estimação

$$\widehat{\text{CPO}}_i = \left(\frac{1}{G} \sum_{g=1}^G \frac{1}{\pi(U_i | \varrho^{(g)}, w_i^{(g)})} \right)^{-1}, \quad (3.1.27)$$

onde $(\varrho^{(1)}, w_i^{(1)}, \dots, \varrho^{(G)}, w_i^{(G)})^T$ é uma amostra da posteriori conjunta de ϱ e \underline{w} , gerada por exemplo de $p(\varrho, \underline{x}, \underline{w} | D_{\text{obs}})$ pelo algoritmo de Gibbs da seção 3.1.2 ou de $p(\varrho, \underline{w} | D_{\text{obs}})$ pelo algoritmo de M-H em Gibbs (desde que eles convirjam).

Fórmula 3

$$\begin{aligned}
\text{CPO}_i &= \int_{\mathcal{X}} \int_{\mathcal{W}} \int_{\Theta} \pi(U_i, \varrho, \underline{x}, \underline{w} | D_{\text{obs}(-i)}) d\varrho d\underline{w} d\underline{x} \\
&= \int_{\mathcal{X}} \int_{\mathcal{W}} \int_{\Theta} \pi(U_i | x_i, w_i, \varrho) \pi(\varrho, \underline{x}, \underline{w} | D_{\text{obs}(-i)}) d\varrho d\underline{w} d\underline{x} \\
&= \left(\int_{\mathcal{X}} \int_{\mathcal{W}} \int_{\Theta} \frac{1}{\pi(U_i | x_i, w_i, \varrho)} \pi(\varrho, \underline{x}, \underline{w} | D_{\text{obs}}) d\varrho d\underline{w} d\underline{x} \right)^{-1} \quad (3.1.28)
\end{aligned}$$

onde

$$U_i | x_i, w_i, \varrho \sim N \left[\begin{pmatrix} \alpha + \beta x_i \\ x_i \end{pmatrix}; \begin{pmatrix} \lambda \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} w_i \right].$$

Estimação

$$\widehat{\text{CPO}}_i = \left(\frac{1}{G} \sum_{g=1}^G \frac{1}{\pi(U_i | x_i^{(g)}, w_i^{(g)}, \varrho^{(g)})} \right)^{-1}, \quad (3.1.29)$$

onde $(\varrho^{(1)}, \underline{x}^{(1)}, w^{(1)}, \dots, \varrho^{(G)}, \underline{x}^{(G)}, w^{(G)})$ é uma amostra da posteriori conjunta

$p(\varrho, \underline{x}, \underline{w} | D_{\text{obs}})$ gerada por exemplo pelo algoritmo de Gibbs da seção 3.1.2, desde que haja convergência.

Observação: Note que o cálculo de $\widehat{\text{CPO}}_i$ nesse caso poderia eventualmente ser obtido gerando $(\theta^{(g)}, w^{(g)})$ de $\pi(\theta, w|D_{\text{obs}})$ pelo algoritmo de M-H em Gibbs (desde que este convirja) e depois gerando $x_i^{(g)}$ de $\pi(x_i|\theta^{(g)}, w_i^{(g)}) \sim N(\mu^{(g)}, \sigma_x^{2(g)} w_i^{(g)})$.

(2) Resíduo de validação cruzada

No modelo estrutural t -independente (com restrição de identificabilidade $\sigma_{e_i}^2 = \lambda\sigma_x^2$), formulado como mistura de escala da distribuição Normal, temos que os resíduos bidimensionais $\begin{pmatrix} e_i \\ u_i \end{pmatrix}$ são tais que

$$\begin{pmatrix} e_i \\ u_i \end{pmatrix} | \theta, x_i, w_i \sim N_2(\underline{0}, V), \text{ onde } V = \begin{pmatrix} \lambda\sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} w_i, \text{ para } i = 1, \dots, n.$$

Os resíduos padronizados são iguais a

$$V^{-\frac{1}{2}} \begin{pmatrix} e_i \\ u_i \end{pmatrix} = \text{Var}^{-\frac{1}{2}}(\underline{U}_i | \theta, x_i, w_i) [\underline{U}_i - E(\underline{U}_i | \theta, x_i, w_i)],$$

$i = 1, \dots, n$, e portanto os resíduos padronizados de validação cruzada nesse modelo são iguais a

$$\text{Var}^{-\frac{1}{2}}(\underline{U}_i | D_{\text{obs}(-i)}) [\underline{U}_i - E(\underline{U}_i | D_{\text{obs}(-i)})], \quad i = 1, \dots, n,$$

onde

- Cálculo de $E(\underline{U}_i | D_{\text{obs}(-i)})$

$$\begin{aligned} E[\underline{U}_i | D_{\text{obs}(-i)}] &= E[E(\underline{U}_i | \theta, \underline{x}, w) | D_{\text{obs}(-i)}] \\ &= \int_{\underline{x}} \int_W \int_{\Theta} E(\underline{U}_i | \theta, \underline{x}, w) \pi(\theta, \underline{x}, w | D_{\text{obs}(-i)}) d\theta d w d \underline{x} \\ &= \text{CPO}_i \int_{\underline{x}} \int_W \int_{\Theta} \frac{E(\underline{U}_i | \theta, \underline{x}, w)}{\pi(\underline{U}_i | \theta, x_i, w_i)} \pi(\theta, \underline{x}, w | D_{\text{obs}}) d\theta d w d \underline{x}, \quad (3.1.30) \end{aligned}$$

onde

$$\text{CPO}_i = \left[\int_{\underline{x}} \int_W \int_{\Theta} \frac{1}{\pi(\underline{U}_i | \theta, x_i, w_i)} \pi(\theta, \underline{x}, w | D_{\text{obs}}) d\theta d w d \underline{x} \right]^{-1} \quad (3.1.31)$$

e onde

$$\underline{U}_i | \underline{\theta}, x_i, w_i \sim N \left[\begin{pmatrix} \alpha + \beta x_i \\ x_i \end{pmatrix}; \begin{pmatrix} \lambda \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} w_i \right].$$

Portanto,

$$\pi(\underline{U}_i | \underline{\theta}, x_i, w_i) = (2\pi\lambda(\sigma^2)^2) \exp \left\{ -\frac{1}{2} \left[\frac{(Y_i - \alpha - \beta x_i)^2}{\lambda \sigma^2} + \frac{(X_i - x_i)^2}{\sigma^2} \right] \right\}. \quad (3.1.32)$$

Estimação

Dada uma amostra de tamanho G de $\pi(\underline{\theta}, w, \underline{x} | D_{\text{obs}})$, obtida por exemplo por meio do algoritmo de Gibbs da seção 3.1.2 (se houver convergência),

$$\hat{E}(\underline{U}_i | D_{\text{obs}(-i)}) = \widehat{\text{CPO}}_i \sum_{g=1}^G \frac{E(\underline{U}_i | \underline{\theta}^{(g)}, x_i^{(g)}, w_i^{(g)})}{\pi(\underline{U}_i | \underline{\theta}^{(g)}, x_i^{(g)}, w_i^{(g)})}, \quad (3.1.33)$$

onde $\pi(\underline{U}_i | \underline{\theta}, x_i, w_i)$ é dada em (3.1.32), $E(\underline{U}_i | \underline{\theta}, x_i, w_i) = \begin{pmatrix} \alpha + \beta x \\ x_i \end{pmatrix}$ e

$$\widehat{\text{CPO}}_i = \left(\frac{1}{G} \sum_{j=1}^G \frac{1}{\pi(\underline{U}_i | \underline{\theta}^{(g)}, x_i^{(g)}, w_i^{(g)})} \right)^{-1}.$$

Observação: Uma forma alternativa de estimar $E(\underline{U}_i | D_{\text{obs}(-i)})$, para $i = 1, \dots, n$, é utilizar a fórmula acima, mas gerando $\underline{\theta}^{(g)}$ e $w^{(g)}$ de $p(\underline{\theta}, w | D_{\text{obs}})$ usando Gibbs ou M-H em Gibbs (desde que converjam) e depois gerando $x_i^{(g)}$ de $\pi(x_i | \underline{\theta}^{(g)}, w_i^{(g)})$, onde $x_i | \underline{\theta}, w_i \sim N(\mu; \sigma_x^2 w_i)$.

Cálculo da $\text{Var}(\underline{U}_i | D_{\text{obs}(-i)})$

$$\begin{aligned} \text{Var}(\underline{U}_i | D_{\text{obs}}) &= E[\text{Var}(\underline{U}_i | \underline{\theta}, \underline{x}, w) | D_{\text{obs}(-i)}] + \text{Var}[E(\underline{U}_i | \underline{\theta}, \underline{x}, w) | D_{\text{obs}(-i)}] \\ &= E_{\underline{\theta}, \underline{x}, w | D_{\text{obs}(-i)}}[\text{Var}(\underline{U}_i | \underline{\theta}, \underline{x}, w)] + E_{\underline{\theta}, \underline{x}, w | D_{\text{obs}(-i)}}\{E(\underline{U}_i | \underline{\theta}, \underline{x}, w)^\top] \\ &\quad - [E_{\underline{\theta}, \underline{x}, w | D_{\text{obs}(-i)}}(E(\underline{U}_i | \underline{\theta}, \underline{x}, w))][E_{\underline{\theta}, \underline{x}, w | D_{\text{obs}(-i)}}(E(\underline{U}_i | \underline{\theta}, \underline{x}, w))]^\top\} \\ &= \int_W \int_{\underline{x}} \int_{\Theta} \text{Var}(\underline{U}_i | \underline{\theta}, \underline{x}, w) \pi(\underline{\theta}, \underline{x}, w | D_{\text{obs}(-i)}) d\underline{\theta} d\underline{x} dw \\ &\quad + \int_W \int_{\underline{x}} \int_{\Theta} (E(\underline{U}_i | \underline{\theta}, \underline{x}, w))(E(\underline{U}_i | \underline{\theta}, \underline{x}, w))^\top \pi(\underline{\theta}, \underline{x}, w | D_{\text{obs}(-i)}) d\underline{\theta} d\underline{x} dw \\ &\quad - [E(\underline{U}_i | D_{\text{obs}(-i)})][E(\underline{U}_i | D_{\text{obs}(-i)})]^\top \end{aligned}$$

$$\begin{aligned}
&= \text{CPO}_i \int_W \int_{\mathcal{X}} \int_{\Theta} \frac{\text{Var}(\underline{U}_i | \varrho, x_i, w_i)}{\pi(\underline{U}_i | \varrho, x_i, w_i)} \pi(\varrho, \underline{x}, w | D_{\text{obs}}) d\varrho d\underline{x} dw \\
&\quad + \text{CPO}_i \int_W \int_{\mathcal{X}} \int_{\Theta} [E(\underline{U}_i | \varrho, x_i, w_i)][E(\underline{U}_i | \varrho, x_i, w_i)]^{\top} \pi(\varrho, w, \underline{x} | D_{\text{obs}}) d\varrho d\underline{x} dw \\
&\quad - [E(\underline{U}_i | D_{\text{obs}(-i)})][E(\underline{U}_i | D_{\text{obs}(-i)})]^{\top}, \tag{3.1.34}
\end{aligned}$$

onde CPO_i é dada por (3.1.31), $E(\underline{U}_i | D_{\text{obs}(-i)})$ é dada por (3.1.30).

Estimação

$$\begin{aligned}
&\widehat{\text{Var}}(\underline{U}_i | D_{\text{obs}(-i)}) = \\
&\widehat{\text{CPO}}_i \sum_{g=1}^G \frac{\text{Var}(\underline{U}_i | \varrho^{(g)}, x_i^{(g)}, w_i^{(g)}) + [E(\underline{U}_i | \varrho^{(g)}, x_i^{(g)}, w_i^{(g)})][E(\underline{U}_i | \varrho^{(g)}, x_i^{(g)}, w_i^{(g)})]^{\top}}{\pi(\underline{U}_i | \varrho^{(g)}, x_i^{(g)}, w_i^{(g)})} \\
&\quad - [\hat{E}(\underline{U}_i | D_{\text{obs}(-i)})][\hat{E}(\underline{U}_i | D_{\text{obs}(-i)})]^{\top}, \tag{3.1.35}
\end{aligned}$$

onde $\widehat{\text{CPO}}_i$ é dada em (3.1.29), $\hat{E}(\underline{U}_i | D_{\text{obs}(-i)})$ é dada em (3.1.33),

$$\underline{U}_i | \varrho, x_i, w_i \sim N \left[\begin{pmatrix} \alpha + \beta x_i \\ x_i \end{pmatrix}; \begin{pmatrix} \lambda \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} w_i \right], \quad i = 1, \dots, n,$$

e $(\varrho^{(1)}, w^{(1)}, \underline{x}^{(1)}, \dots, \varrho^{(G)}, w^{(G)}, \underline{x}^{(G)})$ é uma amostra da posteriori $p(\varrho, w, \underline{x} | D_{\text{obs}})$.

3.2 O modelo estrutural t -independente com graus de liberdade desconhecidos

Nessa seção estudamos o modelo t -independente onde ν é desconhecido. Nesse caso escolhemos uma priori para ν , incorporando-a ao modelo, e com base na distribuição marginal à posteriori de ν obtemos um estimador de Bayes de ν . Na subseção 3.2.2 analisamos o modelo considerando duas prioris próprias para ν , a primeira é uma distribuição uniforme discreta e a segunda é a distribuição exponencial. O modelo de regressão t -independente com ν desconhecido no caso clássico (ou seja quando a variável independente é medida sem erros) foi estudado por Geweke (1993)).

3.2.1 Definição do modelo

Considere o modelo identificável formulado como mistura de escala da distribuição Normal

$$\begin{cases} Y_i = \alpha + \beta x_i + e_i, & i = 1, \dots, n, \\ X_i = x_i + u_i, \end{cases}$$

onde

$$Z_i = \begin{pmatrix} Y_i \\ X_i \\ x_i \end{pmatrix}, \text{ com a suposição } Z_i | \theta, \underline{w}, \nu \stackrel{\text{indep}}{\sim} N_3(\underline{\mu}; \underline{\Sigma} w_i), \quad (3.2.1)$$

onde $\underline{\mu}$ e $\underline{\Sigma}$ são dadas em (3.1.2).

O vetor de parâmetros é $\underline{\theta} = \begin{pmatrix} \theta \\ \underline{w} \\ \nu \end{pmatrix}$, onde $\theta = (\alpha, \beta, \mu, \sigma_x^2, \sigma^2)^T$ e $\underline{w} = (w_1, \dots, w_n)^T$.

Vamos tomar a seguinte distribuição à priori:

$$\pi(\underline{\theta}, \underline{w}, \nu) = \pi(\theta)\pi(\underline{w}|\nu)\pi(\nu), \quad (3.2.2)$$

onde $w_i | \nu \stackrel{\text{iid}}{\sim} IG(\frac{\nu}{2}, \frac{\nu}{2})$, $\pi(\theta)$ é a priori de θ e $\pi(\nu)$ a priori de ν .

Função de verossimilhança

A função de verossimilhança $L(\underline{\theta}, \underline{w}, \nu | D)$, baseada nos dados completos D , é proporcional à função em (3.1.9), e a verossimilhança baseada nos dados observados, $L(\underline{\theta}, \underline{w}, \nu | D_{\text{obs}})$, é proporcional à função em (3.1.10).

Distribuição à posteriori

A distribuição conjunta à posteriori de $(\underline{\theta}, \underline{w}, \nu)$ é dada por

$$p(\underline{\theta}, \underline{w}, \nu | D_{\text{obs}}) = \int_{\underline{x} = \mathbb{R}^n} p(\underline{x}, \underline{\theta}, \underline{w}, \nu | D_{\text{obs}}) d\underline{x},$$

e portanto a distribuição à posteriori

$$p(\underline{\theta}, \nu | D_{\text{obs}}) = \int_{\underline{x}} \int_{W = \mathbb{R}_+^n} p(\underline{x}, \underline{\theta}, \underline{w}, \nu | D_{\text{obs}}) d\underline{w} d\underline{x}, \quad (3.2.3)$$

onde

$$p(\underline{x}, \underline{\theta}, \underline{w}, \nu | D_{\text{obs}}) \propto L(\underline{\theta}, \underline{w}, \nu | D_{\text{obs}}) \pi(\underline{\theta}, \underline{w}, \nu), \quad (3.2.4)$$

onde $L(\underline{\theta}, \underline{w}, \nu | D_{\text{obs}})$ é proporcional à função em (3.1.10).

3.2.2 Implementação de algoritmos MCMC no modelo t -independente para duas particulares escolhas de priori própria

Caso 1: Considere a distribuição à priori

$$\begin{aligned} \pi(\underline{\theta}, \underline{w}, \nu) &= \pi(\underline{\theta}) \pi(\underline{w} | \nu) \pi(\nu) \\ &= \pi(\alpha) \pi(\beta) \pi(\mu) \pi(\sigma_x) \pi(\sigma^2) \left(\prod_{i=1}^n \pi(w_i | \nu) \right) \pi(\nu), \end{aligned} \quad (3.2.5)$$

onde $\alpha \sim N(a, \mu_\alpha)$, $\mu \sim N(m, \sigma_\mu^2)$, $\sigma_x^2 \sim IG(c, d)$, $\sigma^2 \sim IG(f, g)$, $w_i | \nu \stackrel{\text{iid}}{\sim} IG(\frac{\nu}{2}, \frac{\nu}{2})$, $i = 1, \dots, n$, \mathcal{V} é uma v.a. uniforme discreta no intervalo $[c - n_0 + 1, c]$, isto é, $\pi(\nu) = p[\mathcal{V} = \nu] = \begin{cases} \frac{1}{n_0}, & c - n_0 + 1 \leq \nu \leq c, \\ 0, & \text{caso contrário,} \end{cases}$ onde $n_0 = \text{número de elementos do domínio da variável aleatória } \mathcal{V} \text{ com probabilidade positiva e } c \text{ é uma constante inteira } (c > 0)$. (No caso particular em que $n_0 = 1$ e $c = \nu_0$, então \mathcal{V} é uma variável degenerada no ponto ν_0).

Nesse modelo podemos implementar o algoritmo de Gibbs para amostrar da posteriori usando as condicionais completas baseadas na distribuição conjunta à posteriori em (3.2.3). Nesse modelo, as condicionais completas univariadas de x_i , α , β , μ , σ_x^2 , w_i ($i = 1, \dots, n$) coincidem, respectivamente, com as distribuições dadas em (3.1.15) a (3.1.21); a única diferença é que agora devemos acrescentar a cada ciclo do algoritmo de Gibbs a distribuição

$$\pi(\nu | \underline{x}, \alpha, \beta, \mu, \sigma_x^2, \sigma^2, \underline{w}, D_{\text{obs}}) = P(\mathcal{V} = \nu | \underline{x}, \underline{\theta}, \underline{w}, D_{\text{obs}}) = \frac{L(\underline{\theta}, \underline{w}, \nu | D)}{\sum_{\nu=c-n_0-1}^c L(\underline{\theta}, \underline{w}, \nu | D)}, \quad (3.2.6)$$

onde $L(\underline{\theta}, \underline{w}, \nu | D)$ é proporcional à função em (3.1.9).

Caso 2: Tomando a distribuição à priori em (3.2.5) e trocando $\pi(\nu)$ pela distribuição exponencial com parâmetro λ_0 , então as distribuições condicionais completas univariadas à posteriori de $x_i, \alpha, \beta, \mu, \sigma_x^2, \sigma^2$ e w_i , baseadas em (3.2.3) coincidem com as do caso 1 e a condicional completa de ν é também exponencial de parâmetro λ_0 .

3.3 O modelo estrutural t -dependente com o número de graus de liberdade conhecido

Nessa seção apresentamos o modelo de regressão linear simples t -dependente com erros nas variáveis. Da mesma forma que na seção 3.1 vamos considerar duas formulações equivalentes para esse modelo e além disso, discutiremos a implementação de algoritmos MCMC.

3.3.1 Definição do modelo

Formulação 1: Sejam as equações

$$\begin{cases} Y_i = \alpha + \beta x_i + e_i, & i = 1, \dots, n, \\ X_i = x_i + u_i, \end{cases}$$

onde (Y_i, X_i) são as variáveis aleatórias observáveis, x_i a variável latente medida com erro e u_i o erro na medição de x_i com a suposição de que

$$\underline{Z} | \theta \sim t_{3n}(\underline{1}_n \otimes \underline{\mu}; I_{n \times n} \otimes \underline{\Sigma}, \nu),$$

onde

$$\underline{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} \quad \text{e} \quad Z_i = \begin{pmatrix} Y_i \\ X_i \\ x_i \end{pmatrix}$$

e onde

$$\underline{\mu} = \begin{pmatrix} \alpha + \beta\mu \\ \mu \\ \mu \end{pmatrix}, \quad \underline{\Sigma} = \begin{pmatrix} \beta^2\sigma_x^2 + \sigma_e^2 & \beta\sigma_x^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma_u^2 & \sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 & \sigma_x^2 \end{pmatrix}, \quad (3.3.1)$$

com $\sigma_e^2 = \lambda\sigma_u^2$ para λ conhecida, $\nu > 0$ conhecido, e $|\underline{\Sigma}| \neq 0$.

O espaço paramétrico nesse caso é

$$\Theta = \{\theta = (\alpha, \beta, \mu, \sigma_x^2, \sigma^2)^T : \alpha, \beta, \mu \in \mathbb{R} \text{ e } \sigma_x^2, \sigma^2 \in \mathbb{R}_+\}.$$

A distribuição à priori de θ nesse modelo é denotada por $\pi(\theta)$.

Observação 1: Note que $Z|\theta \sim t_{3n}(\underline{1}_n \otimes \mu : I_{n \times n} \otimes \Sigma, \nu) \not\stackrel{\text{iid}}{\sim} Z_i|\theta \sim t_3(\mu, \Sigma, \nu)$, para $i = 1, \dots, n$.

Observação 2: O modelo t -dependente é não identificável pela mesma razão que o modelo t -independente também é.

Observação 3: Se $\nu \geq 3$, Z tem matriz de covariâncias ($\text{Var}(Z) = I_{n \times n} \otimes \Sigma \frac{\nu}{\nu-2}$) finita.

Função de verossimilhança baseada nos dados completos

A função de verossimilhança baseada nos dados completos $D = Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$ é dada por

$$\begin{aligned} L(\theta|D) &\propto |I_{n \times n} \otimes \Sigma|^{-\frac{1}{2}} \left[\nu + (Z - \underline{1}_n \otimes \mu)^\top (I_{n \times n} \otimes \Sigma)^{-1} (Z - \underline{1}_n \otimes \mu) \right]^{-\frac{(3n+\nu)}{2}} \\ &= |\Sigma|^{-\frac{n}{2}} \left[\nu + \sum_{i=1}^n (Z_i - \mu)^\top \Sigma^{-1} (Z_i - \mu) \right]^{-\frac{(3n+\nu)}{2}}, \end{aligned} \quad (3.3.2)$$

onde μ e Σ são dados em (3.3.1).

Função de verossimilhança baseada nos dados observados

A função de verossimilhança baseada nos dados observados $D_{\text{obs}} = \begin{pmatrix} U_1 \\ \vdots \\ U_n \end{pmatrix}$, onde $U_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix}$, é dada por

$$L(\theta|D_{\text{obs}}) \propto |\Sigma^*|^{-\frac{n}{2}} \left[\nu + \sum_{i=1}^n (U_i - \mu^*)^\top \Sigma^{*-1} (U_i - \mu^*) \right]^{-\frac{(2n+\nu)}{2}}, \quad (3.3.3)$$

onde

$$\mu^* = \begin{pmatrix} \alpha + \beta\mu \\ \mu \end{pmatrix} \quad \text{e} \quad \Sigma^* = \begin{pmatrix} \beta^2\sigma_x^2 + \lambda\sigma^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma_u^2 \end{pmatrix}.$$

Distribuição à posteriori

A posteriori $p(\theta|D_{\text{obs}})$ é dada por

$$p(\theta|D_{\text{obs}}) = \int_{\mathcal{X}=\mathbb{R}^n} p(\theta, \mathcal{X}|D_{\text{obs}})d\mathcal{X} = \int_{\mathcal{X}} L(\theta|D)\pi(\theta)d\mathcal{X}, \quad (3.3.4)$$

onde $L(\theta|D)$ é dada por (3.3.2), ou então

$$p(\theta|D_{\text{obs}}) \propto L(\theta|D_{\text{obs}})\pi(\theta), \quad (3.3.5)$$

onde $L(\theta|D_{\text{obs}})$ é dada por (??).

Formulação 2: o modelo de mistura de escala da distribuição Normal

Sejam as equações

$$\begin{cases} Y_i = \alpha + \beta x_i + e_i, & i = 1, \dots, n, \\ X_i = x_i + e_i, \end{cases}$$

e seja $Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$, onde $Z_i = \begin{pmatrix} Y_i \\ X_i \\ x_i \end{pmatrix}$, com a suposição

$$Z|\theta, w \sim N_{3n}(\mathbf{1} \otimes \mu; I_{n \times n} \otimes \Sigma w),$$

onde μ e Σ são definidas em (3.3.1) e priori dada por $\pi(\theta, w) = \pi(\theta)\pi(w)$, onde $w \sim IG(\frac{\nu}{2}, \frac{\nu}{2})$, isto é,

$$\pi(w) = w^{-(\frac{\nu}{2}+1)} \exp\left\{-\frac{\nu}{2w}\right\}, \quad w > 0. \quad (3.3.6)$$

Essa suposição é equivalente a

$$Z_i|\theta, w, \stackrel{\text{iid}}{\sim} N(\mu, \Sigma, w), \quad \text{para } i = 1, \dots, n,$$

onde $\pi(\vartheta, w) = \pi(\vartheta)\pi(w)$, com $w \sim IG(\frac{\nu}{2}, \frac{\nu}{2})$.

Função de verossimilhança baseada nos dados completos

A função de verossimilhança baseada nos dados completos $D = \underline{Z}$ é dada por

$$L(\vartheta, w|D) \propto |\underline{\Sigma}|^{-\frac{n}{2}} w^{-\frac{3n}{2}} \exp \left\{ -\frac{1}{2w} \sum_{i=1}^n (\underline{Z}_i - \underline{\mu})^\top \underline{\Sigma}^{-1} (\underline{Z}_i - \underline{\mu}) \right\} \quad (3.3.7)$$

$$\begin{aligned} \propto w^{-\frac{3n}{2}} ((\sigma^2)^2 \sigma_x^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2w} \left[\sum_{i=1}^n \frac{(Y_i - \alpha - \beta x_i)^2}{\lambda \sigma^2} \right. \right. \\ \left. \left. + \sum_{i=1}^n \frac{(X_i - x_i)^2}{\sigma^2} + \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma_x^2} \right] \right\}. \end{aligned} \quad (3.3.8)$$

Vemos que (3.3.8) coincide com (3.1.9) quando fazemos $w_i = w$, $i = 1, \dots, n$.

Função de verossimilhança baseada nos dados observados

A função de verossimilhança baseada nos dados observados $D_{\text{obs}} = \begin{pmatrix} \underline{U}_1 \\ \vdots \\ \underline{U}_n \end{pmatrix}$,

onde $\underline{U}_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix}$, é dada por

$$L(\vartheta, w|D_{\text{obs}}) \propto |\underline{\Sigma}^*|^{-\frac{n}{2}} w^{-\frac{2n}{2}} \exp \left\{ -\frac{1}{2w} \sum_{i=1}^n (\underline{U}_i - \underline{\mu}^*)^\top \underline{\Sigma}^{*-1} (\underline{U}_i - \underline{\mu}^*) \right\}, \quad (3.3.9)$$

onde $\underline{\mu}^*$ e $\underline{\Sigma}^*$ são definidos em (3.1.3) (temos que (3.3.9) coincide com (3.1.10) quando $w_i = w$, $i = 1, \dots, n$).

Distribuição à posteriori

A posteriori $p(\vartheta|D_{\text{obs}})$ é dada por

$$\begin{aligned} p(\vartheta|D_{\text{obs}}) &= \int_{\mathcal{X}=\mathbb{R}^n} \int_{W=\mathbb{R}_+} p(\vartheta, w, \underline{x}|D_{\text{obs}}) dw d\underline{x} \\ &\propto \int_{\mathcal{X}=\mathbb{R}^n} \int_{W=\mathbb{R}_+} L(\vartheta, w|D) \pi(\vartheta, w) dw d\underline{x}, \end{aligned} \quad (3.3.10)$$

onde $L(\vartheta, w|D)$ é dada em (3.3.8) e $\pi(\vartheta, w)$ é a priori de (ϑ, w) dada em (3.3.6) (vemos que (3.3.10) coincide com (3.1.11) quando $w_i = w$, $i = 1, \dots, n$).

Outra forma de escrever a posteriori é

$$p(\theta|D_{\text{obs}}) \propto \int_{W=\mathbb{R}_+} L(\theta, w|D_{\text{obs}})\pi(\theta, w)dw, \quad (3.3.11)$$

onde $L(\theta, w|D_{\text{obs}})$ é dada em (3.3.9) e $\pi(\theta, w)$ é dada em (3.3.6) (vemos que (3.3.11) coincide com (3.1.12) quando $w_i = w, i = 1, \dots, n$).

Observação: A posteriori de θ em (3.3.10), obtida a partir do modelo formulado como mistura de escala da distribuição normal coincide com a posteriori em (3.3.4), obtida sob o modelo na formulação 1 (o que implica que as duas formulações são equivalentes). A justificativa desse fato segue abaixo.

Demonstração:

$$\begin{aligned} p(\theta|D_{\text{obs}}) &= \int_{\mathfrak{X}=\mathbb{R}^n} \int_{W=\mathbb{R}_+} p(\theta, w, \mathfrak{z}|D_{\text{obs}})dwd\mathfrak{z} \\ &\propto \int_{\mathfrak{X}} \int_W L(\theta, w|D)\pi(\theta)\pi(w)dwd\mathfrak{z} \\ &= \int_{\mathfrak{X}} \int_W |\Sigma|^{-\frac{n}{2}} w^{-\frac{3n}{2}} \exp\left\{-\frac{1}{2w} \sum_{i=1}^n (Z_i - \mu)^\top \Sigma^{-1} (Z_i - \mu)\right\} \\ &\quad \cdot \pi(\theta)w^{-(\frac{\nu}{2}+1)} \exp\left\{-\frac{\nu}{2w}\right\} dwd\mathfrak{z} \\ &= |\Sigma|^{-\frac{n}{2}} \pi(\theta) \int_{\mathfrak{X}} \int_W w^{-(\frac{3n}{2}+\frac{\nu}{2}+1)} \exp\left\{\frac{-\frac{1}{2} \left[\sum_{i=1}^n (Z_i - \mu)^\top \Sigma^{-1} (Z_i - \mu) + \nu\right]}{w}\right\} dwd\mathfrak{z} \\ &= |\Sigma|^{-\frac{n}{2}} \pi(\theta) \int_{\mathfrak{X}} \frac{\Gamma\left(\frac{3n+\nu}{2}\right)}{\frac{1}{2} \left[\sum_{i=1}^n (Z_i - \mu)^\top \Sigma^{-1} (Z_i - \mu) + \nu\right]^{\frac{3n+\nu}{2}}} d\mathfrak{z} \\ &\propto \int_{\mathfrak{X}} |\Sigma|^{-\frac{n}{2}} \left[\sum_{i=1}^n (Z_i - \mu)^\top \Sigma^{-1} (Z_i - \mu) + \nu\right]^{-\left(\frac{3n+\nu}{2}\right)} \pi(\theta)d\mathfrak{z}, \end{aligned}$$

que coincide com (3.3.4).

3.3.2 Implementação de algoritmos MCMC no modelo com particular priori própria de componentes independentes

Tomando-se a distribuição à priori própria

$$\pi(\theta, w) = \pi(\alpha)\pi(\beta)\pi(\mu)\pi(\sigma_x^2)\pi(\sigma^2)\pi(w),$$

onde $\alpha \sim N(a, \sigma_x^2)$, $\beta \sim N(b, \sigma_\beta^2)$, $\mu \sim N(m, \sigma_\mu^2)$, $\sigma_x^2 \sim IG(c, d)$, $\sigma^2 \sim IG(f, g)$, $w \sim IG(\frac{\nu}{2}, \frac{\nu}{2})$ e utilizando o modelo t -dependente formulado como mistura de escala da distribuição normal dado em (3.3.6) podemos implementar o algoritmo de Gibbs usando as distribuições condicionais completas baseadas na posteriori conjunta $p(\theta, \underline{x}, w | D_{\text{obs}})$ dada em (3.3.10). Observamos entretanto que as condicionais completas à posteriori de α , β , μ , σ_x^2 e σ^2 coincidem com as do modelo t -independente da seção 3.1.2 quando fazemos $w_i = w$, $i = 1, \dots, n$.

A distribuição condicional completa de w , entretanto, é dada por

$$w | \underline{x}, \alpha, \beta, \mu, \sigma_x^2, \sigma^2, D_{\text{obs}} \sim IG \left[\frac{3n + \nu}{2}; \frac{\nu}{2(\nu - 2)} \left(\frac{\sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2}{\lambda \sigma_1^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma_1^2} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_x^2} + \nu \right) \right].$$

É importante lembrar que qualquer algoritmo utilizado para o modelo t -independente (na formulação 2) pode ser utilizado para o modelo t -dependente, bastando que $w_i = w$, para $i = 1, \dots, n$.

3.3.3 Distribuição preditiva

Seja $\underline{U}_f = \begin{pmatrix} Y_f \\ X_f \end{pmatrix}$ uma observação futura de $\begin{pmatrix} Y \\ X \end{pmatrix}$. Então a distribuição preditiva à posteriori no modelo t -dependente é dada por

$$\begin{aligned} \pi(\underline{U}_f | D_{\text{obs}}) &= \int_{\Theta} \int_{W=\mathbb{R}_+} \pi(\underline{U}_f, \theta, w | D_{\text{obs}}) dw d\theta \\ &= \int_{\Theta} \int_{W=\mathbb{R}_+} \pi(\underline{U}_f | \theta, w) \pi(\theta, w | D_{\text{obs}}) dw d\theta \\ &= \int_{\Theta} \int_{W=\mathbb{R}_+} \pi(Y_f | X_f, \theta, w) \pi(X_f | \theta, w) \pi(\theta, w | D_{\text{obs}}) dw d\theta, \end{aligned}$$

onde $\pi(\theta, w | D_{\text{obs}})$ é a distribuição à posteriori de θ e w ,

$$X_f | \theta, w \sim N(\mu; (\sigma_x^2 + \sigma^2)w)$$

$$\text{e } Y_f | X_f, \theta, w \sim N\left(\alpha + \beta \mu + \frac{\beta \sigma_x^2}{\sigma_x^2 + \sigma^2} (X_f - \mu); \left[\beta^2 \sigma_x^2 + \lambda \sigma^2 - \frac{\beta^2 (\sigma_x^2)^2}{\sigma_x^2 + \sigma^2} \right] w\right).$$

Para gerar uma observação U_f basta gerar (ϱ, w) da posteriori $\pi(\varrho, w|D_{\text{obs}})$, depois gerar X_f de $\pi(X_f|\varrho, w)$ e finalmente Y_f de $\pi(Y_f|X_f, \varrho, w)$.

Observação: A posteriori preditiva pode ser escrita também de outro modo:

$$\begin{aligned}\pi(U_f|D_{\text{obs}}) &= \int_{\mathfrak{X}_f} \int_{\Theta} \int_{W=\mathbb{R}_+} \pi(U_f, \varrho, w, x_f|D_{\text{obs}}) dw d\varrho dx_f \\ &= \int_{\mathfrak{X}_f} \int_W \int_{\Theta} \pi(U_f|\varrho, w, x_f) \pi(x_f|\varrho, w) \pi(\varrho, w|D_{\text{obs}}) d\varrho dw dx_f \\ &= \int_{\mathfrak{X}_f} \int_W \int_{\Theta} \pi(Y_f|X_f, \varrho, w, x_f) \pi(X_f|\varrho, x_f, w) \pi(x_f|\varrho, w) \pi(\varrho, w|D_{\text{obs}}) d\varrho dw dx_f,\end{aligned}$$

onde $x_f|\varrho, w \sim N(\mu, \sigma_x^2 w)$, $X_f|x_f, \varrho, w \sim N(x_f, \sigma_u^2 w)$,

$Y_f|X_f, x_f, \varrho, w \sim N(\alpha + \beta x_f, \sigma_e^2 w)$.

3.3.4 Avaliação do ajuste

A avaliação da qualidade do ajuste no modelo t -dependente, definido por (3.3.1) ou por (3.3.6), pode ser feita de forma semelhante à do modelo t -independente descrita na seção 3.1.4, fazendo-se as devidas adaptações nas fórmulas de CPO_i , $E(U_i|D_{\text{obs}(-i)})$ e $\text{Var}(U_i|D_{\text{obs}(-i)})$.

3.4 O modelo estrutural t -dependente com graus de liberdade desconhecidos

Nessa seção definimos o modelo t -dependente com erros nas variáveis e número de graus de liberdade desconhecido, formulado como mistura de escala da distribuição Normal.

Considere as equações

$$\begin{cases} Y_i = \alpha + \beta x_i + e_i, \\ X_i = x_i + u_i, \end{cases}$$

e seja $\underline{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$, onde $Z_i = \begin{pmatrix} Y_i \\ X_i \\ x_i \end{pmatrix}$ com a suposição de que

$$\underline{Z} \sim N_{3n}(\underline{1} \otimes \mu; I_{n \times n} \otimes \Sigma w),$$

onde μ e Σ são dadas em (3.2.1), com priori $\pi(\theta, w, \nu)$ dada por

$$\pi(\theta, w, \nu) = \pi(\theta)\pi(w|\nu)\pi(\nu), \text{ onde } w|\nu \sim IG\left(\frac{\nu}{2}, \frac{\nu}{2}\right).$$

Esta suposição é equivalente a $Z_i \stackrel{\text{iid}}{\sim} N_3(\mu, \Sigma w)$, com priori $\pi(\theta, w, \nu) = \pi(\theta)\pi(w|\nu)\pi(\nu)$, onde $w|\nu \sim IG\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$.

Vemos que no modelo t -dependente temos um único w onde $w|\nu \sim IG\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ (enquanto que no modelo t -independente $w_i|\nu \stackrel{\text{iid}}{\sim} IG\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$, $i = 1, \dots, n$) e portanto a análise do modelo e implementação dos algoritmos pode ser feita de forma semelhante à descrita na seção 3.2.

3.5 Condições para a existência da posteriori no modelo estrutural t -Student com erros dependentes e número de graus de liberdade conhecido

Inicialmente apresentamos um resultado que será útil para demonstrar os teoremas.

Lema 3.5.1 *Sejam a e p constantes positivas. Então a condição necessária e suficiente para que a integral $\int_0^\infty \frac{e^{-a/x}}{x^p} dx$ seja finita é que $p > 1$.*

Demonstração: Se $0 < p \leq 1$, então

$$\int_0^\infty \frac{e^{-a/x}}{x^p} dx > \int_1^\infty \frac{e^{-a/x}}{x^p} dx > \int_1^\infty \frac{e^{-a}}{x^p} dx \propto \int_1^\infty \frac{1}{x^p} dx = \infty$$

e portanto, $\int_0^\infty \frac{e^{-a/x}}{x^p} dx$ diverge se $0 < p \leq 1$.

Se $p > 1$, usamos a transformação $t = \frac{a}{x}$ e portanto a integral

$$\int_0^\infty \frac{e^{-a/x}}{x^p} dx = a^{1-p} \int_0^\infty e^{-t} t^{p-2} dt = a^{1-p} \int_0^\infty e^{-t} t^{(p-1)-1} dt,$$

Fazendo $r = p - 1$, então para todo $p > 1$ (isto é, para todo $r > 0$), temos que

$$\int_0^\infty \frac{e^{-a/x}}{x^p} dx = a^{-r} \int_0^\infty e^{-t} t^{r-1} dt = a^{-r} \Gamma(r),$$

onde $\Gamma(r)$ é a função Gama com parâmetro r . Portanto,

$$\begin{aligned} \int_0^\infty \frac{e^{-a/x}}{x^p} dx &= a^{-r}(r-1)\Gamma(r-1) = a^{-r}(r-1)(r-2)\Gamma(r-2) = \dots = \\ &= a^{-r}(r-1)(r-2)\dots(r-k) \int_0^\infty e^{-t} t^{(r-k)-1} dt, \end{aligned}$$

onde k é o maior inteiro menor que r e portanto $0 < r - k < 1$. Além disso,

$$\begin{aligned} \Gamma(r-k) &= \int_0^\infty e^{-t} t^{(r-k)-1} dt = \int_0^1 \frac{e^{-t}}{t^{1-(r-k)}} dt + \int_1^\infty \frac{e^{-t}}{t^{1-(r-k)}} dt \\ &\leq \int_0^1 1 dt + \int_1^\infty e^{-t} dt = \frac{1}{r-k} + e^{-1} < \infty. \end{aligned}$$

Portanto, para todo $r > 0$, isto é, para todo $p > 1$, $\int_0^\infty \frac{e^{-a/x}}{x^p} dx < \infty$.

Concluimos então que $\int_0^\infty \frac{e^{-a/x}}{x^p} dx$ é finita se e somente se $p > 1$. ■

Teoremas

Teorema 3.5.2 *No modelo t -Student com erros dependentes descrito na seção 3.4, com priori $\pi(\theta, w) = \pi(w) \cdot \pi(\theta)$, onde $\pi(w) \sim IG(\frac{\nu}{2}, \frac{u}{2})$ e $\pi(\theta) \propto \pi(\sigma_x^2) \cdot \pi(\sigma^2)$, onde $\pi(\sigma_x^2) \sim IG(c, d)$, $\pi(\sigma^2) \sim IG(f, g)$, então uma condição suficiente para a existência da posteriori é que $n \geq 3$ e $0 < c < n + \frac{(\nu-3)}{2}$ (outra condição suficiente é que $n \geq 3$ e $0 < f < \frac{n+\nu-1}{2}$).*

Demonstração: O objetivo é mostrar que a integral $\int_W \int_\Theta p(\theta, w | D_{\text{obs}}) d\theta dw$ é finita, onde $\int_W \int_\Theta p(\theta, w | D_{\text{obs}}) d\theta dw = \int_W \int_{\mathcal{X}} \int_\Theta p(\theta, w, \mathcal{X} | D_{\text{obs}}) d\theta d\mathcal{X} dw$, onde $p(\theta, w, \mathcal{X} | D_{\text{obs}}) \propto L(\theta, w | D) \pi(\theta, w)$, que é proporcional a

$$\begin{aligned} &w^{-(\frac{3n+\nu}{2}+1)} (\sigma_x^2)^{-(\frac{n}{2}+c+1)} (\sigma^2)^{-(n+f+1)} \cdot \exp \left\{ -\frac{1}{2} \left[\frac{\nu}{w} + \frac{2d}{\sigma_x^2} + \frac{2g}{\sigma^2} + \frac{\sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2}{\lambda \sigma^2 w} \right. \right. \\ &\left. \left. + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma^2 w} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_x^2 w} \right] \right\}. \end{aligned} \quad (3.5.1)$$

Primeiramente, fazemos a integração em relação a Θ , que é bastante semelhante à do modelo Normal da seção anterior, e obtemos

$$\int_\Theta p(\theta, w, \mathcal{X} | D_{\text{obs}}) d\theta$$

$$\propto \frac{w^{-(\frac{3n+\nu-1}{2})} \exp\left\{-\frac{\nu}{2w}\right\}}{S_{xx}^{\frac{1}{2}} \left[2d + \frac{S_{xx}}{w}\right]^{\frac{n+2c-1}{2}} \left[\frac{\sum_{i=1}^n (X_i - x_i)^2}{w} + \frac{(S_{YY}S_{xx} - S_{xY}^2)}{\lambda S_{xx}w} + 2g\right]^{n+f-1}}.$$

Portanto,

$$\int_{\mathfrak{X}} \int_{\Theta} p(\varrho, w, \mathfrak{x} | D_{\text{obs}}) d\varrho d\mathfrak{x} \propto \int_{\mathfrak{X}} \frac{w^{-(\frac{3n+\nu-1}{2})} \exp\left\{-\frac{\nu}{2w}\right\}}{S_{xx}^{\frac{1}{2}} \left[2d + \frac{S_{xx}}{w}\right]^{\frac{n+2c-1}{2}} \left[\frac{\sum_{i=1}^n (X_i - x_i)^2}{w} + \frac{(S_{YY}S_{xx} - S_{xY}^2)}{\lambda S_{xx}w} + 2g\right]^{n+f-1}} dx. \quad (3.5.2)$$

Fazendo a transformação que a cada $(\mathfrak{x}, \underline{X}, \underline{Y})$ associa $(z_{n \times 1}, \underline{b}_{n \times 1}, \underline{\xi}_{n \times 1}) = (H_{n \times n} \mathfrak{x}, H_{n \times n} \underline{X}, H_{n \times n} \underline{Y})$, onde H é a matriz de Helmert (Searle, 1971) e definindo $z_x, \underline{b}_X, \underline{\xi}_Y$ como sendo os vetores formados pelas primeiras $n-1$ componentes de z, \underline{b} e $\underline{\xi}$, respectivamente, então a integral em (3.5.2) torna-se igual à integral

$$\int_{Z=\{z \in \mathbb{R}^n\}} \frac{w^{-(\frac{3n+\nu-1}{2})} \exp\left\{-\frac{\nu}{2w}\right\}}{\|z_x\| \left[\frac{\|z_x\|^2}{w} + 2d\right]^{\frac{n+2c-1}{2}} \left[\frac{\|z_x - \underline{b}_X\|^2}{w} + \frac{\|\underline{\xi}_Y\|^2 \text{sen}^2 \theta}{\lambda w} + \frac{(\sqrt{n}X - z_n)^2}{w} + 2g\right]^{n+f-1}} dz, \quad (3.5.3)$$

onde $z = (z_1, \dots, z_n)^T$.

Para eliminar a variável z_n , fazemos o cálculo da integral apenas em relação a z_n de forma semelhante à apresentada da demonstração do Teorema 2.8.2. Portanto, se a desigualdade $n + f > \frac{3}{2}$ é válida, então a integral em (3.5.3) torna-se proporcional a

$$\int_{Z_x} \frac{w^{-(\frac{3n+\nu-1}{2})} \exp\left\{-\frac{\nu}{2w}\right\}}{\|z_x\| \left[\frac{\|z_x\|^2}{w} + 2d\right]^{\frac{n+2c-1}{2}} \left[\frac{\|z_x - \underline{b}_X\|^2}{w} + \frac{\|\underline{\xi}_Y\|^2 \text{sen}^2 \theta}{\lambda w} + 2g\right]^{n+f-\frac{3}{2}}} dz_x,$$

onde $Z_x = \{z_x = (z_1, \dots, z_{n-1})^T \in \mathbb{R}^{n-1}\}$,

A função dentro do sinal de integral é

$$f(z_x, w) = \frac{w^{-(\frac{3n+\nu-1}{2})} \exp\left\{-\frac{\nu}{2w}\right\}}{\|z_x\| \left[\frac{\|z_x\|^2}{w} + 2d\right]^{\frac{n+2c-1}{2}} \left[\frac{\|z_x - \underline{b}_X\|^2}{w} + \frac{\|\underline{\xi}_Y\|^2 \text{sen}^2 \theta}{\lambda w} + 2g\right]^{n+f-\frac{3}{2}}},$$

que é igual a

$$\frac{\exp\left\{-\frac{\nu}{2w}\right\}}{w^{\frac{n}{2}+1-c-f} \|z_x\| [2dw + \|z_x\|^2]^{\frac{n+2c-1}{2}} \left[2gw + \|z_x - \underline{b}_X\|^2 + \frac{\|\underline{\xi}_Y\|^2 \text{sen}^2 \theta}{\lambda}\right]^{n+f-\frac{3}{2}}}. \quad (3.5.4)$$

Portanto, $\int_W \left[\int_{\mathcal{X}} \int_{\Theta} p(\varrho, w, \mathcal{X} | D_{\text{obs}}) d\varrho d\mathcal{X} \right] dw \propto \int_W \int_{\mathcal{Z}_x} f(z_x, w) dz_x dw$. Para analisar a integral $\int_W \int_{\mathcal{Z}_x} f(z_x, w) dz_x dw$, vamos particionar a região de integração $W \times \mathcal{Z}_x$ em dois conjuntos, $W \times R_1 = \mathbb{R}^+ \times \{z_x \in \mathbb{R}^{n-1} : \|z_x\| < \varepsilon\}$ e $W \times R_2 = \mathbb{R}^+ \times \{z_x \in \mathbb{R}^{n-1} : \|z_x\| \geq \varepsilon\}$, onde ε é fixado arbitrariamente tão pequeno quanto se queira.

AVALIAÇÃO DA INTEGRAL NA REGIÃO $W \times R_1$

Nessa região a integral

$$\int_W \int_{R_1} f(z_x, w) dz_x dw = \int_{\mathbb{R}^+} \int_{\{z_x \in \mathbb{R}^{n-1} : \|z_x\| < \varepsilon\}} f(z_x, w) dz_x dw$$

pode ser majorada pela seguinte integral

$$\begin{aligned} & \int_W \int_{\|z_x\| < \varepsilon} \frac{\exp\left\{-\frac{\nu}{2w}\right\}}{w^{\frac{\nu}{2}+1-c-f} \|z_x\| [2dw+0]^{\frac{n+2c-1}{2}} [2gw+0]^{n+f-\frac{3}{2}}} dz_x dw \\ & \propto \int_W \frac{\exp\left\{-\frac{\nu}{2w}\right\}}{w^{\frac{3n+\nu}{2}-1}} \left[\int_{\|z_x\| < \varepsilon} \frac{1}{\|z_x\|} dz_x \right] dw = \int_W \frac{\exp\left\{-\frac{\nu}{2w}\right\}}{w^{\frac{3n+\nu}{2}-1}} dw \int_{\|z_x\| < \varepsilon} \frac{1}{\|z_x\|} dz_x, \end{aligned}$$

onde pelo Lema 3.5.1, a integral $\int_0^\infty \frac{\exp\left\{-\frac{\nu}{2w}\right\}}{w^{\frac{3n+\nu}{2}-1}} dw$ é finita se e somente se $\frac{3n+\nu}{2} - 1 > 1$ (ou seja, $3n + \nu > 4$) e, pelo Lema 2.8.1, a integral $\int_{\|z_x\| < \varepsilon} \frac{1}{\|z_x\|} dz_x$ é finita se e somente se $1 < n - 1$.

Portanto, na região $W \times R_1$ uma condição suficiente para que a integral convirja é que $n \geq 3$ e $3n + \nu > 4$, ou seja, simplesmente $n \geq 3$.

AVALIAÇÃO DA INTEGRAL NA REGIÃO $W \times R_2$

Nessa região, a integral $\int_W \int_{R_2} f(z_x, w) dz_x dw = \int_{W=\mathbb{R}^+} \int_{\|z_x\| \geq \varepsilon} f(z_x, w) dz_x dw$, pode ser majorada pela seguinte integral:

$$\begin{aligned} & \int_W \int_{\|z_x\| \geq \varepsilon} \frac{\exp\left\{-\frac{\nu}{2w}\right\}}{w^{\frac{\nu}{2}+1-c-f} \|z_x\| [0 + \|z_x\|^2]^{\frac{n+2c-1}{2}} [2gw+0]^{n+f-\frac{3}{2}}} dz_x dw \\ & = \int_W \frac{\exp\left\{-\frac{\nu}{2w}\right\}}{w^{\frac{\nu}{2}+1-c-f} [2gw+0]^{n+f-\frac{3}{2}}} dw \cdot \int_{\|z_x\| \geq \varepsilon} \frac{1}{\|z_x\| [0 + \|z_x\|^2]^{\frac{n+2c-1}{2}}} dz_x, \end{aligned}$$

que é proporcional a

$$\int_W \frac{\exp\left\{-\frac{\nu}{2w}\right\}}{w^{\frac{2n+\nu-2c-1}{2}}} dw \cdot \int_{\|z_x\| \geq \varepsilon} \frac{1}{\|z_x\|^{n+2c}} dz_x.$$

Pelo Lema 3.5.1, a integral em w é finita se e somente se $\frac{2n+\nu-2c-1}{2} > 1$ (isto é, $2n + \nu > 2c + 3$). E pelo Lema 2.8.1 a integral em z_x é finita se e somente se $n + 2c > n - 1$ (ou seja, sempre).

Portanto, na região $W \times R_2$ uma condição suficiente para haver convergência da integral $\int_W \int_{R_2} f(z_x, w) dz_x dw$ é que $2n + \nu > 2c + 3$.

Conclusão: Analisando conjuntamente todas as informações, concluímos que uma condição suficiente para a existência da distribuição a posteriori é que $n \geq 3$ e $0 < c < n + \frac{\nu-3}{2}$.

Observação: Uma outra condição suficiente igualmente interessante é obtida quando na região $W \times R_2$ a integral $\int_{W=\mathbb{R}} \int_{R_2} f(z_x, w) dz dw$ é majorada pela seguinte integral:

$$\begin{aligned} & \int_W \int_{\|z_x\| \geq \varepsilon} \frac{\exp\left\{-\frac{\nu}{2w}\right\}}{w^{\frac{\nu}{2}+1-c-f} \|z_x\| [2dw+0]^{\frac{n+2c-1}{2}} \left[0 + \|z_x - \underline{b}_X\|^2 + \frac{\|\underline{c}_Y\|^2 \sin^2 \theta}{2}\right]^{n+f-\frac{3}{2}}} dz_x dw \\ & \propto \int_0^\infty \frac{\exp\left\{-\frac{\nu}{2w}\right\}}{w^{\frac{n+\nu-2f+1}{2}}} dw \cdot \int_{\|z_x\| \geq \varepsilon} \frac{1}{\|z_x\| \left[\|z_x - \underline{b}_X\|^2 + \frac{\|\underline{c}_Y\|^2 \sin^2 \theta}{\lambda}\right]^{n+f-\frac{3}{2}}} dz_x. \end{aligned}$$

Pelo Lema 3.5.1, a integral em w é finita se e somente se $\frac{n+\nu-2f+1}{2} > 1$ (isto é, $n + \nu > 2f + 1$). A integral em z_x (que é semelhante à integral em R_3 na demonstração do Teorema 2.8.5) converge se e somente se $1 + 2(n + f - \frac{3}{2}) > n - 1$, isto é, $n + 2f > 1$ e portanto vale sempre. Assim, a integral $\int_W \int_{R_2} f(z_x, w) dz_x dw < \infty$ se $n + \nu > 2f + 1$.

Portanto, uma outra condição suficiente para que a posteriori seja própria é que $n \geq 3$, $n + \nu > 2f + 1$ e $n + f > \frac{3}{2}$ (ou seja, $n \geq 3$ e $0 < f < \frac{n+\nu-1}{2}$). ■

Teorema 3.5.3 *No modelo do Teorema 3.5.2, se substituirmos a distribuição $\pi(\sigma_x^2)$ por $\pi(\sigma_x^2) \propto (\sigma_x^2)^{-r}$, $r > 0$, então uma condição suficiente para a existência da posteriori correspondente é que $n \geq 2$, $0 < r < \frac{1}{2}$ e $0 < f < \frac{\nu}{2} + 1 - r$.*

Demonstração: Como a distribuição à priori neste teorema é igual à priori no Teorema 3.5.2 quando substituimos os hiperparâmetros c e d , respectivamente, por $(r - 1)$ e zero, então os cálculos feitos na demonstração do Teorema 3.5.2 podem ser reaproveitados. De fato,

$$\int_W \int_{\Theta} p(\theta, w | D_{\text{obs}}) d\theta dw \propto \int_W \int_{\mathcal{Z}_x = \{z_x \in \mathbb{R}^{n-1}\}} g(z_x, w) dz_x dw, \text{ onde} \quad (3.5.5)$$

$$g(z_x, w) = \frac{\exp\left\{-\frac{\nu}{2w}\right\}}{w^{\frac{\nu}{2}-r-f+2} \|z_x\|^{n+2r-2} \left[\|z_x - \tilde{b}_X\|^2 + \frac{\|C_Y\|^2 \sin^2 \theta}{\lambda} + 2gw\right]^{n+f-\frac{3}{2}}}$$

e $g(z_x, w)$ coincide com a função $f(z_x, w)$ dada em (3.5.4) quando substituimos c por $r - 1$ e d por zero (desde que a desigualdade $n + f > \frac{3}{2}$ seja válida).

Do mesmo modo que no Teorema 3.5.2, a região de integração $W \times \mathcal{Z}_x$ é particionada em duas regiões: $W \times R_1$ e $W \times R_2$, onde

$$W \times R_1 = \mathbb{R}^+ \times \{z_x \in \mathbb{R}^{n-1} : \|z_x\| < \varepsilon\} \quad \text{e} \quad W \times R_2 = W \times R_1^c.$$

AValiação da Integral na Região $W \times R_1$

$$\int_{W \times R_1} g(z_x, w) dz_x dw = \int_W \int_{\|z_x\| < \varepsilon} g(z_x, w) dz_x dw,$$

que pode ser majorada pela integral

$$\int_W \int_{\|z_x\| < \varepsilon} \frac{\exp\left\{-\frac{\nu}{2w}\right\}}{w^{\frac{\nu}{2}-r+2-f} \|z_x\|^{n+2r-2} [0 + 0 + 2gw]^{n+f-\frac{3}{2}}} dz_x dw,$$

que é proporcional a

$$\int_0^\infty \frac{\exp\left\{-\frac{\nu}{2w}\right\}}{w^{\frac{2n+\nu-2r+1}{2}}} dw \int_{\|z_x\| < \varepsilon} \frac{1}{\|z_x\|^{n+2r-2}} dz_x,$$

onde a integral em z_x é finita se e somente se $n + 2r - 2 < n - 1$ (isto é, $0 < r < \frac{1}{2}$). Além disso, pelo Lema 3.5.1 a integral em w é finita se e somente se $\frac{2n+\nu-2r+1}{2} > 1$ (ou seja, $2n + \nu > 2r + 1$). Mas, como esta última condição está satisfeita sempre que $0 < r < \frac{1}{2}$, concluímos que uma condição suficiente para que a integral na região $W \times R_1$ convirja é que $0 < r < \frac{1}{2}$.

AVALIAÇÃO DA INTEGRAL NA REGIÃO $W \times R_2$

Nessa região, a integral $\int_W \int_{R_2} g(z_x, w) dz_x dw$ pode ser majorada por

$$\begin{aligned} & \int_W \int_{\|z_x\| > \varepsilon} \frac{\exp\left\{-\frac{\nu}{2w}\right\}}{w^{\frac{\nu}{2}-r-f+2} \|z_x\|^{n+2r-2} \left[\|z_x - \underline{b}_X\|^2 + \frac{\|\underline{\xi}_Y\|^2 \sin^2 \theta}{\lambda} + 0\right]^{n+f-\frac{3}{2}}} dz_x dw \\ &= \int_W \frac{\exp\left\{-\frac{\nu}{2w}\right\}}{w^{\frac{\nu}{2}-r-f+2}} dw \cdot \int_{\|z_x\| > \varepsilon} \frac{1}{\|z_x\|^{n+2r-2} \left[\|z_x - \underline{b}_X\|^2 + \frac{\|\underline{\xi}_Y\|^2 \sin^2 \theta}{\lambda}\right]^{n+f-\frac{3}{2}}} dz_x, \end{aligned}$$

onde a integral em w converge se e somente se $\frac{\nu}{2} - r - f + 2 > 1$ (isto é, $f < \frac{\nu}{2} + 1 - r$) e a integral em z_x converge se e somente se $n + 2r - 2 + 2(n + f - \frac{3}{2}) > n - 1$ (isto é, $n + r + f > 2$). Entretanto, pelo fato que n deve ser maior ou igual a 2 (ver observação abaixo), a condição $n + r + f > 2$ é sempre verdadeira e portanto nessa região, uma condição suficiente para que a integral $\int_W \int_{R_2} g(z_x, w) dz_x dw$ seja finita é que $f < \frac{\nu}{2} + 1 - r$ e $n \geq 2$.

Analisando simultaneamente essas informações, concluímos então que uma condição suficiente para que exista a posteriori nesse teorema é que $0 < r < \frac{1}{2}$, $0 < f < \frac{\nu}{2} + 1 - r$ e $n \geq 2$. ■

Observação: Quando $n = 1$, a distribuição a posteriori nesse modelo não está bem definida.

De fato, substituindo na integral em (3.5.2) os valores $d = 0$ e $c = r - 1$ e fazendo $S_{xx} = S_{x_1 x_1}$, $S_{xY} = S_{x_1 Y_1}$ e $S_{YY} = S_{Y_1 Y_1}$, então a integral $\int_{\mathcal{X}=\mathbb{R}} \int_{\Theta} p(\vartheta, w, x_1 | D_{\text{obs}}) d\vartheta dx_1$

é proporcional a

$$\int_{\mathfrak{X}=\mathbb{R}} \frac{\left\{-\frac{\nu}{2w}\right\}}{w^{\frac{\nu+2}{2}} S_{x_1 x_1}^{\frac{1}{2}} \left[\frac{S_{x_1 x_1}}{w}\right]^{r-1} \left[\frac{(X_1-x_1)^2}{w} + \frac{S_{x_1 x_1} S_{Y_1 Y_1} - S_{x_1 Y_1}^2}{\lambda S_{x_1 x_1} w} + 2g\right]^f} dx_1.$$

Portanto, $\int_{W=\mathbb{R}^+} \int_{\mathfrak{X}=\mathbb{R}^+} \int_{\Theta} p(\vartheta, w, x_1 | D_{\text{obs}}) d\vartheta dx_1 dw$ é proporcional a

$$\int_{W=\mathbb{R}^+} \int_{\mathfrak{X}=\mathbb{R}} \frac{\exp\left\{-\frac{\nu}{2w}\right\} S_{x_1 x_1}^{f-r+\frac{1}{2}}}{w^{\frac{\nu}{2}-r-f+2} \left[S_{x_1 x_1} (X_1-x_1)^2 + \frac{(S_{x_1 x_1} S_{Y_1 Y_1} - S_{x_1 Y_1}^2)}{\lambda} + 2gw\right]^f} dx_1 dw.$$

Como $S_{x_1 x_1}$ e $S_{x_1 Y_1}$ são nulos, $\forall x_1 \in \mathbb{R}$, esta integral está bem definida e é finita se e somente se $f = 0$ e $r < \frac{1}{2}$ (e nesse caso ela é nula), mas por hipótese f é positivo e portanto a posteriori não está bem definida para $n = 1$.

Teorema 3.5.4 *No modelo do Teorema 3.5.2, se substituirmos a distribuição $\pi(\sigma^2)$ por $\pi(\sigma^2) \propto (\sigma^2)^{-s}$, $s > 0$, e se os vetores de dados \underline{X} e \underline{Y} são linearmente independentes e as componentes de \underline{X} (como as de \underline{Y}) não são todas iguais então uma condição suficiente para a existência da posteriori é que $n + \nu > 2s - 1$ e $n \geq 3$:*

Demonstração: A demonstração desse teorema é semelhante à do Teorema 3.5.5 a seguir e por este motivo não a apresentamos aqui. ■

Teorema 3.5.5 *No modelo do Teorema 3.5.2 se substituirmos as distribuições $\pi(\sigma_x^2)$ e $\pi(\sigma^2)$, respectivamente, por $\pi(\sigma_x^2) \propto (\sigma_x^2)^{-r}$ e $\pi(\sigma^2) \propto (\sigma^2)^{-s}$, $r, s > 0$, e se os vetores de dados \underline{X} e \underline{Y} são linearmente independentes e as componentes de \underline{X} (como também as de \underline{Y}) não são todas iguais, então a condição necessária e suficiente para que a posteriori correspondente seja própria é que $n + r + s > 3$, $0 < r < \frac{1}{2}$ e $0 < s < \frac{\nu}{2} + 2 - r$.*

Demonstração: Como a distribuição à posteriori coincide com a posteriori do Teorema 3.5.2 quando substituimos c por $r - 1$, f por $s - 1$, d por zero e g por zero, então

muito dos cálculos necessários para demonstrar esse teorema são reaproveitados, fazendo-se as devidas substituições. De fato, a integral $\int_{W=\mathbb{R}^+} \int_{\Theta} p(\theta, w | D_{\text{obs}}) d\theta dw$ é proporcional a $\int_{W=\mathbb{R}^+} \int_{Z_x=\{z_x \in \mathbb{R}^{n-1}\}} h(z_x, w) dz_x dw$, onde

$$h(z_x, w) = \frac{\exp\left\{-\frac{\nu}{2w}\right\}}{w^{\frac{\nu}{2}-r-s+3} \|z_x\|^{n+2r-2} \left[\|z_x - \hat{b}_X\|^2 + \frac{\|\zeta_Y\|^2 \text{sen}^2 \theta}{\lambda}\right]^{n+s-\frac{5}{2}}}, \quad (3.5.6)$$

desde que a condição $n+s > \frac{5}{2}$ (que é igual à (2.8.12) quando $f = s-1$) seja válida. (Note que $h(z_x, w)$ é igual à função em (3.5.4) quando são feitas as substituições acima.)

Para analisar a integral dessa função, vamos particionar a região de integração Z_x de forma semelhante à usada na demonstração do Teorema 2.8.5 para o modelo Normal, isto é, $W \times Z_x = W \times [R_1 \cup R_2 \cup R_3]$, onde $R_2 = R_{2a} \cup R_{2b} \cup R_{2c}$ (onde os conjuntos $R_1, R_2, R_3, R_{2a}, R_{2b}$ e R_{2c} são definidos da mesma forma que na demonstração do Teorema 2.8.5).

AVALIAÇÃO DA INTEGRAL NA REGIÃO $W \times R_1$

A integral de $h(z_x, w)$ em $W \times R_1$ é igual ao produto

$$\int_{\mathbb{R}^+} \frac{\exp\left\{-\frac{\nu}{2w}\right\}}{w^{\frac{\nu}{2}-r-s+3}} dw \cdot \int_{\|z_x\| < \varepsilon_1} \frac{1}{\|z_x\|^{n+2r-2} \left[\|z_x - \hat{b}_X\|^2 + \frac{\|\zeta_Y\|^2 \text{sen}^2 \theta}{\lambda}\right]^{n+s-\frac{5}{2}}} dz_x,$$

onde pelo Lema 3.5.1 a integral em w é finita se e somente se $\frac{\nu}{2} - r - s + 3 > 1$ (isto é, $s < \frac{\nu}{2} + 2 - r$). A integral em z_x (que coincide com a integral da função $g(z_x)$ em (2.8.14) na região R_1) é finita se e somente se $n+2r-2 < n-1$ (isto é, $r < \frac{1}{2}$) (ver demonstração do Teorema 2.8.5 para o modelo Normal). Portanto, em $W \times R_1$, a integral de $h(z_x, w)$ é finita se e somente se $0 < r < \frac{1}{2}$ e $0 < s < \frac{\nu}{2} + 2 - r$.

AVALIAÇÃO DA INTEGRAL NA REGIÃO $W \times R_2$

A integral $\int_W \int_{R_2} h(z_x, w) dz_x dw$ é igual a

$$\int_{W=\mathbb{R}^+} \frac{\exp\left\{\frac{\nu}{2w}\right\}}{w^{\frac{\nu}{2}-r-s+3}} dw \int_{R_2=R_{2a} \cup R_{2b} \cup R_{2c}} \frac{1}{\|z_x\|^{n+2r-2} \left[\|z_x - \underline{b}_X\|^2 + \frac{\|\underline{c}_Y\|^2 \sin^2 \theta}{\lambda} \right]^{n+s-\frac{5}{2}}} dz_x,$$

onde a integral em w é finita se e somente se $\frac{\nu}{2} - r - s + 3 > 1$ (isto é, $0 < s < \frac{\nu}{2} + 2 - r$). As integrais em z_x nas regiões R_{2a} , R_{2b} e R_{2c} são sempre finitas, como já foi demonstrado no Teorema 2.8.5 da seção anterior. Portanto, em $W \times R_2$ a integral de $h(z_x, w)$ é finita se e somente se $0 < s < \frac{\nu}{2} + 2 - r$.

AVALIAÇÃO DA INTEGRAL NA REGIÃO $W \times R_3$

$$\int_{w=\mathbb{R}^+} \int_{R_3} h(z_x, w) dz_x dw$$

$$= \int_{\mathbb{R}^+} \frac{\exp\left\{-\frac{\nu}{2w}\right\}}{w^{\frac{\nu}{2}-r-s+3}} dw \int_{\|z_x\|>k} \frac{1}{\|z_x\|^{n+2r-2} \left[\|z_x - \underline{b}_X\|^2 + \frac{\|\underline{c}_Y\|^2 \sin^2 \theta}{\lambda} \right]^{n+s-\frac{5}{2}}} dz_x,$$

onde a integral em w converge se e somente se $0 < s < \frac{\nu}{2} + 2 - r$ e a integral em z_x (na região R_3) é finita se e somente se $n + 2r - 2 + 2(n + s - \frac{5}{2}) > n - 1$ (isto é, $n + r + s > 3$), como já foi demonstrado no Teorema 2.8.5 para o modelo Normal. Portanto, a integral $\int_W \int_{R_3} h(z_x, w) dz_x dw$ é finita se e somente se $0 < s < \frac{\nu}{2} + 2 - r$ e $n + r + s > 3$.

Conclusão: Analisando-se simultaneamente todas as informações, concluímos que a condição necessária e suficiente para que a posteriori seja própria é que $0 < r < \frac{1}{2}$, $0 < \frac{\nu}{2} + 2 - r$ e $n + r + s > 3$. ■

3.6 Sensibilidade do modelo a diferentes valores do número de graus de liberdade

Nessa seção estudamos o efeito nas inferências à posteriori à medida que fixamos diferentes valores para o número de graus de liberdade do modelo estrutural t -Student com ν conhecido (mantendo-se todos os demais hiperparâmetros fixados).

Sabemos que a distribuição t aproxima-se da Normal quando $\nu \rightarrow \infty$ e portanto uma maneira de avaliar a sensibilidade à suposição de Normalidade (ou seja, à verossimilhança Normal) é verificar se as inferências à posteriori mudam muito quando o inverso do número de graus de liberdade, $\frac{1}{\nu}$, varia entre 0 e 1. Uma maneira descritiva de fazer isso é construir um gráfico onde no eixo das abscissas colocamos os valores de $\frac{1}{\nu}$ e no das ordenadas, os valores correspondentes da média (ou mediana, moda, variância, quartis, etc.) da posteriori marginal de cada parâmetro.

O procedimento acima pode ser feito para os modelos t -dependente e t -independente, com ou sem restrição de identificabilidade e com priori própria ou imprópria (desde que a posteriori seja própria).

3.7 Algumas medidas de comparação de modelos

(1) O critério da medida L

A seguir apresentamos a medida L para o modelo de regressão estrutural simples t -independente com número de graus de liberdade fixado e com restrição de identificabilidade $\sigma_e^2 = \lambda\sigma_u^2 = \lambda\sigma^2$.

$$L_2(D_{\text{obs}}, \delta) = \sum_{i=1}^n \text{Var}(\underline{U}_i^* | D_{\text{obs}}) + \delta \sum_{i=1}^n (E(\underline{U}_i^* | D_{\text{obs}}) - \underline{U}_i)(E(\underline{U}_i^* | D_{\text{obs}}) - \underline{U}_i)^T,$$

onde $D_{\text{obs}} = \begin{pmatrix} \underline{U}_1 \\ \vdots \\ \underline{U}_n \end{pmatrix}$ com $\underline{U}_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix}$.

Seja $\underline{U}_1^*, \dots, \underline{U}_n^*$ n vetores de observações futuras com a mesma distribuição dos dados $\left(\underline{U}_i^* = \begin{pmatrix} Y_{j_i} \\ X_{j_i} \end{pmatrix}\right)$. Seja $\boldsymbol{\theta} = (\alpha, \beta, \mu, \sigma_x^2, \sigma^2)^T$ e $\boldsymbol{w} = (w_1, \dots, w_n^T)$, onde

$$E(\underline{U}_i^* | D_{\text{obs}}) = E[E(\underline{U}_i^* | \boldsymbol{\theta}, \boldsymbol{w}) | D_{\text{obs}}] = E_{\boldsymbol{\theta}, \boldsymbol{w} | D_{\text{obs}}}[E(\underline{U}_i^* | \boldsymbol{\theta}, \boldsymbol{w})],$$

$$\text{Var}(\underline{U}_i^* | D_{\text{obs}}) = E[\text{Var}(\underline{U}_i^* | \boldsymbol{\theta}, \boldsymbol{w}) | D_{\text{obs}}] + \text{Var}[E(\underline{U}_i^* | \boldsymbol{\theta}, \boldsymbol{w}) | D_{\text{obs}}]$$

$$\begin{aligned}
&= E_{\underline{\theta}, \underline{w} | D_{\text{obs}}} [\text{Var}(\underline{U}_i^* | \underline{\theta}, \underline{w})] + E_{\underline{\theta}, \underline{w} | D_{\text{obs}}} [(E(\underline{U}_i^* | \underline{\theta}, \underline{w}))(E(\underline{U}_i^* | \underline{\theta}, \underline{w}))^\top] \\
&\quad - [E_{\underline{\theta}, \underline{w} | D_{\text{obs}}}(E[\underline{U}_i^* | \underline{\theta}, \underline{w}])][E_{\underline{\theta}, \underline{w} | D_{\text{obs}}}(E[\underline{U}_i^* | \underline{\theta}, \underline{w}])]^\top,
\end{aligned}$$

onde

$$\begin{aligned}
E(\underline{U}_i^* | \underline{\theta}, \underline{w}) &= \begin{pmatrix} \alpha + \beta\mu \\ \mu \end{pmatrix}, \\
\text{Var}(\underline{U}_i^* | \underline{\theta}, \underline{w}) &= \begin{bmatrix} \beta^2\sigma_x^2 + \lambda\sigma^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma^2 \end{bmatrix} w_i, \quad i = 1, \dots, n.
\end{aligned}$$

Dada uma amostra $(\underline{\theta}^{(1)}, \underline{w}^{(1)}, \dots, \underline{\theta}^{(G)}, \underline{w}^{(G)})$ da posteriori $\pi(\underline{\theta}, \underline{w}, \underline{x} | D_{\text{obs}})$ definida em (3.1.11), podemos estimar a medida L usando o método simples de Monte Carlo por

$$\begin{aligned}
\widehat{L}_2(D_{\text{obs}}, \delta) &= n \left\{ \frac{1}{G} \sum_{g=1}^G (\text{Var}(\underline{U}_i^* | \underline{\theta}^{(g)}, \underline{w}^{(g)}) + [E(\underline{U}_i^* | \underline{\theta}^{(g)}, \underline{w}^{(g)})][E(\underline{U}_i^* | \underline{\theta}^{(g)}, \underline{w}^{(g)})]^\top) \right. \\
&\quad \left. - \left[\frac{1}{G} \sum_{g=1}^G E(\underline{U}_i^* | \underline{\theta}^{(g)}, \underline{w}^{(g)}) \right] \left[\frac{1}{G} \sum_{g=1}^G E(\underline{U}_i^* | \underline{\theta}^{(g)}, \underline{w}^{(g)}) \right]^\top \right\} \\
&\quad + \delta \sum_{i=1}^n \left[\underline{U}_i - \frac{1}{G} \sum_{g=1}^G E(\underline{U}_i^* | \underline{\theta}^{(g)}, \underline{w}^{(g)}) \right] \left[\underline{U}_i - \frac{1}{G} \sum_{g=1}^G E(\underline{U}_i^* | \underline{\theta}^{(g)}, \underline{w}^{(g)}) \right]^\top,
\end{aligned}$$

onde $\delta \in [0, 1]$.

Observação: No caso do modelo t -dependente, o cálculo da medida L é semelhante, bastando substituir \underline{w} por w . Portanto, se tomarmos uma amostra de tamanho G , $(\underline{\theta}^{(1)}, w^{(1)}, \dots, \underline{\theta}^{(G)}, w^{(G)})$ da posteriori $\pi(\underline{\theta}, w, \underline{x} | D_{\text{obs}})$ definida em (3.3.10) podemos obter $\widehat{L}_2(D_{\text{obs}}, \delta)$ de forma semelhante (substituindo $\underline{w}^{(g)}$ por $w^{(g)}$).

(2) O critério da informação (BIC)

$$\text{BIC}_{M_i} = -2 \log(\sup \pi_{M_i}(\underline{U}_1, \dots, \underline{U}_n | \underline{\theta})) + p_i \log n,$$

onde

$$\begin{aligned}
p_i &= \text{número de parâmetros do modelo } M_i, \\
n &= \text{tamanho da amostra,} \\
\underline{U}_i &= \begin{pmatrix} Y_i \\ X_i \end{pmatrix} \text{ é a } i\text{-ésima observação, } i = 1, \dots, n.
\end{aligned}$$

No modelo estrutural t -dependente com número de graus de liberdade, ν , conhecido,

$$\pi_{M_i}(\underline{U}_1, \dots, \underline{U}_n | \varrho) = K(2n, \nu) |\hat{\Sigma}|^{-\frac{1}{2}} (\nu + (\underline{U} - \underline{\mu}^*)^\top \hat{\Sigma}^{-1} (\underline{U} - \underline{\mu}^*))^{-\frac{(2n+\nu)}{2}},$$

onde $K(2n, \nu) = \frac{\Gamma(\frac{2n+\nu}{2}) \nu^{\nu/2}}{\Gamma(\frac{\nu}{2}) \pi^{2n/2}}$, $\underline{U} = \begin{pmatrix} \underline{U}_1 \\ \vdots \\ \underline{U}_n \end{pmatrix}$, $\underline{\mu}^* = \mathbf{1}_n \otimes \underline{\mu}$ e $\hat{\Sigma} = I_{n \times n} \otimes \hat{\Sigma}$ e onde $\underline{\mu}$ e $\hat{\Sigma}$ são definidas em (3.3.1). Portanto,

$$\pi_{M_i}(\underline{U}_1, \dots, \underline{U}_n | \varrho) = k(2n, \nu) |\Sigma|^{-\frac{n}{2}} \left(\nu + \sum_{i=1}^n (\underline{U}_i - \underline{\mu})^\top \Sigma^{-1} (\underline{U}_i - \underline{\mu}) \right)^{-\frac{(2n+\nu)}{2}}.$$

Os estimadores de máxima verossimilhança de $\underline{\mu}$ e $\hat{\Sigma}$ são dados, respectivamente, por

$$\hat{\underline{\mu}} = \frac{1}{n} \sum_{i=1}^n \underline{U}_i \quad \text{e} \quad \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (\underline{U}_i - \hat{\underline{\mu}})(\underline{U}_i - \hat{\underline{\mu}})^\top$$

(ver Arellano-Valle, 1994, página 200).

No modelo t -independente com graus de liberdade conhecidos,

$$\begin{aligned} \pi_{M_i}(\underline{U}_1, \dots, \underline{U}_n | \varrho) &= \prod_{i=1}^n K(2, \nu) |\Sigma|^{-\frac{1}{2}} (\nu + (\underline{U}_i - \underline{\mu})^\top \Sigma^{-1} (\underline{U}_i - \underline{\mu})) \\ &= (2\pi)^{-n} \nu^{n(\frac{\nu}{2}+1)} \prod_{i=1}^n (\nu + (\underline{U}_i - \underline{\mu})^\top \Sigma^{-1} (\underline{U}_i - \underline{\mu}))^{-(1+\frac{\nu}{2})}. \end{aligned}$$

Os EMV de $\underline{\mu}$ e $\hat{\Sigma}$ podem ser obtidos por meio do algoritmo EM aplicado no modelo reparametrizado ortogonalmente (ver Arellano-Valle, 1994, página 206).

(3) Algumas medidas alternativas para a seleção de modelos

(a) O produto $\prod_{i=1}^n \text{CPO}_i$.

Para cada modelo competitivo M calculamos o produto $\prod_{i=1}^n \text{CPO}_i^{(M)}$ e escolhemos o modelo cujo produto é máximo.

No caso do modelo estrutural t -independente com número de graus de liberdade conhecido, podemos estimar CPO_i por meio das fórmulas (3.1.25) ou (3.1.27) ou

(3.1.29). No caso do modelo t -dependente, o cálculo do CPO_i pode ser obtido de forma semelhante fazendo-se as devidas adaptações.

(b) A soma dos resíduos de validação cruzada $\sum_{i=1}^n \left| \frac{U_i - E(U_i | D_{\text{obs}(-i)})}{\sqrt{\text{Var}(U_i | D_{\text{obs}(-i)})}} \right|$.

Para cada modelo M calculamos essa soma e escolhemos aquele cuja soma é mínima. No modelo estrutural t -independente com número de graus de liberdade conhecido, o valor esperado $E(U_i | D_{\text{obs}(-i)})$ e a variância, $\text{Var}(U_i | D_{\text{obs}(-i)})$, podem ser estimados, respectivamente, pelas fórmulas (3.1.33) e (3.1.35).

No caso do modelo t -dependente, $\hat{E}(U_i | D_{\text{obs}})$ e $\widehat{\text{Var}}(U_i | D_{\text{obs}})$ são obtidos de forma semelhante.

Capítulo 4

Uma análise do modelo estrutural Normal com erros nas variáveis com um ponto de mudança desconhecido

Introdução

Neste capítulo vamos propor uma metodologia bayesiana para analisar o modelo de regressão linear simples com erros nas variáveis e mudança nos parâmetros após um determinado ponto da seqüência de observações indexadas pelo tempo (ou espaço ou outra variável qualquer); as mudanças podendo ocorrer em todos os parâmetros, como também em qualquer subconjunto dos parâmetros.

Nesse capítulo estudamos o modelo onde os erros (e a covariável) são normalmente distribuídos e no Capítulo 5, estudaremos o caso em que os erros seguem uma distribuição *t*-Student.

Na literatura, Chang e Huang (1997) estudaram o modelo estrutural Normal com mudança somente na média da covariável e sob o ponto de vista da estatística clássica. Sua metodologia, entretanto, aparentemente pode ser estendida somente para testar mudança apenas no intercepto da reta de regressão no caso dos erros normais e nada mais.

A análise inferencial desenvolvida nesse capítulo engloba tanto o problema de detectar a existência de mudança, como também o de estimar a posição do(s) ponto(s) de mudança (caso haja) e estimar os demais parâmetros de interesse (como por exemplo a magnitude da diferença entre os parâmetros antes e depois do ponto de mudança), como também as próprias distribuições marginais de cada parâmetro, além disso engloba também a análise preditiva.

O problema de detectar mudança pode ser formulado como um problema de teste de hipóteses ou um problema de escolha de modelos. No teste de hipótese testamos a hipótese nula de que não há mudança ao longo da seqüência de n observações, isto é, $H_0 : k = n$ (onde k é o ponto de mudança), versus a hipótese alternativa de que ocorre mudança após o k -ésimo ponto da seqüência, isto é, $H_1 : 1 \leq k \leq n - 1$ (ou equivalentemente $H_0 : \alpha_1 = \alpha_2, \beta_1 = \beta_2, \mu_1 = \mu_2, \sigma_{x_1}^2 = \sigma_{x_2}^2, \sigma_1^2 = \sigma_2^2$ versus H_1 : pelo menos uma diferença).

O problema de detectar múltiplos pontos de mudança pode ser resolvido de forma muito simples aplicando-se o método da “segmentação binária” utilizado em Chen e Gupta (1997). Este método consiste, no primeiro passo, em testar a hipótese nula (H_0) de que não há mudança versus a hipótese alternativa de que há uma única mudança. Se não há mudança, H_0 é aceita. Caso contrário, o ponto de mudança divide a seqüência original de observações em duas subseqüências. Para cada subseqüência repetimos o processo de testar H_0 versus H_1 da mesma forma que no primeiro passo, e continuamos o processo até que nenhuma mudança seja encontrada em quaisquer das subseqüências.

Nas seções 4.1 a 4.6 aparece a fórmulação envolvendo apenas o modelo identificável onde $\sigma_e^2 = \lambda \sigma_u^2$, já que a fórmulação para o caso não identificável é muito semelhante e tornaria o estudo muito repetitivo. Entretanto na seção 4.7 temos um exemplo de aplicação envolvendo o modelo não identificável e que utilizada toda a a

metodologia da seção.

4.1 O modelo estrutural com um ponto de mudança e priori com componentes independentes

Nessa seção inicialmente apresentamos o modelo linear estrutural que possui um único ponto de mudança (desconhecido) e que permite que a mudança possa ocorrer em todos os parâmetros após esse ponto. Em seguida apresentamos o modelo que permite mudança apenas em um subconjunto estritamente contido no conjunto de parâmetros.

O modelo que permite mudança em todos os parâmetros

O modelo que permite mudança em todos os parâmetros (e pelo menos um deles muda) após a k -ésima observação da sequência ordenada de dados (onde k é desconhecido) e tem a restrição de identificabilidade $\sigma_{e_i}^2 = \lambda\sigma_i^2$, (onde $\sigma_i^2 = \sigma_{u_i}^2$, para $i = 1, 2$, com λ conhecido), é definido por

$$\begin{cases} Y_i = \begin{cases} \alpha_1 + \beta_1 x_i + e_i, & i = 1, \dots, k, \\ \alpha_2 + \beta_2 x_i + e_i, & i = k + 1, \dots, n, \end{cases} & 1 \leq k \leq n - 1, \\ X_i = x_i + u_i, & i = 1, \dots, n, \end{cases}$$

onde Y_i e X_i são as variáveis aleatórias observáveis, x_i é a variável aleatória latente e e_i , u_i , os erros aleatórios, com a suposição que $\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix}$ são independentes, para $i = 1, \dots, n$, e

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \sim \begin{cases} N \left(\begin{pmatrix} 0 \\ 0 \\ \mu_1 \end{pmatrix}; \begin{pmatrix} \lambda\sigma_1^2 & 0 & 0 \\ 0 & \sigma_1^2 & 0 \\ 0 & 0 & \sigma_{x(1)}^2 \end{pmatrix} \right), & i = 1, \dots, k, \\ N \left(\begin{pmatrix} 0 \\ 0 \\ \mu_2 \end{pmatrix}; \begin{pmatrix} \lambda\sigma_2^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_{x(2)}^2 \end{pmatrix} \right), & i = k + 1, \dots, n, \end{cases} \quad (4.1.1)$$

Nesse modelo o vetor de parâmetros é $\underline{\theta}^* = \begin{pmatrix} k \\ \underline{\theta}_1 \\ \underline{\theta}_2 \end{pmatrix}$, onde $\underline{\theta}_i = (\alpha_i, \beta_i, \mu_i, \sigma_{x(i)}^2, \sigma_i^2)^T$, para $i = 1, 2$ e o espaço paramétrico é $\Theta^* = \{\underline{\theta}^* : k \in \mathcal{K}, (\alpha_i, \beta_i, \mu_i) \in \mathbb{R}^3 \text{ e } (\sigma_{x(i)}^2, \sigma_i^2) \in \mathbb{R}_+^2, i = 1, 2\}$.

Nesse modelo estamos assumindo que $(\alpha_1, \beta_1, \mu_1, \sigma_{x(1)}^2, \sigma_1^2) \neq (\alpha_2, \beta_2, \mu_2, \sigma_{x(2)}^2, \sigma_2^2)$, isto é, existe pelo menos uma diferença.

Distribuição à priori

A distribuição à priori de θ^* escolhida tem a forma

$$\pi(\underline{\theta}^*) = p(k)\pi(\underline{\theta}_1)\pi(\underline{\theta}_2) = p(k) \prod_{i=1}^2 \pi(\alpha_i)\pi(\beta_i)\pi(\mu_i)\pi(\sigma_{x(i)}^2)\pi(\sigma_i^2), \quad (4.1.2)$$

$$\text{onde } p(k) = \begin{cases} P(K = k), & k \in \mathcal{K} = \{1, \dots, n-1\}, \\ 0, & \text{caso contrário.} \end{cases}$$

Observações:

- Nesse modelo estamos assumindo à priori que existe mudança em pelo menos um parâmetro. Entretanto se à priori não sabemos se existe mudança em algum parâmetro, então devemos desconsiderar a restrição $\underline{\theta}_1 \neq \underline{\theta}_2$ acima.
- Na estatística bayesiana, os parâmetros $\underline{\theta}_1$ e $\underline{\theta}_2$ podem ser estimados qualquer que seja $k, (1 \leq k \leq n-1)$. Entretanto para permitir estimativas “um pouco melhores” dos parâmetros $\underline{\theta}_1$ e $\underline{\theta}_2$, principalmente quando a distribuição à priori não for “muito informativa”, podemos considerar k variando por exemplo no conjunto $\mathcal{K} = \{3, \dots, n-3\}$, ou seja podemos assumir à priori que a probabilidade de k pertencer ao conjunto $\{1, 2, n-2, n-1, n\}$ é zero.

A função de verossimilhança baseada nos dados observados

Seja $D_{\text{obs}} = \{(Y_i, X_i)^\top, i = 1, \dots, n\}$ o conjunto dos dados observados e $D_{\text{obs1}}^{(k)} = \{(Y_i, X_i)^\top, i = 1, \dots, k\}$, $D_{\text{obs2}}^{(n-k)} = \{(Y_i, X_i)^\top, i = k+1, \dots, n\}$, respectivamente, os dados observados até a mudança e os dados observados após a mudança.

A função de verossimilhança $L(\hat{\theta}|D_{\text{obs}}) = L(k, \theta_1, \theta_2|D_{\text{obs}})$ é proporcional a

$$\begin{aligned} & \frac{1}{|\Sigma_1|^{k/2}} \exp \left\{ -\frac{1}{2|\Sigma_1|} \left[(\beta_1^2 \sigma_{x(1)}^2 + \lambda \sigma_1^2) \sum_{i=1}^k (X_i - \mu_1)^2 \right. \right. \\ & \left. \left. - 2\beta_1 \sigma_{x(1)}^2 \sum_{i=1}^k (Y_i - \alpha_1 - \beta_1 \mu_1)(X_i - \mu_1) + (\sigma_{x(1)}^2 + \sigma_1^2) \sum_{i=1}^k (Y_i - \alpha_1 - \beta_1 \mu_1)^2 \right] \right\} \\ & \cdot \frac{1}{|\Sigma_2|^{(n-k)/2}} \exp \left\{ -\frac{1}{2|\Sigma_2|} \left[(\beta_2^2 \sigma_{x(2)}^2 + \lambda \sigma_2^2) \sum_{i=k+1}^n (X_i - \mu_2)^2 \right. \right. \\ & \left. \left. - 2\beta_2 \sigma_{x(2)}^2 \sum_{i=k+1}^n (Y_i - \alpha_2 - \beta_2 \mu_2)(X_i - \mu_2) \right. \right. \\ & \left. \left. + (\sigma_{x(2)}^2 + \sigma_2^2) \sum_{i=k+1}^n (Y_i - \alpha_2 - \beta_2 \mu_2)^2 \right] \right\}, \end{aligned} \quad (4.1.3)$$

onde $|\Sigma_i| = \beta_i^2 \sigma_{x(i)}^2 \sigma_i^2 + \lambda \sigma_i^2 (\sigma_{x(i)}^2 + \sigma_i^2)$, $i = 1, 2$.

Ou seja,

$$L(\hat{\theta}|D_{\text{obs}}) \propto L_1(\theta_1|D_{\text{obs}1}^{(k)}) L_2(\theta_2|D_{\text{obs}2}^{(n-k)}),$$

onde $L_1(\theta_1|D_{\text{obs}1}^{(k)})$ é a verossimilhança baseada nas k primeiras observações e $L_2(\theta_2|D_{\text{obs}2}^{(n-k)})$ a verossimilhança baseada nas observações após o ponto de mudança.

A função de verossimilhança baseada nos dados completos

Sejam $D = \{(Y_i, X_i, x_i)^T, i = 1, \dots, n\}$, $D^{(k)} = \{(Y_i, X_i, x_i)^T, i = 1, \dots, k\}$ e $D^{(n-k)} = \{(Y_i, X_i, x_i)^T, i = k+1, \dots, n\}$. A função de verossimilhança baseada nos dados completos, $L(\hat{\theta}|D) = L(k, \theta_1, \theta_2|D)$ é proporcional a

$$\begin{aligned} & [\sigma_{x(1)}^2 (\sigma_1^2)^2]^{-k/2} \exp \left\{ -\frac{1}{2} \left[\frac{\sum_{i=1}^k (Y_i - \alpha_1 - \beta_1 x_i)^2}{\lambda \sigma_1^2} + \frac{\sum_{i=1}^k (X_i - x_i)^2}{\sigma_1^2} \right. \right. \\ & \left. \left. + \frac{\sum_{i=1}^k (x_i - \mu_1)^2}{\sigma_{x(1)}^2} \right] \right\} [\sigma_{x(2)}^2 (\sigma_2^2)^2]^{-(n-k)/2} \exp \left\{ -\frac{1}{2} \left[\frac{\sum_{i=k+1}^n (Y_i - \alpha_2 - \beta_2 x_i)^2}{\lambda \sigma_2^2} \right. \right. \\ & \left. \left. + \frac{\sum_{i=k+1}^n (X_i - x_i)^2}{\sigma_2^2} + \frac{\sum_{i=k+1}^n (x_i - \mu_2)^2}{\sigma_{x(2)}^2} \right] \right\}, \end{aligned} \quad (4.1.4)$$

ou seja,

$$L(\hat{\theta}|D) \propto L_1(\theta_1|D^{(k)}) L_2(\theta_2|D^{(n-k)}),$$

onde $L_1(\theta_1|D^{(k)})$ é a verossimilhança baseada nos dados completos até a mudança e $L_2(\theta_2|D^{(n-k)})$ a verossimilhança baseada nos dados completos após a mudança.

Distribuição à posteriori

A distribuição à posteriori pode ser expressa de duas formas (ou usando a verossimilhança baseada nos dados observados, ou a verossimilhança baseada nos dados completos), da mesma forma que no Capítulo 2. Ou seja,

$$p(\check{\theta}|D_{\text{obs}}) \propto L(\check{\theta}|D_{\text{obs}})\pi(\check{\theta}),$$

onde $L(\check{\theta}|D_{\text{obs}})$ é proporcional a (4.1.3) e $\pi(\check{\theta})$ é dada por (4.1.2), isto é,

$$L(\check{\theta}|D_{\text{obs}})\pi(\check{\theta}) \propto p(k)L_1(\theta_1|D_{\text{obs}_1}^{(k)})\pi(\theta_1)L_2(\theta_2|D_{\text{obs}_2}^{(n-k)})\pi(\theta_2), \quad (4.1.5)$$

ou então

$$p(\check{\theta}|D_{\text{obs}}) = \int_{\mathcal{X}} p(\underline{x}, \theta^*|D_{\text{obs}})d\underline{x} \propto \int_{\mathcal{X}} L(\check{\theta}|D)\pi(\check{\theta})d\underline{x}, \quad (4.1.6)$$

onde $\mathcal{X} = \{\underline{x} = (x_1, \dots, x_n)^T : x_i \in \mathbb{R}\}$ e $L(\check{\theta}|D)$ é proporcional à expressão (4.1.4) e $\pi(\check{\theta})$ é dada em (4.1.2).

O modelo que permite mudanças num subconjunto de parâmetros

O modelo que permite mudança em apenas um subconjunto fixado, estritamente contido no conjunto de parâmetros, pode ser definido usando o modelo em (4.1.1) como referência, bastando incorporar as devidas restrições. Notemos que existem $\binom{5}{1} + \binom{5}{2} + \binom{5}{3} + \binom{5}{4} = 30$ possíveis modelos desse tipo. Um caso particular é o do modelo que permite mudança somente na média μ , isto é, quando assumimos que os demais parâmetros $(\alpha, \beta, \sigma_x^2, \sigma^2)$ não mudam. Nesse caso, devemos impor ao modelo em (4.1.1) as seguintes restrições: $\alpha_1 = \alpha_2 = \alpha$, $\beta_1 = \beta_2 = \beta$, $\sigma_{x(1)}^2 = \sigma_{x(2)}^2 = \sigma_x^2$ e $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Assim, este modelo é dado por

$$\begin{cases} Y_i = \alpha + \beta x_i + e_i, \\ X_i = x_i + u_i, \end{cases}$$

onde

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \stackrel{\text{iid}}{\sim} \begin{cases} N \left(\begin{pmatrix} 0 \\ 0 \\ \mu_1 \end{pmatrix}; \begin{pmatrix} \lambda\sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma_x^2 \end{pmatrix} \right), & i = 1, \dots, k, \\ N \left(\begin{pmatrix} 0 \\ 0 \\ \mu_2 \end{pmatrix}; \begin{pmatrix} \lambda\sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma_x^2 \end{pmatrix} \right), & i = k + 1, \dots, n, \end{cases} \quad (4.1.7)$$

onde o espaço paramétrico é dado por

$$\Theta^* = \{ \check{\theta} = (k, \alpha, \beta, \mu_1, \mu_2, \sigma_x^2, \sigma^2)^T : k \in \mathcal{K}, (\alpha, \beta, \mu_1, \mu_2) \in \mathbb{R}^4, (\sigma_x^2, \sigma^2) \in \mathbb{R}_+^2 \}$$

e onde a distribuição à priori é dada por

$$\pi(\check{\theta}) = p(k)\pi(\alpha)\pi(\beta)\pi(\mu_1)\pi(\mu_2)\pi(\sigma_x^2)\pi(\sigma^2), \quad (4.1.8)$$

com

$$p(k) = \begin{cases} P[K = k], & k \in \mathcal{K}, \\ 0, & \text{caso contrário.} \end{cases}$$

A função de verossimilhança $L(\check{\theta}|D_{\text{obs}})$ é proporcional à expressão em (4.1.3) quando substituímos α_i por α , β_i por β , $\sigma_{x(i)}^2$ por σ_x^2 e σ_i^2 por σ^2 , para $i = 1, 2$, ou seja,

$$\begin{aligned} L(\check{\theta}|D_{\text{obs}}) &\propto \frac{1}{|\Sigma|^{n/2}} \exp \left\{ -\frac{1}{2|\Sigma|} \left[(\beta^2\sigma_x^2 + \lambda\sigma^2) \left(\sum_{i=1}^k (X_i - \mu_1)^2 \right. \right. \right. \\ &\quad \left. \left. + \sum_{i=k+1}^n (X_i - \mu_2)^2 \right) - 2\beta^2\sigma_x^2 \left(\sum_{i=1}^k (Y_i - \alpha - \beta\mu_1)(X_i - \mu_1) \right. \right. \\ &\quad \left. \left. + \sum_{i=k+1}^n (Y_i - \alpha - \beta\mu_2)(X_i - \mu_2) \right) + (\sigma_x^2 + \sigma^2) \left(\sum_{i=1}^k (Y_i - \alpha - \beta\mu_1)^2 \right. \right. \\ &\quad \left. \left. + \sum_{i=k+1}^n (Y_i - \alpha - \beta\mu_2)^2 \right) \right] \right\}, \quad (4.1.9) \end{aligned}$$

onde $|\Sigma| = \beta^2\sigma_x^2\sigma^2 + \lambda\sigma^2(\sigma_x^2 + \sigma^2)$.

Da mesma forma a função de verossimilhança baseada nos dados completos $L(\check{\theta}|D)$ é proporcional à expressão dada em (4.1.4), fazendo $(\alpha_i, \beta_i, \sigma_{x(i)}^2, \sigma_i^2) = (\alpha, \beta, \sigma_x^2, \sigma^2)$, para $i = 1, 2$.

De forma semelhante define-se qualquer modelo que permite mudança apenas em um subconjunto de parâmetros.

Observação: Poderíamos considerar o conjunto \mathcal{K} (dos possíveis valores de k) variando dependendo do modelo. Por exemplo, no modelo que permite mudança apenas em α e β então $\mathcal{K} = \{2, \dots, n-2\}$ e no modelo que permite mudança apenas em μ , $\mathcal{K} = \{1, \dots, n-1\}$ (já que o tamanho mínimo de amostra para estimar μ na estatística clássica é 1). Entretanto, como já mencionamos anteriormente, na estatística bayesiana isso não é necessário.

4.2 A análise do modelo que permite mudança em todos os parâmetros com particular priori de componentes independentes

Nessa seção o objetivo é fazer inferência sobre o modelo estrutural normal com priori própria de componentes independentes supondo que há mudança (em pelo menos um dos parâmetros) e esta mudança pode eventualmente ocorrer em todos eles.

O modelo é dado em (4.1.1) e a priori escolhida é da forma

$$\pi(\underline{\theta}) = p(k) \prod_{i=1}^2 \pi(\alpha_i) \pi(\beta_i) \pi(\mu_i) \pi(\sigma_{x(i)}^2) \pi(\sigma_i^2), \quad (4.2.1)$$

onde

$$p(k) = \begin{cases} P[K = k] = \frac{1}{\#\mathcal{K}}, & k \in \mathcal{K} = \{1, \dots, n-1\}, \\ 0, & \text{caso contrário,} \end{cases}$$

onde $\#\mathcal{K}$ é o número de elementos do conjunto \mathcal{K}

$\pi(\alpha_i) \sim N(a_i, \sigma_{\alpha_i}^2)$, $\pi(\beta_i) \sim N(b_i, \sigma_{\beta_i}^2)$, $\pi(\mu_i) \sim N(m_i, \sigma_{\mu_i}^2)$, $\pi(\sigma_{x(i)}^2) \sim IG(c_i, d_i)$ e $\pi(\sigma_i^2) \sim IG(f_i, g_i)$, onde $a_i, b_i, m_i, c_i, d_i, f_i, g_i, \sigma_{\alpha_i}^2, \sigma_{\beta_i}^2, \sigma_{\mu_i}^2$ são hiperparâmetros conhecidos para $i = 1, 2$.

4.2.1 Distribuição à posteriori

Como já vimos, a distribuição à posteriori pode ser escrita como

$$p(\underline{\theta} | D_{\text{obs}}) = \int_{\underline{x}} p(\underline{x}, \underline{\theta} | D_{\text{obs}}) d\underline{x} \propto \int_{\underline{\theta}} L(\theta^* | D) \pi(\theta^*) d\underline{\theta}$$

e nesse caso

$$\begin{aligned} p(\underline{x}, \underline{\theta} | D_{\text{obs}}) &\propto (\sigma_1^2)^{-(k+f_1+1)} (\sigma_{x(1)}^2)^{-(\frac{k}{2}+c_1+1)} (\sigma_2^2)^{-(n-k+f_2+1)} (\sigma_{x(2)}^2)^{-(\frac{n-k}{2}+c_2+1)} \\ &\cdot \exp \left\{ -\frac{1}{2} \left[\frac{\sum_{i=1}^k (Y_i - \alpha_1 - \beta_1 x_i)^2 + \lambda (X_i - x_i)^2}{\lambda \sigma_1^2} + \sum_{i=1}^k \frac{(x_i - \mu_1)^2}{\sigma_{x(1)}^2} \right. \right. \\ &+ \frac{\sum_{i=k+1}^n (Y_i - \alpha_2 - \beta_2 x_i)^2 + \lambda (X_i - x_i)^2}{\lambda \sigma_2^2} + \frac{\sum_{i=k+1}^n (x_i - \mu_2)^2}{\sigma_{x(2)}^2} + \frac{(\alpha_1 - a_1)^2}{\sigma_{\alpha_1}^2} \\ &+ \frac{(\alpha_2 - a_2)^2}{\sigma_{\alpha_2}^2} + \frac{(\beta_1 - b_1)^2}{\sigma_{\beta_1}^2} + \frac{(\beta_2 - b_2)^2}{\sigma_{\beta_2}^2} + \frac{(\mu_1 - m_1)^2}{\sigma_{\mu_1}^2} + \frac{(\mu_2 - m_2)^2}{\sigma_{\mu_2}^2} \\ &\left. \left. + \frac{2d_1}{\sigma_{x(1)}^2} + \frac{2d_2}{\sigma_{x(2)}^2} + \frac{2g_1}{\sigma_1^2} + \frac{2g_2}{\sigma_2^2} \right] \right\} p(k). \end{aligned} \quad (4.2.2)$$

4.2.2 Distribuições condicionais completas à posteriori baseadas nos dados completos

Com base em $p(\underline{x}, \underline{\theta} | D_{\text{obs}})$ em (4.2.2) (onde $\underline{\theta} = \begin{pmatrix} k \\ \underline{\theta}_1 \\ \underline{\theta}_2 \end{pmatrix}$), obtemos as condicionais completas de $x_1, x_2, \dots, x_n, \alpha_i, \beta_i, \mu_i, \sigma_{x(i)}^2, \sigma_i^2$ e k , para $i = 1, 2$, dadas pelas fórmulas (4.2.3) a (4.2.14) a seguir.

$$\begin{aligned} &\bullet \quad x_i | x_{(-i)}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, D_{\text{obs}} \\ &\sim \left\{ \begin{array}{l} N \left(\frac{\sigma_{x(1)}^2 [\lambda X_i + \beta_1 (Y_i - \alpha_1) + \lambda \sigma_1^2 \mu_1]}{\sigma_{x(1)}^2 [\lambda + \beta_1^2] + \lambda \sigma_1^2}, \frac{\lambda \sigma_1^2 \sigma_{x(1)}^2}{\sigma_{x(1)}^2 [\lambda + \beta_1^2] + \lambda \sigma_1^2} \right), \\ i = 1, \dots, k \\ \\ N \left(\frac{\sigma_{x(2)}^2 [\lambda X_i + \beta_2 (Y_i - \alpha_2) + \lambda \sigma_2^2 \mu_2]}{\sigma_{x(2)}^2 [\lambda + \beta_2^2] + \lambda \sigma_2^2}, \frac{\lambda \sigma_2^2 \sigma_{x(2)}^2}{\sigma_{x(2)}^2 [\lambda + \beta_2^2] + \lambda \sigma_2^2} \right), \\ i = k + 1, \dots, n \end{array} \right. \end{aligned} \quad (4.2.3)$$

- $\alpha_1 | \mathfrak{X}, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, D_{\text{obs}}$
 $\sim N\left(\frac{\sigma_{\alpha_1}^2 \sum_{i=1}^k (Y_i - \beta_1 x_i) + \lambda \sigma_1^2 a_1}{k \sigma_{\alpha_1}^2 + \lambda \sigma_1^2}; \frac{\lambda \sigma_1^2 \sigma_{\alpha_1}^2}{k \sigma_{\alpha_1}^2 + \lambda \sigma_1^2}\right),$ (4.2.4)

- $\alpha_2 | \mathfrak{X}, \alpha_1, \beta, \beta_2, \mu, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, D_{\text{obs}}$
 $\sim N\left(\frac{\sigma_{\alpha_2}^2 \sum_{i=k+1}^n (Y_i - \beta_2 x_i) + \lambda \sigma_2^2 a_2}{(n-k) \sigma_{\alpha_2}^2 + \lambda \sigma_2^2}; \frac{\lambda \sigma_2^2 \sigma_{\alpha_2}^2}{(n-k) \sigma_{\alpha_2}^2 + \lambda \sigma_2^2}\right),$ (4.2.5)

- $\beta_1 | \mathfrak{X}, \alpha_1, \alpha_2, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, D_{\text{obs}}$
 $\sim N\left(\frac{\sigma_{\beta_1}^2 \sum_{i=1}^k (Y_i - \alpha_1) x_i + \lambda b_1 \sigma_1^2}{\sigma_{\beta_1}^2 \sum_{i=1}^k x_i^2 + \lambda \sigma_1^2}; \frac{\lambda \sigma_1^2 \sigma_{\beta_1}^2}{\sigma_{\beta_1}^2 \sum_{i=1}^k x_i^2 + \lambda \sigma_1^2}\right),$ (4.2.6)

- $\beta_2 | \mathfrak{X}, \alpha_1, \alpha_2, \beta_1, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, D_{\text{obs}}$
 $\sim N\left(\frac{\sigma_{\beta_2}^2 \sum_{i=k+1}^n (Y_i - \alpha_2) x_i + \lambda b_2 \sigma_2^2}{\sigma_{\beta_2}^2 \sum_{i=k+1}^n x_i^2 + \lambda \sigma_2^2}; \frac{\lambda \sigma_2^2 \sigma_{\beta_2}^2}{\sigma_{\beta_2}^2 \sum_{i=k+1}^n x_i^2 + \lambda \sigma_2^2}\right),$ (4.2.7)

- $\mu_1 | \mathfrak{X}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, D_{\text{obs}}$
 $\sim N\left(\frac{\sigma_{\mu_1}^2 \sum_{i=1}^k x_i + \sigma_{x(1)}^2 m_1}{k \sigma_{\mu_1}^2 + \sigma_{x(1)}^2}; \frac{\sigma_{\mu_1}^2 \sigma_{x(1)}^2}{k \sigma_{\mu_1}^2 + \sigma_{x(1)}^2}\right),$ (4.2.8)

- $\mu_2 | \mathfrak{X}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, D_{\text{obs}}$
 $\sim N\left(\frac{\sigma_{\mu_2}^2 \sum_{i=k+1}^n x_i + \sigma_{x(2)}^2 m_2}{(n-k) \sigma_{\mu_2}^2 + \sigma_{x(2)}^2}; \frac{\sigma_{\mu_2}^2 \sigma_{x(2)}^2}{(n-k) \sigma_{\mu_2}^2 + \sigma_{x(2)}^2}\right),$ (4.2.9)

- $\sigma_{x(1)}^2 | \mathfrak{X}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, D_{\text{obs}}$
 $\sim IG\left(\frac{k}{2} + c_1; \frac{1}{2} \sum_{i=1}^k (x_i - \mu_1)^2 + d_1\right),$ (4.2.10)

- $\sigma_{x(2)}^2 | \mathfrak{X}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_1^2, \sigma_2^2, k, D_{\text{obs}}$

$$\sim IG \left(\frac{n-k}{2} + c_2; \frac{1}{2} \sum_{i=k+1}^n (x_i - \mu_2)^2 + d_2 \right), \quad (4.2.11)$$

- $\sigma_1^2 | \mathcal{X}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_2^2, k, D_{\text{obs}}$

$$\sim IG \left(k + f_1; \frac{1}{2} \left[\frac{\sum_{i=1}^k (Y_i - \alpha_1 - \beta_1 x_i)^2}{\lambda} + \sum_{i=1}^k (X_i - x_i)^2 \right] + g_1 \right), \quad (4.2.12)$$

- $\sigma_2^2 | \mathcal{X}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, k, D_{\text{obs}}$

$$\sim IG \left(n-k + f_2; \frac{1}{2} \left[\frac{\sum_{i=k+1}^n (Y_i - \alpha_2 - \beta_2 x_i)^2}{\lambda} + \sum_{i=k+1}^n (X_i - x_i)^2 \right] + g_2 \right), \quad (4.2.13)$$

$$P[K = k | \mathcal{X}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, D_{\text{obs}}] = \frac{L(k, \vartheta_1, \vartheta_2 | D) p(k)}{\sum_{k=1}^{n-1} L(k, \vartheta_1, \vartheta_2 | D) p(k)},$$

$$= \frac{L(k, \vartheta_1, \vartheta_2 | D)}{\sum_{k \in \mathcal{K}} L(k, \vartheta_1, \vartheta_2 | D)}, k \in \mathcal{K}, \quad (4.2.14)$$

onde $L(k, \vartheta_1, \vartheta_2 | D)$ é dada por (4.1.4)

Observação 1: Usando as distribuições condicionais completas acima podemos implementar o algoritmo de Gibbs descrito na seção 2.5.3 com o objetivo de obter uma amostra da distribuição à posteriori cujo núcleo é dado em (4.2.2).

Observação 2: A análise do modelo com priori de componentes condicionalmente independentes, como as do Capítulo 2, pode ser feita de forma semelhante.

4.2.3 Estimação do modelo que permite mudança em todos os parâmetros

Semelhantemente à seção 2.9, podemos obter as estimativas das densidades marginais à posteriori de cada parâmetro do modelo em (4.1.1), bem como as estimativas pontuais da média, mediana, quantis e variância dessas distribuições e os intervalos

de densidade máxima (HPD).

A distribuição marginal à posteriori de K , $P[K = k|D_{\text{obs}}]$ será estimada por

$$\frac{1}{(T - T_0)} \sum_{j=T_0+1}^T P(K = k|\vartheta^{(j)}, \underline{x}^{(j)}, D_{\text{obs}}), \quad k \in \mathcal{K},$$

onde $\vartheta^{(j)} = \begin{pmatrix} \vartheta_1^{(j)} \\ \vartheta_2^{(j)} \end{pmatrix}$, $\vartheta_i^{(j)} = (\alpha_i, \beta_i, \mu_i, \sigma_{x^{(i)}}^2, \sigma_i^2)^T$ e $\underline{x}^{(j)} = (x_1^{(j)}, \dots, x_n^{(j)})^T$,

T = número de iterações,

T_0 = número de iterações iniciais descartadas referentes ao período “burn in”,

$$e \quad P(K = k|\vartheta^{(j)}, \underline{x}^{(j)}, D_{\text{obs}}) = \frac{L(k, \vartheta^{(j)}|D^{(j)})}{\sum_{k \in \mathcal{K}} L(k, \vartheta^{(j)}|D^{(j)}), \quad k \in \mathcal{K},$$

onde $D^{(j)} = \{(Y_i, X_i, x_i^{(j)})^T, i = 1, \dots, n\}$ e $L(k, \vartheta^{(j)}|D^{(j)})$ é dada em (4.1.4).

No caso de m cadeias paralelas o estimador será

$$\frac{1}{m(T - T_0)} \sum_{i=1}^m \sum_{j=T_0+1}^T P(K = k|\vartheta_{(i)}^{(j)}, \underline{x}_{(i)}^{(j)}, D_{\text{obs}})$$

onde $P(K = k|\vartheta_{(i)}^{(j)}, \underline{x}_{(i)}^{(j)}, D_{\text{obs}}) = P(K = k|\vartheta_{(i)}^{(j)}, D_{(i)}^{(j)})$

e $D_{(i)}^{(j)} = \{(Y_k, X_k, x_{k(i)}^{(j)}), k = 1, \dots, n\}$

O ponto de mudança pode ser estimado também pela moda da distribuição acima (que corresponde ao estimador de Bayes quando se adota a função perda 0-1).

A média da posteriori de k pode ser estimada por

$$\frac{1}{m(T - T_0)} \sum_{i=1}^m \sum_{j=T_0+1}^T E[K|\vartheta_i^{(j)}, \underline{x}_i^{(j)}, D_{\text{obs}}],$$

onde

$$E(K|\vartheta_i^{(j)}, \underline{x}_i^{(j)}, D_{\text{obs}}) = \sum_{k \in \mathcal{K}} k P[K = k|\vartheta_i^{(j)}, \underline{x}_i^{(j)}, D_{\text{obs}}]$$

e a variância da posteriori pode ser estimada por

$$\begin{aligned} & \frac{1}{m(T - T_0)} \sum_{i=1}^m \sum_{j=T_0+1}^T \text{Var}(K|\vartheta_i^{(j)}, \underline{x}_i^{(j)}, D_{\text{obs}}) \\ & + \frac{1}{m(T - T_0)} \sum_{i=1}^m \sum_{j=T_0+1}^T \left(E[K|\vartheta_i^{(j)}, \underline{x}_i^{(j)}, D_{\text{obs}}] \right)^2 \\ & - \left(\frac{1}{m(T - T_0)} \sum_{i=1}^m \sum_{j=T_0+1}^T E(K|\vartheta_i^{(j)}, \underline{x}_i^{(j)}, D_{\text{obs}}) \right)^2, \end{aligned}$$

onde

$$\text{Var}(K|\vartheta_i^{(j)}, \mathbf{x}_i^{(j)}, D_{\text{obs}}) = \sum_{k \in \mathcal{K}} k^2 P[K = k|\vartheta_i^{(j)}, \mathbf{x}_i^{(j)}, D_{\text{obs}}] - \left(\sum_{k \in \mathcal{K}} k P[K = k|\vartheta_i^{(j)}, \mathbf{x}_i^{(j)}, D_{\text{obs}}] \right)^2.$$

As distribuições marginais à posteriori dos demais parâmetros do modelo, bem como suas médias e variâncias, podem ser estimados de forma semelhante à da seção 2.9 utilizando as distribuições condicionais completas dadas em (4.2.4) a (4.2.13).

4.3 Implementação do algoritmo de Gibbs no modelo que permite mudança em um subconjunto fixado de parâmetros

Nesta seção o objetivo é implementar o algoritmo de Gibbs para amostrar da posteriori do modelo estrutural com priori própria de componentes independentes e que permite mudança apenas em um subconjunto não vazio contido estritamente em $\{\alpha, \beta, \mu, \sigma_x^2, \sigma^2\}$.

Temos 30 possíveis modelos diferentes $\left(\binom{5}{1} + \binom{5}{2} + \binom{5}{3} + \binom{5}{4}\right)$ que permitem mudanças em até quatro parâmetros simultaneamente. Entretanto percebemos que se as distribuições condicionais completas do modelo que permite mudança em todos os cinco parâmetros forem conhecidas, isto é, se conhecermos primeiramente as seguintes distribuições de (a₁) a (g) abaixo:

$$(a_1) [x_i | x_{(-i)}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, D_{\text{obs}}], \text{ para } i = 1, \dots, k,$$

$$(a_2) [x_i | x_{(-i)}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, D_{\text{obs}}], \text{ para } i = k+1, \dots, n,$$

$$(b_1) \text{ e } (b_2) [\alpha_i | \mathbf{x}, \alpha_{(-i)}, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, D_{\text{obs}}], i = 1, 2,$$

$$(c_1) \text{ e } (c_2) [\beta_i | \mathbf{x}, \alpha_1, \alpha_2, \beta_{(-i)}, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, D_{\text{obs}}], i = 1, 2.$$

(d₁) e (d₂) $[\mu_i | \underline{x}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_{(-i)}, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, D_{\text{obs}}], i = 1, 2,$

(e₁) e (e₂) $[\sigma_{x(i)}^2 | \underline{x}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(-i)}^2, \sigma_1^2, \sigma_2^2, k, D_{\text{obs}}], i = 1, 2,$

(f₁) e (f₂) $[\sigma_i^2 | \underline{x}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_{(-i)}^2, k, D_{\text{obs}}], i = 1, 2,$

(g) $P[K = k | \underline{x}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, D_{\text{obs}}], k \in \mathcal{K}$

e além disso, se conhecermos também as distribuições condicionais completas de (h) a (l) abaixo,

(h) $[\alpha | \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, \underline{x}, k, D_{\text{obs}}]$ do modelo que assume que o intercepto α não muda (mas os demais parâmetros podem mudar),

(i) $[\beta | \alpha_1, \alpha_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, \underline{x}, k, D_{\text{obs}}]$ do modelo que assume que β não muda (mas os demais parâmetros podem mudar),

(j) $[\mu | \alpha_1, \alpha_2, \beta_1, \beta_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, \underline{x}, k, D_{\text{obs}}]$ do modelo que assume que μ não muda (mas os demais parâmetros podem mudar),

(k) $[\sigma_x^2 | \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \underline{x}, k, D_{\text{obs}}]$ do modelo que assume que σ_x^2 não muda (mas os demais podem mudar),

(l) $[\sigma^2 | \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \underline{x}, k, D_{\text{obs}}]$ do modelo que assume que σ^2 não muda (mas os demais podem mudar),

então seremos capazes de obter quaisquer distribuições condicionais completas que quisermos, bastando fazer as restrições convenientes, e portanto do ponto de vista computacional teremos apenas o trabalho de escrever basicamente um único programa que servirá para todos os 31 possíveis modelos.

Por exemplo, para implementar o algoritmo de Gibbs no modelo que permite mudança em pelo menos um dos parâmetros $\{\alpha, \beta, \sigma^2\}$, as distribuições condicionais

completas necessárias são obtidas a partir de (a_1) , (a_2) , (b_1) , (b_2) , (f_1) , (f_2) , (g) , (j) e (k) impondo-se as restrições $\mu_i = \mu$ e/ou $\sigma_{x(i)}^2 = \sigma_x^2$ nas distribuições que envolverem os parâmetros μ_i e/ou $\sigma_{x(i)}^2$, para $i = 1, 2$, como também no valor inicial do algoritmo.

As distribuições condicionais de (a_1) a (g) são dadas, respectivamente, por (4.2.3), (4.2.4), ..., (4.2.14) e as distribuições de (h) até (l) são dadas respectivamente por (4.3.1) a (4.3.5) a seguir:

$$\alpha|\beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, \underline{x}, D_{\text{obs}} \sim N\left(\frac{\sigma_2^2 \sigma_\alpha^2 \sum_{i=1}^k (Y_i - \beta_1 x_i) + \sigma_1^2 \sigma_\alpha^2 \sum_{i=k+1}^n (Y_i - \beta_2 x_i) + \lambda \sigma_1^2 \sigma_2^2 a}{k \sigma_2^2 \sigma_\alpha^2 + (n-k) \sigma_1^2 \sigma_\alpha^2 + \lambda \sigma_1^2 \sigma_2^2}; \frac{\lambda \sigma_1^2 \sigma_2^2 \sigma_\alpha^2}{k \sigma_2^2 \sigma_\alpha^2 + (n-k) \sigma_1^2 \sigma_\alpha^2 + \lambda \sigma_1^2 \sigma_2^2}\right), \quad (4.3.1)$$

$$\beta|\alpha_1, \alpha_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, \underline{x}, D_{\text{obs}} \sim N\left(\frac{\sigma_2^2 \sigma_\beta^2 \sum_{i=1}^k (Y_i - \alpha_1) x_i + \sigma_1^2 \sigma_\beta^2 \sum_{i=k+1}^n (Y_i - \alpha_2) x_i + \lambda \sigma_1^2 \sigma_2^2 b}{\sigma_2^2 \sigma_\beta^2 \sum_{i=1}^k x_i^2 + \sigma_1^2 \sigma_\beta^2 \sum_{i=k+1}^n x_i^2 + \lambda \sigma_1^2 \sigma_2^2}; \frac{\sigma_1^2 \sigma_2^2 \sigma_\beta^2}{\sigma_2^2 \sigma_\beta^2 \sum_{i=1}^k x_i^2 + \sigma_1^2 \sigma_\beta^2 \sum_{i=k+1}^n x_i^2 + \lambda \sigma_1^2 \sigma_2^2}\right), \quad (4.3.2)$$

$$\mu|\alpha_1, \alpha_2, \beta_1, \beta_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, \underline{x}, D_{\text{obs}} \sim N\left(\frac{\sigma_{x(2)}^2 \sigma_\mu^2 \sum_{i=1}^k x_i + \sigma_{x(1)}^2 \sigma_\mu^2 \sum_{i=k+1}^n x_i + \sigma_{x(1)}^2 \sigma_{x(2)}^2 m}{k \sigma_{x(2)}^2 \sigma_\mu^2 + (n-k) \sigma_{x(1)}^2 \sigma_\mu^2 + \sigma_{x(1)}^2 \sigma_{x(2)}^2}; \frac{\sigma_{x(1)}^2 \sigma_{x(2)}^2 \sigma_\mu^2}{k \sigma_{x(2)}^2 \sigma_\mu^2 + (n-k) \sigma_{x(1)}^2 \sigma_\mu^2 + \sigma_{x(1)}^2 \sigma_{x(2)}^2}\right), \quad (4.3.3)$$

$$\sigma_x^2|\alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, k, \underline{x}, D_{\text{obs}} \sim IG\left(\frac{n}{2} + c; \frac{1}{2} \left(\sum_{i=1}^k (x_i - \mu_1)^2 + \sum_{i=k+1}^n (x_i - \mu_2)^2 \right) + d\right), \quad (4.3.4)$$

$$\sigma_x^2|\alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, k, \underline{x}, D_{\text{obs}} \sim \quad (4.3.5)$$

$$IG\left(n+f; \frac{1}{2\lambda}\left(\sum_{i=1}^k (Y_i - \alpha_1 - \beta_1 x_i)^2 + \sum_{i=k+1}^n (Y_i - \alpha_2 - \beta_2 x_i)^2 + \lambda \sum_{i=1}^n (X_i - x_i)^2\right) + g\right).$$

Observações:

1. As distribuições em (4.3.1) a (4.3.5) coincidem com as distribuições em (5.1.22) a (5.1.26) quando substituimos w_i , por 1, para $i = 1, \dots, n$.
2. Na distribuição de K dada em (g), o conjunto \mathcal{K} pode diferir dependendo do modelo, entretanto \mathcal{K} pode ser sempre igual a $\{1, \dots, n - 1\}$, como já foi discutido anteriormente.
3. A estimação dos parâmetros do modelo que admite mudança apenas em um subconjunto de parâmetros pode ser feita de forma semelhante à da seção 4.2.3.
4. A análise do modelo com priori análoga à da seção 2.7 (isto é priori de componentes condicionalmente independentes), pode ser feita também de forma análoga.

4.4 Detecção de mudança

Nas seções anteriores foram considerados modelos com um único ponto de mudança e o problema era estimar a localização desse ponto bem como estimar os demais parâmetros. Entretanto, em muitas situações nas aplicações não sabemos se há ou não mudança nos parâmetros. Nessa seção apresentamos três procedimentos distintos para analisar o problema, que serão descritos nas subseções 4.4.1, 4.4.2 e 4.4.3, respectivamente.

4.4.1 Procedimento 1 – Um teste baseado na distribuição à posteriori de k

Uma maneira de analisar o problema é reformular o modelo permitindo que k possa assumir o valor n , ou seja, permitir que o modelo inclua a possibilidade de que não haja mudança nos parâmetros (via $k = n$) e então testar a hipótese nula de que não há mudança ($H_0 : k = n$) versus a alternativa (H_1) de que há exatamente um ponto de mudança (isto é, $H_1 : k \neq n$).

Nessa subseção, apresentamos um teste bayesiano informal baseado na distribuição marginal à posteriori de k , que consiste simplesmente em comparar a probabilidade à posteriori $P[k = n|D_{\text{obs}}]$ de que não há mudança com a probabilidade $1 - P(k = n|D_{\text{obs}})$ de que há uma mudança. (Broemeling e Choy, 1981, desenvolveram esse teste para o modelo de regressão clássico que permite mudança no coeficiente angular e no intercepto.

É interessante notar que nesse caso o teste de hipóteses e a estimação dos parâmetros, são feitos simultaneamente.

A seguir, apresentamos o modelo estendido que nos permite realizar o teste bayesiano mencionado, e depois apresentamos os passos para implementar o algoritmo de Gibbs, com o objetivo de amostrar da posteriori desse modelo estendido.

O modelo estendido que permite mudança em todos os parâmetros:

$$\begin{cases} Y_i = \begin{cases} \alpha_1 + \beta_1 x_i + e_i, & i = 1, 2, \dots, k, \\ \alpha_2 + \beta_2 x_i + e_i, & i = k + 1, \dots, n, \\ \alpha_1 + \beta_1 x_i + e_i, & i = 1, \dots, n, \end{cases} & \begin{array}{l} \text{se } 1 \leq k \leq n - 1 \\ \text{se } k = n \end{array} \\ X_i = x_i + u_i, & i = 1, \dots, n, \end{cases}$$

onde $(e_i, u_i, x_i)^T$ são independentes e

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \sim \begin{cases} N \left(\begin{pmatrix} 0 \\ 0 \\ \mu_1 \end{pmatrix}; \begin{pmatrix} \lambda\sigma_1^2 & 0 & 0 \\ 0 & \sigma_1^2 & 0 \\ 0 & 0 & \sigma_{x(1)}^2 \end{pmatrix} \right) & i = 1, \dots, k, \\ N \left(\begin{pmatrix} 0 \\ 0 \\ \mu_2 \end{pmatrix}; \begin{pmatrix} \lambda\sigma_2^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_{x(2)}^2 \end{pmatrix} \right) & i = k + 1, \dots, n, \\ N \left(\begin{pmatrix} 0 \\ 0 \\ \mu_1 \end{pmatrix}; \begin{pmatrix} \lambda\sigma_1^2 & 0 & 0 \\ 0 & \sigma_1^2 & 0 \\ 0 & 0 & \sigma_{x(1)}^2 \end{pmatrix} \right) & i = 1, \dots, n, \quad \text{se } k = n \end{cases} \quad \text{se } 1 \leq k \leq n \quad (4.4.1)$$

Notemos que para $1 \leq k \leq n - 1$ então $\theta_1 \neq \theta_2$, onde $\theta_i = (\alpha_i, \beta_i, \mu_i, \sigma_{x(i)}^2, \sigma_i^2)^T$ e para $k = n$ então $\theta_1 = \theta_2$.

Portanto, esse modelo permite que haja mudança em todos os parâmetros como também que não haja mudança em nenhum.

Nesse modelo o vetor de parâmetros é $\check{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ k \end{pmatrix}$ e, se $1 \leq k \leq n - 1$, a função de verossimilhança $L(\check{\theta}|D)$ é proporcional a

$$\begin{aligned} & (\sigma_{x(1)}^2)^{-k/2} (\sigma_1^2)^{-k} \exp \left\{ -\frac{1}{2} \left[\frac{\sum_{i=1}^k (Y_i - \alpha_1 - \beta_1 x_i)^2}{\lambda\sigma_1^2} + \frac{\sum_{i=1}^k (X_i - x_i)^2}{\sigma_1^2} \right. \right. \\ & \left. \left. + \frac{\sum_{i=1}^k (x_i - \mu_1)^2}{\sigma_{x(1)}^2} \right] \right\} (\sigma_{x(2)}^2)^{-(n-k)/2} (\sigma_2^2)^{-(n-k)} \exp \left\{ -\frac{1}{2} \left[\frac{\sum_{i=k+1}^n (Y_i - \alpha_2 - \beta_2 x_i)^2}{\lambda\sigma_2^2} \right. \right. \\ & \left. \left. + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma_2^2} + \frac{\sum_{i=k+1}^n (x_i - \mu_2)^2}{\sigma_{x(2)}^2} \right] \right\}. \end{aligned} \quad (4.4.2)$$

Se $k = n$, a função de verossimilhança é proporcional a

$$\begin{aligned} & (\sigma_{x(1)}^2)^{-n/2} (\sigma_1^2)^{-n} \exp \left\{ -\frac{1}{2} \left[\frac{\sum_{i=1}^n (Y_i - \alpha_1 - \beta_1 x_i)^2}{\lambda\sigma_1^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma_1^2} \right. \right. \\ & \left. \left. + \frac{\sum_{i=1}^n (x_i - \mu_1)^2}{\sigma_{x(1)}^2} \right] \right\}. \end{aligned} \quad (4.4.3)$$

Uma priori conveniente nesse caso é dada por

$$\pi(\check{\theta}) = \begin{cases} p(k)\pi(\theta_1), & \text{se } k = n, \\ p(k)\pi(\theta_1)\pi(\theta_2), & \text{se } k \in \mathcal{K} = \{1, \dots, n - 1\}, \\ 0, & \text{caso contrário,} \end{cases} \quad (4.4.4)$$

onde

$$p(k) = \begin{cases} c, & \text{se } k = n, \\ (1-c)\frac{1}{\#\mathcal{K}}, & \text{se } k \in \mathcal{K}, \\ 0, & \text{caso contrário,} \end{cases} \quad (4.4.5)$$

e

$$\pi(\boldsymbol{\theta}_i) = \pi(\alpha_i)\pi(\beta_i)\pi(\mu_i)\pi(\sigma_{x(i)}^2)\pi(\sigma_i^2), \quad (4.4.6)$$

onde $\alpha_i \sim N(a_i, \sigma_{\alpha_i}^2)$, $\beta_i \sim N(b_i, \sigma_{\beta_i}^2)$, $\mu_i \sim N(m_i, \sigma_{\mu_i}^2)$, $\sigma_{x(i)}^2 \sim IG(c_i, d_i)$, $\sigma_i^2 \sim IG(f_i, g_i)$, $i = 1, 2$.

Na próxima seção apresentamos as distribuições condicionais completas à posteriori que serão utilizadas para implementar o algoritmo de Gibbs.

Implementação do algoritmo de Gibbs:

A seguir apresentamos as distribuições condicionais completas baseadas na posteriori do modelo estendido usadas para implementar o algoritmo de Gibbs.

Passo 0: Fornecer um vetor inicial $\boldsymbol{\theta}^{(0)} = (k^{(0)}, \boldsymbol{\theta}_1^{(0)}, \boldsymbol{\theta}_2^{(0)})^T$, onde $\boldsymbol{\theta}_i^{(0)} = (\alpha_i^{(0)}, \beta_i^{(0)}, \mu_i^{(0)}, \sigma_{x(i)}^{(0)}, \sigma_i^{(0)})^T$, $i = 1, 2$.

Passo 1:

- Se $k = n$:

então gere $\boldsymbol{x} = (x_1, \dots, x_n)^T$, $\alpha_1, \beta_1, \mu_1, \sigma_{x(1)}^2$ e σ_1^2 , onde:

$$x_i | x_{(-i)} \boldsymbol{\theta}, D_{\text{obs}} \sim N\left(\sigma_{x(1)}^2 \left[\frac{\lambda X_i + \beta_1(Y_i - \alpha_1) + \lambda \sigma_1^2 \mu_1}{\sigma_{x(1)}^2(\lambda + \beta_1^2) + \lambda \sigma_1^2} \right]; \frac{\lambda \sigma_1^2 \sigma_{x(1)}^2}{\sigma_{x(1)}^2(\lambda + \beta_1^2) + \lambda \sigma_1^2} \right), \quad (4.4.7)$$

para $i = 1, \dots, n$,

$$\alpha_1 | \boldsymbol{x}, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, D_{\text{obs}} \sim N\left(\frac{\sigma_{\alpha_1}^2 \sum_{i=1}^k (Y_i - \beta_1 x_i) + \lambda \sigma_1^2 a_1}{n \sigma_{\alpha_1}^2 + \lambda \sigma_1^2}; \frac{\lambda \sigma_1^2 \sigma_{\alpha_1}^2}{n \sigma_{\alpha_1}^2 + \lambda \sigma_1^2} \right), \quad (4.4.8)$$

$$\beta_1 | \underline{x}, \alpha_1, \alpha_2, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, D_{\text{obs}} \sim N \left(\frac{\sigma_{\beta_1}^2 \sum_{i=1}^n (Y_i - \alpha_1) x_i + \lambda b_1 \sigma_1^2}{\sigma_{\beta_1}^2 \sum_{i=1}^n x_i^2 + \lambda \sigma_1^2}, \frac{\lambda \sigma_1^2 \sigma_{\beta_1}^2}{\sigma_{\beta_1}^2 \sum_{i=1}^n x_i^2 + \lambda \sigma_1^2} \right), \quad (4.4.9)$$

$$\mu_1 | \underline{x}, \alpha_1, \alpha_2, \beta_1, \beta_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, D_{\text{obs}} \sim N \left(\frac{\sigma_{\mu_1}^2 \sum_{i=1}^n x_i + \sigma_{x(1)}^2 m_1}{n \sigma_{\mu_1}^2 + \sigma_{x(1)}^2}, \frac{\sigma_{\mu_1}^2 \sigma_{x(2)}^2}{n \sigma_{\mu_1}^2 + \sigma_{x(1)}^2} \right), \quad (4.4.10)$$

$$\sigma_{x(1)}^2 | \underline{x}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, D_{\text{obs}} \sim IG \left(\frac{n}{2} + c_1; \frac{1}{2} \sum_{i=1}^n (x_i - \mu_1)^2 + d_1 \right), \quad (4.4.11)$$

$$\sigma_1^2 | \underline{x}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_2^2, D_{\text{obs}} \sim IG \left(n + f_1; \frac{1}{2} \left[\frac{\sum_{i=1}^n (Y_i - \alpha_1 - \beta_1 x_i)^2}{\lambda} + \sum_{i=1}^n (X_i - x_i)^2 \right] + g_1 \right), \quad (4.4.12)$$

Faça:

$$\alpha_2 = \alpha_1, \beta_2 = \beta_1, \mu_2 = \mu_1, \sigma_{x(2)}^2 = \sigma_{x(1)}^2, \sigma_2^2 = \sigma_1^2.$$

- Caso contrário, se $1 \leq k \leq n - 1$:

então gere $\underline{x}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2$ e σ_2^2 usando as fórmulas (4.2.3)

a (4.2.13), respectivamente.

Passo 2: Gerar um valor de K da distribuição discreta

$$P(K = k | \underline{x}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, D_{\text{obs}}) = \begin{cases} \frac{L_1(\theta_1 | D) p(n) \pi(\theta_1)}{L_1(\theta_1 | D) p(n) \pi(\theta_1) + \sum_{k=1}^{n-1} L(k, \theta_1, \theta_2 | D) p(k) \pi(\theta_1) \pi(\theta_2)}, & \text{se } k = n, \\ \frac{L(k, \theta_1, \theta_2 | D) p(k) \pi(\theta_1) \pi(\theta_2)}{L_1(\theta_1 | D) p(n) \pi(\theta_1) + \sum_{k=1}^{n-1} L(k, \theta_1, \theta_2 | D) p(k) \pi(\theta_1) \pi(\theta_2)}, & 1 \leq k \leq n - 1, \end{cases}$$

onde:

$p(k)$ = probabilidade à priori de que $K = k$ ($k = 1, 2, \dots, n$),

dada por exemplo por (4.4.4) e (4.4.5),

$L(k, \theta_1, \theta_2|D)$ é proporcional a (4.4.2),

$L_1(\theta_1|D_{\text{obs}})$ é proporcional a (4.4.3),

$$\pi(\theta_2) = M(\sigma_{x(2)}^2)^{-(c_2+1)}(\sigma_2^2)^{-(f_2+1)} \exp\left\{-\frac{1}{2}\left[\frac{(\alpha_2 - a_2)^2}{\sigma_{\alpha_2}^2} + \frac{(\beta_2 - b_2)^2}{\sigma_{\beta_2}^2} + \frac{(\mu_2 - m_2)^2}{\sigma_{\mu_2}^2} + \frac{2d_2}{\sigma_{x(2)}^2} + \frac{2g_2}{\sigma_2^2}\right]\right\},$$

A constante de proporcionalidade M é igual a

$$\frac{d_2^{c_2} g_2^{f_2}}{\Gamma(c_2)\Gamma(f_2)2\pi\sqrt{2\pi\sigma_{\alpha_2}^2\sigma_{\beta_2}^2\sigma_{\mu_2}^2}}.$$

Passo 3: volte ao Passo 1.

Usando as distribuições condicionais completas acima podemos implementar o algoritmo de Gibbs com o objetivo de amostrar da posteriori do modelo dado por (4.4.1), (4.4.4), (4.4.5) e (4.4.6).

Se observações geradas pelo algoritmo constituem de fato uma amostra da posteriori, podemos estimar a distribuição à posteriori marginal de k da mesma forma que na subseção 4.2.3 (só que agora k pode assumir o valor n) e se $P(k = n|D_{\text{obs}}) > 1 - P(k = n|D_{\text{obs}})$, então aceitamos H_0 ; caso contrário, rejeitamos H_0 .

Observação 1: No caso do modelo que permite mudança em um subconjunto de até quatro parâmetros podemos testar a hipótese H_0 versus H_1 usando a distribuição marginal à posteriori de k de forma semelhante à descrita acima, fazendo uma analogia com o procedimento da seção 4.3.

Observação 2: É importante que se faça uma análise de sensibilidade da distribuição à posteriori de k a diferentes prioris. Uma idéia bem simples é considerar uma grade de vários possíveis valores para o hiperparâmetro c (ond $c = P(k = n)$ à priori) e ver como isso afeta a posteriori marginal de k .

Segundo Kim (1991) que estudou o modelo de regressão clássico (cuja covariável não tem erros de medida), o teste de Bayes tradicional que usa a posteriori de k para detectar mudança é bastante sensível à escolha da priori na maioria dos casos.

4.4.2 Procedimento 2 – Uso da distribuição à posteriori da diferença (ou quociente) entre os parâmetros.

Uma outra forma de detetar se existe mudança nos parâmetros é examinar as distribuições marginais à posteriori da diferença (ou quociente) entre os parâmetros “antes e depois da mudança” no modelo que permite mudança em todos os parâmetros. A hipótese H_0 pode ser dividida nas sub-hipóteses: $H_{01} : \alpha_1 - \alpha_2 = 0$, $H_{02} : \beta_1 - \beta_2 = 0$, $H_{03} : \mu_1 - \mu_2 = 0$, $H_{04} : \frac{\sigma_{x(2)}^2}{\sigma_{x(1)}^2} = 1$, $H_{05} : \frac{\sigma_2^2}{\sigma_1^2} = 1$, sendo que H_0 é rejeitada quando pelo menos uma das sub-hipóteses é rejeitada.

Em princípio a separação da hipótese nula H_0 nas cinco sub-hipóteses H_{01}, \dots, H_{05} é útil para determinarmos quais parâmetros mudam (isto é, quais parâmetros são responsáveis pela mudança).

Conceitualmente a decisão, por exemplo, de aceitar a sub-hipótese H_{02} seria razoável se $P(\beta_2 - \beta_1 > 0 | D_{\text{obs}}) = P(\beta_2 - \beta_1 < 0 | D_{\text{obs}}) = 0.5$. Assim se observarmos por exemplo que $\hat{P}(\beta_2 - \beta_1 > 0 | D_{\text{obs}}) \approx 0.80$, ficaríamos razoavelmente confiantes de “rejeitar” essa hipótese.

Obs: A análise fica bastante simplificada quando a distribuição da diferença de $\theta_2 - \theta_1$ é analisada apenas marginalmente (em vez de conjuntamente) entretanto o preço é que esta aproximação pode ser bem ruim.

4.4.3 Procedimento 3 – Uso de critérios de seleção ou escolha de modelos

Outra forma de detectar a existência de mudança é tratar o problema de ponto de mudança como um problema de seleção de modelos, isto é utilizar métodos

bayesianos de comparação (ou escolha) de modelos, para comparar o modelo sem nenhuma mudança com o modelo que permite mudança.

Existem vários critérios de escolha de modelos descritos na literatura para essa finalidade. Alguns dos mais populares são, por exemplo, o método do “Fator de Bayes”, o “Pseudo Fator de Bayes”, a “Medida L” e o BIC (“Bayesian Information Criterion”).

Uma maneira de tratar o problema é comparar o modelo com mudança após o ponto k desconhecido (aleatório) com o modelo sem nenhuma mudança (usando obviamente o mesmo conjunto de dados). Nesse caso, checka-se a existência de mudança e estima-se o ponto de mudança simultaneamente.

Uma alternativa para tratar o problema é para cada k fixado, comparar o modelo com mudança em k com o modelo sem mudança.

A seguir, apresentamos os critérios de escolha mencionados acima para o problema de mudança no modelo estrutural normal com erros nas variáveis.

(1°) O critério do Fator de Bayes

O Fator de Bayes é um critério de comparação de dois modelos que se aplica somente no caso de prioris próprias. Na seção [4.4.3] do Apêndice, apresentamos a definição dessa medida.

O Fator de Bayes para comparar os modelos $M_{1,k}$ e M_2 (onde $M_{1,k}$ é o modelo com mudança nos parâmetros após a k -ésima observação e M_2 é o modelo sem mudanças) é definido por

$$\text{BF}_{(1,k),2} = \frac{\pi(\underline{U}_1, \dots, \underline{U}_n | M_{1,k})}{\pi(\underline{U}_1, \dots, \underline{U}_n | M_2)}, \quad (4.4.13)$$

onde $\underline{U}_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix}$ e onde $\pi(\underline{U}_1, \dots, \underline{U}_n | M_{1,k})$ e $\pi(\underline{U}_1, \dots, \underline{U}_n | M_2)$ são as densidades marginais dos dados sob os modelos $M_{1,k}$ e M_2 , respectivamente.

Na literatura existem várias técnicas que podem ser usadas para estimar o fator individual de Bayes $B_{(1,k),2}$. Um método direto proposto por Newton e Raftery (1994), estima o numerador $\pi(\underline{U}_1, \dots, \underline{U}_n | M_{1,k})$ por

$$\hat{\pi}(\underline{U}_1, \dots, \underline{U}_n | M_{1,k}) = \left(\frac{1}{G} \sum_{g=1}^G \frac{1}{\pi(\underline{U}_1, \dots, \underline{U}_n | \theta_1^{(g)}, \theta_2^{(g)}, k^{(g)})} \right)^{-1},$$

onde $\left(\left(\begin{matrix} \theta_1^{(1)} \\ \theta_2^{(1)} \\ k^{(1)} \end{matrix} \right), \dots, \left(\begin{matrix} \theta_1^{(G)} \\ \theta_2^{(G)} \\ k^{(G)} \end{matrix} \right) \right)$ é uma amostra da posteriori sob o modelo $M_{1,k}$ e $\pi(\underline{U}_1, \dots, \underline{U}_n | \theta_1, \theta_2, k)$ é a densidade das observações sob o modelo $M_{1,k}$, descrito em (4.4.1).

O denominador $\pi(\underline{U}_1, \dots, \underline{U}_n | M_2)$ é estimado por

$$\hat{\pi}(\underline{U}_1, \dots, \underline{U}_n | M_2) = \left(\frac{1}{G} \sum_{g=1}^G \frac{1}{\pi(\underline{U}_1, \dots, \underline{U}_n | \theta^{(g)})} \right)^{-1},$$

onde $(\theta^{(1)}, \dots, \theta^{(G)})$ é uma amostra da posteriori sob o modelo M_2 definido em (2.3.1) e $\pi(\underline{U}_1, \dots, \underline{U}_n | \theta)$ é a densidade das observações sob esse modelo.

Outra maneira de tratar o problema é usar o método de Dey (1997).

Dey (1997) mostrou que o Fator de Bayes que compara um modelo sem mudança (M_2) com um modelo com mudança (M_1), onde o ponto de mudança k não é especificado, pode ser escrito como uma combinação linear convexa de fatores de Bayes individuais, isto é,

$$\text{BF}_{1,2} = \sum_{k=1}^{n-1} p_k \text{BF}_{(1,k),2} \quad \text{onde} \quad p_k = \begin{cases} (1-c)p[K=k], & 1 \leq k \leq n-1, \\ c, & k=n. \end{cases}$$

é a distribuição à priori de k .

Esse resultado pode ser usado, por exemplo, para comparar o modelo estrutural Normal que permite mudança em todos os parâmetros (que está definido em (4.1.1) e tem priori em (4.1.2)) com o modelo estrutural Normal sem mudança que está definido em (2.3.1) e tem priori em (2.3.2). Nesse exemplo, o Fator de Bayes é dado

por

$$\text{BF}_{1,2} = \sum_{k \in \mathcal{K} = \{1, \dots, n-1\}} p_k \text{BF}_{(1,k),2}, \quad (4.4.14)$$

onde $p_k = \frac{1}{n-1}(1-c)$ (onde $c = P(K = n) = 0$) e $\text{BF}_{(1,k),2}$ é o fator individual de Bayes que compara o modelo com mudança após o ponto k fixado com o modelo sem mudança,

$$\text{BF}_{(1,k),2} = \frac{\pi(\underline{U}_1, \dots, \underline{U}_n | M_{1,k})}{\pi(\underline{U}_1, \dots, \underline{U}_n | M_2)} = \frac{\int_{\Theta_1} \int_{\Theta_2} \pi(\underline{U}_1, \dots, \underline{U}_n | \underline{\theta}_1, \underline{\theta}_2, k) \pi(\underline{\theta}_1, \underline{\theta}_2, k) d\underline{\theta}_1 d\underline{\theta}_2}{\int_{\Theta} \pi(\underline{U}_1, \dots, \underline{U}_n | \underline{\theta}) \pi(\underline{\theta}) d\underline{\theta}}$$

onde $\underline{\theta} = (\alpha, \beta, \mu, \sigma_x^2, \sigma^2)^T$ e $\underline{\theta}_i = (\alpha_i, \beta_i, \mu_i, \sigma_{x(i)}^2, \sigma_i^2)^T$, $i = 1, 2$,

$\pi(\underline{U}_1, \dots, \underline{U}_n | \underline{\theta}_1, \underline{\theta}_2, k)$ é a densidade das observações sob o modelo $M_{1,k}$ definido em (4.1.1) (com k fixado),

$\pi(\underline{\theta}_1, \underline{\theta}_2, k) = \pi(\underline{\theta}_1) \pi(\underline{\theta}_2)$, é a distribuição à priori dada em (4.2.1) quando $p(k) = 1$,

$\pi(\underline{U}_1, \dots, \underline{U}_n | \underline{\theta})$ é a densidade das observações sob o modelo M_2 definido em (2.3.1) e $\pi(\underline{\theta})$ é a priori dada em (2.3.2).

O resultado em (4.4.14) será verificado abaixo usando a justificativa dada em Chung e Dey (1996).

Justificativa: O Fator de Bayes para comparar o modelo M_1 com o modelo M_2 é

$$\text{BF}_{1,2} = \frac{\pi(\underline{U}_1, \dots, \underline{U}_n | M_1)}{\pi(\underline{U}_1, \dots, \underline{U}_n | M_2)}, \text{ onde}$$

$$\pi(\underline{U}_1, \dots, \underline{U}_n | M_1) = \pi(\underline{U}_1, \dots, \underline{U}_n | \text{mudança}) = \frac{\pi(\underline{U}_1, \dots, \underline{U}_n, 1 \leq k \leq n-1)}{P(1 \leq k \leq n-1)},$$

onde $P(1 \leq k \leq n-1) = 1-c$ (onde $c = p[K = n]$) e portanto

$$\begin{aligned} \pi(\underline{U}_1, \dots, \underline{U}_n | M_1) &= \frac{1}{1-c} \int_{\Theta_1} \int_{\Theta_2} \pi(\underline{U}_1, \dots, \underline{U}_n, \underline{\theta}_1, \underline{\theta}_2, 1 \leq k \leq n-1) d\underline{\theta}_1 d\underline{\theta}_2 \\ &= \frac{1}{1-c} \int_{\Theta_1} \int_{\Theta_2} \pi(\underline{U}_1, \dots, \underline{U}_n | \underline{\theta}_1, \underline{\theta}_2, 1 \leq k \leq n-1) \pi(\underline{\theta}_1, \underline{\theta}_2, 1 \leq k \leq n-1) d\underline{\theta}_1 d\underline{\theta}_2 \\ &= \frac{1}{1-c} \sum_{k=1}^{n-1} \int_{\Theta_1} \int_{\Theta_2} \pi(\underline{U}_1, \dots, \underline{U}_n | \underline{\theta}_1, \underline{\theta}_2, k) \pi(\underline{\theta}_1, \underline{\theta}_2, k) d\underline{\theta}_1 d\underline{\theta}_2 \\ &= \frac{1}{1-c} \sum_{k=1}^{n-1} \frac{(1-c)}{n-1} \int_{\Theta_1} \int_{\Theta_2} \pi(\underline{U}_1, \dots, \underline{U}_n | \underline{\theta}_1, \underline{\theta}_2, k) \pi(\underline{\theta}_1, \underline{\theta}_2) d\underline{\theta}_1 d\underline{\theta}_2 \end{aligned}$$

$$= \frac{1}{n-1} \sum_{k=1}^{n-1} \pi(\underline{U}_1, \dots, \underline{U}_n | M_{1,k}).$$

Portanto,

$$\text{BF}_{1,2} = \frac{1}{n-1} \sum_{k=1}^{n-1} \frac{\pi(\underline{U}_1, \dots, \underline{U}_n | M_{1,k})}{\pi(\underline{U}_1, \dots, \underline{U}_n | M_2)} = \sum_{k=1}^{n-1} \frac{1}{n-1} \text{BF}_{(1,k),2}.$$

Observação: Utilizando o resultado de Dey (1997) podemos checar se existe mudança e obter uma estimativa do ponto de mudança simultaneamente. Se houver mudança, o ponto de mudança será o valor de k cujo Fator de Bayes individual é máximo.

O Fator de Bayes individual nesse caso é dado por

$$\widehat{\text{BF}}_{(1,k),2} = \frac{\hat{\pi}(\underline{U}_1, \dots, \underline{U}_n | M_{1,k})}{\hat{\pi}(\underline{U}_1, \dots, \underline{U}_n | M_2)},$$

que pode ser estimado pelo método de Newton e Raftery descrito anteriormente fazendo $k^{(g)} = k$.

Outro método relativamente simples de ser implementado é o método de Chib (1995), descrito na seção [4.4.3a] do Apêndice. A seguir, vamos apresentar os cálculos para a estimação do fator individual de Bayes pelo método de Chib (1995).

Estimação de $\pi(\underline{U}_1, \dots, \underline{U}_n | M_2) = \pi(D_{\text{obs}} | M_2)$

Usando o método de Chib descrito na seção [4.4.3 a] do apêndice, temos que

$$\log \hat{\pi}(\underline{U}_1, \dots, \underline{U}_n | M_2) = \log \pi(\underline{U}_1, \dots, \underline{U}_n | \vartheta') + \log \pi(\vartheta') - \log \hat{\pi}(\vartheta' | D_{\text{obs}}),$$

onde ϑ' pode ser uma estimativa da média da posteriori $p(\vartheta | D_{\text{obs}})$. Além disso,

$$\begin{aligned} \hat{\pi}(\vartheta' | D_{\text{obs}}) &= \pi(\sigma^2 | \underline{x}', \alpha', \beta', \mu', \sigma_x^2, D_{\text{obs}}) \hat{\pi}(\sigma_x^2 | \underline{x}', \alpha', \beta', \mu', D_{\text{obs}}). \\ &\quad \hat{\pi}(\mu' | \underline{x}', \alpha', \beta', D_{\text{obs}}) \hat{\pi}(\beta' | \underline{x}', \alpha', D_{\text{obs}}) \pi(\alpha' | \underline{x}', D_{\text{obs}}) \pi(\underline{x}' | D_{\text{obs}}). \end{aligned}$$

A seguir damos os passos para a estimação dessas quantidades.

(a) $\hat{\pi}(\underline{x}'|D_{\text{obs}})$ é calculada usando o algoritmo de Gibbs seguinte:

Passo 1: fornecer um valor inicial $(\alpha^{(0)}, \beta^{(0)}, \mu^{(0)}, \sigma_x^{2(0)}, \sigma^2(0))$, faça $i = 0$.

Passo 2: gerar $x_i^{(i+1)}$ de $\pi(x_i|\alpha^{(i)}, \beta^{(i)}, \mu^{(i)}, \sigma_x^{2(i)}, \sigma^2(i), D_{\text{obs}})$, $i = 1, 2, \dots, n$, dada em (2.5.14),

gerar $\alpha^{(i+1)}$ de $\pi(\alpha|\underline{x}^{(i+1)}, \beta^{(i)}, \mu^{(i)}, \sigma_x^{2(i)}, \sigma^2(i), D_{\text{obs}})$, dada em (2.5.9),

gerar $\beta^{(i+1)}$ de $\pi(\beta|\underline{x}^{(i+1)}, \alpha^{(i+1)}, \mu^{(i)}, \sigma_x^{2(i)}, \sigma^2(i), D_{\text{obs}})$, dada em (2.5.10),

gerar $\mu^{(i+1)}$ de $\pi(\mu|\underline{x}^{(i+1)}, \alpha^{(i+1)}, \beta^{(i+1)}, \sigma_x^{2(i)}, \sigma^2(i), D_{\text{obs}})$, dada em (2.5.11),

gerar $\sigma_x^{2(i+1)}$ de $\pi(\sigma_x^2|\underline{x}^{(i+1)}, \alpha^{(i+1)}, \beta^{(i+1)}, \mu^{(i+1)}, \sigma^2(i), D_{\text{obs}})$, dada em (2.5.12),

gerar $\sigma^{2(i+1)}$ de $\pi(\sigma^2|\underline{x}^{(i+1)}, \alpha^{(i+1)}, \beta^{(i+1)}, \mu^{(i+1)}, \sigma_x^{2(i+1)}, D_{\text{obs}})$, dada em (2.5.13),

Passo 3: fazer $i = i + 1$ e voltar ao passo 2, até convergir. Calcule

$$\hat{\pi}(\underline{x}'|D_{\text{obs}}) = \frac{1}{G} \sum_{g=1}^G \prod_{i=1}^n \pi(x'_i|\alpha^{(g)}, \beta^{(g)}, \mu^{(g)}, \sigma_x^{2(g)}, \sigma^{2(g)}, D_{\text{obs}}),$$

onde $(\theta^{(1)}, \dots, \theta^{(G)})$ é uma amostra de G elementos gerados pelo algoritmo.

(b) $\hat{\pi}(\alpha'|x', D_{\text{obs}})$ é calculada usando o seguinte algoritmo de Gibbs:

Passo 1: fornecer um valor inicial $(\beta^{(0)}, \mu^{(0)}, \sigma_x^{2(0)}, \sigma^2(0))$, fazer $i = 0$,

Passo 2: gerar

$\alpha^{(i+1)}$ de $\pi(\alpha|x = x', \beta^{(i)}, \mu^{(i)}, \sigma_x^{2(i)}, \sigma^2(i), D_{\text{obs}})$,

$\beta^{(i+1)}$ de $\pi(\beta|x = x', \alpha^{(i+1)}, \mu^{(i)}, \sigma_x^{2(i)}, \sigma^2(i), D_{\text{obs}})$,

$\mu^{(i+1)}$ de $\pi(\mu|x = x', \alpha^{(i+1)}, \beta^{(i+1)}, \sigma_x^{2(i)}, \sigma^2(i), D_{\text{obs}})$,

$\sigma_x^{2(i+1)}$ de $\pi(\sigma_x^2|x = x', \alpha^{(i+1)}, \beta^{(i+1)}, \mu^{(i+1)}, \sigma^2(i), D_{\text{obs}})$,

$\sigma^{2(i+1)}$ de $\pi(\sigma^2|x = x', \alpha^{(i+1)}, \beta^{(i+1)}, \mu^{(i+1)}, \sigma_x^{2(i+1)}, D_{\text{obs}})$,

Passo 3: fazer $i = i + 1$ e voltar ao passo 2, até convergir. Calcule

$$\hat{\pi}(\alpha'|x', D_{\text{obs}}) = \frac{1}{G} \sum_{g=1}^G \pi(\alpha'|x = x', \beta^{(g)}, \mu^{(g)}, \sigma_x^{2(g)}, \sigma^{2(g)}, D_{\text{obs}}).$$

(c) $\hat{\pi}(\beta' | \underline{x}', \alpha', D_{\text{obs}})$ é calculada usando o seguinte algoritmo de Gibbs:

Passo 1: fornecer um valor inicial $(\mu^{(0)}, \sigma_x^{2(0)}, \sigma^{2(0)})$, fazer $i = 0$,

Passo 2: gerar

$$\beta^{(i+1)} \text{ de } \pi(\beta | \underline{x} = \underline{x}', \alpha = \alpha', \mu^{(i)}, \sigma_x^{2(i)}, \sigma^{2(i)}, D_{\text{obs}}),$$

$$\mu^{(i+1)} \text{ de } \pi(\mu | \underline{x} = \underline{x}', \alpha = \alpha', \beta^{(i+1)}, \sigma_x^{2(i)}, \sigma^{2(i)}, D_{\text{obs}}),$$

$$\sigma_x^{2(i+1)} \text{ de } \pi(\sigma_x^2 | \underline{x} = \underline{x}', \alpha = \alpha', \beta^{(i+1)}, \mu^{(i+1)}, \sigma^{2(i)}, D_{\text{obs}}),$$

$$\sigma^{2(i+1)} \text{ de } \pi(\sigma^2 | \underline{x} = \underline{x}', \alpha = \alpha', \beta^{(i+1)}, \mu^{(i+1)}, \sigma^{2(i+1)}, D_{\text{obs}}),$$

Passo 3: fazer $i = i + 1$ e voltar ao passo 2, até convergir. Calcule

$$\hat{\pi}(\beta' | \underline{x}', \alpha', D_{\text{obs}}) = \frac{1}{G} \sum_{g=1}^G \pi(\beta' | \underline{x} = \underline{x}', \alpha = \alpha', \mu^{(g)}, \sigma_x^{2(g)}, \sigma^{2(g)}, D_{\text{obs}}).$$

(d) $\hat{\pi}(\mu' | \underline{x}', \alpha', \beta', D_{\text{obs}})$ é calculado usando o seguinte algoritmo de Gibbs:

Passo 1: fornecer um valor inicial $(\sigma_x^{2(0)}, \sigma^{2(0)})$, fazer $i = 0$,

Passo 2: gerar

$$\mu^{(i+1)} \text{ de } \pi(\mu | \underline{x} = \underline{x}', \alpha = \alpha', \beta = \beta', \sigma_x^{2(i)}, \sigma^{2(i)}, D_{\text{obs}}),$$

$$\sigma_x^{2(i+1)} \text{ de } \pi(\sigma_x^2 | \underline{x} = \underline{x}', \alpha = \alpha', \beta = \beta', \mu^{(i+1)}, \sigma^{2(i+1)}, D_{\text{obs}}),$$

$$\sigma^{2(i+1)} \text{ de } \pi(\sigma^2 | \underline{x} = \underline{x}', \alpha = \alpha', \beta = \beta', \mu^{(i+1)}, \sigma_x^{2(i+1)}, D_{\text{obs}}),$$

Passo 3: fazer $i = i + 1$ e voltar ao passo 2, até convergir. Calcule

$$\hat{\pi}(\mu' | \underline{x}', \alpha', \beta', D_{\text{obs}}) = \frac{1}{G} \sum_{g=1}^G \pi(\mu' | \underline{x} = \underline{x}', \alpha = \alpha', \beta = \beta', \sigma_x^{2(g)}, \sigma^{2(g)}, D_{\text{obs}}).$$

(e) $\hat{\pi}(\sigma_x^{2'} | \underline{x}', \alpha', \beta', \mu', D_{\text{obs}})$ é calculado usando o seguinte algoritmo de Gibbs:

Passo 1: fornecer um valor inicial $\sigma^{2(0)}$, fazer $i = 0$,

Passo 2: gerar

$$\sigma_x^{2(i+1)} \text{ de } \pi(\sigma_x^2 | \underline{x} = \underline{x}', \alpha = \alpha', \beta = \beta', \mu = \mu', \sigma^{2(i)}, D_{\text{obs}}),$$

$$\sigma^{2(i+1)} \text{ de } \pi(\sigma^2 | \underline{x} = \underline{x}', \alpha = \alpha', \beta = \beta', \mu = \mu', \sigma_x^{2(i+1)}, D_{\text{obs}}).$$

Passo 3: fazer $i = i + 1$ e voltar ao passo 2, até convergir. Calcule

$$\hat{\pi}(\sigma_x^2 | \underline{x}', \alpha', \beta', \mu', D_{\text{obs}}) = \frac{1}{G} \sum_{g=1}^G \pi(\sigma_x^2 | \underline{x} = \underline{x}', \alpha = \alpha', \beta = \beta', \mu = \mu', \sigma^{2(g)}, D_{\text{obs}}).$$

(f) Calcular $\pi(\sigma_x^2 | \underline{x}', \alpha', \beta', \mu', \sigma_x^2, D_{\text{obs}})$ que é dada em (2.5.13) quando substituímos $\underline{\theta}$ por $\underline{\theta}'$.

Estimação de $\pi(\underline{U}_1, \dots, \underline{U}_n | M_{1,k})$, onde $M_{1,k}$ é o modelo definido em (4.1.1) que permite mudança em todos os parâmetros após o ponto k .

No modelo $M_{1,k}$ com k fixado, o vetor de parâmetros é $\begin{pmatrix} \underline{x}^{(1)} \\ \underline{\theta}^{(1)} \\ \underline{x}^{(2)} \\ \underline{\theta}^{(2)} \end{pmatrix}$, onde $\underline{\theta}_i = (\alpha_i, \beta_i, \mu_i, \sigma_{x(i)}^2, \sigma_i^2)^T$, $\underline{x}^{(1)} = (x_1, \dots, x_k)^T$, $\underline{x}^{(2)} = (x_{k+1}, \dots, x_n)^T$

e $\hat{\pi}(\underline{U}_1, \dots, \underline{U}_n | M_{1,k})$ pode ser obtido de forma totalmente semelhante a $\hat{\pi}(\underline{U}_1, \dots, \underline{U}_n | M_2)$, descrito acima.

(2°) O critério do Pseudo Fator de Bayes

Na seção [4.4.3b] do Apêndice damos uma breve introdução sobre essa medida no caso geral.

O Pseudo Fator de Bayes para comparar o modelo $M_{1,k}$ com o modelo M_2 (onde $M_{1,k}$ é o modelo estrutural Normal que permite mudança em todos os parâmetros e M_2 é o modelo sem mudança) é dado por

$$\text{PBF}_{(1,k)2} = \frac{\prod_{i=1}^n \text{CPO}_i^{(M_{1,k})}}{\prod_{i=1}^n \text{CPO}_i^{(M_2)}},$$

onde $\text{CPO}_i^{(M_2)} = \pi_{M_2}(\underline{U}_i | D_{\text{obs}(-i)})$ pode ser estimada usando as fórmulas (2.10.5) ou (2.10.9) ou (2.10.11) e $\text{CPO}_i^{(M_{1,k})} = \pi_{M_{1,k}}(\underline{U}_i | D_{\text{obs}(-i)})$ pode ser estimada usando por exemplo as fórmulas (4.5.5) ou (4.5.8), dadas mais adiante, para todo $i = 1, \dots, n$.

Observação: Se o modelo $M_{1,k}$ à priori permite mudança em apenas um subconjunto de parâmetros após o ponto k , os cálculos podem ser refeitos facilmente, de

forma totalmente semelhante, apenas incorporando as devidas restrições.

(2º) A medida L no modelo com ponto de mudança

Na seção 2.11, a medida L foi obtida para o modelo estrutural Normal sem mudança. Nesta seção obtemos a medida L aplicada ao modelo estrutural Normal que permite mudança em todos os parâmetros após o ponto k . O modelo escolhido será aquele com o menor valor dessa medida.

• Formulação

Seja $D_{\text{obs}} = \begin{pmatrix} \underline{U}_1 \\ \vdots \\ \underline{U}_n \end{pmatrix}$ onde $\underline{U}_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix}$ é tal que

$$\underline{U}_i | \underline{\theta} \sim \begin{cases} N \left[\begin{pmatrix} \alpha_1 + \beta_1 \mu_1 \\ \mu_1 \end{pmatrix}; \begin{pmatrix} \beta_1^2 \sigma_{x(1)}^2 + \lambda \sigma_1^2 & \beta_1 \sigma_{x(1)}^2 \\ \beta_1 \sigma_{x(1)}^2 & \sigma_{x(1)}^2 + \sigma_{u_1}^2 \end{pmatrix} \right], i = 1, \dots, k, \\ N \left[\begin{pmatrix} \alpha_2 + \beta_2 \mu_2 \\ \mu_2 \end{pmatrix}; \begin{pmatrix} \beta_2^2 \sigma_{x(2)}^2 + \lambda \sigma_2^2 & \beta_2 \sigma_{x(2)}^2 \\ \beta_2 \sigma_{x(2)}^2 & \sigma_{x(2)}^2 + \sigma_{u_2}^2 \end{pmatrix} \right], i = k + 1, \dots, n, \end{cases}$$

onde $\underline{\theta} = \begin{pmatrix} k \\ \theta_1 \\ \theta_2 \end{pmatrix}$ e $\theta_i = (\alpha_i, \beta_i, \mu_i, \sigma_{x(i)}^2, \sigma_i^2)^T$, $i = 1, 2$. Além disso, $\underline{U}_i | \underline{\theta}$ são independentes para $i = 1, \dots, n$.

Seja \underline{U}_i^* , $i = 1, \dots, n$, os vetores de observações futuras, com a mesma distribuição de $\underline{U}_i | \underline{\theta}$. Então a medida L é dada por

$$L_2(D_{\text{obs}}, \delta) = \sum_{i=1}^n \text{Var}(\underline{U}_i^* | D_{\text{obs}}) + \delta \sum_{i=1}^n [\underline{U}_i - E(\underline{U}_i^* | D_{\text{obs}})][\underline{U}_i - E(\underline{U}_i^* | D_{\text{obs}})]^T,$$

onde $0 \leq \delta \leq 1$ e onde

$$E(\underline{U}_i^* | D_{\text{obs}}) = E_{\underline{\theta} | D_{\text{obs}}} [E(\underline{U}_i^* | \underline{\theta})], \quad e$$

$$\text{Var}(\underline{U}_i^* | D_{\text{obs}}) = E_{\underline{\theta} | D_{\text{obs}}} [\text{Var}(\underline{U}_i^* | \underline{\theta})] + \text{Var}_{\underline{\theta} | D_{\text{obs}}} [E(\underline{U}_i^* | \underline{\theta})],$$

onde

$$\text{Var}_{\underline{\theta} | D_{\text{obs}}} [E(\underline{U}_i^* | \underline{\theta})] = E_{\underline{\theta} | D_{\text{obs}}} ([E(\underline{U}_i^* | \underline{\theta})][E(\underline{U}_i^* | \underline{\theta})]^T)$$

$$-[E_{\check{\theta}^*|D_{\text{obs}}} (E(\underline{U}_i^*|\check{\theta}^*))][E_{\check{\theta}^*|D_{\text{obs}}} (E(\underline{U}_i^*|\check{\theta}^*))]^\top,$$

Portanto, a medida L pode ser facilmente calculada usando Monte Carlo da seguinte maneira.

Seja $(\check{\theta}^{(1)}, \dots, \check{\theta}^{(G)})$, onde $\check{\theta}^{(g)} = \begin{pmatrix} k^{(g)} \\ \theta_1^{(g)} \\ \theta_2^{(g)} \end{pmatrix}$ uma amostra da posteriori $p(\check{\theta}^*|D_{\text{obs}})$,

gerada por exemplo pelo algoritmo de Gibbs da seção 4.2.2, então

$$\begin{aligned} \widehat{L}_2(D_{\text{obs}}, \delta) &= \sum_{i=1}^n \left\{ \frac{1}{G} \sum_{g=1}^G \text{Var}(\underline{U}_i^*|\check{\theta}^{(g)}) + \frac{1}{G} \sum_{g=1}^G E(\underline{U}_i^*|\check{\theta}^{(g)}) E(\underline{U}_i^*|\check{\theta}^{(g)})^\top \right. \\ &\quad \left. - \left(\frac{1}{G} \sum_{g=1}^G E(\underline{U}_i^*|\check{\theta}^{(g)}) \right) \left(\frac{1}{G} \sum_{g=1}^G E(\underline{U}_i^*|\check{\theta}^{(g)}) \right)^\top \right\} \\ &\quad - \delta \sum_{i=1}^n \left[\underline{U}_i - \frac{1}{G} \sum_{g=1}^G E(\underline{U}_i^*|\check{\theta}^{(g)}) \right] \left[\underline{U}_i - \frac{1}{G} \sum_{g=1}^G E(\underline{U}_i^*|\check{\theta}^{(g)}) \right]^\top \end{aligned}$$

onde

$$\begin{aligned} E(\underline{U}_i^*|\check{\theta}^{(g)}) &= \begin{cases} E^{(1)}(\underline{U}_i^*|\theta_1^{(g)}) = \begin{pmatrix} \alpha_1^{(g)} + \beta_1^{(g)} \mu_1^{(g)} \\ \mu_1^{(g)} \end{pmatrix}, & \text{se } 1 \leq i \leq k^{(g)}, \\ E^{(2)}(\underline{U}_i^*|\theta_2^{(g)}) = \begin{pmatrix} \alpha_2^{(g)} + \beta_2^{(g)} \mu_2^{(g)} \\ \mu_2^{(g)} \end{pmatrix}, & \text{se } k^{(g)} + 1 \leq i \leq n \end{cases} \\ \text{Var}(\underline{U}_i^*|\check{\theta}^{(g)}) &= \begin{cases} \text{Var}^{(1)}(\underline{U}_i^*|\theta_1^{(g)}) = \begin{pmatrix} \beta_1^{(g)} \sigma_{x(1)}^{2(g)} + \lambda \sigma_1^{2(g)} & \beta_1^{(g)} \sigma_{x(1)}^{2(g)} \\ \beta_1^{(g)} \sigma_{x(1)}^{2(g)} & \sigma_{x(1)}^{2(g)} + \sigma_1^{2(g)} \end{pmatrix}, & \text{se } 1 \leq i \leq k^{(g)}, \\ \text{Var}^{(2)}(\underline{U}_i^*|\theta_2^{(g)}) = \begin{pmatrix} \beta_2^{(g)} \sigma_{x(2)}^{2(g)} + \lambda \sigma_2^{2(g)} & \beta_2^{(g)} \sigma_{x(2)}^{2(g)} \\ \beta_2^{(g)} \sigma_{x(2)}^{2(g)} & \sigma_{x(2)}^{2(g)} + \sigma_2^{2(g)} \end{pmatrix}, & \text{se } k^{(g)} + 1 \leq i \leq n, \end{cases} \end{aligned}$$

onde $\delta \in [0, 1]$.

Note que $\widehat{L}_2(D_{\text{obs}}, \delta)$ é uma matriz 2×2 e uma possível medida resumo seria obter o determinante ou traço dessa matriz, isto é, $\det(\widehat{L}_2(D_{\text{obs}}, \delta))$ ou $\text{tr}(\widehat{L}_2(D_{\text{obs}}, \delta))$.

Observação: No modelo com mudança apenas num subconjunto de parâmetros, o cálculo da medida L é totalmente semelhante, bastando adaptar as fórmulas (incorporando-se as devidas restrições de igualdade).

(4°) O BIC

Seja $M_{1,k}$ o modelo estrutural Normal com restrição de identificabilidade e que permite mudança em todos os parâmetros após o ponto k e seja M_2 o modelo sem mudança.

Na seção 2.11 vimos como calcular o BIC_{M_2} . O cálculo do $BIC_{M_{1,k}}$ também é simples já que no modelo identificável com ponto de mudança k , os estimadores de máxima verossimilhança são conhecidos explicitamente.

Observação: É interessante ressaltar que esses métodos podem ser utilizados para escolher o “melhor modelo” dentre os 2^5 possíveis modelos (isto é, aqueles que permitem mudança em até 5 ou 4 ou 3 ou 2 ou 1 ou nenhum parâmetro).

4.5 Distribuição preditiva e avaliação do ajuste

A avaliação da qualidade do ajuste do modelo estrutural Normal com um ponto de mudança pode ser feita de forma análoga à apresentada na seção 2.10 para o modelo sem mudança. Para isso, vamos obter as fórmulas de cálculo da densidade condicional preditiva de validação cruzada CPO e dos resíduos de validação cruzada correspondentes.

Nessa seção vamos trabalhar apenas com o modelo que permite mudança em todos os parâmetros, lembrando que o procedimento para o caso de um modelo que permite mudança em apenas um subconjunto de parâmetros é totalmente semelhante, bastando fazer as devidas adaptações.

Na subseção 4.5.1, a seguir, apresentamos a distribuição preditiva, na subseção 4.5.2 as fórmulas de CPO e na subseção 4.5.3 os resíduos de validação cruzada.

4.5.1 Distribuição preditiva

A distribuição preditiva sob o modelo estrutural Normal que permite mudança em todos os parâmetros, dado em (4.1.1), é dada por $\pi(\underline{U}_f|D_{\text{obs}})$, onde \underline{U}_f é uma observação futura $\begin{pmatrix} Y_f \\ X_f \end{pmatrix}$, e D_{obs} é a amostra $(\underline{U}_1, \underline{U}_2, \dots, \underline{U}_n)$, onde $\underline{U}_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix}$.

Cálculo de $\pi(\underline{U}_f|D_{\text{obs}})$

$$\begin{aligned} \pi(\underline{U}_f|D_{\text{obs}}) &= \int_{\Theta^*} \pi(\underline{U}_f, \check{\theta}|D_{\text{obs}})d\check{\theta} = \int_{\Theta^*} \pi(\underline{U}_f|\check{\theta})\pi(\check{\theta}|D_{\text{obs}})d\check{\theta} \\ &= \int_{\Theta^*} \pi(Y_f|X_f, \check{\theta})\pi(X_f|\check{\theta})\pi(\check{\theta}|D_{\text{obs}})d\check{\theta} \\ &\quad \left(= \int_{\Theta^*} \pi(X_f|Y_f, \check{\theta})\pi(Y_f|\check{\theta})\pi(\check{\theta}|D_{\text{obs}})d\check{\theta} \right), \end{aligned}$$

onde $\check{\theta} = \begin{pmatrix} k \\ \theta_1 \\ \theta_2 \end{pmatrix}$ e $\pi(\check{\theta}|D_{\text{obs}})$ é a posteriori de $\check{\theta}$ e onde

$$\underline{U}_f|\check{\theta} \stackrel{\text{indep}}{\sim} \begin{cases} N\left(\begin{pmatrix} \alpha_1 + \beta_1\mu_1 \\ \mu_1 \end{pmatrix}; \begin{pmatrix} \beta_1^2\sigma_{x(1)}^2 + \lambda\sigma_1^2 & \beta_1\sigma_{x(1)}^2 \\ \beta_1\sigma_{x(1)}^2 & \sigma_{x(1)}^2 + \sigma_1^2 \end{pmatrix}\right), & \text{se } f \leq k \\ N\left(\begin{pmatrix} \alpha_2 + \beta_2\mu_2 \\ \mu_2 \end{pmatrix}; \begin{pmatrix} \beta_2^2\sigma_{x(2)}^2 + \lambda\sigma_2^2 & \beta_2\sigma_{x(2)}^2 \\ \beta_2\sigma_{x(2)}^2 & \sigma_{x(2)}^2 + \sigma_2^2 \end{pmatrix}\right), & \text{se } f > k \end{cases} \quad (4.5.1)$$

Geração de uma observação futura

Para gerar, por exemplo, uma observação \underline{U}_f após a mudança, basta seguir por exemplo os seguintes passos:

- (1) gerar $\check{\theta}^{(g)}$ da posteriori $\pi(\check{\theta}|D_{\text{obs}})$ (ou de $\pi(\check{\theta}, \underline{x}|D_{\text{obs}})$),
- (2) gerar $X_f^{(g)}$ de $\pi(X_f|\check{\theta}^{(g)})$, onde $X_f|\check{\theta} \sim N(\mu_2; \sigma_{x(2)}^2 + \sigma_2^2)$,
- (3) gerar $Y_f^{(g)}$ de $\pi(Y_f|X_f^{(g)}, \check{\theta}^{(g)})$, onde

$$Y_f|X_f, \check{\theta} \sim N\left(\alpha_2 + \beta_2\mu_2 + \frac{\beta_2^2\sigma_{x(2)}^2}{\sigma_{x(2)}^2 + \sigma_2^2}(X_f - \mu_2); \beta_2^2\sigma_{x(2)}^2 + \lambda\sigma_2^2 - \frac{\beta_2^2(\sigma_{x(2)}^2)^2}{\sigma_{x(2)}^2 + \sigma_2^2}\right).$$

Portanto, para gerar Y_f , após a mudança, dado que $X_f = X_0$, basta

- (1) gerar $\check{\theta}^{(g)}$ de $\pi(\check{\theta}|D_{\text{obs}})$ (ou $\pi(\check{\theta}, \underline{x}|D_{\text{obs}})$),
- (2) gerar $Y_f^{(g)}$ de $\pi(Y_f|X_f = X_0, \check{\theta}^{(g)})$.

(b) Para gerar uma observação futura \underline{U}_f (ou $\underline{Y}_f|X_f = X_0$), até a mudança, basta seguir os mesmos passos, porém agora

$$X_f|\check{\theta} \sim N(\mu_1; \sigma_{x(1)}^2 + \sigma_1^2) \quad \text{e}$$

$$Y_f|X_f, \check{\theta} \sim N\left(\alpha_1 + \beta_1\mu_1 + \frac{\beta_1^2\sigma_{x(1)}^2}{\sigma_{x(1)}^2 + \sigma_1^2}(X_f - \mu_1); \beta_1^2\sigma_{x(1)}^2 + \lambda\sigma_1^2 - \frac{\beta_1^2(\sigma_{x(1)}^2)^2}{\sigma_{x(1)}^2 + \sigma_1^2}\right).$$

Observação: A distribuição preditiva $\pi(\underline{U}_f|D_{\text{obs}})$ também pode ser escrita como

$$\int_{\Theta^*} \int_{\mathfrak{X}_f} \pi(\underline{U}_f, \check{\theta}, x_f|D_{\text{obs}}) d\check{\theta} dx_f = \int_{\Theta^*} \int_{\mathfrak{X}_f} \pi(\underline{U}_f|x_f, \check{\theta})\pi(x_f|\check{\theta})\pi(\check{\theta}|D_{\text{obs}}) d\check{\theta} dx_f, \quad (4.5.2)$$

onde

$$x_f|\check{\theta} \stackrel{\text{indep}}{\sim} \begin{cases} N(\mu_1, \sigma_{x(1)}^2), & \text{se } f \leq k, \\ N(\mu_2, \sigma_{x(2)}^2), & \text{se } f > k, \end{cases}$$

e

$$\underline{U}_f|x_f, \check{\theta} \stackrel{\text{indep}}{\sim} \begin{cases} N\left[\begin{pmatrix} \alpha_1 + \beta_1 x_f \\ x_f \end{pmatrix}; \begin{pmatrix} \lambda\sigma_1^2 & 0 \\ 0 & \sigma_1^2 \end{pmatrix}\right], & f \leq k, \\ N\left[\begin{pmatrix} \alpha_2 + \beta_2 x_f \\ x_f \end{pmatrix}; \begin{pmatrix} \lambda\sigma_2^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right], & f > k. \end{cases}$$

4.5.2 Densidade condicional preditiva de validação cruzada

A seguir apresentamos três maneiras alternativas de estimar a densidade preditiva de validação cruzada (CPO_i) para o modelo estrutural Normal com mudança em todos os parâmetros após a k -ésima observação.

Seja $\underline{U}_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix}$ a i -ésima observação da amostra, $D_{\text{obs}(-i)} = \underline{U}_{(-i)}$, e $\text{CPO}_i = \pi(\underline{U}_i|D_{\text{obs}(-i)})$.

Fórmulas de cálculo da CPO_i

Fórmula 1

$$\text{CPO}_i = \pi(\underline{U}_i | D_{\text{obs}(-i)}) = \left(\int_{\Theta} \frac{1}{\pi(\underline{U}_i | \underline{\theta}^*)} \pi(\underline{\theta}^* | D_{\text{obs}}) d\underline{\theta}^* \right)^{-1}, \quad (4.5.3)$$

$i = 1, \dots, n$, onde $\underline{\theta}^* = \begin{pmatrix} \theta_1 \\ \theta_2 \\ k \end{pmatrix}$, $\theta_i = (\alpha_i, \beta_i, \mu_i, \sigma_{x(i)}^2, \sigma_i^2)^T$, $i = 1, 2$ (com $\sigma_i^2 = \sigma_{u_i}^2$ e $\sigma_{e_i}^2 = \lambda \sigma_i^2$), onde

$$\underline{U}_i | \underline{\theta}^* \stackrel{\text{indep}}{\sim} \begin{cases} N \left(\begin{pmatrix} \alpha_1 + \beta_1 \mu_1 \\ \mu_1 \end{pmatrix}; \begin{pmatrix} \beta_1^2 \sigma_{x(1)}^2 + \lambda \sigma_1^2 & \beta_1^2 \sigma_{x(1)}^2 \\ \beta_1^2 \sigma_{x(1)}^2 & \sigma_{x(1)}^2 + \sigma_1^2 \end{pmatrix} \right), \\ \text{se } 1 \leq i \leq k, \\ N \left(\begin{pmatrix} \alpha_2 + \beta_2 \mu_2 \\ \mu_2 \end{pmatrix}; \begin{pmatrix} \beta_2^2 \sigma_{x(2)}^2 + \lambda \sigma_2^2 & \beta_2^2 \sigma_{x(2)}^2 \\ \beta_2^2 \sigma_{x(2)}^2 & \sigma_{x(2)}^2 + \sigma_2^2 \end{pmatrix} \right), \\ \text{se } k+1 \leq i \leq n. \end{cases} \quad (4.5.4)$$

Estimação

Dada uma amostra $(\underline{\theta}^{(1)}, \dots, \underline{\theta}^{(G)})$ da posteriori $p(\underline{\theta}^* | D_{\text{obs}})$, onde $\underline{\theta}^{(g)} = \begin{pmatrix} \theta_1^{(g)} \\ \theta_2^{(g)} \\ k^{(g)} \end{pmatrix}$

então uma estimativa de Monte Carlo de CPO_i é dada por

$$\widehat{\text{CPO}}_i = \left(\frac{1}{G} \sum_{g=1}^G \frac{1}{\pi(\underline{U}_i | \underline{\theta}^{(g)})} \right)^{-1};$$

onde

$$\pi(\underline{U}_i | \underline{\theta}^{(g)}) = \begin{cases} \pi_1(\underline{U}_i | \theta_1^{(g)}), & \text{se } 1 \leq i \leq k^{(g)}, \\ \pi_2(\underline{U}_i | \theta_2^{(g)}), & \text{se } k^{(g)} < i \leq n, \end{cases} \quad (4.5.5)$$

para $g = 1, \dots, G$, onde,

$$\begin{aligned} \pi_j(\underline{U}_i | \theta_j^{(g)}) &= (2\pi | \sum_j^{(g)} |)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2 | \sum_j^{(g)} |} \left[((\beta_j^{(g)})^2 \sigma_{x(j)}^{2(g)} + \lambda \sigma_j^{2(g)}) (X_i - \mu_j^{(g)})^2 \right. \right. \\ &\quad \left. \left. - 2\beta_j^{(g)} \sigma_{x(j)}^{2(g)} (Y_i - \alpha_j^{(g)} - \beta_j^{(g)} \mu_j^{(g)}) (X_i - \mu_j^{(g)}) + (\sigma_{x(j)}^{2(g)} + \sigma_j^{2(g)}) (Y_i - \alpha_j^{(g)} - \beta_j^{(g)} \mu_j^{(g)})^2 \right] \right\} \\ e \sum_j^{(g)} &= (\beta_j^{(g)})^2 \sigma_j^{2(g)} \sigma_{x(j)}^{2(g)} + \lambda \sigma_{x(j)}^{2(g)} \sigma_j^{2(g)} + \lambda (\sigma_{x(j)}^{2(g)})^2, j = 1, 2. \end{aligned} \quad (4.5.6)$$

Observação: A amostra $(\check{\theta}^{(1)}, \dots, \check{\theta}^{(G)})$ pode ser gerada da posteriori $p(\theta|D_{\text{obs}})$ por exemplo pelo algoritmo de Metropolis-Hastings em Gibbs (desde que este convirja), ou então da posteriori conjunta $p(\check{\theta}, \underline{x}|D_{\text{obs}})$, por exemplo pelo algoritmo de Gibbs (desde que este convirja).

Fórmula 2

$$\text{CPO}_i = \pi(\underline{U}_i|D_{\text{obs}(-i)}) = \left(\int_{\underline{x}} \int_{\Theta^*} \frac{1}{\pi(\underline{U}_i|\underline{x}, \check{\theta})} \pi(\underline{x}, \check{\theta}|D_{\text{obs}}) d\underline{x} d\check{\theta} \right)^{-1}, \quad (4.5.7)$$

$i = 1, \dots, n$, onde

$$\underline{U}_i|\underline{x}, \check{\theta} \overset{\text{indep}}{\sim} \begin{cases} N\left(\begin{pmatrix} \alpha_1 + \beta_1 x_i \\ x_i \end{pmatrix}; \begin{pmatrix} \lambda\sigma_1^2 & 0 \\ 0 & \sigma_1^2 \end{pmatrix}\right), & 1 \leq i \leq k, \\ N\left(\begin{pmatrix} \alpha_2 + \beta_2 x_i \\ x_i \end{pmatrix}; \begin{pmatrix} \lambda\sigma_2^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right), & k+1 \leq i \leq n. \end{cases}$$

Estimação

Dada uma amostra $\left(\left(\frac{\check{\theta}^{(1)}}{\check{\underline{x}}^{(1)}}\right), \dots, \left(\frac{\check{\theta}^{(G)}}{\check{\underline{x}}^{(G)}}\right)\right)$ da posteriori conjunta $\pi(\underline{x}, \check{\theta}|D_{\text{obs}})$, uma estimativa de Monte Carlo de CPO_i é dada por

$$\widehat{\text{CPO}}_i = \left(\frac{1}{G} \sum_{g=1}^G \frac{1}{\pi(\underline{U}_i|\check{\theta}^{(g)}, x_i^{(g)}, k^{(g)})} \right)^{-1},$$

onde

$$\pi(\underline{U}_i|\check{\theta}^{(g)}, x_i^{(g)}, k^{(g)}) = \begin{cases} \pi_1(\underline{U}_i|\check{\theta}_1^{(g)}, x_i^{(g)}), & \text{se } 1 \leq i \leq k^{(g)}, \\ \pi_2(\underline{U}_i|\check{\theta}_2^{(g)}, x_i^{(g)}), & \text{se } k^{(g)} + 1 \leq i \leq n, \end{cases} \quad (4.5.8)$$

onde

$$\pi_j(\underline{U}_i|\check{\theta}_j^{(g)}, x_i^{(g)}) = (2\pi\lambda(\sigma_j^{2(g)})^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[\frac{(Y_i - \alpha_j^{(g)} - \beta_j^{(g)} x_i^{(g)})^2}{\lambda\sigma_j^{2(g)}} + \frac{(X_i - x_i^{(g)})^2}{\sigma_j^{2(g)}} \right] \right\}, \quad (4.5.9)$$

$g = 1, \dots, G, j = 1, 2$.

Observação: Se não for possível obter uma amostra $\left(\left(\frac{\check{\theta}^{(1)}}{\check{\underline{x}}^{(1)}}\right), \dots, \left(\frac{\check{\theta}^{(G)}}{\check{\underline{x}}^{(G)}}\right)\right)$ da posteriori $p(\check{\theta}, \underline{x}|D_{\text{obs}})$ pelo algoritmo de Gibbs (por exemplo, porque não há convergência), mas for possível amostrar de $p(\check{\theta}|D_{\text{obs}})$ (por exemplo, pelo algoritmo

de M-H em Gibbs), então podemos ainda utilizar a expressão em (4.5.8) para estimar a CPO_i. A diferença, entretanto, está no processo de geração. Nesse caso, dada a amostra $(\check{\theta}^{(1)}, \dots, \check{\theta}^{(G)})$ de $\pi(\check{\theta}|D_{\text{obs}}) = \pi(\theta, k|D_{\text{obs}})$, devemos ainda gerar $x_i^{(g)}$ de $\pi(x_i|\check{\theta}^{(g)})$, onde

$$x_i|\check{\theta}^{(g)}, k^{(g)} \sim \begin{cases} N(\mu_1^{(g)}, \sigma_{x(1)}^{2(g)}), & 1 \leq k^{(g)} \leq n, \\ N(\mu_2^{(g)}, \sigma_{x(2)}^{2(g)}), & k^{(g)} + 1 \leq i \leq n. \end{cases}$$

A justificativa desse procedimento segue do fato que

$$\pi(\underline{U}_i|D_{\text{obs}(-i)}) = \left[\int_{\mathfrak{X}_i} \int_{\Theta^*} \frac{1}{\pi(\underline{U}_i|x_i, \check{\theta})} \pi(x_i|\check{\theta}) \pi(\check{\theta}|D_{\text{obs}}) d\check{\theta} dx_i \right]^{-1},$$

onde

$$x_i|\check{\theta} \sim \begin{cases} N(\mu_1, \sigma_{x(1)}^2), & 1 \leq i \leq k, \\ N(\mu_2, \sigma_{x(2)}^2), & k + 1 \leq i \leq n. \end{cases}$$

4.5.3 Resíduos de validação cruzada

O resíduo padronizado de validação cruzada do modelo com mudança é dado por

$$\check{d}_i = \text{Var}^{-\frac{1}{2}}(\underline{U}_i|D_{\text{obs}(-i)})[\underline{U}_i - E(\underline{U}_i|D_{\text{obs}(-i)})], \quad i = 1, \dots, n.$$

Cálculo de $E(\underline{U}_i|D_{\text{obs}(-i)})$

$$\begin{aligned} E(\underline{U}_i|D_{\text{obs}(-i)}) &= E[E(\underline{U}_i|\check{\theta}, \mathfrak{X})|D_{\text{obs}(-i)}] \\ &= \int_{\mathfrak{X}} \int_{\Theta^*} E(\underline{U}_i|\check{\theta}, x_i) \pi(\check{\theta}, \mathfrak{X}|D_{\text{obs}(-i)}) d\check{\theta} d\mathfrak{X} \quad (4.5.10) \\ &= \left(\int_{\mathfrak{X}} \int_{\Theta^*} \frac{1}{\pi(\underline{U}_i|\check{\theta}, x_i)} \pi(\check{\theta}, \mathfrak{X}|D_{\text{obs}}) d\check{\theta} d\mathfrak{X} \right)^{-1} \int_{\mathfrak{X}} \int_{\Theta^*} \frac{E(\underline{U}_i|\check{\theta}, x_i)}{\pi(\underline{U}_i|\check{\theta}, x_i)} \pi(\check{\theta}, \mathfrak{X}|D_{\text{obs}}) d\check{\theta} d\mathfrak{X} \end{aligned}$$

onde

$$\underline{U}_i|\check{\theta}, x_i \stackrel{\text{indep}}{\sim} \begin{cases} N\left(\begin{pmatrix} \alpha_1 + \beta_1 x_i \\ x_i \end{pmatrix}; \begin{pmatrix} \lambda \sigma_1^2 & 0 \\ 0 & \sigma_1^2 \end{pmatrix}\right), & i = 1, \dots, k, \\ N\left(\begin{pmatrix} \alpha_2 + \beta_2 x_i \\ x_i \end{pmatrix}; \begin{pmatrix} \lambda \sigma_2^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right), & i = k + 1, \dots, n. \end{cases}$$

Estimação de $E(\underline{U}_i|D_{\text{obs}(-i)})$

Dada uma amostra $\left(\left(\frac{\hat{\theta}^{(1)}}{\hat{x}^{(1)}}\right), \dots, \left(\frac{\hat{\theta}^{(G)}}{\hat{x}^{(G)}}\right)\right)$ de $\pi(\underline{\theta}, \underline{x}|D_{\text{obs}})$, uma estimativa de Monte Carlo de $E(\underline{U}_i|D_{\text{obs}(-i)})$ é dada por

$$\hat{E}(\underline{U}_i|D_{\text{obs}(-i)}) = \widehat{\text{CPO}}_i \frac{1}{G} \sum_{g=1}^G \frac{1}{\pi(\underline{U}_i|\underline{\theta}^{(g)}, x_i^{(g)}, k^{(g)})} E(\underline{U}_i|\underline{\theta}^{(g)}, x_i^{(g)}, k^{(g)}), \quad (4.5.11)$$

onde $\widehat{\text{CPO}}_i$ é dada por (4.5.8), $\pi(\underline{U}_i|\underline{\theta}^{(g)}, x_i^{(g)})$ é dada por (4.5.9),

$$E(\underline{U}_i|\underline{\theta}^{(g)}, x_i^{(g)}, k^{(g)}) = \begin{cases} \begin{pmatrix} \alpha_1^{(g)} + \beta_1^{(g)} x_1^{(g)} \\ x_i^{(g)} \end{pmatrix}, & i = 1, \dots, k^{(g)}, \\ \begin{pmatrix} \alpha_2^{(g)} + \beta_2^{(g)} x_i^{(g)} \\ x_i^{(g)} \end{pmatrix}, & i = k^{(g)} + 1, \dots, n, \end{cases} \quad (4.5.12)$$

Estimação da $\text{Var}(\underline{U}_i|D_{\text{obs}(-i)})$

$$\begin{aligned} \widehat{\text{Var}}(\underline{U}_i|D_{\text{obs}(-i)}) &= \widehat{\text{CPO}}_i \left[\frac{1}{G} \sum_{g=1}^G \frac{\text{Var}(\underline{U}_i|\underline{\theta}^{(g)}, x_i^{(g)}, k^{(g)})}{\pi(\underline{U}_i|\underline{\theta}^{(g)}, x_i^{(g)}, k^{(g)})} + \right. \\ &\quad \left. \frac{1}{G} \sum_{j=1}^G \frac{E(\underline{U}_i|\underline{\theta}^{(g)}, x_i^{(g)}, k^{(g)}) E(\underline{U}_i|\underline{\theta}^{(j)}, x_i^{(j)}, k^{(j)})^\top}{\pi(\underline{U}_i|\underline{\theta}^{(g)}, x_i^{(g)}, k^{(g)})} \right] \\ &\quad - [\hat{E}(\underline{U}_i|D_{\text{obs}(-i)})][\hat{E}(\underline{U}_i|D_{\text{obs}(-i)})]^\top, \end{aligned}$$

onde

$\widehat{\text{CPO}}_i$ é dada em (4.5.8), $\pi(\underline{U}_i|\underline{\theta}^{(g)}, x_i^{(g)})$ é dada em (4.5.9),

$\hat{E}(\underline{U}_i|D_{\text{obs}(-i)})$ é dada em (4.5.11), $E(\underline{U}_i|\underline{\theta}^{(g)}, x_i^{(g)})$ é dada em (4.5.12), e

$$\text{Var}(\underline{U}_i|\underline{\theta}^{(g)}, x_i^{(g)}, k^{(g)}) = \begin{cases} \begin{pmatrix} \lambda \sigma_1^{2(g)} & 0 \\ 0 & \sigma_1^{2(g)} \end{pmatrix}, & i = 1, \dots, k^{(g)}, \\ \begin{pmatrix} \lambda \sigma_2^{2(g)} & 0 \\ 0 & \sigma_2^{2(g)} \end{pmatrix}, & i = k^{(g)} + 1, \dots, n. \end{cases}$$

4.6 Estabelecimento das condições para a existência da posteriori sob algumas escolhas de prioris impróprias

Nessa seção vamos estudar a existência da distribuição à posteriori no modelo de regressão estrutural Normal com ponto de mudança e com restrição de identificabilidade (de que a razão das variâncias é conhecida) considerando prioris impróprias análogas as usadas na seção 2.8. Inicialmente, na subseção 4.6.1, estudamos o modelo que permite mudança em todos os parâmetros e depois, na subseção 4.6.2, o modelo que permite mudança apenas em σ_x^2 com priori do tipo $\pi(\varrho, k) \propto p(k)\pi(\sigma^2) \prod_{i=1}^2 \pi(\sigma_{x(i)}^2)$, onde $\pi(\sigma_{x(i)}^2) \sim IG(c_i, d_i)$, para $i = 1, 2$.

4.6.1 Condições necessárias e suficientes no modelo que permite mudança em todos os parâmetros

Nesta seção apresentamos alguns teoremas que garantem a existência da posteriori no modelo de regressão linear simples com erros normais que permite mudança em todos os parâmetros quando certas prioris impróprias do tipo $\pi(k, \varrho) = p(k)\pi(\varrho_1)\pi(\varrho_2)$ são utilizadas.

No modelo com mudança em todos os parâmetros e priori de componentes independentes, vimos em (4.1.5) que a posteriori $p(k, \varrho|D_{\text{obs}})$ é proporcional a

$$p(k) \cdot L_1(\varrho_1|D_{\text{obs}_1}^{(k)}) \cdot \pi(\varrho_1) \cdot L_2(\varrho_2|D_{\text{obs}_2}^{(n-k)}) \pi(\varrho_2)$$

onde $L_1(\varrho_1|D_{\text{obs}_1}^{(k)})$ e $L_2(\varrho_2|D_{\text{obs}_2}^{(n-k)})$ são, respectivamente, as verossimilhanças antes e depois da mudança dadas em (4.1.3) e $\pi(\varrho_1)$ e $\pi(\varrho_2)$ são, respectivamente, as prioris antes e depois da mudança. O espaço paramétrico é $\mathcal{K} \times \Theta$, onde \mathcal{K} é um conjunto finito dos possíveis valores de k e $\Theta = \{\varrho = \begin{pmatrix} \varrho_1 \\ \varrho_2 \end{pmatrix} : \varrho_i \in \Theta_i\}$.

Portanto, para mostrar que a posteriori é própria (isto é, para mostrar que

$\sum_{k \in \mathcal{K}} \int_{\Theta} p(k, \vartheta | D_{\text{obs}}) d\vartheta < \infty$), devemos mostrar que

$$\sum_{k \in \mathcal{K}} p(k) \int_{\Theta_1} L_1(\vartheta_1 | D_{\text{obs}_1}^{(k)}) \pi(\vartheta_1) d\vartheta_1 \cdot \int_{\Theta_2} L_2(\vartheta_2 | D_{\text{obs}_2}^{(n-k)}) \pi(\vartheta_2) d\vartheta_2 < \infty,$$

onde $L_1(\vartheta_1 | D_{\text{obs}_1}^{(k)}) \pi(\vartheta_1)$ é o núcleo da posteriori $p_1(\vartheta_1 | D_{\text{obs}_1}^{(k)})$ (baseada nos dados até a mudança) e $L_2(\vartheta_2 | D_{\text{obs}_2}^{(n-k)}) \pi(\vartheta_2)$ é o núcleo da posteriori $p_2(\vartheta_2 | D_{\text{obs}_2}^{(n-k)})$ (baseada nos dados após a mudança).

Como \mathcal{K} é um conjunto finito, então basta mostrar que qualquer parcela dessa soma é finita, ou seja, que as posteriores $p_1(\vartheta_1 | D_{\text{obs}_1}^{(k)})$ e $p_2(\vartheta_2 | D_{\text{obs}_2}^{(n-k)})$ são próprias para k fixado ($k \in \mathcal{K}$).

Teorema 4.6.1 *No modelo Normal com ponto de mudança dado em (4.1.1), com priori da forma $\pi(\vartheta) = \pi(k, \vartheta_1, \vartheta_2) = p(k) \prod_{i=1}^2 \pi(\vartheta_i)$, onde*

$$p(k) = \begin{cases} P[K=k], & k \in \mathcal{K}, \\ 0, & \text{caso contrário,} \end{cases} \quad \text{e } \pi(\vartheta_i) \propto \pi(\sigma_{x^{(i)}}^2) \pi(\sigma_i^2), \text{ onde } \pi(\sigma_{x^{(i)}}^2) \sim IG(c_i, d_i)$$

e $\pi(\sigma_i^2) \sim IG(f_i, g_i)$, $i = 1, 2$, então a condição necessária e suficiente para a existência da posteriori é que $k \geq 3$ e $n - k \geq 3$.

Demonstração: Para cada k fixado, o Teorema 2.8.2 garante que a integral

$\int_{\Theta_1} p_1(\vartheta_1 | D_{\text{obs}_1}^{(k)}) d\vartheta_1$ é finita se e somente se $k \geq 3$, e a integral $\int_{\Theta_2} p_2(\vartheta_2 | D_{\text{obs}_2}^{(n-k)}) d\vartheta_2$ é finita se e somente se $(n - k) \geq 3$. Portanto, concluímos que a posteriori do modelo é própria se e somente se $3 \leq k \leq n - 3$ e $n \geq 6$. ■

Teorema 4.6.2 *No modelo do Teorema 4.6.1, se a distribuição $\pi(\sigma_{x^{(i)}}^2)$ for substituída por $\pi(\sigma_{x^{(i)}}^2) \propto (\sigma_{x^{(i)}}^2)^{-r_i}$, $r_i > 0$, $i = 1, 2$, então a condição necessária e suficiente para que a posteriori correspondente seja própria é que $k \geq 2$, $n - k \geq 2$ e $0 < r_i < \frac{1}{2}$, $i = 1, 2$.*

Demonstração: Para cada k fixado, o Teorema 2.8.3 garante que a integral

$\int_{\Theta_1} p_1(\vartheta_1 | D_{\text{obs}_1}^{(k)}) d\vartheta_1$ é finita se e somente se $k \geq 2$ e $0 < r_1 < \frac{1}{2}$ e a integral

$\int_{\Theta_2} p_2(\theta_2 | D_{\text{obs}_2}^{(n-k)}) d\theta_2$ é finita se e somente se $(n-k) \geq 2$ e $0 < r_2 < \frac{1}{2}$. Portanto, a posteriori sob esse modelo é própria se e somente se $k \geq 2$, $n-k \geq 2$ e $0 < r_i < \frac{1}{2}$, $i = 1, 2$. ■

Teorema 4.6.3 *Se no modelo do Teorema 4.6.1 a distribuição de $\pi(\sigma_i^2)$ for substituída por $\pi(\sigma_i^2) \propto (\sigma_i^2)^{-s_i}$, $s_i > 0$, $i = 1, 2$, e além disso os vetores de dados $\underline{X}^{(i)}$ e $\underline{Y}^{(i)}$ forem linearmente independentes e as componentes de $\underline{X}^{(i)}$ (bem como as de $\underline{Y}^{(i)}$) não são todas iguais (para $i = 1, 2$), então a condição necessária e suficiente para que a posteriori sob esse modelo seja própria é que $k \geq 3$ e $n-k \geq 3$.*

Demonstração: A demonstração segue usando o mesmo argumento do teorema anterior e o Teorema 2.8.4 do Capítulo 2. ■

Teorema 4.6.4 *No modelo do Teorema 4.6.1, se as distribuições de $\pi(\sigma_{x(i)}^2)$ e $\pi(\sigma_i^2)$ forem substituídas, respectivamente, por $\pi(\sigma_{x(i)}^2) \propto (\sigma_{x(i)}^2)^{-r_i}$ e $\pi(\sigma_i^2) \propto (\sigma_i^2)^{-s_i}$, $r_i, s_i > 0$, $i = 1, 2$, e além disso $\underline{X}^{(i)}$ e $\underline{Y}^{(i)}$ forem linearmente independentes e tais que as componentes de $\underline{X}^{(i)}$ (bem como as de $\underline{Y}^{(i)}$) não são todas iguais, então a condição necessária e suficiente para que a distribuição à posteriori sob esse modelo seja própria é que $k + r_1 + s_1 > 3$ e $(n-k) + r_2 + s_2 > 3$ e $0 < r_i < \frac{1}{2}$, $i = 1, 2$.*

Demonstração: Usando o mesmo argumento do teorema anterior e o Teorema 2.8.5, concluímos que para cada k fixado as integrais

$$\int_{\Theta_1} p_1(\theta_1 | D_{\text{obs}_1}^{(k)}) d\theta_1 \quad \text{e} \quad \int_{\Theta_2} p_2(\theta_2 | D_{\text{obs}_2}^{(n-k)}) d\theta_2$$

são finitas se e somente se $k + r_1 + s_1 > 3$, $(n-k) + r_2 + s_2 > 3$ e $0 < r_i < \frac{1}{2}$, $i = 1, 2$, e portanto a posteriori existe se e somente se estas condições forem satisfeitas. ■

4.6.2 Condições suficientes no modelo que permite mudança apenas em σ_x^2

Teorema 4.6.5 *No modelo estrutural Normal que permite mudança apenas no parâmetro σ_x^2 após o ponto k , com priori $\pi(\underline{\theta}, k) \propto p(k)\pi(\sigma^2) \prod_{i=1}^2 \pi(\sigma_{x(i)}^2)$, onde*

$\underline{\theta} = (\alpha, \beta, \mu, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma^2)^T$, $p(k) = \begin{cases} P[K=k], & k \in \mathcal{K}, \text{ onde } \mathcal{K} \text{ é um conjunto finito,} \\ 0, & \text{caso contrário,} \end{cases}$
 $\pi(\sigma^2) \sim IG(f, g)$ e $\pi(\sigma_{x(i)}^2) \sim IG(c_i, d_i)$, $i = 1, 2$, então uma condição suficiente para que a posteriori seja própria é que $n \geq 3$.

Demonstração: O objetivo é mostrar que $\sum_{k \in \mathcal{K}} \int_{\Theta} p(k, \underline{\theta} | D_{\text{obs}}) d\underline{\theta} < \infty$,

onde $p(k, \underline{\theta} | D_{\text{obs}}) = \int_{\mathcal{X}} p(k, \underline{\theta}, \underline{x} | D_{\text{obs}}) d\underline{x}$ e $p(k, \underline{\theta}, \underline{x} | D_{\text{obs}}) \propto L(k, \underline{\theta} | D) \pi(k, \underline{\theta})$,

onde $D_{\text{obs}} = \{(Y_i, X_i)^T, i = 1, \dots, n\}$, $D = \{(Y_i, X_i, x_i)^T, i = 1, \dots, n\}$, $\Theta = \{\underline{\theta} = (\alpha, \beta, \mu, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma^2)^T \in \mathbb{R}^3 \times \mathbb{R}_+^3\}$, \mathcal{K} é o conjunto finito dos possíveis valores de k e $\mathcal{X} = \{\underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^n\}^T$.

Portanto, para mostrar que a posteriori é própria, basta mostrar que

$$\int_{\mathcal{X}} \int_{\Theta} L(k, \underline{\theta} | D) \pi(\underline{\theta}) d\underline{\theta} d\underline{x} < \infty \text{ para } k \text{ fixado, onde}$$

$$\begin{aligned} L(k, \underline{\theta} | D) \pi(\underline{\theta}, k) &\propto (\sigma^2)^{-(n+f+1)} (\sigma_{x(1)}^2)^{-(\frac{k}{2}+c_1+1)} (\sigma_{x(2)}^2)^{-(\frac{n-k}{2}+c_2+1)} \\ &\cdot \exp \left\{ -\frac{1}{2} \left[\frac{2d_1}{\sigma_{x(1)}^2} + \frac{2d_2}{\sigma_{x(2)}^2} + \frac{2g}{\sigma^2} + \frac{\sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2}{\lambda \sigma^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma^2} \right. \right. \\ &\left. \left. + \frac{\sum_{i=1}^k (x_i - \mu)^2}{\sigma_{x(1)}^2} + \frac{\sum_{i=k+1}^n (x_i - \mu)^2}{\sigma_{x(2)}^2} \right] \right\}. \end{aligned}$$

Esta função pode ser majorada pela seguinte função:

$$\begin{aligned} f(k, \underline{\theta}, \underline{x}, D_{\text{obs}}) &= (\sigma^2)^{-(n+f+1)} (\sigma_{x(1)}^2)^{-(\frac{k}{2}+c_1+1)} (\sigma_{x(2)}^2)^{-(\frac{n-k}{2}+c_2+1)} \cdot \exp \left\{ -\frac{1}{2} \left[\frac{2d_1}{\sigma_{x(1)}^2} \right. \right. \\ &\left. \left. + \frac{2d_2}{\sigma_{x(2)}^2} + \frac{2g}{\sigma^2} + \frac{\sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2}{\lambda \sigma^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma^2} + \frac{\sum_{i=1}^k (x_i - \mu)^2}{\sigma_{x(1)}^2} \right] \right\} \end{aligned}$$

e portanto

$$\int_{\mathcal{X}} \int_{\Theta} L(k, \underline{\theta} | D) \pi(\underline{\theta}, k) d\underline{\theta} d\underline{x} \leq \int_{\mathcal{X}} \int_{\Theta} f(k, \underline{\theta}, \underline{x}, D_{\text{obs}}) d\underline{\theta} d\underline{x}.$$

Integrando inicialmente a função $f(k, \theta, \underline{x}, D_{\text{obs}})$ em relação a $\theta =$

$(\alpha, \beta, \mu, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma^2)^T$, obtemos que $\int_{\Theta} f(k, \theta, \underline{x}, D_{\text{obs}}) d\theta$ é proporcional a

$$1 \left/ \left[S_{xx}^{\frac{1}{2}} (2d_1 + S_{x(1)x(1)})^{\frac{k+2c_1+1}{2}} \left(2g + \sum_{i=1}^n (X_i - x_i)^2 + \frac{S_{xx} S_{YY} - S_{xY}^2}{\lambda S_{xx}} \right)^{n+f-1} \right] \right.,$$

onde $S_{x(1)x(1)} = \sum_{i=1}^k (x_i - \bar{x}_1)^2$.

$$\begin{aligned} & \text{Portanto, } \int_{\mathfrak{X}} \int_{\Theta} f(k, \theta, \underline{x}, D_{\text{obs}}) d\theta d\underline{x} \\ & \leq (2d_1)^{-\frac{(k+2c_1+1)}{2}} \int_{\mathfrak{X}} \frac{1}{S_{xx}^{\frac{1}{2}} \left(2g + \sum_{i=1}^n (X_i - x_i)^2 + \frac{S_{xx} S_{YY} - S_{xY}^2}{\lambda S_{xx}} \right)^{n+f-1}} d\underline{x}. \end{aligned}$$

A integral $\int_{\mathfrak{X}} \frac{1}{S_{xx}^{\frac{1}{2}} \left(2g + \sum_{i=1}^n (X_i - x_i)^2 + \frac{S_{xx} S_{YY} - S_{xY}^2}{\lambda S_{xx}} \right)^{n+f-1}} d\underline{x}$ pode ser analisada mais facilmente fazendo uma transformação de variáveis (usada na seção 2.8) que a cada $(\underline{x}_{n \times 1}, \underline{X}_{n \times 1}, Y_{n \times 1})$ associa $(\underline{z}_{n \times 1}, \underline{b}_{n \times 1}, \underline{\xi}_{n \times 1}) = (H \underline{x}_{n \times 1}, H \underline{X}_{n \times 1}, H Y_{n \times 1})$, onde H é a matriz de Helmert.

De fato, com esta mudança de variáveis, a integral acima fica igual a

$$\int_Z \frac{1}{\|\underline{z}_x\| \left(2g + \|\underline{z}_x - \underline{b}_X\|^2 + (\sqrt{n} \bar{X} - z_n)^2 + \frac{\|\underline{\xi}_Y\|^2 \sin^2 \theta}{\lambda} \right)^{n+f-1}} d\underline{z},$$

onde \underline{z}_x , \underline{b}_x e $\underline{\xi}_Y$ são os vetores formados pelas primeiras $n-1$ componentes de \underline{z} , \underline{b} e $\underline{\xi}$, respectivamente. θ é o ângulo entre $\underline{\xi}_Y$ e \underline{z}_x e $Z = \{ \underline{z} = (z_1, \dots, z_n)^T : z_i \in \mathbb{R} \}$.

Essa integral pode ser escrita como

$$\int_{Z_x} \int_{\{z_n \in \mathbb{R}\}} \frac{1}{\|\underline{z}_x\| \left(2g + \|\underline{z}_x - \underline{b}_X\|^2 + (\sqrt{n} \bar{X} - z_n)^2 + \frac{\|\underline{\xi}_Y\|^2 \sin^2 \theta}{\lambda} \right)^{n+f-1}} dz_n d\underline{z}_x,$$

onde $Z_x = \{ \underline{z}_x = (z_1, \dots, z_{n-1}) : z_i \in \mathbb{R} \}$ e portanto, desde que $n+f > \frac{3}{2}$, ela é proporcional a

$$\int_{Z_x} \frac{1}{\|\underline{z}_x\| \left(2g + \|\underline{z}_x - \underline{b}_X\|^2 + \frac{\|\underline{\xi}_Y\|^2 \sin^2 \theta}{\lambda} \right)^{n+f-\frac{3}{2}}} d\underline{z}_x.$$

Essa última integral (que é totalmente semelhante àquela que aparece na demonstração do Teorema 2.8.4) é finita se $1 < n - 1$ e $1 + 2(n + f - \frac{3}{2}) > n - 1$, isto é, se $n \geq 3$. Concluimos portanto que a posteriori é finita se $n \geq 3$. ■

Teorema 4.6.6 *No modelo do Teorema 4.6.1, se trocarmos a distribuição $\pi(\sigma^2) \sim IG(f, g)$ por $\pi(\sigma^2) \propto (\sigma^2)^{-s}$, $s > 0$, e se \underline{X} e \underline{Y} forem linearmente independentes, onde as componentes de \underline{X} , e as de \underline{Y} , não são todas iguais, então uma condição suficiente para a existência da posteriori é que $n \geq 3$.*

Demonstração: Substituindo f por $(s - 1)$ e g por zero na demonstração anterior, vemos que $\int_{\mathcal{X}} \int_{\Theta} p(\theta, k, \underline{x} | D_{\text{obs}}) d\theta d\underline{x}$ é majorada por uma integral que é proporcional a

$$\int_{\mathcal{Z}_x} \frac{1}{\|z_x\| \left(\|z_x - \underline{b}_X\|^2 + \frac{\|\underline{z}_Y\|^2 \sin^2 \theta}{\lambda} \right)^{n+s-\frac{5}{2}}} dz_x,$$

que é totalmente semelhante àquela que aparece na demonstração do Teorema 2.8.5 e portanto é finita se e somente se $1 < n - 1$ e $1 + 2(n + s - \frac{5}{2}) > n - 1$. Concluimos que para cada k fixado em \mathcal{K} , a integral $\int_{\mathcal{X}} \int_{\Theta} p(\theta, k, \underline{x} | D_{\text{obs}}) d\theta d\underline{x}$ é finita se $n \geq 3$ e portanto, se $n \geq 3$, então

$$\sum_{k \in \mathcal{K}} \int_{\mathcal{X}} \int_{\Theta} p(\theta, k, \underline{x} | D_{\text{obs}}) d\theta d\underline{x} < \infty. \quad \blacksquare$$

4.7 EXEMPLOS

Nessa seção são apresentados exemplos de aplicação da teoria das seções 4.1 a 4.4, envolvendo os modelos com restrição e sem restrição de identificabilidade, utilizando uma amostra de dados simulados.

4.7.1 Análise do modelo estrutural normal identificável

1. O MODELO que permite mudança em todos os parâmetros:

$$\begin{cases} Y_i = \begin{cases} \alpha_1 + \beta_1 x_i + e_i, & i = 1, \dots, k, \quad 1 \leq k \leq n-1, \\ \alpha_2 + \beta_2 x_i + e_i, & i = k+1, \dots, n, \end{cases} \\ X_i = x_i + u_i, \quad i = 1, \dots, n, \end{cases}$$

onde

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \stackrel{\text{iid}}{\sim} \begin{cases} N \left(\begin{pmatrix} 0 \\ 0 \\ \mu_1 \end{pmatrix}; \begin{pmatrix} \lambda\sigma_1^2 & 0 & 0 \\ 0 & \sigma_1^2 & 0 \\ 0 & 0 & \sigma_{x(1)}^2 \end{pmatrix} \right), & i = 1, \dots, k, \\ N \left(\begin{pmatrix} 0 \\ 0 \\ \mu_2 \end{pmatrix}; \begin{pmatrix} \lambda\sigma_2^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_{x(2)}^2 \end{pmatrix} \right), & i = k+1, \dots, n, \end{cases} \quad (4.7.1)$$

Distribuição à priori:

$$\pi(\varrho_1, \varrho_2, k) = \pi(k)\pi(\varrho_1)\pi(\varrho_2), \text{ onde } \pi(\varrho_i) = \pi(\alpha_i)\pi(\beta_i)\pi(\mu_i)\pi(\sigma_{x_i}^2)\pi(\sigma_i^2),$$

$$i = 1, 2, \text{ onde } \alpha_i \sim N(0; 1000), \beta_i \sim N(0; 1000), \mu_i \sim N(0; 1000),$$

$$\sigma_{x_i}^2 \sim IG(2.001; 1.0), \sigma_i^2 \sim IG(2.001; 1.0) \text{ e}$$

$$p(k) = \begin{cases} P(K = k) = \frac{1}{\#\mathcal{K}}, & k \in \mathcal{K} = \{1, \dots, n-1\}, \\ 0, & \text{caso contrário.} \end{cases} \quad \text{⊕}$$

e $\lambda = 1$ com probabilidade 1.

2. DADOS

A Tabela 4.1 apresenta uma amostra de tamanho $n=50$ de dados gerados a partir do modelo em (4.7.1), com mudança em todos os parâmetros, onde $\alpha_1=2.0$,

$\beta_1=1.0$, $\mu_1=2.0$, $\sigma_{x(1)}^2=0.6$, $\sigma_1^2=0.1$, $\alpha_2=1.0$, $\beta_2=2.0$, $\mu_2=4.0$, $\sigma_{x(2)}^2=1.2$, $\sigma_2^2=0.2$, $\lambda=1.0$, $k = 20$, utilizando-se um programa elaborado na linguagem OX (usando a semente “ranseed=17”).

Tabela 4.1: Amostra de tamanho $n=50$ de dados simulados a partir do modelo (4.7.1), onde $\alpha_1=2.0$, $\beta_1=1.0$, $\mu_1=2.0$, $\sigma_{x(1)}^2=0.6$, $\sigma_1^2=0.1$, $\alpha_2=1.0$, $\beta_2=2.0$, $\mu_2=4.0$, $\sigma_{x(2)}^2=1.2$, $\sigma_2^2=0.2$, $\lambda=1.0$, $k = 20$.

i	Y_i	X_i									
1	4.3092	1.7503	14	4.1315	2.3448	27	8.7985	3.6264	40	11.355	5.4473
2	3.4970	1.2028	15	4.6854	2.1908	28	13.962	6.3921	41	8.7974	2.9984
3	3.6984	1.8785	16	3.0402	1.0815	29	7.0030	2.9641	42	7.9116	4.0154
4	5.1943	2.4628	17	3.8688	2.0400	30	8.3874	4.1586	43	7.2250	4.0347
5	5.1489	2.8678	18	5.5450	3.5886	31	11.844	6.2737	44	11.923	5.7843
6	2.3622	0.7882	19	4.3375	1.6627	32	10.348	3.9149	45	5.3486	2.1688
7	4.2616	1.8405	20	4.2863	2.1210	33	8.1851	3.7377	46	10.021	4.5871
8	3.5888	2.5897	21	4.6501	2.8548	34	6.7018	2.7880	47	8.2339	3.5194
9	3.6661	2.5529	22	9.2052	4.0509	35	9.3974	4.4571	48	3.2243	0.7888
10	4.8814	3.6252	23	4.6988	1.9951	36	10.736	3.7249	49	8.8972	3.3560
11	4.0650	2.5828	24	12.144	5.0546	37	8.3573	3.4238	50	9.8834	4.3421
12	3.9926	2.5085	25	8.7343	3.4862	38	12.353	5.4157			
13	4.2061	2.4555	26	5.2684	2.2406	39	7.8066	3.1820			

OBSERVAÇÃO: Nesse estudo, pelo fato de que a distribuição à priori escolhida é praticamente não informativa e o conjunto de dados foi simulado a partir do modelo em (4.7.1), onde $\alpha_1=2.0$, $\beta_1=1.0$, $\mu_1=2.0$, $\sigma_{x(1)}^2=0.6$, $\sigma_1^2=0.1$, $\alpha_2=1.0$, $\beta_2=2.0$, $\mu_2=4.0$, $\sigma_{x(2)}^2=1.2$, $\sigma_2^2=0.2$, $\lambda=1.0$, $k = 20$ então, se a metodologia empregada é adequada, devemos esperar que as estimativas e os resultados à posteriori sejam compatíveis com esse modelo.

3. AJUSTE DO MODELO

Os algoritmos de Gibbs (simples) e “Modified and Collapsed Gibbs” foram utilizados com o objetivo de amostrar da distribuição à posteriori do modelo.

De cada algoritmo geramos $M = 4$ seqüências paralelas de comprimento $T = 10.000$ partindo de pontos iniciais obtidos pelo seguinte procedimento:

Obtivemos as estimativas de máximaverossimilhança de $\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$, para cada valor de k fixado (onde $k = 3, \dots, n - 3$). Aplicamos um algoritmo de otimização

de Quase-Newton (denominado BFGS) usando cada uma dessas estimativas como ponto inicial e localizamos n-5 pontos estacionários da posteriori. Depois disso geramos 1000 pontos $\hat{\theta}^* = \begin{pmatrix} k \\ \theta_1 \\ \theta_2 \end{pmatrix}$ da distribuição formada pela mistura de distribuições $t_4(\hat{\theta}_i^*, \hat{V}_i^*)$, $i = 1, \dots, n - 5$, onde $\hat{\theta}_i^*$ é o i-ésimo ponto estacionário e $\hat{V}_i^* = - \left(\frac{\partial^2 \log p(\theta | D_{\text{obs}})}{\partial \theta \partial \theta^\top} \right)^{-1} \Big|_{\theta = \hat{\theta}_i^*}$, onde os pesos w_i de cada distribuição são dados por $w_i = q(\hat{\theta}_i^* | D_{\text{obs}}) |\hat{V}_i^*|^{-\frac{1}{2}}$, onde $q(\hat{\theta}_i^* | D_{\text{obs}})$ é a posteriori não normalizada calculada no ponto $\hat{\theta}_i^*$.

Finalmente reamostramos apenas 10 pontos (utilizando re-amostragem ponderada), dos quais apenas os seguintes quatro pontos foram usados para inicializar o algoritmo de Gibbs:

$$\hat{\theta}_1^{(0)} = (21, 2.153, 0.828, 2.608, 0.393, 0.216, 0.265, 2.281, 3.335, 0.862, 0.215)^T$$

$$\hat{\theta}_2^{(0)} = (21, 1.466, 1.197, 2.286, 0.256, 0.185, 0.331, 2.149, 3.813, 1.299, 0.202)^T$$

$$\hat{\theta}_3^{(0)} = (21, 3.140, 0.393, 2.522, 0.582, 0.102, 2.299, 1.716, 3.998, 1.726, 0.084)^T$$

$$\hat{\theta}_4^{(0)} = (21, 1.924, 0.967, 1.950, 0.404, 0.064, 1.758, 2.013, 3.791, 1.638, 0.306)^T$$

3.a Diagnósticos de convergência

A seguir são apresentados alguns dispositivos gráficos para o diagnóstico de possíveis falhas na convergência dos algoritmos de Gibbs e Modified Collapsed Gibbs.

Para cada um dos parâmetros do modelo proposto temos os gráficos das séries temporais, das funções de auto-correlação e da medida de Gelman e Rubin baseados nos algoritmos de Gibbs e Modified Collapsed Gibbs. Temos também os gráficos da medida de Brooks e Gelman obtidos a partir desses dois algoritmos.

As figuras 4.1 a 4.8 não sugerem que exista problema na convergência dos algoritmos de Gibbs e Modified Collapsed Gibbs. Além disso o algoritmo M.Collap.Gibbs parece ter uma melhor performance que o Gibbs (como pode ser visto principalmente pela comparação das funções de auto-correlação amostral correspondentes).

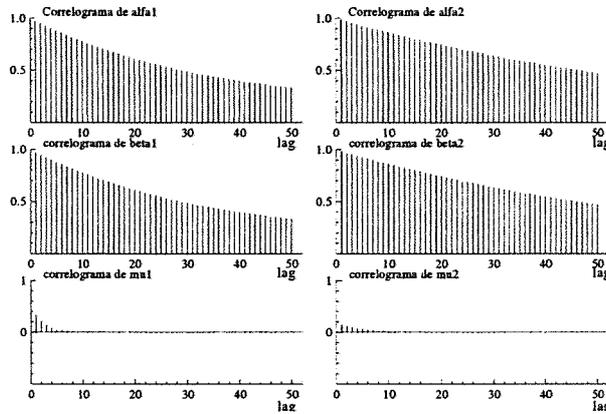


Figura 4.1: Gráficos das funções de auto-correlação de $\alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1$ e μ_2 , baseados em uma única seqüência de Gibbs de comprimento $T = 50.000$, partindo do ponto inicial $\hat{\theta}_1^{(0)}$.

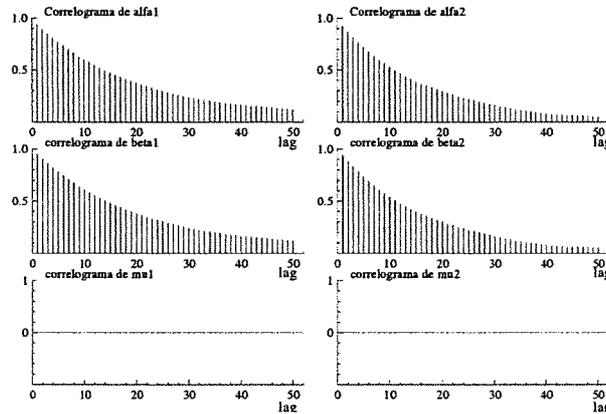


Figura 4.2: Gráficos das funções de auto-correlação de $\alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1$ e μ_2 , baseados em uma única seqüência do algoritmo M.Collapsd Gibbs de comprimento $T = 50.000$.

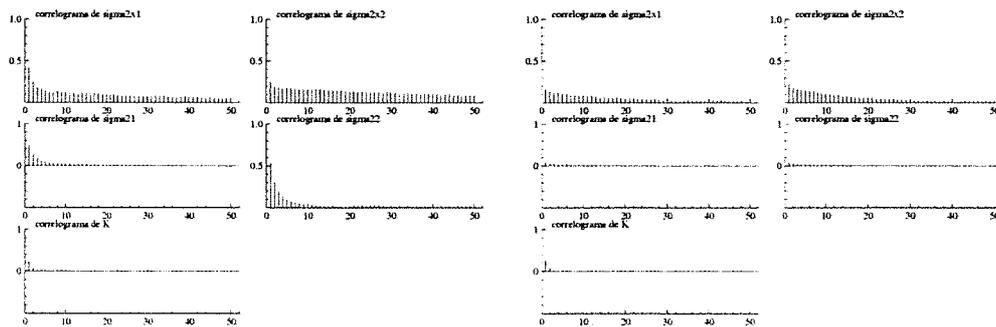


Figura 4.3: Gráficos das funções de auto-correlação de $\sigma_{x_1}^2, \sigma_{x_2}^2, \sigma_1^2, \sigma_2^2$ e k . Os 4 do lado esquerdo baseiam-se na mesma seq. de Gibbs da figura 4.1 e os 4 do lado direito na mesma seq. do alg. "M.Collapsd Gibbs" da fig 4.2.

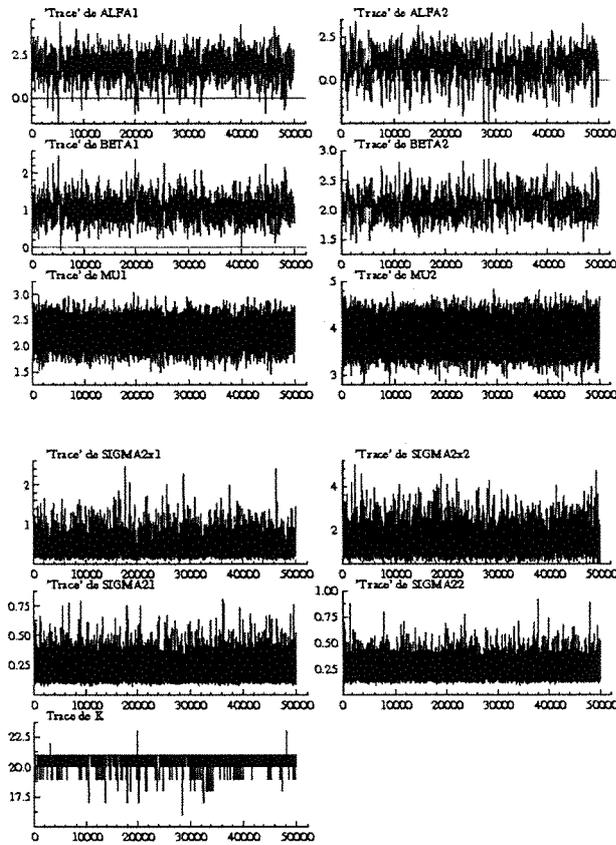


Figura 4.4: Gráficos das séries temporais (“trace”) de $\alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{r1}^2, \sigma_{r2}^2, \sigma_{l1}^2, \sigma_{l2}^2$ e k versus “número de iterações” baseados em uma única seqüência de comprimento $T = 50.000$ do algoritmo de Gibbs.

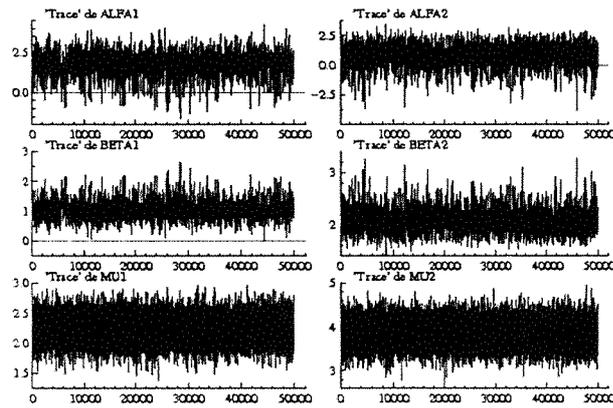


Figura 4.5: “Trace” de $\alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2$ baseados em uma seq. de $T=50000$ do alg. M.Collap.Gibbs.

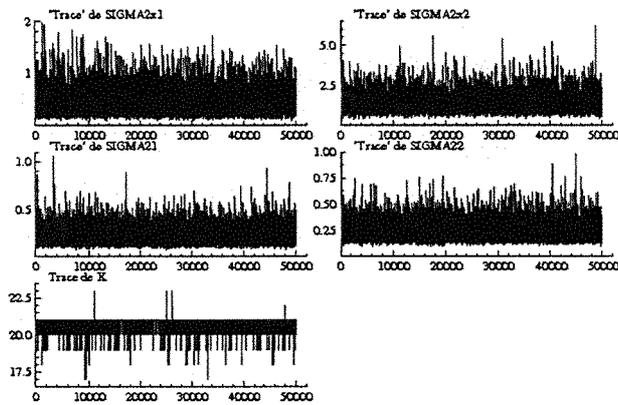
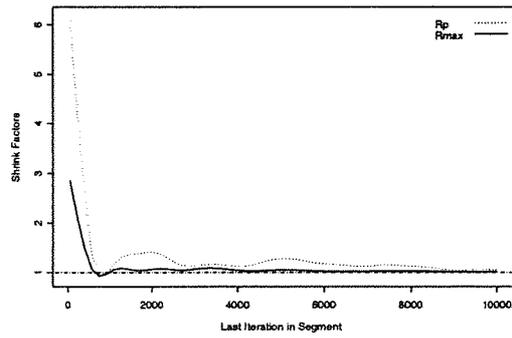


Figura 4.6: Gráficos das séries temporais (“Trace”) de $\sigma_{x_1}^2, \sigma_{x_2}^2, \sigma_1^2, \sigma_2^2$ e k versus “nro. de iterações” baseados em uma única seqüência de comprimento $T = 50.000$ do algoritmo M.Collapsd Gibbs .

Brooks & Gelman Multivariate Shrink Factors



Brooks & Gelman Multivariate Shrink Factors

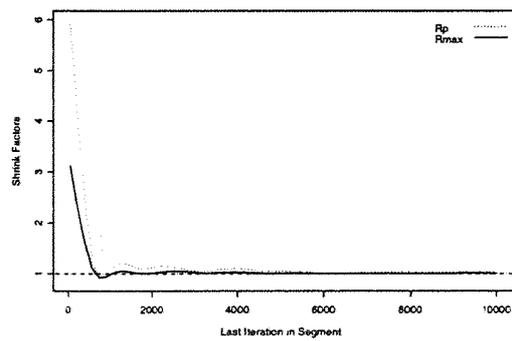
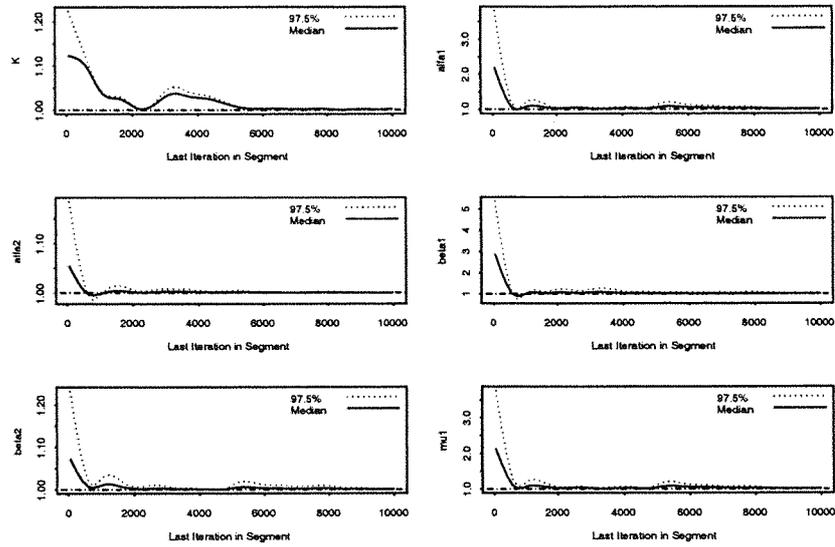


Figura 4.7: “Multiv. Potential Scale Reduction” de Brooks e Gelman (1998) baseados em $M = 4$ seqüências de comprimento $T = 10.000$. O gráfico superior é baseado no alg. de Gibbs e o inferior no alg. M.Collapsd Gibbs.

Gelman & Rubin Shrink Factors



Gelman & Rubin Shrink Factors

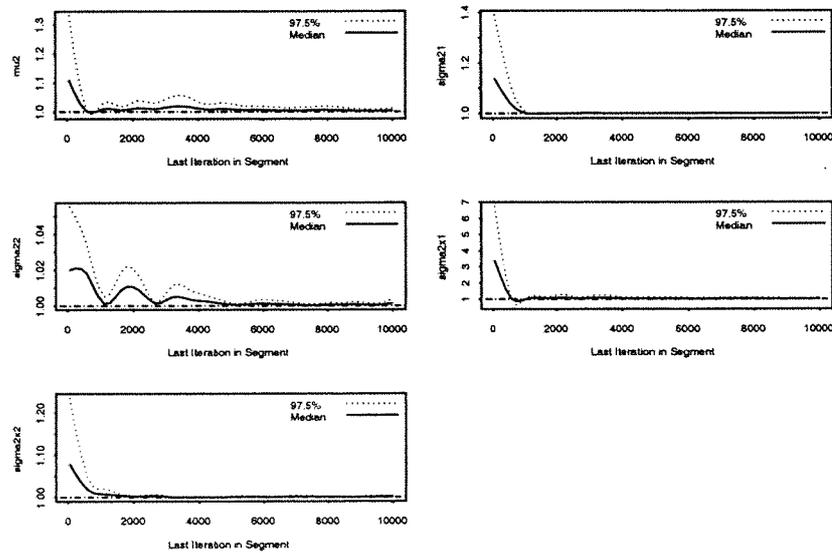
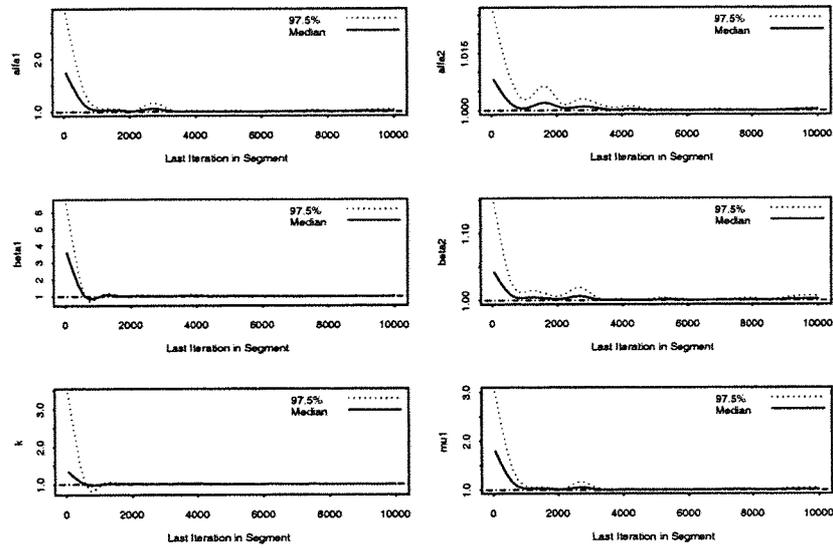


Figura 4.8: Gráfico da medida de Gelman e Rubin corrigida “Corrected potential scale reduction factors” (CSRF) para cada parâmetro $k, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{21}^2, \sigma_{22}^2, \sigma_{11}^2, \sigma_{21}^2$, baseado em $M = 4$ seqüências de comprimento $T = 10.000$ de Gibbs.

Gelman & Rubin Shrink Factors



Gelman & Rubin Shrink Factors

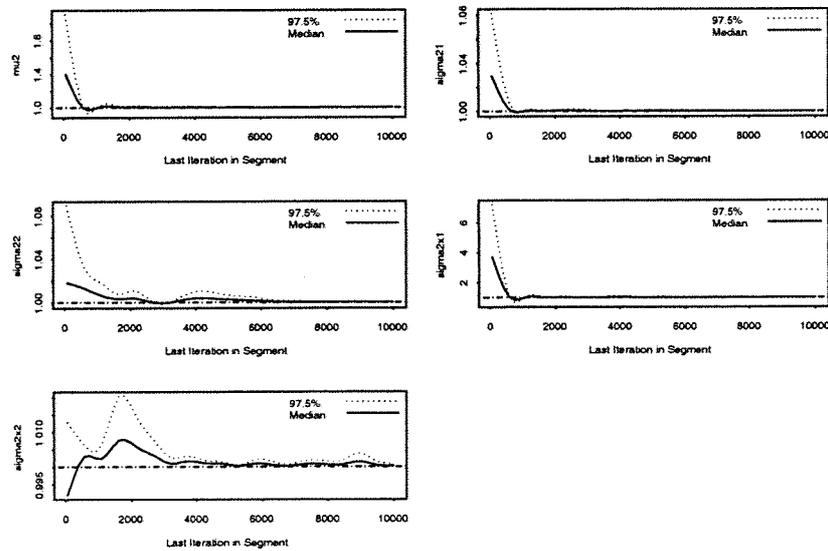


Figura 4.9: Gráfico da medida de Gelman e Rubin corrigida “Corrected potencial scale reduction factors” (CSRF) para cada parâmetro $k, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x_1}^2, \sigma_{x_2}^2, \sigma_1^2, \sigma_2^2$, baseado em $M = 4$ seqüências de comprimento $T = 10.000$ do algoritmo M.Collapsd Gibbs.

3.b Estimação:

Para verificar se as estimativas à posteriori do ponto de mudança k diferem muito do valor verdadeiro, aplicamos o algoritmo M.Collapsd Gibbs $n - 1$ vezes, cada vez usando um conjunto diferente de dados simulados (cada conjunto foi gerado a partir do modelo em (4.7.1) com um valor de k diferente (k entre 1 e $n-1$), mas com os mesmos valores de θ_1 e θ_2 usados na geração da tabela 4.1).

A tabela 4.2 apresenta os valores verdadeiros de k e as estimativas correspondentes, \hat{k} , obtidas utilizando-se o algoritmo “M. Collapsd Gibbs” partindo-se do ponto inicial $(k^{(0)}, \theta_1^{(0)}, \theta_2^{(0)})$, onde $k^{(0)} = 25$ e $\theta_i^{(0)} = (\beta_i^{(0)}, \sigma_{x_i}^{2(0)}, \sigma_i^{2(0)}) = (1, 1, 1)$, $i=1,2$, e descartando-se a primeira metade do total de 10.000 iterações.

Tabela 4.2: Valores verdadeiros de k e respectivos valores estimados(\hat{k}).

k	\hat{k}								
1	1.11	11	10.95	21	20.99	31	30.96	41	40.97
2	13.97	12	11.97	22	22.84	32	31.97	42	41.90
3	2.58	13	12.86	23	22.96	33	32.98	43	42.94
4	3.03	14	14.95	24	23.88	34	33.94	44	43.70
5	4.80	15	14.74	25	25.28	35	34.95	45	44.90
6	5.89	16	15.97	26	25.95	36	35.71	46	45.33
7	6.75	17	16.92	27	26.98	37	36.92	47	47.99
8	8.87	18	17.95	28	27.91	38	37.96	48	47.90
9	8.99	19	18.96	29	28.73	39	38.97	49	48.97
10	10.00	20	20.97	30	29.93	40	39.64	50	

Na tabela 4.2 observa-se que a única estimativa discrepante ocorre quando o verdadeiro valor de k é 2.

Observou-se nesse exemplo que para valores de k bem pequenos há uma tendência de que as estimativas de $\theta_1 = (\alpha_1, \beta_1, \mu_1, \sigma_{x_1}^2, \sigma_1^2)$ não sejam muito boas (em termos do viés) quando comparadas com o verdadeiro valor de θ_1 e que as estimativas de $\theta_2 = (\alpha_2, \beta_2, \mu_2, \sigma_{x_1}^2, \sigma_1^2)$ sejam melhores (ou seja, mais próximas do verdadeiro valor

de θ_2). Semelhantemente, quando o verdadeiro valor de k está próximo de n , as estimativas de θ_1 são melhores do que as de θ_2 . Esse fato já era esperado pela própria estrutura do problema.

A seguir apresentamos as estimativas de várias medidas resumo da distribuição à posteriori bem como as estimativas das distribuições posterioris marginais. Essas estimativas basearam-se em $M=4$ seqüências do algoritmo M. Collapsed Gibbs de comprimento $T = 10.000$, após descartar-se a primeira metade de cada uma. Os pontos iniciais dessas seqüências foram obtidos apartir dos pontos $\theta_1^{(0)}$ a $\theta_4^{(0)}$ (dados no início da seção) após descartar-se as componentes $\alpha_1^{(0)}, \alpha_2^{(0)}, \mu_1^{(0)}, \mu_2^{(0)}$ de cada um.

A tabela 4.4 apresenta as estimativas à posteriori de algumas medidas resumo da distribuição à posteriori (quando os dados observados são os da tabela 4.1 e a distribuição à priori é dada no início da seção). Observa-se nessa tabela que os valores da média aritmética e da média Rao-Blackelizada estão muito próximas. Observa-se também que tanto a média como a mediana amostral superestimam o verdadeiro valor do parâmetro σ_1^2 , entretanto, de maneira geral, elas estimam cada um dos demais parâmetros apenas com um pequeno viés. Mais importante ainda é que os intervalos de credibilidade estimados contém os valores verdadeiros dos parâmetros (no caso de σ_1^2 apenas o I.C. de 95% com duas casas decimais satisfaz isso).

A tabela 4.3 apresenta a distribuição de freqüências à posteriori de k .

Tabela 4.3: Distribuição de freqüência relativa à posteriori de k :

k	1 a 18	19	20	21	22 a 50	
$\hat{P}(K = k \text{Dobs})$	0.000	0.003	0.029	0.968	0.000	1

A Figura 4.10 apresenta os gráficos das estimativas “Rao-Blackwelizadas” das distribuições posterioris marginais de $\alpha_1, \alpha_2, \beta_1, \beta_2, \dots, \sigma_1^2, \sigma_2^2$ e k .

Tabela 4.4: Estimativas do valor esperado, quantis, desvio padrão, máximo, mínimo e intervalos de credibilidade de comprimento mínimo das distribuições marginais à posteriori de $\alpha_1, \alpha_2, \beta_1, \beta_2, \dots, \sigma_1^2, \sigma_2^2$ e k , usando o algoritmo M.Collap.Gibbs. (Os dados simulados são os da tabela 4.1 e a distribuição à priori é dada em (4.7.1))

ESTIMATIVAS	α_1	α_2	β_1	β_2	μ_1	μ_2
Média aritm.	1.7799	0.8341	1.0649	2.0774	2.2372	3.8593
Média R-Blackw.	1.7800	0.8329	1.0646	2.0774	2.2366	3.8582
Quantis						
0.025	0.3712	-0.7153	0.6040	1.7608	1.8894	3.3937
.05	0.6654	-0.4000	0.6673	1.8051	1.9503	3.4707
.25	1.4302	0.3768	0.8828	1.9534	2.1218	3.7050
.5	1.8420	0.8873	1.0406	2.0649	2.2367	3.8604
.75	2.1985	1.3374	1.2176	2.1911	2.3532	4.0149
.95	2.6901	1.9396	1.5450	2.3865	2.5238	4.2427
.975	2.8424	2.1221	1.6810	2.4640	2.5814	4.3267
Int.Credib. HPD						
$p = 0.95$ Lim. inf.	0.5146	-0.5797	0.5561	1.7445	1.8964	1.8785
Lim sup.	2.9365	2.2255	1.6076	2.4371	2.5854	4.3136
$p = 0.90$ Lim. inf.	0.7975	-0.3160	0.6229	1.7889	1.9570	1.9490
Lim sup.	2.7811	2.0026	1.4830	2.3606	2.5296	4.2402
desv.padrão (dp)	0.6201	0.7277	0.2689	0.1811	0.1751	0.2349
dp R-Blackweliz.	0.6198	0.7307	0.2687	0.1811	0.1742	0.2345
"Naive SE"	0.0044	0.0051	0.0019	0.0013	0.0012	0.0017
"Batch SE"	0.0207	0.0017	0.0232	0.0026	0.0090	0.0062
Mínimo	-1.6637	-2.8830	0.1122	1.4197	1.4755	2.6634
Máximo	4.0242	3.3749	2.6588	3.1043	2.9229	4.8612
ESTIMATIVAS	$\sigma_{x_1}^2$	$\sigma_{x_2}^2$	σ_1^2	σ_2^2	k	
Média aritm.	0.41950	1.3540	0.2156	0.2410	20.9650	
Média R-Blackw.	0.4199	1.3608	0.2158	0.2408	20.9670	
Quantis						
0.025	0.1828	0.7446	0.1189	0.1434	20	
.05	0.2059	0.8114	0.1291	0.1544	21	
.25	0.2981	1.0601	0.1678	0.1944	21	
.5	0.3878	1.2860	0.2034	0.2306	21	
.75	0.5037	1.5760	0.2495	0.2756	21	
.95	0.7417	2.1165	0.3422	0.3633	21	
.975	0.8412	2.3523	0.3815	0.3989	21	
Int.Credib. HPD						
$p = 0.95$ Lim. inf.	0.1575	0.6745	0.1040	0.1346	21	
Lim sup.	0.7629	2.2009	0.3502	0.3766	21	
$p = 0.90$ Lim. inf.	0.1632	0.7479	0.1164	0.1418	21	
Lim sup.	0.6613	1.9915	0.3172	0.3402	21	
desv.padrão (dp)	0.1722	0.4161	0.0685	0.0663	0.2117	
dp R-Blackweliz.	0.1736	0.4192	0.0685	0.0663	0.2044	
"Naive SE"	0.0012	0.0030	0.0005	0.0005	0.0014	
"Batch SE"	0.0059	0.0007	0.0012	0.0005	0.0022	
Mínimo	0.1043	0.4343	0.0642	0.0915	17	
Máximo	1.8396	5.5860	0.8925	0.7710	23	

NOTA: Os intervalos de credibilidade nesse exemplo coincidem com os intervalos HPD.
 "Batch SE" é o desvio padrão amostral das médias de grupos consecutivos de tamanho 50, dividido pela raiz quadrada do número de grupos (calculado por meio do BOA ("Bayesian Output Analysis")).

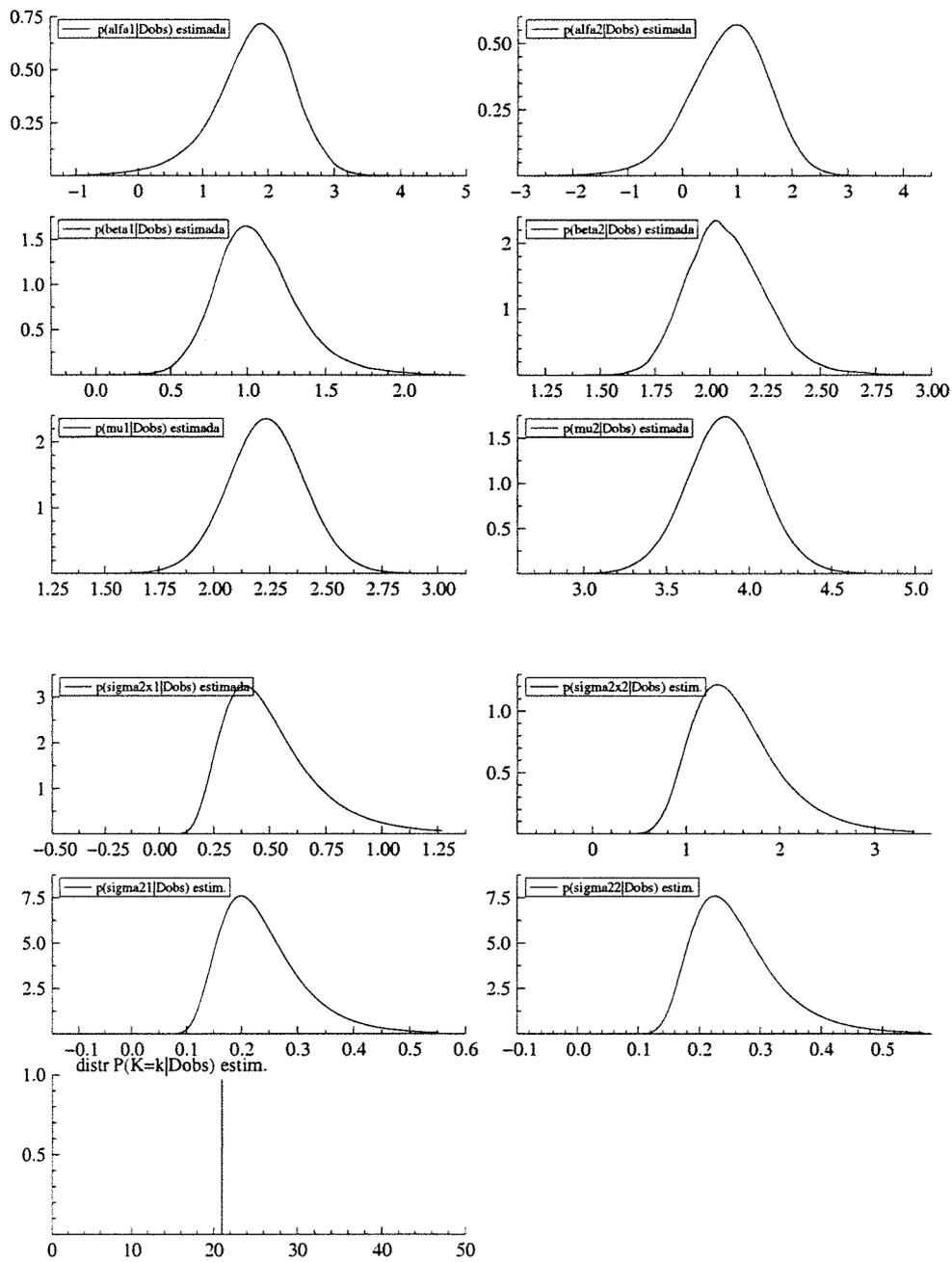


Figura 4.10: Gráficos das estimativas “Rao-Blackwelizadas” das densidades marginais à posteriori $p(\alpha_1|D_{obs}), p(\alpha_2|D_{obs}), \dots, p(\sigma_1^2|D_{obs}), p(\sigma_2^2|D_{obs})$ e $p(k|D_{obs})$.

3.c Avaliação do ajuste

Para avaliar a qualidade do ajuste primeiramente apresentamos a figura 4.11 com os gráficos dos valores da densidade condicional preditiva de validação cruzada CPO_i (onde $CPO_i = \pi \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{obs} \right)$) primeiramente em função de i , depois em função de X_i , e depois em função de Y_i , para $i = 1, \dots, 50$, onde CPO_i é estimada por meio da fórmula (4.5.8). As estimativas foram baseadas na segunda metade de uma seqüência de comprimento $T=10.000$ do algoritmo M.Collap.Gibbs que parte do ponto inicial $\theta_1^{(0)}$ dado no início da seção (descartando as componentes $\alpha_1^{(0)}, \alpha_2^{(0)}, \mu_1^{(0)}, \mu_2^{(0)}$).

Em seguida apresentamos a figura 4.12 com os gráficos dos resíduos bayesianos de validação cruzada.

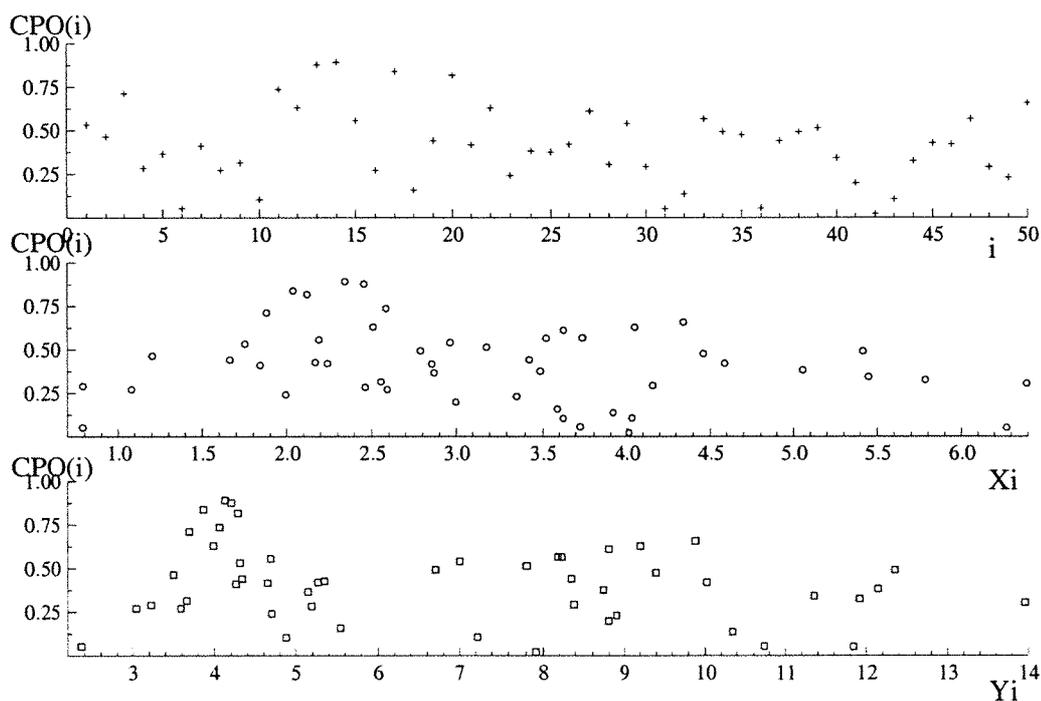


Figura 4.11: Gráfico dos valores da densidade condicional preditiva, $CPO(i)$, versus i (posição superior), versus X_i (posição do meio) e versus Y_i (posição inferior.)

Na seção 4.5.3 vimos que os resíduos bayesianos bidimensionais de validação cruzada podem ser estimados por

$$d_i = \begin{pmatrix} d_{1i} \\ d_{2i} \end{pmatrix} = \widehat{\text{Var}}^{-\frac{1}{2}} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{\text{obs}(-i)} \right) \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} - \hat{E} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{\text{obs}(-i)} \right) \right),$$

$i = 1, \dots, n$, onde $\hat{E} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{\text{obs}(-i)} \right)$ é dada em (4.5.11)

e $\widehat{\text{Var}}^{-\frac{1}{2}} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{\text{obs}(-i)} \right)$ é dada em (4.5.12), para $i = 1, \dots, n$.

A Figura 4.12 apresenta do lado esquerdo os gráficos da componente d_{1i} , e do lado direito os gráficos da componente d_{2i} , do vetor de resíduos.

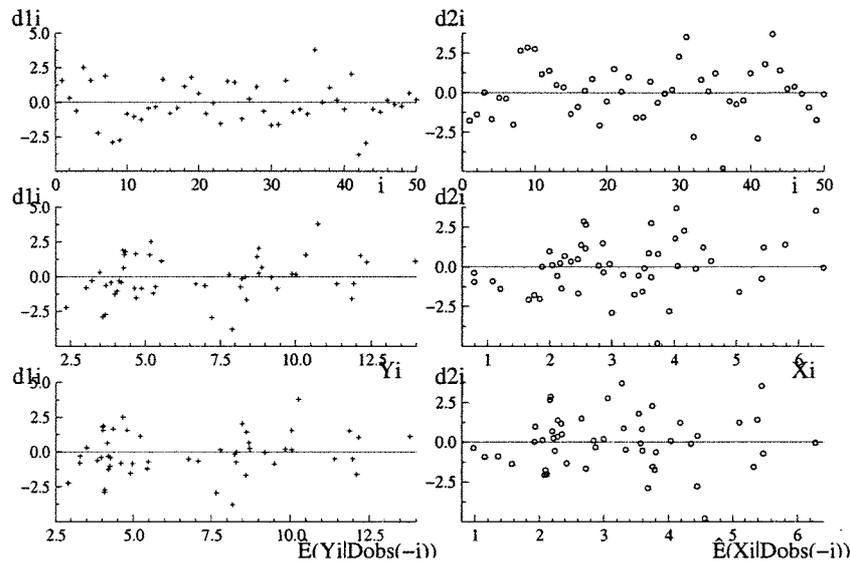


Figura 4.12: Do lado esquerdo os gráficos de d_{1i} , versus i , versus Y_i e versus $E(Y_i | \hat{D}_{\text{obs}}(-i))$. Do lado direito os gráficos de (d_{2i}) versus i , versus X_i e versus $E(X_i | \hat{D}_{\text{obs}}(-i))$.

Os gráficos acima não sugerem que exista grandes problemas de falta de ajustamento. Entretanto na figura 4.12 existem 3 pontos nos gráficos de ordenada d_{1i} , (do lado esquerdo) e 4 pontos nos gráficos de ordenada d_{2i} , (do lado direito) cujos valores são maiores do que 3 e portanto, aparentemente, poderiam ser vistos como pontos aberrantes, entretanto uma análise de resíduos mais aprofundada deve ser conduzida futuramente para examinar melhor a qualidade desse ajuste.

4.7.2 Análise do modelo estrutural não identificável

Nesta seção utilizamos uma amostra de tamanho $n = 50$ de dados gerados a partir do modelo que permite mudança em todos os parâmetros dado a seguir:

1. MODELO

$$\begin{cases} Y_i = \begin{cases} \alpha_1 + \beta_1 x_i + e_i, & i = 1, \dots, k, \quad 1 \leq k \leq n-1, \\ \alpha_2 + \beta_2 x_i + e_i, & i = k+1, \dots, n, \end{cases} \\ X_i = x_i + u_i, \quad i = 1, \dots, n, \end{cases}$$

onde

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \stackrel{\text{iid}}{\sim} \begin{cases} N \left(\begin{pmatrix} 0 \\ 0 \\ \mu_1 \end{pmatrix}; \begin{pmatrix} \sigma_{e_1}^2 & 0 & 0 \\ 0 & \sigma_1^2 & 0 \\ 0 & 0 & \sigma_{x(1)}^2 \end{pmatrix} \right), & i = 1, \dots, k, \\ N \left(\begin{pmatrix} 0 \\ 0 \\ \mu_2 \end{pmatrix}; \begin{pmatrix} \sigma_{e_2}^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_{x(2)}^2 \end{pmatrix} \right), & i = k+1, \dots, n, \end{cases} \quad (4.7.2)$$

Distribuição à priori

$\pi(\hat{\theta}) = \pi(\theta_1, \theta_2, k) = \pi(k)\pi(\theta_1)\pi(\theta_2)$, onde $\pi(\theta_i) = \pi(\alpha_i)\pi(\beta_i)\pi(\mu_i)\pi(\sigma_{x_i}^2)\pi(\sigma_{e_i}^2)\pi(\sigma_i^2)$, $i = 1, 2$, onde $\alpha_i \sim N(0; 1000)$, $\beta_i \sim N(0; 1000)$, $\mu_i \sim N(0; 1000)$, $\sigma_{x_i}^2 \sim IG(2.001; 1.0)$, $\sigma_{e_i}^2 \sim IG(2.001; 1.0)$, $\sigma_i^2 \sim IG(2.001; 1.0)$ e

$$p(k) = \begin{cases} P(K = k) = \frac{1}{\#\mathcal{K}}, & k \in \mathcal{K} = \{1, \dots, n-1\}, \\ 0, & \text{caso contrário.} \end{cases}$$

2. DADOS

Os dados foram simulados a partir do modelo em (4.7.2) com mudança em todos os parâmetros onde $\alpha_1=2.0$, $\beta_1=1.0$, $\mu_1=2.0$, $\sigma_{x(1)}^2=0.6$, $\sigma_{e_1}^2=0.1$, $\sigma_1^2=0.1$, $\alpha_2=1.0$, $\beta_2=2.0$, $\mu_2=4.0$, $\sigma_{x(2)}^2=1.2$, $\sigma_2^2=0.2$, $\sigma_{e_2}^2=0.2$, $k = 20$. Usando a mesma semente da seção anterior, o conjunto de dados é o mesmo que o da Tabela 4.1.

OBSERVAÇÃO: Nesse estudo, como os dados utilizados foram simulados a partir do modelo em (4.7.2) e a distribuição à priori escolhida acima é praticamente não informativa, devemos esperar que as estimativas e conclusões à posteriori sejam compatíveis com esse modelo (4.7.2).

3. AJUSTE DO MODELO

Os algoritmos de Gibbs e “Modified and Collapsed Gibbs” foram utilizados com o objetivo de amostrar da distribuição à posteriori do modelo. De cada algoritmo foram geradas $M = 4$ seqüências paralelas de comprimento $T = 10.000$ partindo de pontos iniciais obtidos pelo procedimento descrito a seguir:

Inicialmente para cada valor fixado de k ($3 \leq k \leq n-3$) foram fixados 40 valores de η , onde η é tal que $\sigma_{e_i}^2 = \eta\sigma_i^2$, $i = 1, 2$, ($\eta = \frac{1}{20}, \dots, \frac{1}{3}, \frac{1}{2}, 1, 2, 3, \dots, 20$). Para cada uma dessas possíveis combinações foram obtidas (analiticamente) as estimativas de máximaverossimilhança $\hat{\theta}_i$ de $\theta_i = (\alpha_i, \beta_i, \mu_i, \sigma_{x_i}^2, \sigma_i^2, \sigma_{e_i}^2)$, para $i = 1, 2$.

A seguir foi aplicado o algoritmo de otimização BFGS tomando-se cada um dos pontos $\begin{pmatrix} k \\ \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix}$ acima como ponto inicial do processo iterativo, tendo sido localizados $n-5$ pontos estacionários da distribuição à posteriori.

O próximo passo foi gerar 1000 pontos do tipo $\hat{\theta} = \begin{pmatrix} k \\ \theta_1 \\ \theta_2 \end{pmatrix}$ da distribuição formada pela mistura de distribuições $t_4(\hat{\theta}_i^*, \hat{V}_i^*)$, $i = 1, \dots, n-5$, onde $\hat{\theta}_i^*$ é o i -ésimo ponto estacionário e $\hat{V}_i^* = - \left(\frac{\partial^2 \log p(\theta | D_{\text{obs}})}{\partial \theta \partial \theta^T} \right)^{-1} \Big|_{\theta = \hat{\theta}_i^*}$, onde os pesos w_i de cada distribuição são dados por $w_i = q(\hat{\theta}_i^* | D_{\text{obs}}) |\hat{V}_i^*|^{\frac{1}{2}}$, onde $q(\hat{\theta}_i^* | D_{\text{obs}})$ é a posteriori não normalizada calculada no ponto $\hat{\theta}_i^*$.

Finalmente, (utilizando-se re-amostragem ponderada) foram reamostrados 10 pontos dos quais apenas os seguintes quatro foram usados para o algoritmo de Gibbs:

$$\hat{\theta}_1^{(0)} = (21, 2.26, 0.85, 2.38, 0.37, 0.19, 0.214, 0.10, 2.37, 3.66, 1.54, 0.25, 0.25)^T$$

$$\hat{\theta}_2^{(0)} = (21, 0.92, 1.21, 2.54, 0.37, 0.23, 0.19, 0.98, 2.11, 3.35, 0.62, 0.36, 0.17)^T$$

$$\hat{\theta}_3^{(0)} = (21, 2.13, 0.92, 2.07, 0.41, 0.25, 0.19, 0.64, 2.02, 4.04, 1.33, 0.28, 0.22)^T$$

$$\hat{\theta}_4^{(0)} = (21, 0.35, 1.63, 2.31, -0.10, 0.46, 0.11, -2.08, 2.69, 4.33, 0.49, 0.24, 0.35)^T$$

A seguir são apresentados alguns dispositivos gráficos para o diagnóstico de possíveis falhas na convergência dos algoritmos de Gibbs e Modified Collapsed Gibbs.

3.a Diagnósticos de convergência

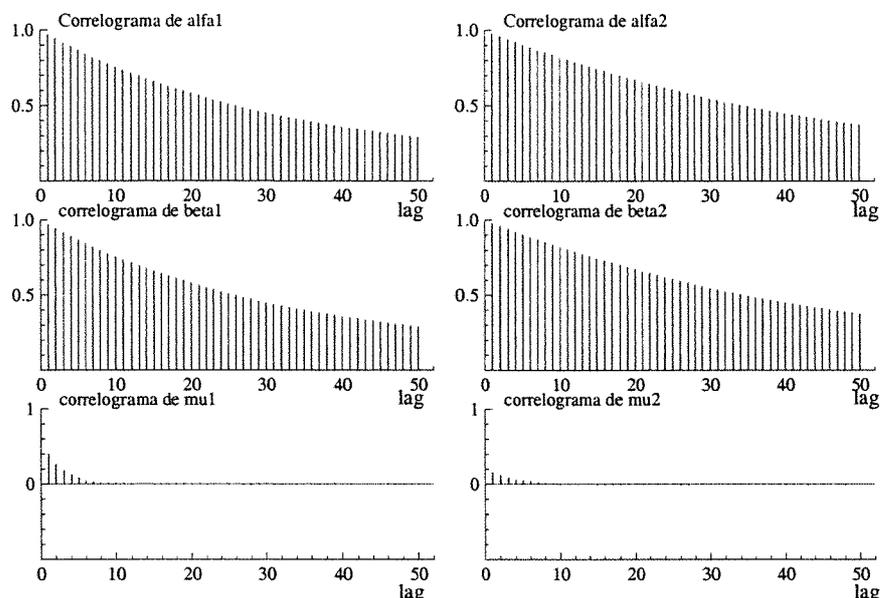


Figura 4.13: Gráficos das funções de auto-correlação de α_1 , $\alpha_2, \beta_1, \beta_2, \mu_1$ e μ_2 , baseados em uma única seqüência de Gibbs de comprimento $T = 50.000$, partindo do ponto inicial $\hat{\theta}_1^{(0)}$.

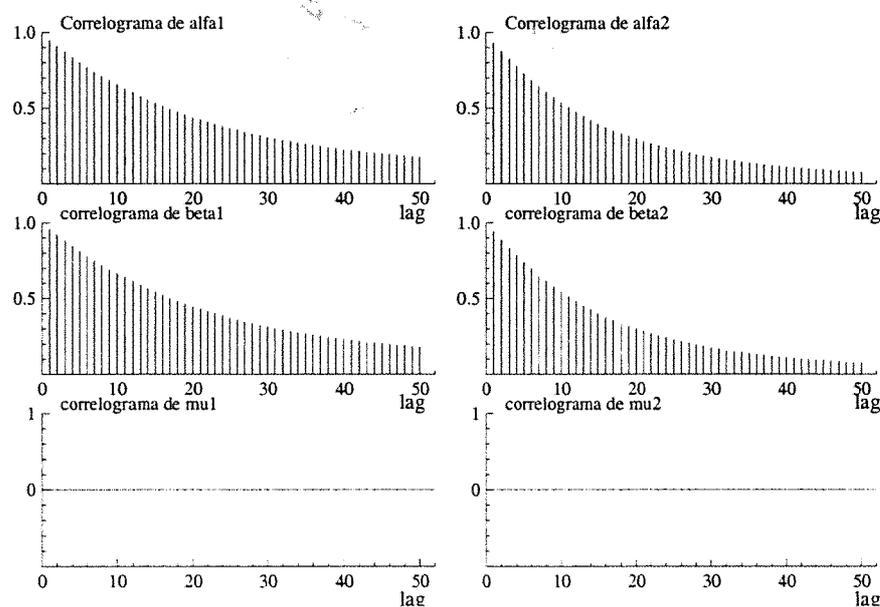


Figura 4.14: Gráficos das funções de auto-correlação de $\alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1$ e μ_2 , baseados em uma única seqüência do algoritmo M.Collap. Gibbs de comprimento $T = 50.000$.

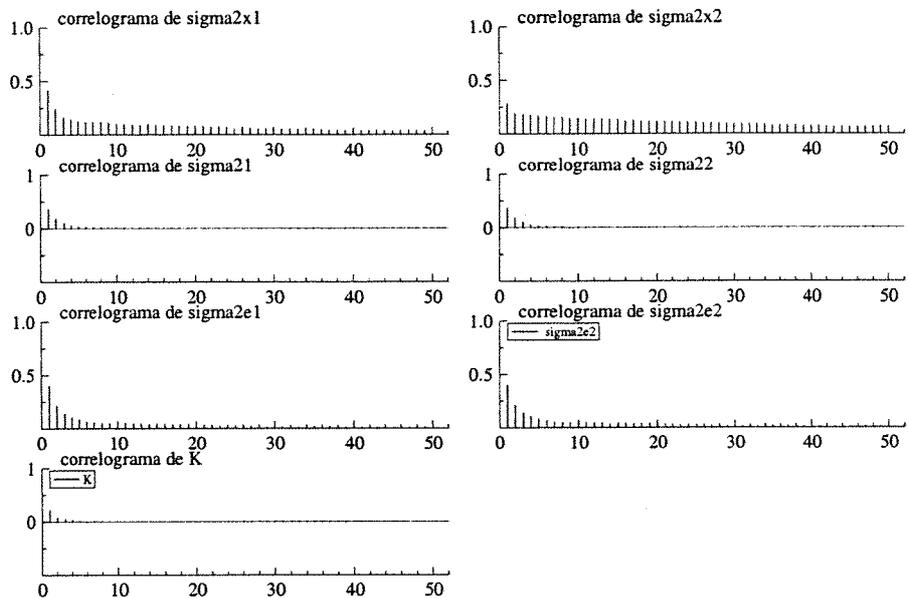


Figura 4.15: Gráficos das funções de auto-correlação de $\sigma_{x_1}^2, \sigma_{x_2}^2, \sigma_{e_1}^2, \sigma_{e_2}^2, \sigma_1^2, \sigma_2^2$ e k baseados em uma seqüência de Gibbs de comprimento $T=50.000$, partindo de $\hat{\theta}_1^{(0)}$.

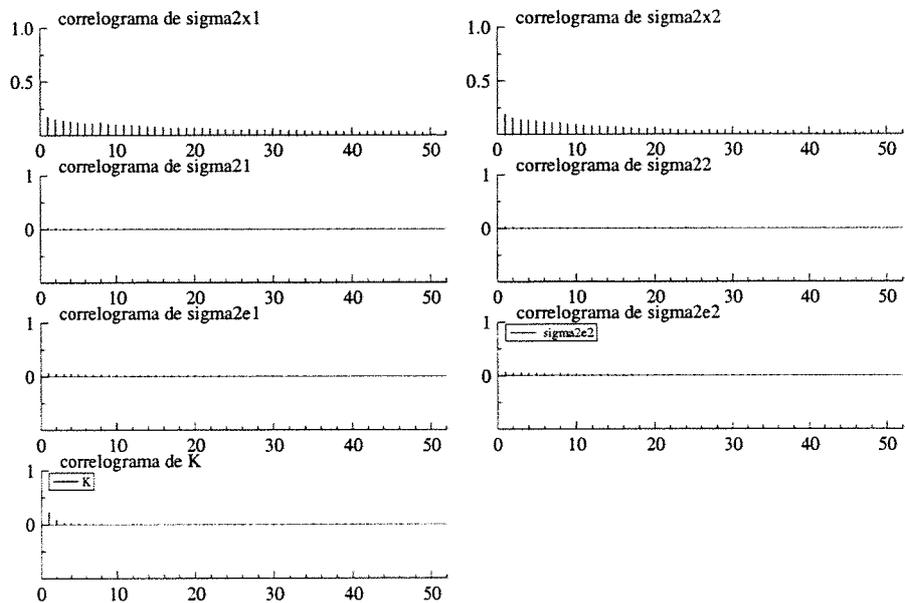


Figura 4.16: Gráficos das funções de auto-correlação de $\sigma_{x_1}^2, \sigma_{x_2}^2, \sigma_{e_1}^2, \sigma_{e_2}^2, \sigma_1^2, \sigma_2^2$ e k , baseados em uma seqüência do algoritmo M.Collap. Gibbs de comprimento $T=50000$.

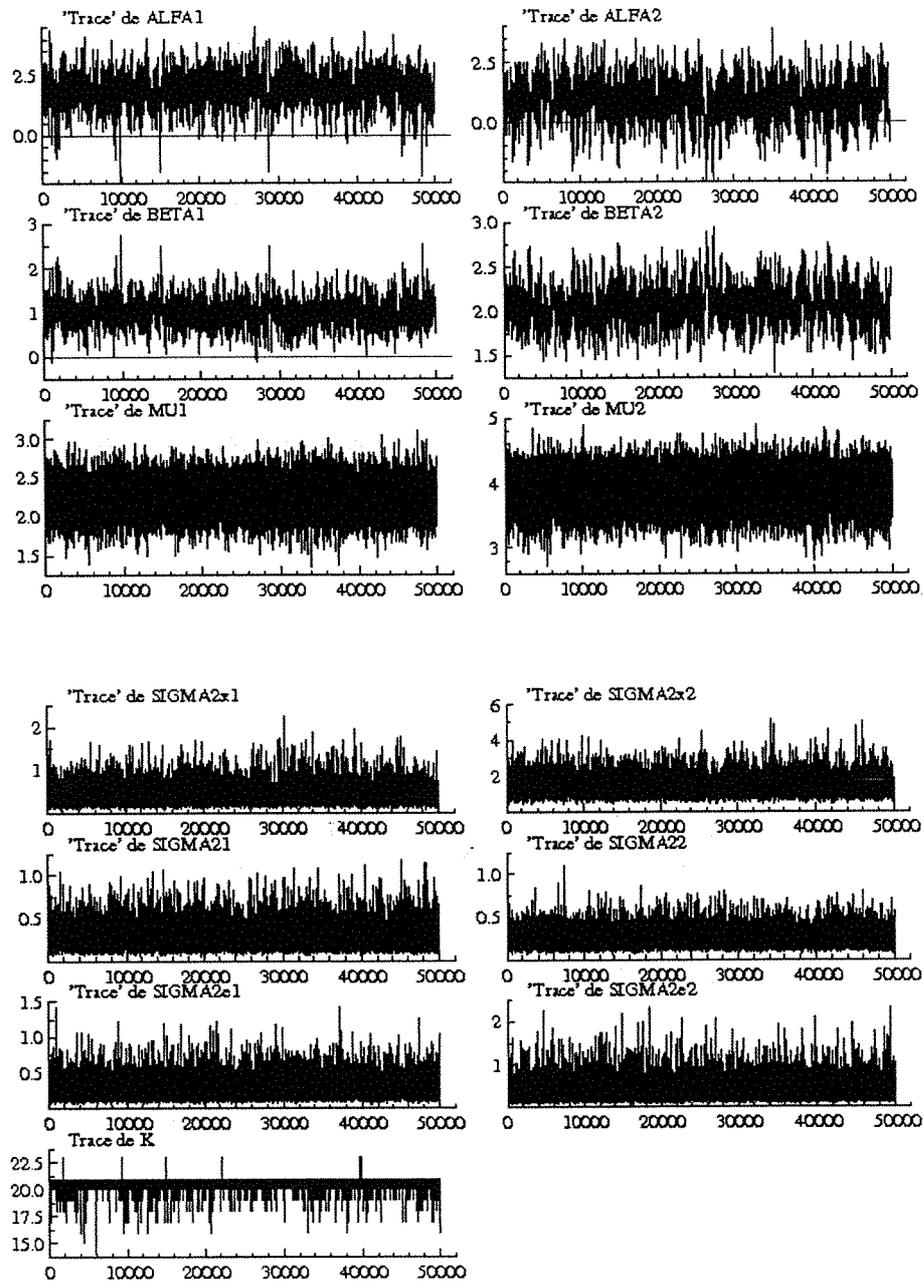


Figura 4.17: Gráficos das séries temporais de $\alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x_1}^2, \sigma_{x_2}^2, \sigma_{e_1}^2, \sigma_{e_2}^2, \sigma_1^2, \sigma_2^2$ e k baseados em uma sequência de comprimento $T = 50.000$ do algoritmo de Gibbs partindo de $\hat{\theta}_1^{(0)}$.

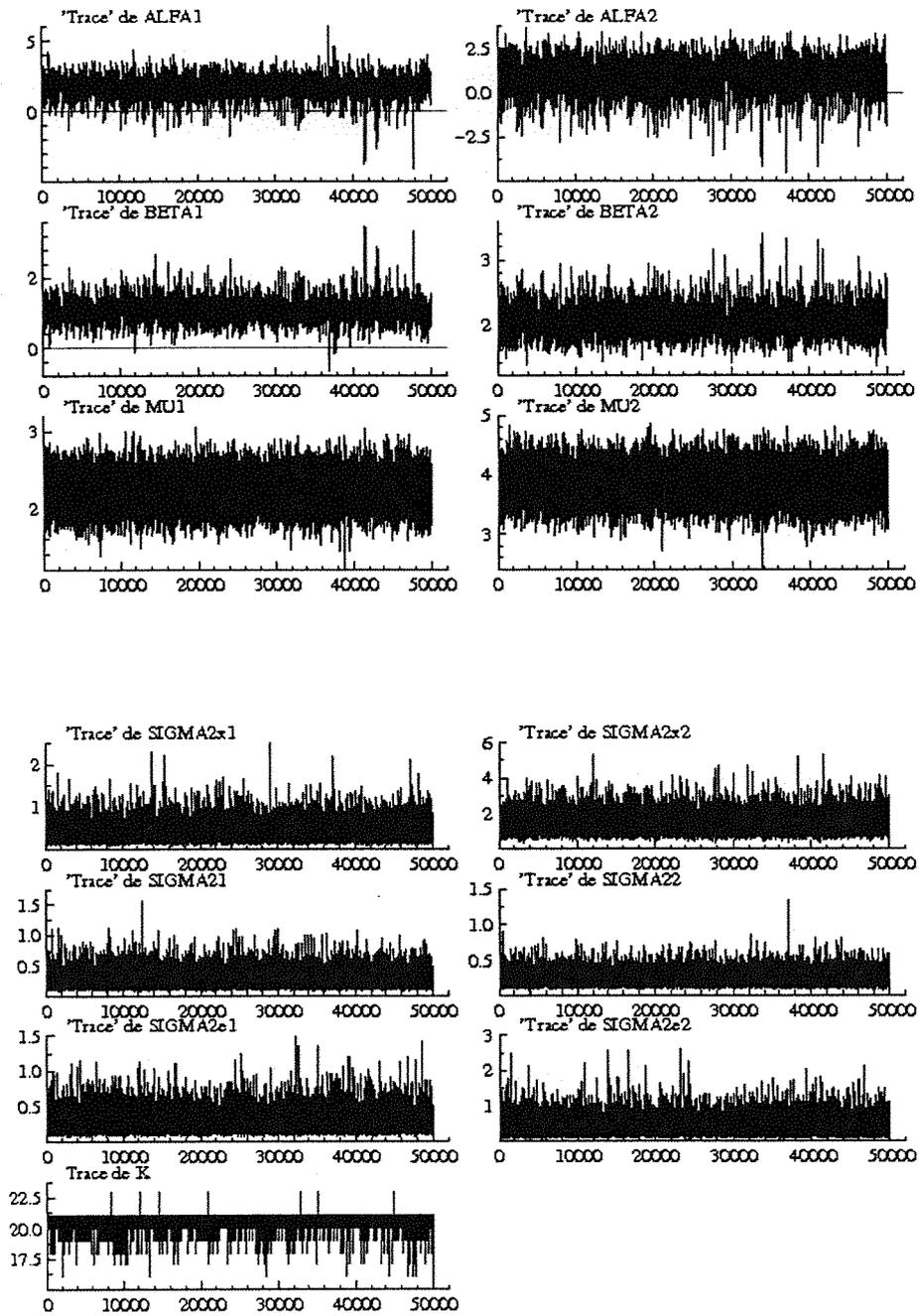


Figura 4.18: Gráficos das séries temporais de $\sigma_{x_1}^2, \sigma_{x_2}^2, \sigma_{e_1}^2, \sigma_{e_2}^2, \sigma_1^2, \sigma_2^2$ e k baseados em uma sequência de comprimento $T = 50.000$ do algoritmo Modified Collapsed Gibbs.

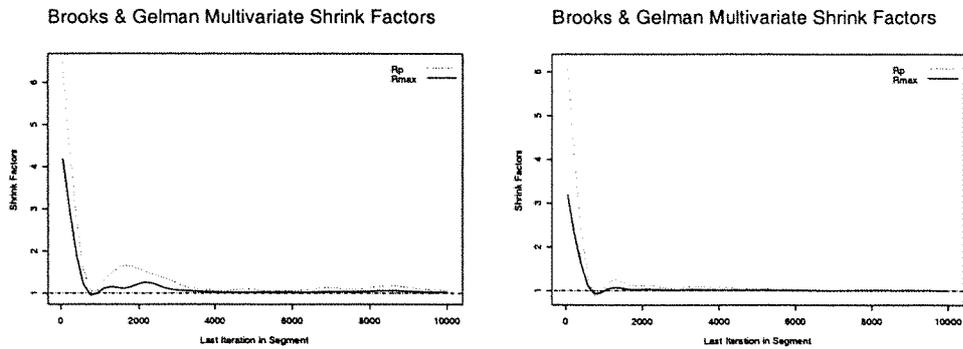


Figura 4.19: “Multiv. Potential Scale Reduction” de Brooks e Gelman (1998). O gráfico à esquerda é baseado em $M = 4$ seqüências de comprimento $T = 10.000$ do algoritmo de Gibbs e o da direita no algoritmo M.Collap.Gibbs.

Gelman & Rubin Shrink Factors

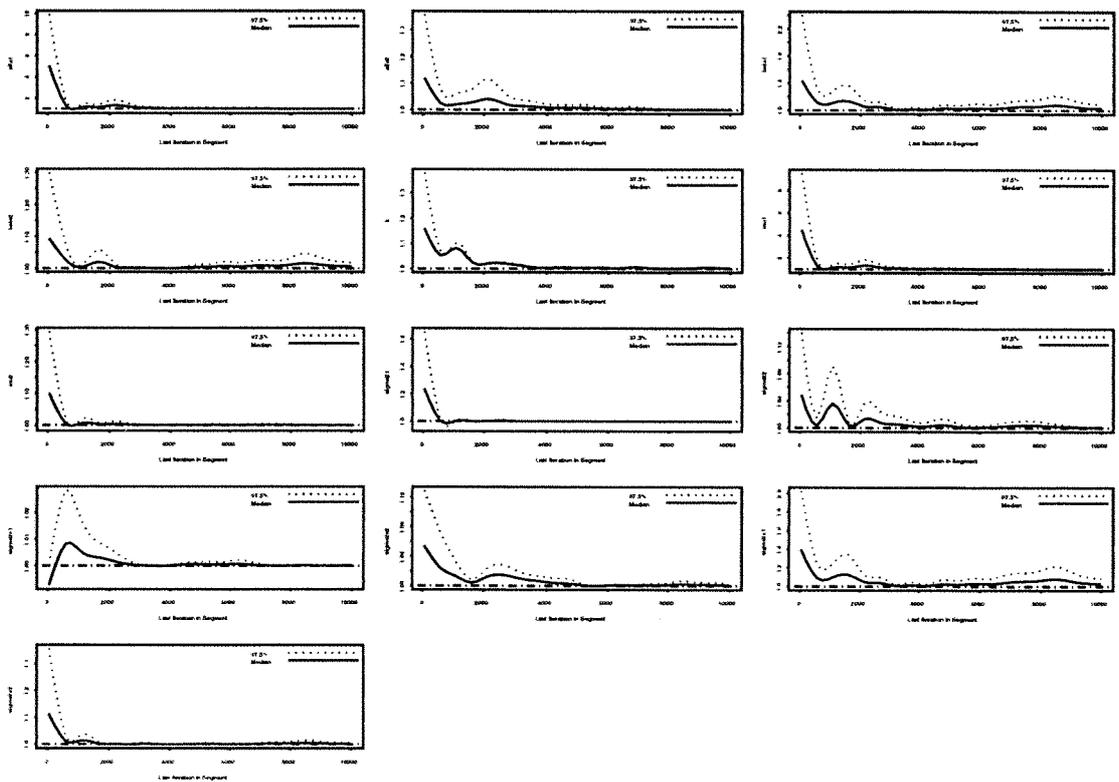


Figura 4.20: Gráfico da medida de Gelman e Rubin corrigida “Corrected potential scale reduction factors” (CSRF) para cada um dos parametros $k, \alpha_1, \alpha_2, \beta_1, \beta_2, \dots, \sigma_1^2, \sigma_2^2$, baseada em $M = 4$ seqüências de comprimento $T = 10.000$ de Gibbs.

Gelman & Rubin Shrink Factors

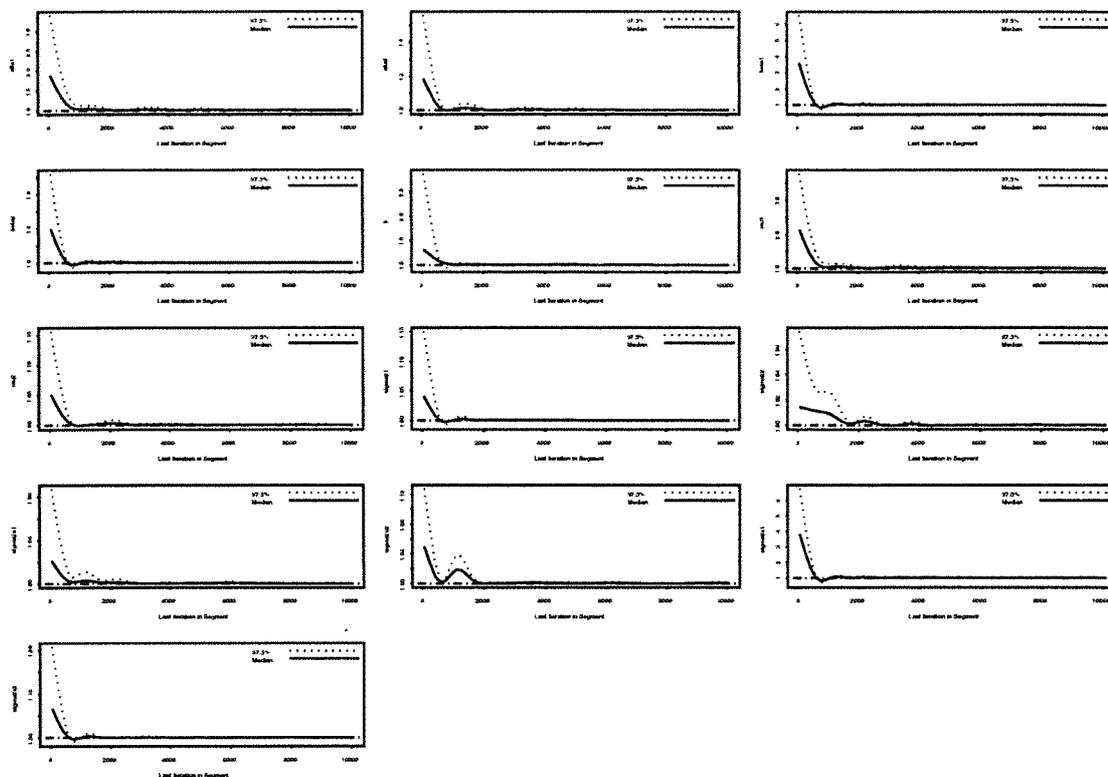


Figura 4.21: Gráfico da medida de Gelman e Rubin corrigida “Corrected potencial scale reduction factors” (CSRF) para cada um dos parâmetros $k, \alpha_1, \alpha_2, \beta_1, \dots, \sigma_1^2, \sigma_2^2$, baseada em $M = 4$ seqüências de comprimento $T = 10.000$ do algoritmo M.Collapsed Gibbs.

As figuras 4.13 a 4.21 não sugerem que exista problemas na convergência dos algoritmos de Gibbs e Modified Collapsed Gibbs. Além disso o algoritmo Modified Collapsed Gibbs parece ter uma melhor performance que o Gibbs (como pode ser visto principalmente pela comparação das funções de auto-correlação amostral correspondentes).

3.b Estimação

Semelhantemente à seção 4.7.1, e com o objetivo de examinar se a estimativa do ponto de mudança k , do modelo em (4.7.2), (obtida por meio do algoritmo M.

Collapsed Gibbs), difere muito do seu valor verdadeiro, apresentamos também nessa seção uma tabela com diferentes valores de k fixados e suas correspondentes estimativas \hat{k} obtidas por esse algoritmo partindo-se sempre do ponto inicial $(k^{(0)}, \theta_1^{(0)}, \theta_2^{(0)})$, onde $k^{(0)} = 25$ e $\theta_i^{(0)} = (\beta_i^{(0)}, \sigma_{x_i}^{2(0)}, \sigma_{e_i}^{2(0)}, \sigma_i^{2(0)}) = (1, 1, 1, 1)$, $i=1,2$, utilizando-se apenas a segunda metade do total de 10.000 iterações.

(Observemos que cada linha da tabela envolveu um conjunto de dados diferente, simulado a partir do modelo (4.7.2) com um valor de k entre 1 e $n - 1$ e com $\theta_1=(2.0, 1.0, 2.0, 0.6, 0.1, 0.1)$ e $\theta_2=(1.0, 2.0, 4.0, 1.2, 0.2, 0.2)$).

Tabela 4.5: Valores verdadeiros de k e correspondentes valores estimados(\hat{k}).

k	\hat{k}								
1	1.11	11	10.92	21	20.99	31	30.93	41	40.94
2	2.67	12	11.97	22	22.80	32	31.95	42	41.76
3	4.13	13	12.80	23	22.92	33	32.97	43	42.90
4	16.37	14	14.93	24	23.79	34	33.91	44	43.45
5	4.52	15	14.67	25	25.39	35	34.93	45	44.82
6	5.83	16	15.95	26	25.93	36	35.46	46	38.17
7	6.37	17	16.85	27	26.98	37	36.86	47	47.99
8	8.76	18	17.91	28	27.77	38	37.95	48	47.93
9	8.92	19	18.94	29	28.71	39	38.95	49	48.78
10	10.00	20	20.93	30	29.90	40	39.56	50	

A seguir apresentamos as estimativas de várias medidas resumo da distribuição à posteriori bem como as estimativas das distribuições posterioris marginais. Essas estimativas basearam-se em $M=4$ seqüências do algoritmo M. Collapsed Gibbs de comprimento $T = 10.000$, após descartar-se a primeira metade de cada uma. Os pontos iniciais dessas seqüências foram obtidos a partir dos pontos $\theta_1^{(0)}$ a $\theta_4^{(0)}$ (dados no início da seção) após descartar-se as componentes $\alpha_1^{(0)}, \alpha_2^{(0)}, \mu_1^{(0)}, \mu_2^{(0)}$ de cada um.

A tabela 4.6 apresenta a distribuição de freqüências à posteriori de k e a tabela 4.7 exhibe as estimativas de várias medidas resumo da distribuição à posteriori.

Tabela 4.6: Distribuição de freqüência relativa à posteriori de k :

k	16	17	18	19	20	21	23	outros
$\hat{P}(K = k Dobs)$	0.0001	0.0006	0.0016	0.0059	0.0504	0.9414	0.0002	0.0000
								1

Tabela 4.7: Estimativas do valor esperado, quantis, desvio padrão, máximo, mínimo e intervalos de credibilidade de comprimento mínimo das distribuições marginais à posteriori de $\alpha_1, \alpha_2, \dots, \sigma_{e_1}^2, \sigma_{e_2}^2, \sigma_1^2, \sigma_2^2$, usando o algoritmo M.Collap.Gibbs. (Os valores verdadeiros dos parâmetros são: $\alpha_1=2.0, \beta_1=1.0, \mu_1=2.0, \sigma_{x(1)}^2=0.6, \sigma_1^2=0.1, \sigma_{e_1}^2=0.1, \alpha_2=1.0, \beta_2=2.0, \mu_2=4.0, \sigma_{x(2)}^2=1.2, \sigma_2^2=0.2, \sigma_{e_2}^2=0.2, k=20$.)

ESTIMATIVAS	α_1	α_2	β_1	β_2	μ_1	μ_2	
Média aritm.	1.8264	0.89619	1.0444	2.0602	2.2337	3.8574	
Média R-Blackw.	1.8267	0.89714	1.0442	2.0601	2.2358	3.8569	
Quantis							
0.025	0.4356	-0.9168	0.5101	1.7019	1.8835	3.3929	
.05	0.6979	-0.5635	0.6044	1.7538	1.9414	3.4698	
.25	1.4138	0.4030	0.8546	1.9184	2.1152	3.7016	
.5	1.8574	0.9589	1.0327	2.0438	2.2329	3.8597	
.75	2.2632	1.4670	1.2229	2.1828	2.3523	4.0117	
.95	2.8488	2.1315	1.5333	2.4200	2.5262	4.2366	
.975	3.0496	2.3402	1.6495	2.5132	2.5859	4.3145	
Int.Credib. HPD							
$p = 0.95$ Lim. inf.	0.5317	-0.7548	0.4852	1.6697	1.8754	3.4102	
Lim sup.	3.1392	2.4604	1.6174	2.4727	2.5751	4.3267	
$p = 0.90$ Lim. inf.	0.7635	-0.3577	0.5839	1.7201	1.9420	3.4695	
Lim sup.	2.8982	2.2916	1.5090	2.3774	2.5265	4.2357	
desv.padrão (dp)	0.6576	0.8348	.2848	0.2079	0.1791	0.2349	
dp R-Blackweliz.	0.6585	0.8340	0.2857	0.8340	0.1802	0.2356	
“Naive SE”	0.0046	0.0059	0.0020	0.0015	0.0013	0.0017	
“Batch SE”	0.0302	0.0034	0.0351	0.0082	0.0130	0.0014	
Mínimo	-1.3483	-4.5125	-0.6558	1.4031	1.2050	2.7809	
Máximo	5.9849	3.4845	2.4486	3.3271	3.0601	4.8526	
ESTIMATIVAS	$\sigma_{x_1}^2$	$\sigma_{x_2}^2$	σ_1^2	σ_2^2	$\sigma_{e_1}^2$	$\sigma_{e_2}^2$	k
Média aritm.	0.41292	1.3674	0.26525	0.24144	0.27657	0.39666	20.93
Média R-Blackw.	0.41219	1.3654	0.26597	0.24224	0.27639	0.39562	20.93
Quantis							
0.025	0.1757	0.7274	0.1228	0.1237	0.1237	0.1477	20
.05	0.1995	0.8011	0.1369	0.1358	0.1385	0.1671	20
.25	0.2924	1.0657	0.1924	0.1848	0.1963	0.2574	21
.5	0.3798	1.2989	0.2459	0.2285	0.2537	0.3540	21
.75	0.4977	1.5906	0.3140	0.2834	0.3305	0.4892	21
.95	0.7361	2.1655	0.4600	0.3900	0.4878	0.7697	21
.975	0.8338	2.4072	0.5246	0.4342	0.5609	0.8942	21
Int.Credib. HPD							
$p = 0.95$ Lim. inf.	0.1409	0.6280	0.0978	0.1099	0.1044	0.1085	20
Lim sup.	0.7509	2.2114	0.4671	0.4032	0.5013	0.7796	21
$p = 0.90$ Lim. inf.	0.1625	0.7182	0.1078	0.1222	0.1154	0.1204	21
Lim sup.	0.6584	2.0073	0.4056	0.3613	0.4375	0.6592	21
desv.padrão (dp)	0.1732	0.4303	0.1045	0.0802	0.1149	0.1953	0.3095
dp R-Blackweliz.	0.1739	0.4254	0.1057	0.0810	0.1146	0.1950	0.3104
“Naive SE”	0.0012	0.0030	0.0007	0.0006	0.0008	0.0014	0.0022
“Batch SE”	0.0088	0.0013	0.0025	0.0015	0.0024	0.0039	0.0029
Mínimo	0.0799	0.3990	0.0663	0.06571	0.0689	0.0777	16
Máximo	2.4958	5.1758	1.1225	1.3458	1.3485	2.5940	23

NOTA: Os intervalos de credibilidade nesse exemplo coincidem com os intervalos HPD.

A Figura 4.22 apresenta os gráficos das estimativas “Rao-Blackwelizadas” das distribuições posterioris marginais de $\alpha_1, \alpha_2, \beta_1, \beta_2, \dots, \sigma_1^2, \sigma_2^2$ e k .

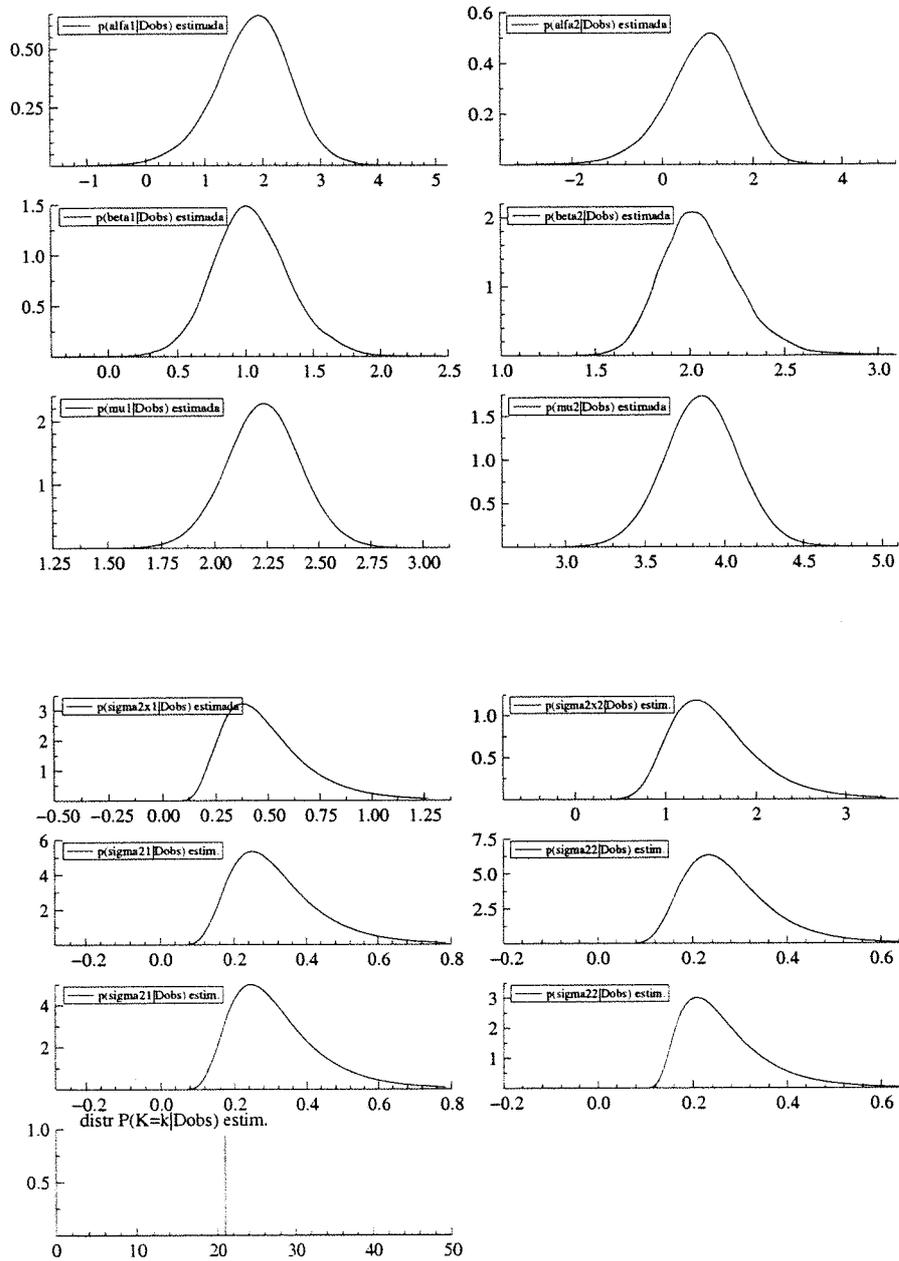


Figura 4.22: Gráficos das estimativas “Rao-Blackwelizadas” das densidades marginais à posteriori $p(\alpha_1|D_{obs}), p(\alpha_2|D_{obs}), \dots, p(\sigma_1^2|D_{obs}), p(\sigma_2^2|D_{obs})$ e $p(k|D_{obs})$.

Observações:

Na tabela 4.5 a única estimativa discrepante ocorre quando o verdadeiro valor de k é 4. Semelhantemente a seção anterior, observou-se também que se o valor verdadeiro de k é bem pequeno, então o vetor de parâmetros θ_1 (onde $\theta_1 = (\alpha_1, \beta_1, \mu_1, \sigma_{x_1}^2, \sigma_1^2, \sigma_{e_1}^2)^T$) tende a não ser bem estimado, e se k é grande (próximo de n), então θ_2 é que tende a ter estimativas pobres. Esse fato já era esperado pela própria estrutura do problema.

Na tabela 4.7 vemos que tanto a média como a mediana superestimam os valores dos parâmetros σ_1^2 , $\sigma_{e_1}^2$ e $\sigma_{e_2}^2$, entretanto o valor do viés estimado é pequeno para os demais parâmetros. Mais importante ainda é que todos os intervalos estimados de credibilidade de 95% contém os verdadeiros valores dos parâmetros (no caso de $\sigma_{e_1}^2$ isso é verdade se considerarmos apenas 2 casas decimais).

É interessante observar que no caso particular em que o conjunto de dados é o da tabela 4.1, os comprimentos dos intervalos de credibilidade no modelo sem restrição de identificabilidade (com priori dada em (4.7.2)) tendem a ser ligeiramente maiores do que no modelo com restrição de identificabilidade (com priori dada em (4.7.1)), entretanto em ambos os modelos esses intervalos estimados contiveram os verdadeiros valores dos parâmetros.

3.c Avaliação do ajuste

A figura 4.23 apresenta os gráficos dos valores da densidade condicional preditiva de validação cruzada CPO_i (onde $CPO_i = \pi\left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{obs}\right)$) primeiramente em função de i , depois em função de X_i , e também em função de Y_i , para $i = 1, \dots, 50$ (onde CPO_i é estimada por meio da fórmula (4.5.8) adaptada para o caso do modelo sem restrições de identificabilidade). As estimativas basearam-se na segunda metade de uma seqüência de comprimento $T=10.000$ do algoritmo M.Collap.Gibbs cujo ponto inicial é formado pelas componentes de $\theta_1^{(0)}$, após descartar $\alpha_1^{(0)}, \alpha_2^{(0)}, \mu_1^{(0)}, \mu_2^{(0)}$.

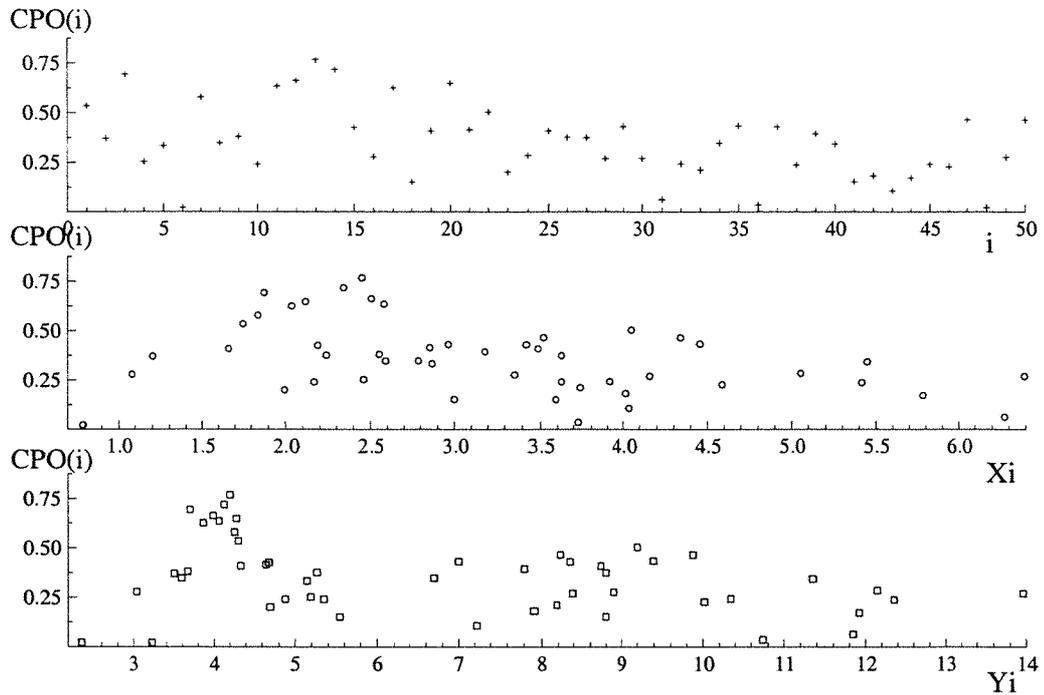


Figura 4.23: Gráficos dos valores da densidade condicional preditiva, $CPO(i)$, versus i (posição superior), versus X_i (posição do meio) e versus Y_i (posição inferior).

A figura 4.24 a seguir apresenta os gráficos das componentes d_{1i} e d_{2i} do vetor de resíduos bayesianos bidimensionais de validação cruzada

$$d_i = \begin{pmatrix} d_{1i} \\ d_{2i} \end{pmatrix} = \widehat{\text{Var}}^{-\frac{1}{2}} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{\text{obs}(-i)} \right) \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} - \hat{E} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{\text{obs}(-i)} \right) \right),$$

$i = 1, \dots, n$, dado na seção 4.5.3, onde as fórmulas de $\hat{E} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{\text{obs}(-i)} \right)$ e de $\widehat{\text{Var}}^{-\frac{1}{2}} \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix} | D_{\text{obs}(-i)} \right)$ foram adaptadas para o modelo sem restrição de identificabilidade.

As figuras 4.23 e 4.24 não sugerem que exista problema de falta de ajustamento. Entretanto nos gráficos de ordenada d_{1i} , da figura 4.24 existem 2 pontos cujos valores são maiores do que 3 (em valor absoluto) e nos gráficos de ordenada d_{2i} , da mesma figura, existem 4 pontos nessas condições. Esses pontos aparentemente poderiam

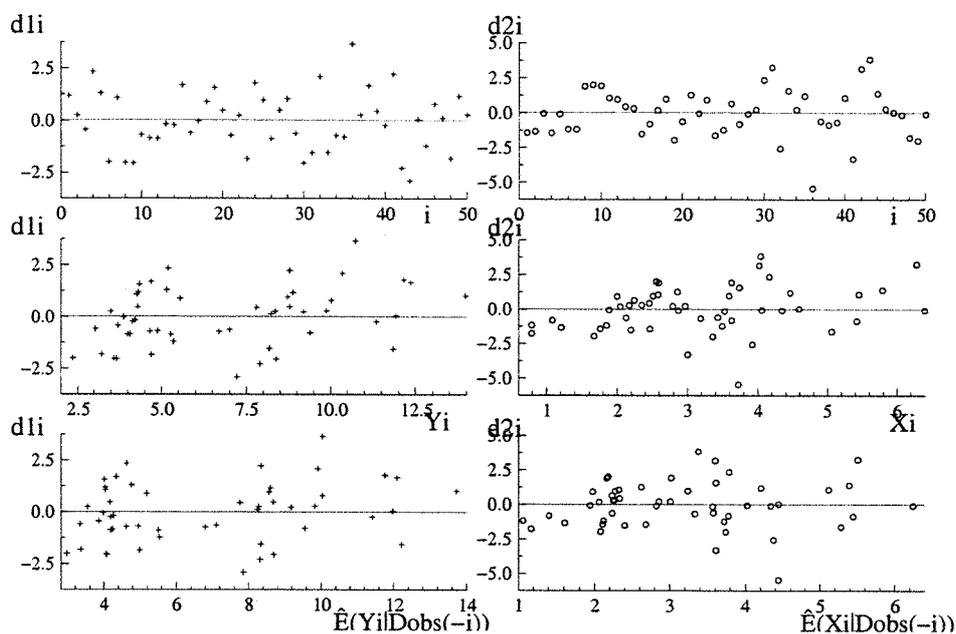


Figura 4.24: Os gráficos do lado esquerdo referem-se à primeira componente dos resíduos (d_{i1}), e os do lado direito à segunda componente (d_{i2}).

ser vistos como aberrantes, entretanto uma análise de resíduos mais aprofundada deve ser conduzida futuramente para examinar melhor a qualidade do ajuste.

Quando o conjunto de dados observados é o da tabela 4.1, a qualidade do ajuste aparentemente não difere muito entre os modelos com e sem restrição de identificabilidade (com suas respectivas prioris dadas em (4.7.1) e (4.7.2)).

Levando em conta também que as estimativas dos parâmetros não diferem muito nos dois modelos, o modelo sem restrição aparentemente é mais vantajoso já que ele não requer o conhecimento do valor de lambda (o quociente entre as variâncias residuais).

4.7.3 Análise do procedimento 1 para detetar mudança baseada na distribuição à posteriori de k .

1. MODELO

Nessa seção utilizamos o modelo com restrição de identificabilidade que permite mudança em todos os parâmetros e onde k pode assumir os valores de 1 a n , definido na seção 4.4.1.

A distribuição à priori escolhida é

$$\pi(\underline{\theta}) = \begin{cases} p(k)\pi(\underline{\theta}_1), & \text{se } k = n, \\ p(k)\pi(\underline{\theta}_1)\pi(\underline{\theta}_2), & \text{se } k \in \mathcal{K} = \{1, \dots, n-1\}, \\ 0, & \text{caso contrário,} \end{cases} \quad (4.7.3)$$

onde $\pi(\underline{\theta}_i) = \pi(\alpha_i)\pi(\beta_i)\pi(\mu_i)\pi(\sigma_{x(i)}^2)\pi(\sigma_i^2)$, $i = 1, 2$ e

$$p(k) = \begin{cases} c, & \text{se } k = n, \\ (1-c)\frac{1}{\#\mathcal{K}}, & \text{se } k \in \mathcal{K}, \\ 0, & \text{caso contrário,} \end{cases} \quad (4.7.4)$$

onde $c = P(K = n) = \frac{1}{n}$, $\alpha_i \sim N(0; 1000)$, $\beta_i \sim N(0; 1000)$, $\mu_i \sim N(0; 1000)$, $\sigma_{x_i}^2 \sim IG(2.001; 1.0)$, $\sigma_{e_i}^2 \sim IG(2.001; 1.0)$, $\sigma_i^2 \sim IG(2.001; 1.0)$.

Nessa seção utilizaremos o conjunto de dados da tabela 4.1.

2. AJUSTE DO MODELO

Para amostrar da distribuição à posteriori foi implementado o algoritmo de Gibbs descrito na seção 4.4.1. Foram geradas 4 seqüências paralelas de comprimento $T = 10.000$ partindo-se dos mesmos pontos iniciais usados na seção 4.7.1.

A seguir nas figuras 4.25 a 4.28 e na tabela 4.8 apresentamos alguns gráficos e medidas de diagnóstico de convergência desse algoritmo.

Os dispositivos apresentados nessas figuras não nos sugeriram que exista algum problema na convergência do algoritmo de Gibbs.

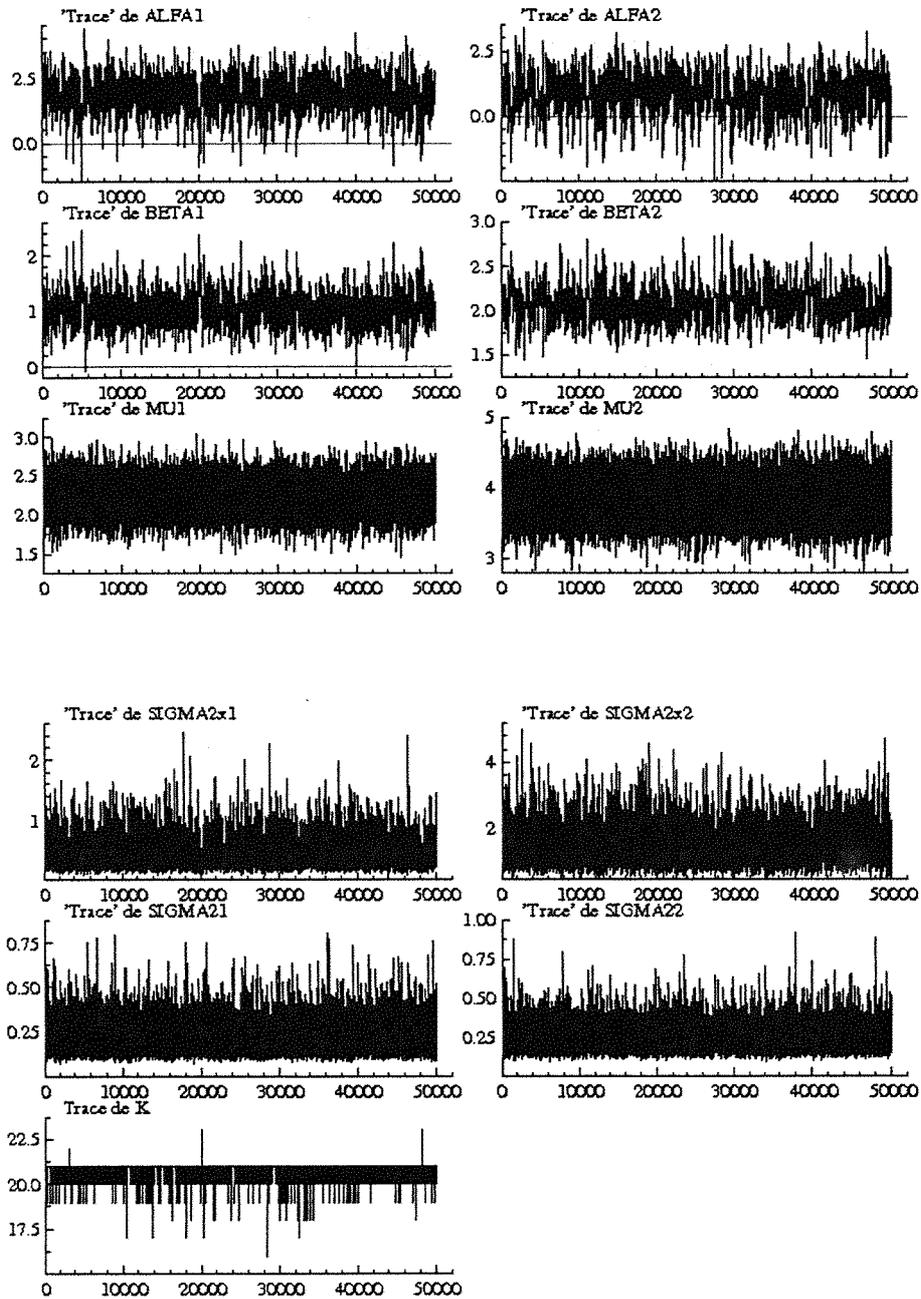


Figura 4.25: Séries temporais de $\alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x_1}^2, \sigma_{x_2}^2, \sigma_1^2, \sigma_2^2$ e k baseadas em uma sequência de comprimento $T = 50.000$ do algoritmo de Gibbs partindo do ponto $\hat{\theta}_{(0)}$.

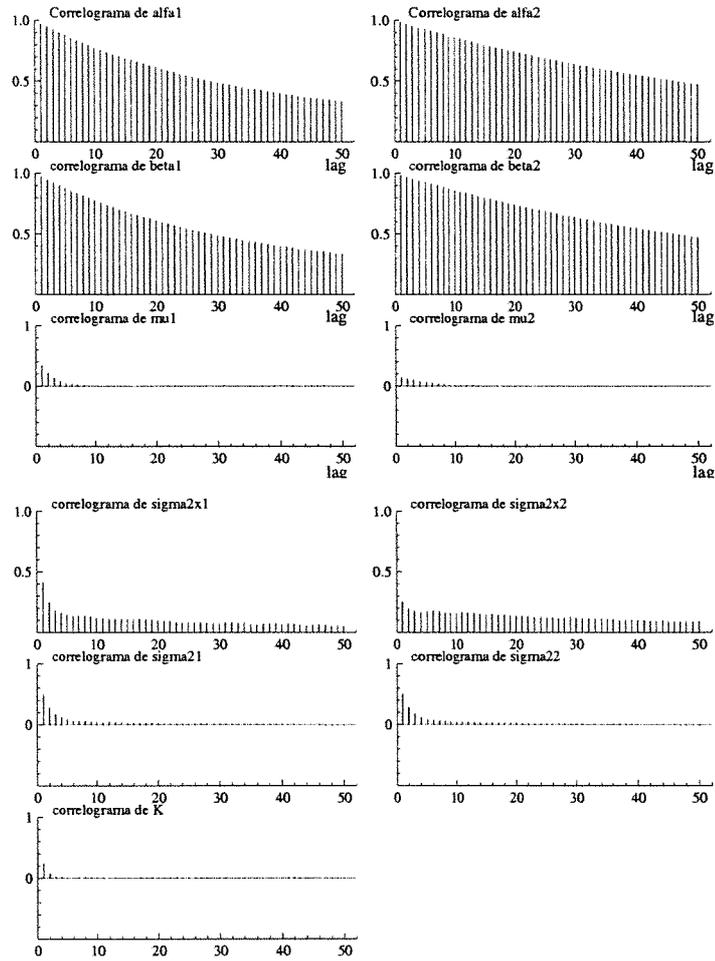


Figura 4.26: Gráficos das funções de auto-correlação de $\alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x_1}^2, \sigma_{x_2}^2, \sigma_1^2, \sigma_2^2$ e k baseados numa seqüência de comprimento $T = 50.000$ de Gibbs, partindo de $\hat{\theta}_{(0)}$.

Tabela 4.8: Medidas de Diagnóstico de convergência de Brooks, Gelman e Rubin

	PSRF	CSRF	Quantil 0.975		PSRF	CSRF	Quantil 0.975
α_1	1.003949	1.017607	1.027021	α_2	1.000187	1.000210	1.000818
β_1	1.017694	1.030607	1.071077	β_2	1.000654	1.000877	1.002472
μ_1	1.003834	1.017390	1.026530	μ_2	1.003358	1.004180	1.011505
$\sigma_{x_1}^2$	1.018132	1.030291	1.071477	$\sigma_{x_2}^2$	1.000985	1.001547	1.003849
σ_1^2	1.000147	1.000228	1.000750	σ_2^2	1.000961	1.001298	1.003543
k	1.000149	1.003085	1.003623				

A medida multivariada (MPSRF) é igual a 1.0576164.

Brooks & Gelman Multivariate Shrink Factors

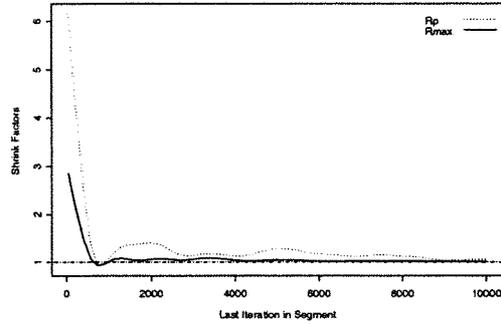


Figura 4.27: “Multiv. Pot.Scale Reduction” de Brooks e Gelman (1998) com base em 4 seg. onde $T=10.000$.

Gelman & Rubin Shrink Factors

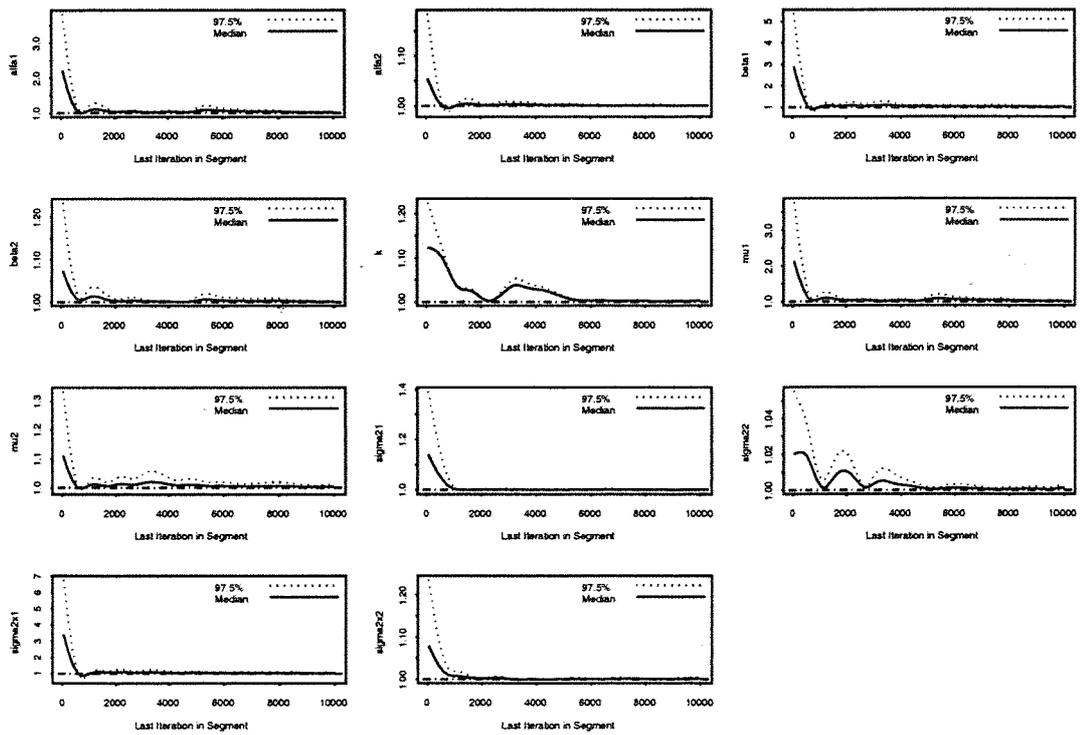


Figura 4.28: Gráfico da medida de Gelman e Rubin corrigida “Corrected potential scale reduction factors” (CSRF) para cada um dos parâmetros $k, \alpha_1, \alpha_2, \beta_1, \beta_2, \dots, \sigma_1^2, \sigma_2^2$, baseada em $M = 4$ seqüências de comprimento $T = 10.000$ de Gibbs.

2.a Estimação

A tabela 4.9 exibe para cada valor de $k = k_0$ (entre 1 e n), a respectiva estimativa (\hat{k}) baseada nos dados gerados apartir do modelo (dessa seção) onde $k = k_0$, $\vartheta_1 = (2.0, 1.0, 2.0, 0.6, 0.1)^T$ e $\vartheta_2 = (1.0, 2.0, 4.0, 1.2, 0.2)^T$ utilizando-se o algoritmo de Gibbs partindo do ponto inicial $(k^{(0)}, \vartheta_1^{(0)}, \vartheta_2^{(0)})$, onde $k^{(0)} = 1$ e $\vartheta_i^{(0)} = (\alpha_i^{(0)}, \beta_i^{(0)}, \mu_i^{(0)}, \sigma_{x_i}^{2(0)}, \sigma_i^{2(0)}) = (1, 1, 1, 1, 1)$, $i = 1, 2$.

Tabela 4.9: Valores verdadeiros de k e respectivos valores estimados(\hat{k}) baseados na segunda metade da seqüência de Gibbs correspondentemente.

k	\hat{k}								
1	12.57	11	10.95	21	20.99	31	30.96	41	40.96
2	1.93	12	11.96	22	22.83	32	31.97	42	41.85
3	50.00	13	12.87	23	22.95	33	32.98	43	42.94
4	31.18	14	14.95	24	23.88	34	33.94	44	43.71
5	4.82	15	14.72	25	25.25	35	34.95	45	44.91
6	5.89	16	15.97	26	25.95	36	35.71	46	45.32
7	6.72	17	16.93	27	26.98	37	36.92	47	49.73
8	8.87	18	17.95	28	27.92	38	37.96	48	50.00
9	8.99	19	18.96	29	28.73	39	38.97	49	50.00
10	10.00	20	20.97	30	29.93	40	39.68	50	50.00

Na tabela 4.9 estimativas extremamente discrepantes ocorreram para $k=1$, $k=3$ e $k=4$. Observou-se também que para valores de k bem pequenos, as estimativas de cada uma das componentes de ϑ_1 , onde $\vartheta_1 = (\alpha_1, \beta_1, \mu_1, \sigma_{x_1}^2, \sigma_1^2, \sigma_{e_1}^2)^T$, são muito viesadas. Também se k é grande (próximo de n), as componentes de $\vartheta_2 = (\alpha_2, \beta_2, \mu_2, \sigma_{x_1}^2, \sigma_1^2, \sigma_{e_2}^2)^T$ tendem a ter um grande viés.

A tabela 4.10 apresenta as estimativas de várias medidas resumo da distribuição à posteriori e a tabela 4.11, a distribuição de freqüências à posteriori de k . A figura 4.24 apresenta os gráficos das estimativas “Rao-blackwelizadas” das distribuições marginais à posteriori. Essas estimativas baseiam-se em 4 seqüências de Gibbs de comprimento $T = 10.000$ que partem respectivamente dos pontos iniciais $\check{\vartheta}_1^{(0)}$, $\check{\vartheta}_2^{(0)}$, $\check{\vartheta}_3^{(0)}$ e $\check{\vartheta}_4^{(0)}$, após descartar-se a primeira metade de cada uma.

Tabela 4.10: Estimativas do valor esperado, quantis, desvio padrão, máximo, mínimo e intervalos de credibilidade de comprimento mínimo das distribuições marginais à posteriori de $\alpha_1, \alpha_2, \dots, \sigma_1^2, \sigma_2^2$, usando o algoritmo de Gibbs. (Os valores verdadeiros dos parâmetros são: $\alpha_1=2.0, \beta_1=1.0, \mu_1=2.0, \sigma_{x(1)}^2=0.6, \sigma_1^2=0.1, \alpha_2=1.0, \beta_2=2.0, \mu_2=4.0, \sigma_{x(2)}^2=1.2, \sigma_2^2=0.2, k = 20$.)

ESTIMATIVAS	α_1	α_2	β_1	β_2	μ_1	μ_2
Média aritm.	1.8140	0.8539	1.0500	2.0720	2.2357	3.8581
Média R-Blackw.	1.8147	0.8540	1.0504	2.0721	2.2357	3.8587
Quantis						
0.025	0.4568	-0.9560	0.5845	1.7635	1.8915	3.3909
.05	0.7484	-0.5399	0.6601	1.8080	1.9472	3.4740
.25	1.4138	0.4030	0.8710	1.9452	2.1181	3.7017
.5	1.8645	0.9417	1.0250	2.0514	2.2361	3.8578
.75	2.2251	1.366	1.2033	2.1737	2.3533	4.0163
.95	2.7121	1.9164	1.5133	2.4214	2.5211	4.2412
.975	2.8967	2.1074	1.6468	2.5069	2.5832	4.3169
Int.Credib. HPD						
$p = 0.95$ Lim. inf.	0.5478	-0.7254	0.5412	1.7412	1.8923	1.8940
Lim sup.	2.9576	2.2576	1.5907	2.4721	2.5839	4.3205
$p = 0.90$ Lim. inf.	0.8242	-0.3584	0.6339	1.7846	1.9516	1.9527
Lim sup.	2.7743	2.0416	1.4828	2.3795	2.5240	4.2456
desv.padrão (dp)	0.6219	0.7461	.2701	0.1851	0.1753	0.2354
dp R-Blackweliz.	0.6212	0.7467	0.2704	0.1851	0.1749	0.2356
"Naive SE"	0.0044	0.0053	0.0019	0.0013	0.0012	0.0017
Mínimo	-1.4821	-2.4096	-0.0792	1.4869	1.5147	2.8388
Máximo	4.3952	3.1949	2.4700	2.8586	3.0517	4.8382
ESTIMATIVAS	$\sigma_{x_1}^2$	$\sigma_{x_2}^2$	σ_1^2	σ_2^2	k	
Média aritm.	0.4254	1.3662	0.2155	0.2412	20.970	
Média R-Blackw.	0.4242	1.3670	0.2159	0.2410	20.968	
Quantis						
0.025	0.1840	0.7387	0.1185	0.1451	20	
.05	0.2085	0.8098	0.1290	0.1552	21	
.25	0.3036	1.0690	0.1671	0.1940	21	
.5	0.3920	1.3003	0.2030	0.2304	21	
.75	0.5090	1.5913	0.2499	0.2746	21	
.95	0.7490	2.1412	0.3437	0.3674	21	
.975	0.8531	2.3643	0.38261	0.4029	21	
Int.Credib. HPD						
$p = 0.95$ Lim. inf.	0.1465	0.6547	0.1047	0.1324	21	
Lim sup.	0.7674	2.2156	0.3507	0.3770	21	
$p = 0.90$ Lim. inf.	0.1643	0.7316	0.1137	0.1422	21	
Lim sup.	0.6661	2.0020	0.3133	0.3396	21	
desv.padrão (dp)	0.1762	0.4212	0.0689	0.0667	0.1862	
dp R-Blackweliz.	0.1746	0.4211	0.0689	0.0667	0.1951	
"Naive SE"	0.0012	0.0030	0.0005	0.0005	0.0013	
Mínimo	0.0799	0.3990	0.0663	0.06571	16	
Máximo	2.4958	5.1758	1.1225	1.3458	21	

Tabela 4.11: Distribuição de frequência relativa à posteriori de k :

k	18	19	20	21	outros
$\hat{P}(K = k \text{Dobs})$	0.0001	0.0019	0.0250	0.9729	0.0000
					1

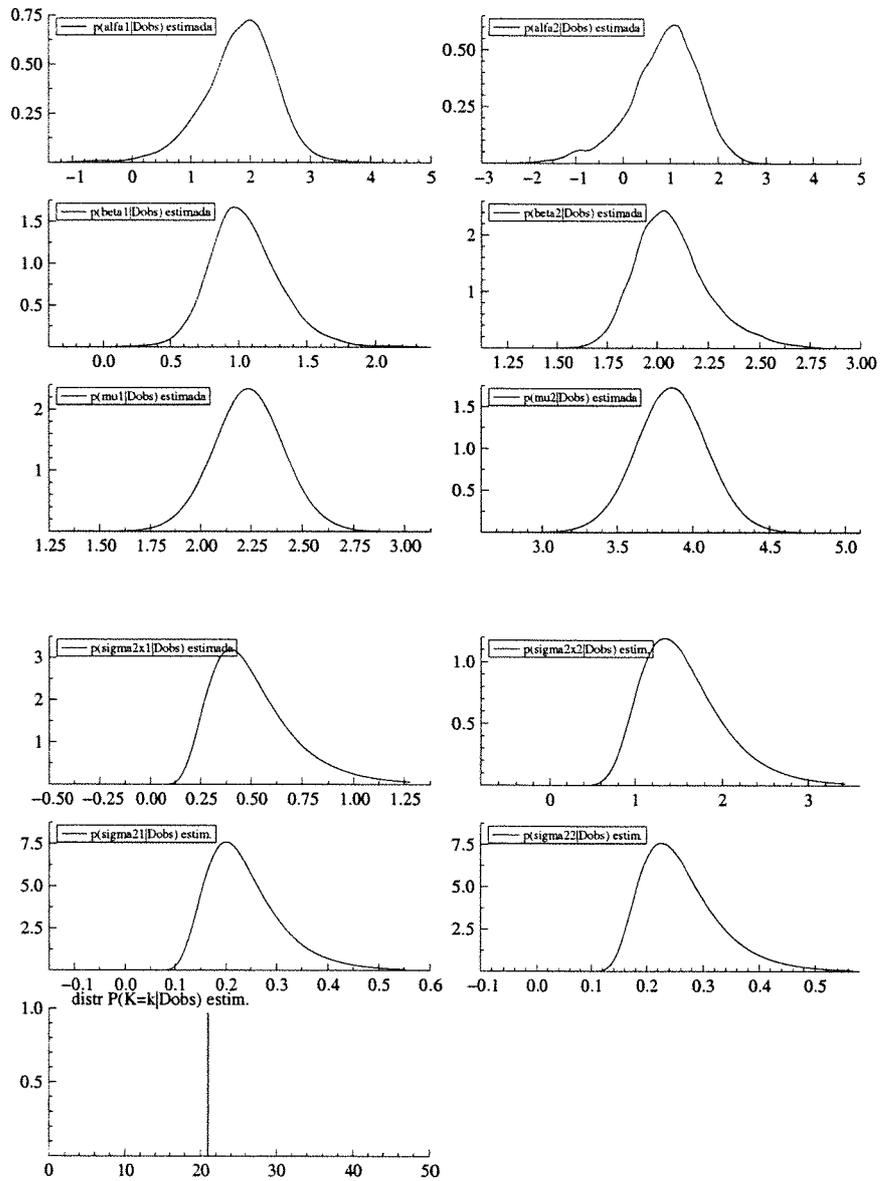


Figura 4.29: Gráficos das estimativas “Rao-Blackwelizadas” das densidades marginais à posteriori $p(\alpha_1|D_{obs}), p(\alpha_2|D_{obs}), \dots, p(\sigma_1^2|D_{obs}), p(\sigma_2^2|D_{obs})$ e $p(k|D_{obs})$.

Observação 1: Da tabela 4.11 vemos que $\hat{P}(k = n) < \hat{P}(k \neq n)$ (já que $\hat{P}(k = n) \approx 0$) e portanto a hipótese nula H_0 (de que não há mudança) é rejeitada, o que era esperado já que a amostra de dados utilizada é a da tabela 4.1 e a distribuição

à priori é praticamente não informativa.

Observação 2: Os comentários sobre a tabela 4.10 são semelhantes àqueles feitos sobre a tabela 4.3.

2.b Sensibilidade das estimativas ao valor do hiperparâmetro c

A seguir analisamos a sensibilidade nas estimativas dos parâmetros do modelo estudado nessa seção quando variamos os valores do hiperparâmetro c à priori, onde $c=P(K=n)$. A tabela 4.12 apresenta, para cada valor de c fixado, a estimativa da média e mediana de cada parâmetro e a tabela 4.13 a distribuição de frequências à posteriori do ponto de mudança k .

A tabela 4.14 apresenta também, para cada valor de c fixado, a estimativa da média de cada parâmetro primeiramente quando o conjunto de dados é gerado a partir do modelo com mudança no ponto $k=7$ e depois a partir do modelo com mudança em $k=43$ (ambos com $\theta_1=(2.0, 1.0, 2.0, 0.6, 0.1)^T$ e $\theta_2=(1.0, 2.0, 4.0, 1.2, 0.2)^T$) e finalmente a partir do modelo sem nenhuma mudança ($k=50$).

Os pontos iniciais das seqüências de Gibbs, em cada caso, foram obtidos pelo mesmo método usado anteriormente para o caso $k=20$.

Observações:

- O algoritmo usado nessa seção pode não funcionar (convergir) dependendo dos valores iniciais. Para ilustrar fizemos a seguinte experiência: geramos 50 conjuntos de dados, cada um a partir do modelo com um valor de k diferente ($1 \leq k \leq 50$) e θ_1 e θ_2 fixados ($\theta_1=(2.0, 1.0, 2.0, 0.6, 0.1)$ e $\theta_2=(1.0, 2.0, 4.0, 1.2, 0.2)$). Para cada um desses conjuntos de dados, o algoritmo foi rodado partindo-se de pontos iniciais distintos da forma $(k^{(0)}, \theta_1^{(0)T}, \theta_2^{(0)T})$, onde $3 \leq k^{(0)} \leq 47$ e $\theta_1^{(0)}$ e $\theta_2^{(0)}$ são os estimadores de máximaveros. de θ_1 e θ_2 dado $k=k^{(0)}$.

Observou-se que se o valor verdadeiro de k está próximo de $\frac{n}{2}$, então partindo-se de qualquer ponto inicial $(k^{(0)}, \theta_1^{(0)T}, \theta_2^{(0)T})$ descrito acima, o algoritmo tende

Tabela 4.12: Estimativas do valor esperado e da mediana dos parâmetros da distribuição à posteriori quando diferentes valores do hiperparâmetro c , onde $c=P(K = n)$, são fixados à priori. Os valores verdadeiros dos parâmetros são: $k=20$, $\alpha_1=2.0$, $\beta_1=1.0$, $\mu_1=2.0$, $\sigma_{x(1)}^2=0.6$, $\sigma_1^2=0.1$, $\alpha_2=1.0$, $\beta_2=2.0$, $\mu_2=4.0$, $\sigma_{x(2)}^2=1.2$, $\sigma_2^2=0.2$.

	c=0.02		c=0.5		c=0.95		c=0.99	
	media	mediana	media	mediana	media	mediana	media	mediana
k	20.97	21	20.97	21	20.97	21	20.97	21
α_1	1.8140	1.8645	1.8140	1.8645	1.8147	1.8645	1.8147	1.8645
α_2	0.8539	0.9417	0.8539	0.9417	0.8540	0.9417	0.8540	0.9417
β_1	1.0500	1.0250	1.0500	1.0250	1.0504	1.0250	1.0504	1.0250
β_2	2.0720	2.0514	2.0720	2.0514	2.0721	2.0514	2.0721	2.0514
μ_1	2.2357	2.2361	2.2357	2.2361	2.2357	2.2361	2.2357	2.2361
μ_2	3.8581	3.8578	3.8581	3.8578	3.6567	3.8578	3.8587	3.8578
$\sigma_{x_1}^2$	0.4254	0.3920	0.4254	0.3920	0.4242	0.3920	0.4242	0.3920
$\sigma_{x_2}^2$	1.3662	1.3003	1.3662	1.3003	1.3670	1.3003	1.3670	1.3003
σ_1^2	0.2155	0.2030	0.2155	0.2030	0.2159	0.2030	0.2159	0.2030
σ_2^2	0.2412	0.2304	0.2412	0.2304	0.2410	0.2304	0.2410	0.2304

Tabela 4.13: Distribuição de frequências à posteriori do ponto de mudança k quando diferentes valores do hiperparâmetro c são fixados.

k	c=0.02	c=0.5	c=0.95	c=0.99
18	0.0001	0.0001	0.0001	0.0001
19	0.0019	0.0019	0.0019	0.0019
20	0.0250	0.0250	0.0250	0.0250
21	0.9729	0.9729	0.9729	0.9729
outros	0.0000	0.0000	0.0000	0.0000

a convergir. Entretanto quando o verdadeiro valor de k está próximo de 1 ou n , o algoritmo pode não convergir dependendo do ponto inicial $(k^{(0)}, \theta_1^{(0)T}, \theta_2^{(0)T})$.

Esse algoritmo aparentemente tem uma certa tendência de retornar ao estado $k = n$ (embora este não seja um ponto absorvente) se alguma vez ele entrou nesse estado, a não ser que o valor de c ($c=P(K=n)=p(n)$) à priori seja bem pequeno. Uma saída foi tomar como valor inicial de k o menor valor possível ($k^{(0)}=3$), e tomar $\theta_1^{(0)}$ e $\theta_2^{(0)}$ como sendo os estimadores de maximaverossimilhança de θ_1 e θ_2 respectivamente. Uma outra saída é tentar utilizar algum método para obter bons valores iniciais como por exemplo a

Tabela 4.14: Estimativas do valor esperado dos parâmetros da distribuição à posteriori quando diferentes conjuntos de dados são utilizados (simulados a partir do modelo com $k=7$, $k=43$ e $k=50$), quando diferentes valores de c são utilizados ($c=0.02$, $c=0.5$ e $c=0.99$).

	k=7			k=43			k=50		
	c=0.02	c=0.5	c=0.99	c=0.02	c=0.5	c=0.99	c=0.02	c=0.5	c=0.99
α_1	2.2919	2.2919	2.2919	2.1052	2.1052	2.1052	2.0357	2.0245	2.0245
α_2	1.4291	1.4291	1.4291	0.6667	0.6667	0.6667	2.0357	2.0245	2.0245
β_1	1.1428	1.1428	1.1428	0.9904	0.9904	0.9904	1.0049	1.0104	1.0104
β_2	1.8845	1.8845	1.8845	2.0705	2.0705	2.0705	1.0049	1.0104	1.0104
μ_1	1.3258	1.3258	1.3258	1.8683	1.8683	1.8683	1.9215	1.9217	1.9217
μ_2	3.8306	3.8306	3.8306	3.6363	3.6363	3.6363	1.9215	1.9217	1.9217
$\sigma_{x_1}^2$	0.4980	0.4980	0.4980	0.6334	0.6334	0.6334	0.6021	0.6000	0.6000
$\sigma_{x_2}^2$	1.4209	1.4209	1.4209	1.5202	1.5202	1.5202	0.6021	0.6000	0.6000
σ_1^2	0.3831	0.3831	0.3831	0.1449	0.1449	0.1449	0.1181	0.1179	0.1179
σ_2^2	0.1823	0.1823	0.1823	0.3746	0.3746	0.3746	0.1181	0.1179	0.1179
k	6.7099	6.7099	6.7099	42.944	42.944	42.944	49.999	50.000	50.000

técnica apresentada na seção 4.7.1 (vale lembrar entretanto que esta técnica não funcionou nesse exemplo quando o verdadeiro valor de k é 5).

- Na tabela 4.14 observa-se que se o verdadeiro valor de k é 7, então a maioria dos parâmetros antes da mudança tende a ter estimativas mais pobres (em termos do viés) do que a maioria dos parâmetros depois da mudança. Semelhantemente quando $k = 43$, os parâmetros após a mudança são pior estimados do que os parâmetros antes da mudança. Isso já era esperado pelo fato de que no primeiro caso o tamanho da amostra antes da mudança é pequeno e no segundo caso o tamanho da amostra depois da mudança é que é pequeno, além disso a distribuição à priori é pouco informativa (embora k tenha uma distribuição de probabilidade discreta com domínio limitado, as distribuições $\pi(\theta_1)$ e $\pi(\theta_2)$ são muito pouco informativas).
- Os dados das tabelas 4.12 a 4.14 sugerem que as estimativas dos valores esperados dos parâmetros da distr. à posteriori, bem como a distrib. de frequências à posteriori do ponto de mudança k , não são sensíveis a variação no valor hiperparâmetro c , onde c é a probabilidade à priori de não existir mudança.

4.7.4 Análise do procedimento 2 para detetar mudança baseada na distribuição à posteriori das diferenças (ou quocientes)

Caso (a): Modelo com restrição de identificabilidade

Nessa seção vamos considerar inicialmente o modelo com restrição de identificabilidade e que permite mudança em todos os parâmetros definido na seção 4.7.1. (É importante lembrar que no caso em que $\alpha_1 = \alpha_2$, $\beta_1 = \beta_2$, $\mu_1 = \mu_2$, $\sigma_{x_1}^2 = \sigma_{x_2}^2$ e $\sigma_1^2 = \sigma_2^2$ então caímos no modelo sem mudança, independentemente do valor de k)

Nesse estudo foram usados dados simulados e a distribuição à priori escolhida (dada na seção 4.7.1) é muito pouco informativa. Então é razoável esperar, se de fato o método utilizado é adequado, que as estimativas à posteriori (bem como os resultados e conclusões) sejam compatíveis com o modelo a partir do qual os dados são gerados.

As figuras 4.30 e 4.31 apresentam os histogramas das densidades marginais à posteriori de $\alpha_i, \beta_i, \mu_i, \sigma_{x_i}^2, \sigma_i^2$ e k para $i = 1, 2$ e também os histogramas das diferenças $\alpha_1 - \alpha_2, \beta_2 - \beta_1, \mu_2 - \mu_1$ e dos quocientes $\sigma_{x_2}^2 / \sigma_{x_1}^2$ e σ_2^2 / σ_1^2 . Os cálculos são obtidos com base nos dados da tabela 4.1 e o algoritmo usado para amostrar da distribuição à posteriori é o M.Collapsed Gibbs (partindo do ponto inicial $\hat{\theta}^{(0)}$ dado na seção 4.7.1).

A partir dessas distribuições as seguintes estimativas foram obtidas: $\hat{P}(\alpha_1 - \alpha_2 > 0 | D_{\text{obs}}) \approx 0.842$, $\hat{P}(\beta_2 - \beta_1 > 0 | D_{\text{obs}}) \approx 0.996$, $\hat{P}(\mu_2 - \mu_1 > 0 | D_{\text{obs}}) \approx 0.993$, $\hat{P}(\sigma_{x_2}^2 / \sigma_{x_1}^2 > 1 | D_{\text{obs}}) \approx 1.000$ e $\hat{P}(\sigma_2^2 / \sigma_1^2 > 1 | D_{\text{obs}}) \approx 0.625$.

Esses resultados sugerem que todos os parâmetros do modelo mudam após o ponto $k = 21$, o que está compatível com o esperado.

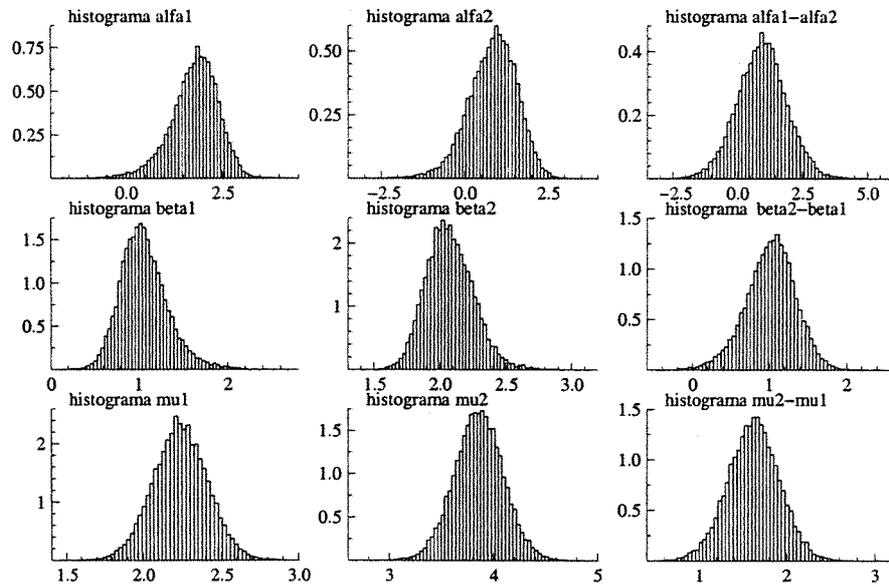


Figura 4.30: Histograma das densidades marginais à posteriori de $p(\alpha_i|D_{\text{obs}})$, $p(\beta_i|D_{\text{obs}})$, $p(\mu_i|D_{\text{obs}})$, $i = 1, 2$ e $p(\alpha_1 - \alpha_2|D_{\text{obs}})$, $p(\beta_2 - \beta_1|D_{\text{obs}})$, $p(\mu_2 - \mu_1|D_{\text{obs}})$.

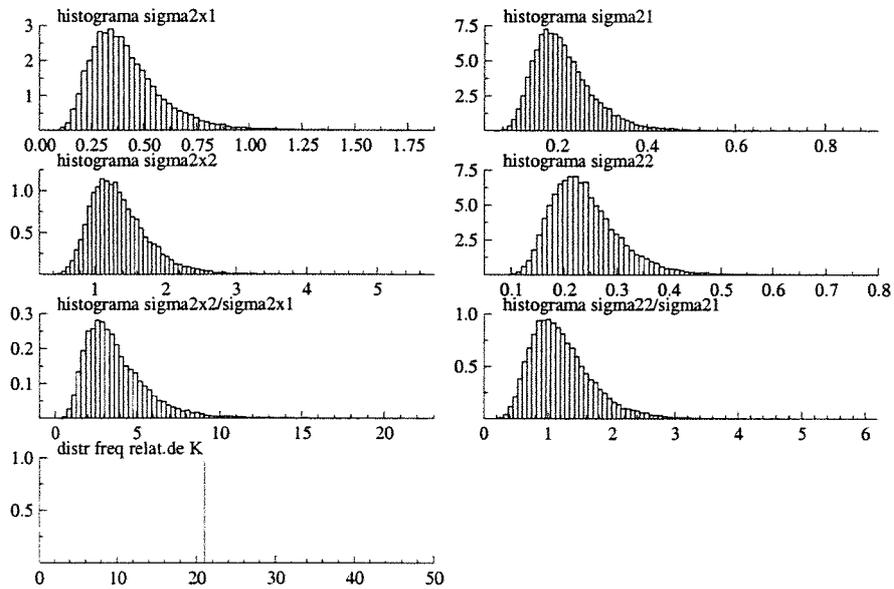


Figura 4.31: Histograma das densidades marginais à posteriori de $p(\sigma_{x_1}^2|D_{\text{obs}})$, $p(\sigma_{x_2}^2|D_{\text{obs}})$ e $p(\sigma_{x_2}^2/\sigma_{x_1}^2|D_{\text{obs}})$ e também de $p(\sigma_1^2|D_{\text{obs}})$, $p(\sigma_2^2|D_{\text{obs}})$ e $p(\sigma_2^2/\sigma_1^2|D_{\text{obs}})$.

Observações: O mesmo modelo (modelo estrutural com restrição de identificabilidade e que permite mudança em todos os parâmetros com distribuição à priori dada em (4.7.1)) foi ajustado a outros dois conjuntos de dados. O primeiro conjunto foi gerado a partir do modelo (com restr. identif.) e sem nenhuma mudança (onde $\alpha_1 = \alpha_2 = 2.0$, $\beta_1 = \beta_2 = 1.0$, $\mu_1 = \mu_2 = 2.0$, $\sigma_{x_1}^2 = \sigma_{x_2}^2 = 0.6$, $\sigma_1^2 = \sigma_2^2 = 0.1$ e $\lambda = 1.0$), o segundo conjunto foi gerado a partir do modelo (com restr. identif.) com mudança apenas na média e na variância de x (onde $\alpha_1 = \alpha_2 = 2.0$, $\beta_1 = \beta_2 = 1.0$, $\mu_1 = 2.0$, $\mu_2 = 4.0$, $\sigma_{x_1}^2 = 0.6$, $\sigma_{x_2}^2 = 2.2$, $\sigma_1^2 = \sigma_2^2 = 0.1$ e $\lambda = 1.0$).

Nos dois casos o algoritmo M.Collapsed Gibbs partiu do ponto inicial ($k^{(0)}=25$, $\beta_i^{(0)}=1$, $\sigma_i^{2(0)}=1$, $\sigma_{x_i}^{2(0)}=1$, $i = 1, 2$). Do primeiro ajuste obtivemos as estimativas: $\hat{P}(\alpha_2 - \alpha_1 > 0|D_{\text{obs}}) \approx 0.51$, $\hat{P}(\beta_2 - \beta_1 > 0|D_{\text{obs}}) \approx 0.50$, $\hat{P}(\mu_2 - \mu_1 > 0|D_{\text{obs}}) \approx 0.51$, $\hat{P}(\sigma_{x_2}^2/\sigma_{x_1}^2 > 1|D_{\text{obs}}) \approx 0.53$ e $\hat{P}(\sigma_2^2/\sigma_1^2 > 1|D_{\text{obs}}) \approx 0.64$. Nesse caso em geral as estimativas estão dentro do esperado, apenas a estimativa da probabilidade $P(\sigma_2^2/\sigma_1^2 > 1|D_{\text{obs}})$ aparentemente está um pouco mais discrepante (em relação ao valor 0.5).

Do segundo ajuste observamos que $\hat{P}(\alpha_1 - \alpha_2 > 0|D_{\text{obs}}) \approx 0.62$, $\hat{P}(\beta_1 - \beta_2 > 0|D_{\text{obs}}) \approx 0.57$, $\hat{P}(\mu_2 - \mu_1 > 0|D_{\text{obs}}) \approx 1.00$, $\hat{P}(\sigma_{x_2}^2/\sigma_{x_1}^2 > 1|D_{\text{obs}}) \approx 1.00$ e $\hat{P}(\sigma_1^2/\sigma_2^2 > 1|D_{\text{obs}}) \approx 0.86$. Esses resultados não sugerem que a mudança ocorre somente na média e na variância de x (o que não estaria compatível com o esperado).

Nesses particulares exemplos, vemos que, embora o método pode não detectar corretamente qual parâmetro está mudando (no modelo que permite mudança em todos os parâmetros), ele forneceu resultado mais ou menos razoável no caso em que não há mudança.

Caso (b) Modelo sem restrição de identificabilidade

Nesse item vamos usar procedimento semelhante ao do caso (a) porém agora

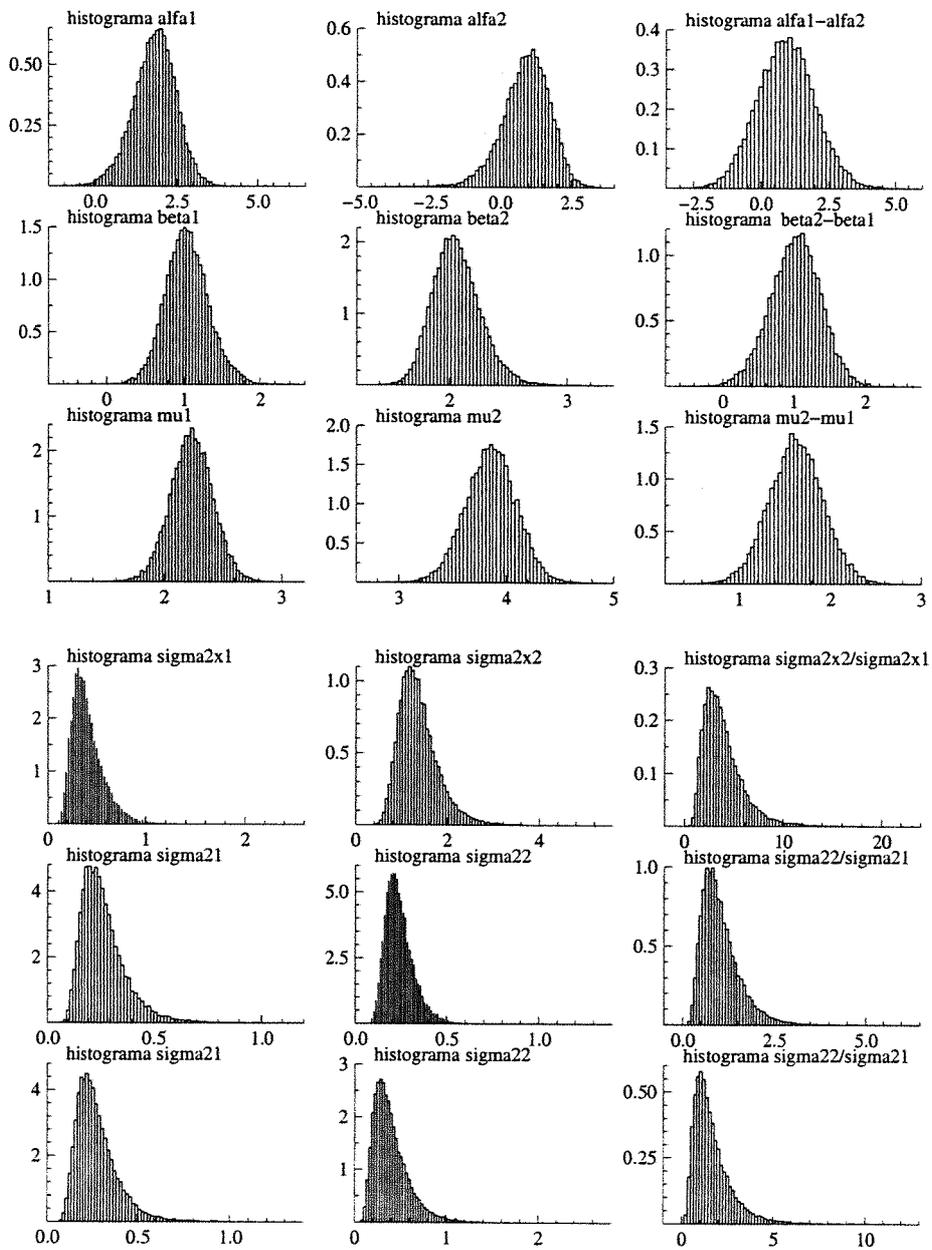


Figura 4.32: Histograma das densidades marginais à posteriori de $p(\alpha_i|D_{\text{obs}})$, $p(\beta_i|D_{\text{obs}})$, $p(\mu_i|D_{\text{obs}})$, $p(\sigma_{x_i}^2|D_{\text{obs}})$, $p(\sigma_{e_i}^2|D_{\text{obs}})$, $p(\sigma_{e_i}^2|D_{\text{obs}})$, $p(K = k|D_{\text{obs}})$ para $i = 1, 2$, e $p(\alpha_1 - \alpha_2|D_{\text{obs}})$, $p(\beta_2 - \beta_1|D_{\text{obs}})$, $p(\mu_2 - \mu_1|D_{\text{obs}})$, $p(\sigma_{x_2}^2/\sigma_{x_1}^2|D_{\text{obs}})$, $p(\sigma_{e_2}^2/\sigma_{e_1}^2|D_{\text{obs}})$, $p(\sigma_{e_2}^2/\sigma_{e_1}^2|D_{\text{obs}})$.

para o modelo sem restrição de identificabilidade e que permite mudança em todos os parâmetros definido na seção 4.7.2.

A figura 4.32 apresenta os histogramas das densidades marginais à posteriori de $\alpha_i, \beta_i, \mu_i, \sigma_{x_i}^2, \sigma_{e_i}^2, \sigma_i^2$ e k e também os histogramas das diferenças $\alpha_1 - \alpha_2, \beta_2 - \beta_1, \mu_2 - \mu_1$ e dos quocientes $\sigma_{x_2}^2/\sigma_{x_1}^2, \sigma_{e_2}^2/\sigma_{e_1}^2$ e σ_2^2/σ_1^2 . Os cálculos foram obtidos com base nos dados da tabela 4.1 e o algoritmo usado para amostrar da distribuição à posteriori é o M.Collapsd Gibbs (o ponto inicial usado é o $\hat{\theta}^{(0)}$ dado na seção 4.7.2).

A partir das distribuições das diferenças e quocientes acima obtivemos as seguintes estimativas : $\hat{P}(\alpha_1 - \alpha_2 > 0 | D_{\text{obs}}) \approx 0.808, \hat{P}(\beta_2 - \beta_1 > 0 | D_{\text{obs}}) \approx 0.997, \hat{P}(\mu_2 - \mu_1 > 0 | D_{\text{obs}}) \approx 1.000, \hat{P}(\sigma_{x_2}^2/\sigma_{x_1}^2 > 1 | D_{\text{obs}}) \approx 0.994, \hat{P}(\sigma_{e_1}^2/\sigma_{e_2}^2 > 1 | D_{\text{obs}}) \approx 0.559, \hat{P}(\sigma_2^2/\sigma_1^2 > 1 | D_{\text{obs}}) \approx 0.706.$

Os resultados acima sugerem que há mudança em todos os parâmetros do modelo (o que está dentro do esperado).

Observações:

O modelo sem restrição foi ajustado ainda a outros dois conjuntos de dados (o primeiro gerado a partir do modelo sem mudança (com $\alpha_1 = \alpha_2 = 2.0, \beta_1 = \beta_2 = 1.0, \mu_1 = \mu_2 = 2.0, \sigma_{x_1}^2 = \sigma_{x_2}^2 = 0.6, \sigma_1^2 = \sigma_2^2 = 0.1$ e $\sigma_{e_1}^2 = \sigma_{e_2}^2 = 0.1$)) e o outro gerado a partir do modelo com mudança na média e na variância de x (com $\alpha_1 = \alpha_2 = 2.0, \beta_1 = \beta_2 = 1.0, \mu_1 = 2.0, \mu_2 = 4.0, \sigma_{x_1}^2 = 0.6, \sigma_{x_2}^2 = 2.2, \sigma_1^2 = \sigma_2^2 = 0.1$ e $\sigma_{e_1}^2 = \sigma_{e_2}^2 = 0.1$). Os pontos iniciais em cada exemplo foram obtidos pelo mesmo procedimento descrito na seção 4.7.2.

Do primeiro ajuste obtivemos as estimativas: $\hat{P}(\alpha_2 - \alpha_1 > 0 | D_{\text{obs}}) \approx 0.53, \hat{P}(\beta_2 - \beta_1 > 0 | D_{\text{obs}}) \approx 0.53, \hat{P}(\mu_2 - \mu_1 > 0 | D_{\text{obs}}) \approx 0.59, \hat{P}(\sigma_{x_1}^2/\sigma_{x_2}^2 > 1 | D_{\text{obs}}) \approx 0.51, \hat{P}(\sigma_1^2/\sigma_2^2 > 1 | D_{\text{obs}}) \approx 0.64, \hat{P}(\sigma_{e_1}^2/\sigma_{e_2}^2 > 1 | D_{\text{obs}}) \approx 0.64.$ Nesse caso os primeiros quatro valores estão dentro do esperado entretanto os dois últimos valores estão um pouco acima do esperado.

Do segundo ajuste obtivemos: $\hat{P}(\alpha_2 - \alpha_1 > 0 | D_{\text{obs}}) \approx 0.54$, $\hat{P}(\beta_1 - \beta_2 > 0 | D_{\text{obs}}) \approx 0.51$, $\hat{P}(\mu_2 - \mu_1 > 0 | D_{\text{obs}}) \approx 0.99$, $\hat{P}(\sigma_{x_2}^2 / \sigma_{x_1}^2 > 1 | D_{\text{obs}}) \approx 0.99$, $\hat{P}(\sigma_1^2 / \sigma_2^2 > 1 | D_{\text{obs}}) \approx 0.74$ $\hat{P}(\sigma_{e_1}^2 / \sigma_{e_2}^2 > 1 | D_{\text{obs}}) \approx 0.75$. Esses resultados não sugerem que a mudança ocorre somente na média e na variância de x (o que não está compatível com o esperado).

Observação:

Dos resultados acima sobre os particulares exemplos, vemos que o método de examinar as distribuições das diferenças ou quocientes dos parâmetros pode não detectar corretamente qual parâmetro está (ou não) mudando.

4.7.5 Análise do procedimento 3 para detetar mudança baseada nos critérios de seleção de modelos.

Uso de critérios de escolha de modelos em modelos identificáveis

A tabela 4.15 a seguir contém as estimativas do Fator de Bayes e Pseudo-Fator de Bayes para comparar dois modelos com restrição de identificabilidade: o modelo que permite mudança em todos os parâmetros (estudado na seção 4.7.1) e o modelo sem nenhuma mudança (estudado na seção 2.12.1), quando os dados observados são os da tabela 4.1. Além disso contém as estimativas da Medida L e da soma das normas dos resíduos padronizados sob os dois modelos. Os cálculos envolvendo cada modelo foram obtidos sempre com base na segunda metade de uma sequência de comprimento $T=10000$ do algoritmo M.Collapsd Gibbs (partindo de um único ponto inicial, $\hat{\theta}_1^{(0)}$, obtido pelo método descrito na seção 4.7.1).

Tabela 4.15: Estimativa do Fator de Bayes (pelo método de Newton e Raftery), do Pseudo-Fator de Bayes, da Medida L e da soma das normas dos resíduos sob cada um dos modelos, com e sem restrição de identificabilidade (utilizando os dados da tabela 4.1).

	medidas	modelo com mudança	modelo sem mudança		
Medida L	$\det(L_2)$	$\delta = 0.0$	3427.99	7764.21	
		$\delta = 0.1$	3986.91	9216.22	
		$\delta = 0.5$	6628.24	16258.00	
		$\delta = 0.9$	9918.61	25273.70	
		$\delta = 1.0$	10842.60	27836.10	
	$\text{tr}(L_2)$	$\delta = 0.0$	256.63	554.12	
		$\delta = 0.1$	280.40	607.91	
		$\delta = 0.5$	375.46	823.06	
		$\delta = 0.9$	470.52	1038.21	
		$\delta = 1.0$	494.29	1092.00	
	Soma das normas dos resíduos		92.47	104.87	
	$\prod_{i=1}^n CPO_i^{M_i}$		5.16×10^{-25}	1.84×10^{-33}	Pseudo-Fator de Bayes= 2.80×10^8
	$\hat{\pi}(\underline{U}_1, \dots, \underline{U}_n M_i)$		2.03×10^{-40}	1.60×10^{-54}	Fator de Bayes= 1.27×10^{14}
	\hat{k}		20.97		

Na figura 4.33 temos um gráfico comparando os modelos com e sem mudança (com restrição de identificabilidade) por meio dos valores da densidade condicional

preditiva (CPO_i) versus i , quando os dados são os da tabela 4.1 e o algoritmo utilizado para amostrar das posteriores é o M.Collap.Gibbs.

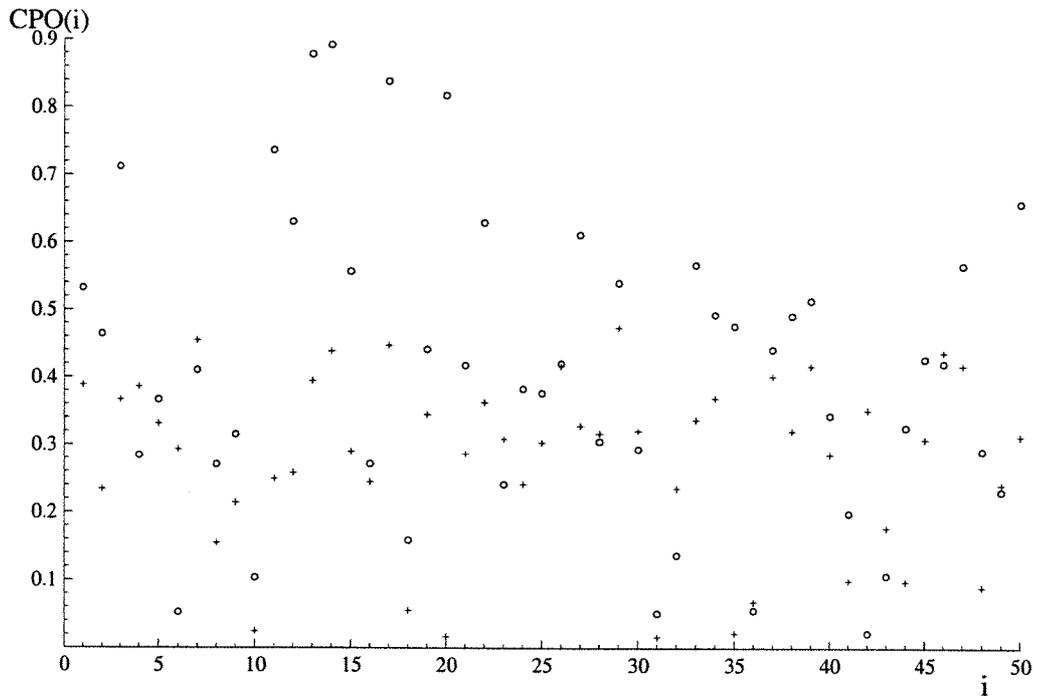


Figura 4.33: Valores da densidade condicional preditiva ($CPO(i)$) versus i , sob o modelo com mudança (símbolo \circ) e sob o modelo sem mudança (símbolo $+$).

Uso de critérios de seleção de modelos em modelos sem restrição de identificabilidade

Da mesma forma que no caso anterior (com restrição de identificabilidade), os dados utilizados para ajustar cada um dos modelos foram os da tabela 4.1 e o algoritmo empregado foi o M.Collapsed Gibbs (partindo em cada caso de um único ponto inicial gerado pela técnica descrita na seção 4.7.2).

A tabela 4.16 descrita à seguir é semelhante à tabela 4.15, porém agora envolvendo os modelos sem restrição de identificabilidade. Da mesma forma, a figura 4.34 é semelhante à figura 4.33 mas agora envolvendo os modelos sem restrição.

Tabela 4.16: Estimativa do Fator de Bayes (pelo método de Newton e Raftery), do Pseudo-Fator de Bayes, da Medida L e da soma das normas dos resíduos sob cada um dos modelos (utilizando os dados da tabela 4.1).

medidas	modelo com mudança	modelo sem mudança			
Medida L	det(L ₂)	δ = 0.0	4033.32	8219.96	
		δ = 0.1	4647.45	9717.95	
		δ = 0.5	7509.92	16943.70	
		δ = 0.9	11021.90	26143.50	
		δ = 1.0	12001.40	28751.90	
	tr(L ₂)	δ = 0.0	262.79	555.08	
		δ = 0.1	286.54	608.86	
		δ = 0.5	381.54	824.02	
		δ = 0.9	476.54	1039.17	
		δ = 1.0	500.29	1092.96	
Soma das normas dos resíduos	92.38	109.90			
$\prod_{i=1}^n CPO_i^{M_1}$	3.04×10^{-27}	3.18×10^{-34}	Pseudo-Fator de Bayes= 9.6×10^6		
$\hat{\pi}(U_1, \dots, U_n M_i)$	1.68×10^{-40}	3.56×10^{-55}	Fator de Bayes= 4.7×10^{14}		
\hat{k}	20.93				

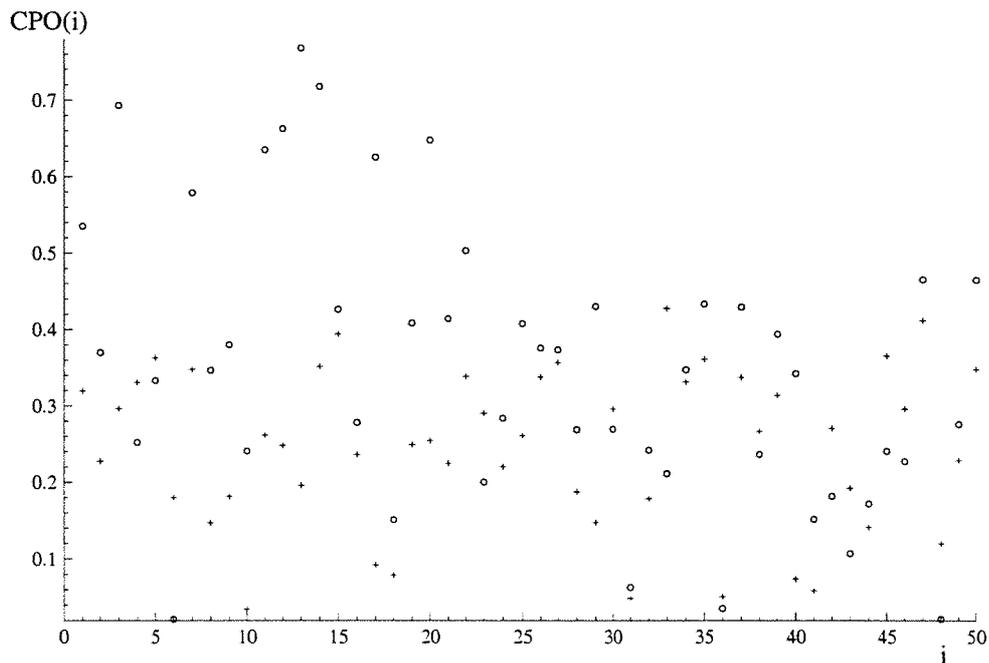


Figura 4.34: Valores da densidade condicional preditiva (CPO(i)) versus i , sob o modelo com mudança (o) e sob o modelo sem mudança(+).

COMENTÁRIOS:

Analisando as tabelas 4.15 e 4.16 e os gráficos 4.33 e 4.34, observamos que há uma forte evidência a favor do modelo que permite mudança em todos os parâmetros (do que o modelo sem nenhuma mudança) tanto no caso com restrição como no caso sem restrição de identificabilidade, ou seja existe evidência de que a hipótese nula (de que não há mudança) deve ser rejeitada nos dois casos.

Isso já era esperado pelo fato de que os dados da tabela 4.1 foram simulados a partir de um modelo onde todos os parâmetros mudam após a vigésima observação da seqüência de dados e a priori envolvida é muito pouco informativa.

É importante observar que os valores do Fator de Bayes e Pseudo-Fator de Bayes do exemplo (tabelas 4.15 e 4.16) nos pareceram extremamente grandes, além disso eles oscilam muito dependendo do algoritmo utilizado como também do ponto inicial escolhido. Nesse caso uma solução alternativa para o cálculo do Fator de Bayes, que deve ser tentada, é o método de Chib (1995) descrito na seção 4.4.3.

4.7.6 Resumo dos resultados

Na seção 4.7 basicamente estimamos os parâmetros dos modelos com e sem restrição de identificabilidade e que permitem mudança em todos os parâmetros após o ponto k . Ilustramos também a aplicação de três métodos para detectar mudança nesses modelos.

As observações a seguir referem-se apenas aos particulares exemplos apresentados nas subseções 4.7.1 a 4.7.5.

(a)- Nos particulares exemplos tratados o algoritmo M.Collap. Gibbs teve um melhor desempenho que o algoritmo de Gibbs simples (principalmente em termos da função de autocorrelação e da matriz de correlações cruzadas).

Os dispositivos de diagnóstico de convergência apresentados nessa seção não

sugerem que haja falhas na convergência dos algoritmos em nenhum dos exemplos.

(b)- Nos modelos com e sem restrição de identificabilidade, os intervalos de credibilidade de 95% contiveram os verdadeiros valores dos parâmetros mas os comprimentos desses intervalos foram em geral ligeiramente menores no modelo com restrição. Entretanto, levando em conta que a qualidade do ajuste aparentemente não diferiu e que na prática dificilmente o valor de λ é conhecido então seria bastante razoável escolher o modelo sem restrição de identificabilidade (quando os dados são os da tabela 4.1 e as priores são as descritas nas seções 4.7.1 e 4.7.2).

Obs: As estimativas dos parâmetros nos modelos das seções 4.7.1 e 4.7.3 foram bastante semelhantes para o particular conjunto de dados.

(c)- No modelo que permite k assumir o valor n (com a particular distr. à priori dada na seção 4.7.3), as estimativas dos parâmetros da distribuição à posteriori, bem como a distribuição de frequências à posteriori do ponto de mudança k , aparentemente não são sensíveis a variação do valor de c (onde c é a prob.à priori de não haver mudança), quando a amostra de dados é a da tabela 4.1.

(d)- O método 1, usado para detetar mudança, quando aplicado ao particular conjunto de dados da tabela 4.1 (e particular priori vaga da seção 4.7.3), detectou corretamente que a hipótese nula (de que não há mudança) deve ser rejeitada e ao mesmo tempo forneceu as estimativas dos parâmetros do modelo.

O método 2 para detetar mudança, aparentemente não é muito preciso para detetar qual parâmetro está mudando, entretanto no particular exemplo envolvendo os dados da tabela 4.1 (e particular distr. à priori da seção 4.7.1), o método conseguiu detetar corretamente que todos os parâmetros mudam (tanto no modelo identificável como no modelo não identificável). Ele também forneceu a decisão correta no caso particular de um conjunto de dados gerados a partir de um modelo identificável sem

nenhuma mudança (com a mesma distribuição à priori).

No método 3 (que usa os critérios de seleção de modelos do Fator de Bayes, Pseudo-Fator de Bayes, Medida L e a soma dos módulos dos resíduos) vimos que todos esses critérios concordaram entre si (em rejeitar a hipótese H_0 de que não há mudança) no particular exemplo apresentado, independentemente do modelo ser ou não identificável.

É importante enfatizar que não estamos concluindo que os critérios de comparação aqui descritos são sempre adequados para detetar mudança e muito menos comparando os diversos critérios entre si, mas apenas exibindo os resultados para um particular conjunto de dados simulados. Um estudo mais completo deve ser conduzido futuramente com essa finalidade.

Observações gerais:

- (1) As médias aritméticas e as médias R-Blackwelizadas ficaram sempre muito próximas independentemente do modelo estudado.
- (2) A medida L baseada no traço aparentemente não serviu para diferenciar os modelos.

Capítulo 5

Uma análise do modelo linear estrutural t -Student com erros nas variáveis com um ponto de mudança desconhecido

Neste capítulo faremos um estudo semelhante ao do capítulo 4, porém agora para os modelos estruturais t -independente e t -dependente com um ponto de mudança desconhecido, considerando a mesma restrição de identificabilidade utilizada no capítulo 4.

5.1 O modelo t com erros independentes, com particular priori de componentes independentes

5.1.1 O modelo que permite mudança em todos os parâmetros após o ponto de mudança k

O modelo t -independente formulado como mistura de escala da distribuição Normal é definido a seguir:

$$\begin{cases} Y_i = \begin{cases} \alpha_1 + \beta_1 x_i + e_i, & i = 1, \dots, k, \text{ onde } 1 \leq k \leq n-1, \\ \alpha_2 + \beta_2 x_i + e_i, & i = k+1, \dots, n, \end{cases} \\ X_i = x_i + u_i, \quad i = 1, \dots, n, \end{cases} \quad (5.1.1)$$

onde Y_i e X_i são as variáveis aleatórias observáveis, x_i é variável aleatória latente e e_i e u_i são os erros aleatórios.

$$\text{Seja } \underline{Z}^{(1)} = \begin{pmatrix} \tilde{Z}_1 \\ \vdots \\ \tilde{Z}_k \end{pmatrix} \text{ e } \underline{Z}^{(2)} = \begin{pmatrix} \tilde{Z}_{k+1} \\ \vdots \\ \tilde{Z}_n \end{pmatrix}, \text{ onde } \underline{Z}_i = \begin{pmatrix} Y_i \\ X_i \\ x_i \end{pmatrix}.$$

O modelo que vamos considerar tem a seguinte suposição:

$$\underline{Z}_i | \underline{\theta} \stackrel{\text{indep}}{\sim} \begin{cases} N_3(\underline{\mu}_1, \underline{\Sigma}_1 w_i), & i = 1, \dots, k, \\ N_3(\underline{\mu}_2, \underline{\Sigma}_2 w_i), & i = k + 1, \dots, n, \end{cases} \quad (5.1.2)$$

onde

$\underline{\theta} = (\theta_1^T, \theta_2^T, k, w^T, \nu_1, \nu_2)^T$ é o vetor de parâmetros, com

$$\theta_i = (\alpha_i, \beta_i, \mu_i, \sigma_{x(i)}^2, \sigma_i^2)^T, \quad i = 1, 2, \quad w = (w_1, \dots, w_n)^T, \quad (\theta_1^T, \nu_1) \neq (\theta_2^T, \nu_2)$$

$$\underline{\mu}_1 = \begin{pmatrix} \alpha_1 + \beta_1 \mu_1 \\ \mu_1 \\ \mu_1 \end{pmatrix}, \quad \underline{\Sigma}_1 = \begin{pmatrix} \beta_1 \sigma_{x(1)}^2 + \lambda \sigma_1^2 & \beta_1 \sigma_{x(1)}^2 & \beta_1 \sigma_{x(1)}^2 \\ \beta_1 \sigma_{x(1)}^2 & \sigma_{x(1)}^2 + \sigma_1^2 & \sigma_{x(1)}^2 \\ \beta_1 \sigma_{x(1)}^2 & \sigma_{x(1)}^2 & \sigma_{x(1)}^2 \end{pmatrix}$$

e

$$\underline{\mu}_2 = \begin{pmatrix} \alpha_2 + \beta_2 \mu_2 \\ \mu_2 \\ \mu_2 \end{pmatrix}, \quad \underline{\Sigma}_2 = \begin{pmatrix} \beta_2 \sigma_{x(2)}^2 + \lambda \sigma_2^2 & \beta_2 \sigma_{x(2)}^2 & \beta_2 \sigma_{x(2)}^2 \\ \beta_2 \sigma_{x(2)}^2 & \sigma_{x(2)}^2 + \sigma_2^2 & \sigma_{x(2)}^2 \\ \beta_2 \sigma_{x(2)}^2 & \sigma_{x(2)}^2 & \sigma_{x(2)}^2 \end{pmatrix} \quad (5.1.3)$$

ν_1 = número de graus de liberdade antes da mudança

ν_2 = número de graus de liberdade após a mudança ($\nu_i > 2, i = 1, 2$),

λ é a razão das variâncias dos erros (conhecida) e $\Sigma_i > 0, i = 1, 2$.

A distribuição à priori é dada por

$$\pi(\underline{\theta}) \propto \pi(k, \theta_1, \theta_2, \nu_1, \nu_2, w) = p(k) \pi(\theta_1) \pi(\theta_2) \pi(w | \nu_1, \nu_2) \pi(\nu_1) \pi(\nu_2),$$

onde

$$w_i | \nu_1, \nu_2 \stackrel{\text{indep}}{\sim} \begin{cases} IG(\frac{\nu_1}{2}, \frac{\nu_1}{2}), & i = 1, \dots, k, \\ IG(\frac{\nu_2}{2}, \frac{\nu_2}{2}), & i = k + 1, \dots, n, \end{cases}$$

$$p(k) = \begin{cases} P[K = k] = \frac{1}{\#\mathcal{K}}, & k \in \mathcal{K}, \\ 0, & \text{caso contrário.} \end{cases}$$

$$\pi(\theta_i) = \pi(\alpha_i) \pi(\beta_i) \pi(\mu_i) \pi(\sigma_{x(i)}^2) \pi(\sigma_i^2), \quad i = 1, 2,$$

onde

$$\begin{aligned} \pi(\alpha_i) &\sim N(a_i, \sigma_{\alpha_i}^2), \quad \pi(\beta_i) \sim N(b_i, \sigma_{\beta_i}^2), \quad \pi(\mu_i) \sim N(m_i, \sigma_{\mu_i}^2), \\ \pi(\sigma_{x(i)}^2) &\sim IG(c_i, d_i), \quad \pi(\sigma_i^2) \sim IG(f_i, g_i), \quad i = 1, 2, \\ \text{e} \quad \pi(\nu_i) = p[\mathcal{V}_i = \nu_i] &= \begin{cases} \frac{1}{n_{0_i}}, & c_i - n_{0_i} + 1 \leq \nu_i \leq c_i, \\ 0, & \text{caso contrário,} \end{cases} \end{aligned} \quad (5.1.4)$$

onde $i = 1, 2$, n_{0_i} = número de elementos do domínio da variável aleatória discreta \mathcal{V}_i cuja probabilidade é positiva e c_i é uma constante inteira (podemos tomar $c_i \geq 3$ para assegurar a existência dos momentos de ordem 1 e 2).

observação: O caso em que os números de graus de liberdade ν_1 e ν_2 são conhecidos é um caso particular de (5.1.4) quando $n_{0_i} = 1$

Distribuição à posteriori

A posteriori pode ser escrita como

$$p(\check{\theta} | D_{\text{obs}}) = \int_{\mathfrak{X}} p(\underline{x}, \check{\theta} | D_{\text{obs}}) d\underline{x},$$

onde $\mathfrak{X} = \{(x_1, \dots, x_n)^T : x_i \in \mathbb{R}\}$ e $p(\underline{x}, \check{\theta} | D_{\text{obs}}) \propto L(\check{\theta} | D) \pi(\check{\theta})$, onde $L(\check{\theta} | D)$ é a verossimilhança baseada nos dados completos $D = \{Z_i : i = 1, 2, \dots, n\}$, dada por

$$\begin{aligned} L(\check{\theta} | D) = L(k, \theta_1, \theta_2, \underline{w} | D) &\propto \frac{1}{\left(\prod_{i=1}^k w_i^{3/2}\right) (\sigma_1^2)^k (\sigma_{x(1)}^2)^{\frac{k}{2}}} \\ &\cdot \exp\left\{-\frac{1}{2} \left[\sum_{i=1}^k \frac{(Y_i - \alpha_1 - \beta_1 x_i)^2}{w_i \lambda \sigma_1^2} + \sum_{i=1}^k \frac{(X_i - x_i)^2}{w_i \sigma_1^2} + \sum_{i=1}^k \frac{(x_i - \mu_1)^2}{w_i \sigma_{x(1)}^2} \right]\right\} \\ &\cdot \frac{1}{\left(\prod_{i=k+1}^n w_i^{3/2}\right) (\sigma_2^2)^{n-k} (\sigma_{x(2)}^2)^{\frac{n-k}{2}}} \exp\left\{-\frac{1}{2} \left[\sum_{i=k+1}^n \frac{(Y_i - \alpha_2 - \beta_2 x_i)^2}{w_i \lambda \sigma_2^2} \right. \right. \\ &\left. \left. + \sum_{i=k+1}^n \frac{(X_i - x_i)^2}{w_i \sigma_2^2} + \sum_{i=k+1}^n \frac{(x_i - \mu_2)^2}{w_i \sigma_{x(2)}^2} \right]\right\} \\ &\propto L_1(k, \theta_1, \nu_1, \underline{w}_1 | D_1) L_2(k, \theta_2, \nu_2, \underline{w}_2 | D_2), \end{aligned} \quad (5.1.5)$$

onde $\underline{w}_1 = (w_1, \dots, w_k)^T$ e $\underline{w}_2 = (w_{k+1}, \dots, w_n)^T$ e onde o primeiro fator é a verossimilhança baseada nos dados completos (D_1) até a mudança e o segundo fator a verossimilhança baseada nos dados completos (D_2) após a mudança.

5.1.2 Implementação do algoritmo de Gibbs no modelo que permite mudança em todos os parâmetros

Usando a distribuição $p(\underline{x}, \underline{\theta} | D_{\text{obs}})$ acima, obtemos as condicionais completas de $x_1, \dots, x_n, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, \nu_1, \nu_2$ e k , que apresentamos a seguir.

$$x_i | x_{(-i)}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, \underline{w}, \nu_1, \nu_2, k, D_{\text{obs}} \sim N \left(\frac{\sigma_{x(1)}^2 [\lambda X_i + \beta_1 (Y_i - \alpha_1) + \lambda \sigma_1^2 \mu_1]}{\sigma_{x(1)}^2 [\lambda + \beta_1^2] + \lambda \sigma_1^2}, \frac{\lambda \sigma_1^2 \sigma_{x(1)}^2 w_i}{\sigma_{x(1)}^2 [\lambda + \beta_1^2] + \lambda \sigma_1^2} \right), \quad (5.1.6)$$

$$i = 1, \dots, k$$

$$N \left(\frac{\sigma_{x(2)}^2 [\lambda X_i + \beta_2 (Y_i - \alpha_2) + \lambda \sigma_2^2 \mu_2]}{\sigma_{x(2)}^2 [\lambda + \beta_2^2] + \lambda \sigma_2^2}, \frac{\lambda \sigma_2^2 \sigma_{x(2)}^2 w_i}{\sigma_{x(2)}^2 [\lambda + \beta_2^2] + \lambda \sigma_2^2} \right), \quad (5.1.7)$$

$$i = k + 1, \dots, n.$$

$\alpha_1 | \underline{x}, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, \underline{w}, \nu_1, \nu_2, k, D_{\text{obs}}$ tem distribuição

$$N \left(\frac{\sigma_{\alpha_1}^2 \sum_{i=1}^k \frac{(Y_i - \beta_1 x_i)}{w_i} + \lambda \sigma_1^2 a_1}{\sigma_{\alpha_1}^2 \sum_{i=1}^k \frac{1}{w_i} + \lambda \sigma_1^2}, \frac{\lambda \sigma_1^2 \sigma_{\alpha_1}^2}{\sigma_{\alpha_1}^2 \sum_{i=1}^k \frac{1}{w_i} + \lambda \sigma_1^2} \right), \quad (5.1.8)$$

$\alpha_2 | \underline{x}, \alpha_1, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, \underline{w}, \nu_1, \nu_2, k, D_{\text{obs}}$ tem distribuição

$$N \left(\frac{\sigma_{\alpha_2}^2 \sum_{i=k+1}^n \frac{(Y_i - \beta_2 x_i)}{w_i} + \lambda \sigma_2^2 a_2}{\sigma_{\alpha_2}^2 \sum_{i=k+1}^n \frac{1}{w_i} + \lambda \sigma_2^2}, \frac{\lambda \sigma_2^2 \sigma_{\alpha_2}^2}{\sigma_{\alpha_2}^2 \sum_{i=k+1}^n \frac{1}{w_i} + \lambda \sigma_2^2} \right), \quad (5.1.9)$$

$\beta_1 | \underline{x}, \alpha_1, \alpha_2, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, \nu_1, \nu_2, D_{\text{obs}}$ tem distribuição

$$N \left(\frac{\sigma_{\beta_1}^2 \sum_{i=1}^k \frac{(Y_i - \alpha_1) x_i}{w_i} + \lambda b_1 \sigma_1^2}{\sigma_{\beta_1}^2 \sum_{i=1}^k \frac{x_i^2}{w_i} + \lambda \sigma_1^2}, \frac{\lambda \sigma_1^2 \sigma_{\beta_1}^2}{\sigma_{\beta_1}^2 \sum_{i=1}^k \frac{x_i^2}{w_i} + \lambda \sigma_1^2} \right), \quad (5.1.10)$$

$$\beta_2 | \underline{x}, \alpha_1, \alpha_2, \beta_1, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, \underline{w}, \nu_1, \nu_2, k, D_{\text{obs}} \sim N \left(\frac{\sigma_{\beta_2}^2 \sum_{i=k+1}^n \frac{(Y_i - \alpha_2) x_i}{w_i} + \lambda b_2 \sigma_2^2}{\sigma_{\beta_2}^2 \sum_{i=k+1}^n \frac{x_i^2}{w_i} + \lambda \sigma_2^2}, \frac{\lambda \sigma_2^2 \sigma_{\beta_2}^2}{\sigma_{\beta_2}^2 \sum_{i=k+1}^n \frac{x_i^2}{w_i} + \lambda \sigma_2^2} \right), \quad (5.1.11)$$

$\mu_1 | \underline{x}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, \underline{w}, \nu_1, \nu_2, k, D_{\text{obs}}$

$$\sim N \left(\frac{\sigma_{\mu_1}^2 \sum_{i=1}^k \frac{x_i}{w_i} + \sigma_{x(1)}^2 m_1}{\sigma_{\mu_1}^2 \sum_{i=1}^k \frac{1}{w_i} + \sigma_{x(1)}^2}; \frac{\sigma_{\mu_1}^2 \sigma_{x(1)}^2}{\sigma_{\mu_1}^2 \sum_{i=1}^k \frac{1}{w_i} + \sigma_{x(1)}^2} \right), \quad (5.1.12)$$

$$\begin{aligned} & \mu_2 | \underline{x}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, \underline{w}, \nu_1, \nu_2, k, D_{\text{obs}} \\ & \sim N \left(\frac{\sigma_{\mu_2}^2 \sum_{i=k+1}^n \frac{x_i}{w_i} + \sigma_{x(2)}^2 m_2}{\sigma_{\mu_2}^2 \sum_{i=k+1}^n \frac{1}{w_i} + \sigma_{x(2)}^2}; \frac{\sigma_{\mu_2}^2 \sigma_{x(2)}^2}{\sigma_{\mu_2}^2 \sum_{i=k+1}^n \frac{1}{w_i} + \sigma_{x(2)}^2} \right), \end{aligned} \quad (5.1.13)$$

$$\begin{aligned} & \sigma_{x(1)}^2 | \underline{x}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, \nu_1, \nu_2, k, \underline{w}, D_{\text{obs}} \\ & \sim IG \left(\frac{k}{2} + c_1; \frac{1}{2} \sum_{i=1}^k \frac{(x_i - \mu_1)^2}{w_i} + d_1 \right), \end{aligned} \quad (5.1.14)$$

$$\begin{aligned} & \sigma_{x(2)}^2 | \underline{x}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_1^2, \sigma_2^2, \nu_1, \nu_2, k, \underline{w}, D_{\text{obs}} \\ & \sim IG \left(\frac{n-k}{2} + c_2; \frac{1}{2} \sum_{i=k+1}^n \frac{(x_i - \mu_2)^2}{w_i} + d_2 \right), \end{aligned} \quad (5.1.15)$$

$$\begin{aligned} & \sigma_1^2 | \underline{x}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_2^2, \nu_1, \nu_2, k, \underline{w}, D_{\text{obs}} \\ & \sim IG \left(k + f_1; \frac{1}{2\lambda} \left(\sum_{i=1}^k \frac{(Y_i - \alpha_1 - \beta_1 x_i)^2}{w_i} + \sum_{i=1}^k \frac{(X_i - x_i)^2}{w_i} \right) + g_1 \right), \end{aligned} \quad (5.1.16)$$

$$\begin{aligned} & \sigma_2^2 | \underline{x}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \nu_1, \nu_2, k, \underline{w}, D_{\text{obs}} \\ & \sim IG \left(n-k + f_2; \frac{1}{2\lambda} \left(\sum_{i=k+1}^n \frac{(Y_i - \alpha_2 - \beta_2 x_i)^2}{w_i} + \sum_{i=k+1}^n \frac{(X_i - x_i)^2}{w_i} \right) + g_2 \right), \end{aligned} \quad (5.1.17)$$

$$\begin{aligned} & w_i | \underline{x}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, \nu_1, \nu_2, w_{(-i)}, k, D_{\text{obs}} \\ & \sim IG \left(\frac{3 + \nu_1}{2}; \frac{1}{2} \left[\left(\frac{(Y_i - \alpha_1 - \beta_1 x_i)^2 + \lambda (X_i - x_i)^2}{\lambda \sigma_1^2} + \frac{(x_i - \mu_1)^2}{\sigma_{x(1)}^2} \right) + \nu_1 \right] \right), \end{aligned} \quad (5.1.18)$$

$$\begin{aligned} & \sim IG \left(\frac{3 + \nu_2}{2}; \frac{1}{2} \left[\left(\frac{(Y_i - \alpha_2 - \beta_2 x_i)^2 + \lambda (X_i - x_i)^2}{\lambda \sigma_2^2} + \frac{(x_i - \mu_2)^2}{\sigma_{x(2)}^2} \right) + \nu_2 \right] \right), \\ & i = k+1, \dots, n, \end{aligned} \quad (5.1.19)$$

$$\begin{aligned}
& P[K=k | \underline{x}, \alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, \nu_1, \nu_2, \underline{w}, D_{\text{obs}}] \\
&= \frac{L(k, \underline{\theta}_1, \underline{\theta}_2, \underline{w}, \nu_1, \nu_2 | D)}{\sum_{k \in \mathcal{K}} L(k, \underline{\theta}_1, \underline{\theta}_2, \underline{w}, \nu_1, \nu_2 | D)}, \tag{5.1.20}
\end{aligned}$$

$k \in \mathcal{K}$, onde $L(k, \underline{\theta}_1, \underline{\theta}_2, \underline{w} | D)$ é dada em (5.1.5).

$$p(\mathcal{V}_i = \nu_i | \underline{x}, \alpha_1, \alpha_2, \dots, \nu_1, \nu_2, \underline{w}, D_{\text{obs}}) = \frac{L_i(k, \underline{\theta}_i, \underline{w}_i, \nu_i | D_i)}{\sum_{\nu_i=c_i-n_{0_i}+1}^{c_i} L_i(k, \underline{\theta}_i, \underline{w}_i, \nu_i | D_i)}, \quad i = 1, 2, \tag{5.1.21}$$

onde $L_i(k, \underline{\theta}_i, \nu_i, \underline{w}_i | D_i)$ é dada em (5.1.5) e n_{0_i} e c_i são definidos em (5.1.4).

Observação 1: No caso em que ν_1 e ν_2 são conhecidos devemos desconsiderar as distribuições em (5.1.21).

Observação 2: Fazendo $w_i = 1$, $i = 1, \dots, n$ e desprezando as distribuições em (5.1.21), obtemos as condicionais completas do modelo Normal dadas na Seção 4.2.2.

5.1.3 Implementação do algoritmo de Gibbs no modelo com mudança em um subconjunto qualquer de parâmetros

Nessa subseção o objetivo é implementar o algoritmo de Gibbs para amostrar da distribuição à posteriori de um modelo que permite mudança em um subconjunto qualquer não vazio de parâmetros estritamente contidos em $\{\alpha, \beta, \mu, \sigma_x^2, \sigma^2\}$. Para esse propósito, utilizamos as distribuições condicionais completas à posteriori do modelo com mudança em todos os parâmetros (que são dadas pelas fórmulas (5.1.6) a (5.1.21), fazendo-se as devidas restrições; utilizamos também as distribuições condicionais completas a seguir (fazendo-se também as devidas restrições):

- $[\alpha | \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, \underline{w}, \nu_1, \nu_2, \underline{x}, D_{\text{obs}}]$

$$\sim N\left(\frac{\sigma_2^2\sigma_\alpha^2\sum_{i=1}^k\frac{(Y_i-\beta_1x_i)}{w_i} + \sigma_1^2\sigma_\alpha^2\sum_{i=k+1}^n\frac{(Y_i-\beta_2x_i)}{w_i} + \lambda\sigma_1^2\sigma_2^2a}{\sigma_2^2\sigma_\alpha^2\sum_{i=1}^k\frac{1}{w_i} + \sigma_1^2\sigma_\alpha^2\sum_{i=k+1}^n\frac{1}{w_i} + \lambda\sigma_1^2\sigma_2^2}, \frac{\lambda\sigma_1^2\sigma_2^2\sigma_\alpha^2}{\sigma_2^2\sigma_\alpha^2\sum_{i=1}^k\frac{1}{w_i} + \sigma_1^2\sigma_\alpha^2\sum_{i=k+1}^n\frac{1}{w_i} + \lambda\sigma_1^2\sigma_2^2}\right), \quad (5.1.22)$$

- $[\beta|\alpha_1, \alpha_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, \underline{w}, \nu_1, \nu_2, \underline{x}, D_{\text{obs}}]$,

$$\sim N\left(\frac{\sigma_2^2\sigma_\beta^2\sum_{i=1}^k\frac{(Y_i-\alpha_1)x_i}{w_i} + \sigma_1^2\sigma_\beta^2\sum_{i=k+1}^n\frac{(Y_i-\alpha_2)x_i}{w_i} + \lambda\sigma_1^2\sigma_2^2b}{\sigma_2^2\sigma_\beta^2\sum_{i=1}^k\frac{x_i^2}{w_i} + \sigma_1^2\sigma_\beta^2\sum_{i=k+1}^n\frac{x_i^2}{w_i} + \lambda\sigma_1^2\sigma_2^2}, \frac{\lambda\sigma_1^2\sigma_2^2\sigma_\beta^2}{\sigma_2^2\sigma_\beta^2\sum_{i=1}^k\frac{x_i^2}{w_i} + \sigma_1^2\sigma_\beta^2\sum_{i=k+1}^n\frac{x_i^2}{w_i} + \lambda\sigma_1^2\sigma_2^2}\right), \quad (5.1.23)$$

- $[\mu|\alpha_1, \alpha_2, \beta_1, \beta_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, k, \nu_1, \nu_2, \underline{w}, D_{\text{obs}}]$,

$$\sim N\left(\frac{\sigma_{x(2)}^2\sigma_\mu^2\sum_{i=1}^k\frac{x_i}{w_i} + \sigma_{x(1)}^2\sigma_\mu^2\sum_{i=k+1}^n\frac{x_i}{w_i} + \sigma_{x(1)}^2\sigma_{x(2)}^2m}{\sigma_{x(2)}^2\sigma_\mu^2\sum_{i=1}^k\frac{1}{w_i} + \sigma_{x(1)}^2\sigma_\mu^2\sum_{i=k+1}^n\frac{1}{w_i} + \sigma_{x(1)}^2\sigma_{x(2)}^2}, \frac{\sigma_{x(1)}^2\sigma_{x(2)}^2\sigma_\mu^2}{\sigma_{x(2)}^2\sigma_\mu^2\sum_{i=1}^k\frac{1}{w_i} + \sigma_{x(1)}^2\sigma_\mu^2\sum_{i=k+1}^n\frac{1}{w_i} + \sigma_{x(1)}^2\sigma_{x(2)}^2}\right), \quad (5.1.24)$$

- $[\sigma_x^2|\alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, k, \underline{w}, \nu_1, \nu_2, \underline{x}, D_{\text{obs}}]$

$$\sim IG\left(\frac{n}{2} + c; \frac{1}{2}\left(\sum_{i=1}^k\frac{(x_i - \mu_1)^2}{w_i} + \sum_{i=k+1}^n\frac{(x_i - \mu_2)^2}{w_i}\right) + d\right), \quad (5.1.25)$$

- $[\sigma^2|\alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, k, \underline{w}, \nu_1, \nu_2, \underline{x}, D_{\text{obs}}]$

$$\sim IG\left[n + f; \frac{1}{2}\left(\sum_{i=1}^k\frac{(Y_i - \alpha_1 - \beta_1x_i)^2}{\lambda w_i} + \sum_{i=k+1}^n\frac{(Y_i - \alpha_2 - \beta_2x_i)^2}{\lambda w_i} + \sum_{i=1}^k\frac{(X_i - x_i)^2}{w_i} + \sum_{i=k+1}^n\frac{(X_i - x_i)^2}{w_i}\right) + g\right], \quad (5.1.26)$$

Assim, por exemplo, para implementar o algoritmo de Gibbs no modelo que permite mudança apenas em $\{\alpha, \beta, \sigma^2\}$, as distribuições condicionais completas necessárias são as distribuições de (5.1.6) a (5.1.11), (5.1.24) a (5.1.25) e (5.1.16) a (5.1.21), impondo-se as restrições $\mu_1 = \mu_2$ e $\sigma_{x(1)}^2 = \sigma_{x(2)}^2$ nas distribuições que envolverem os parâmetros μ_i e $\sigma_{x(i)}^2$, $i = 1, 2$, como também no valor inicial, da mesma forma que na Seção 4.3.

5.2 O modelo t -dependente com particular priori de componentes independentes

5.2.1 O modelo que permite mudança em todos os parâmetros após o ponto k

O modelo t -dependente formulado como mistura de escala de normais é dado por:

$$\begin{cases} Y_i = \begin{cases} \alpha_1 + \beta_1 x_i + e_i, & i = 1, \dots, k, \\ \alpha_2 + \beta_2 x_i + e_i, & i = k+1, \dots, n, \end{cases} \text{ onde } 1 \leq k \leq n-1, \\ X_i = x_i + u_i, \quad i = 1, 2, \dots, n, \end{cases} \quad (5.2.1)$$

onde X_i e Y_i são v.a. observáveis, x_i é v.a. latente e (e_i, u_i) os erros aleatórios.

$$\text{Sejam } \tilde{Z}^{(1)} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_k \end{pmatrix}, \quad \tilde{Z}^{(2)} = \begin{pmatrix} Z_{k+1} \\ \vdots \\ Z_n \end{pmatrix}, \quad \text{onde } \tilde{Z}_i = \begin{pmatrix} Y_i \\ X_i \\ x_i \end{pmatrix}.$$

O modelo assume que $\tilde{Z}^{(1)}$ e $\tilde{Z}^{(2)}$ são independentes com

$$\tilde{Z}^{(1)} | w_1, \vartheta_1, \nu_1 \sim N_{3k} \left(\underset{1}{\mathbb{1}}_{k \times 1} \otimes \mu_1; I_{k \times k} \otimes \Sigma_1 w_1 \right)$$

e

$$\tilde{Z}^{(2)} | w_2, \vartheta_2, \nu_2 \sim N_{3(n-k)} \left(\underset{1}{\mathbb{1}}_{(n-k) \times 1} \otimes \mu_2; I_{(n-k) \times (n-k)} \otimes \Sigma_2 w_2 \right),$$

onde

$$\begin{aligned} \mu_1 &= \begin{pmatrix} \alpha_1 + \beta_1 \mu_1 \\ \mu_1 \\ \mu_1 \end{pmatrix}, & \mu_2 &= \begin{pmatrix} \alpha_2 + \beta_2 \mu_2 \\ \mu_2 \\ \mu_2 \end{pmatrix}, \\ \Sigma_1 &= \begin{pmatrix} \beta_1 \sigma_{x(1)}^2 + \lambda \sigma_1^2 & \beta_1 \sigma_{x(1)}^2 & \beta_1 \sigma_{x(1)}^2 \\ \beta_1 \sigma_{x(1)}^2 & \sigma_{x(1)}^2 + \sigma_1^2 & \sigma_{x(1)}^2 \\ \beta_1 \sigma_{x(1)}^2 & \sigma_{x(1)}^2 & \sigma_{x(1)}^2 \end{pmatrix} \\ \text{e } \Sigma_2 &= \begin{pmatrix} \beta_2 \sigma_{x(2)}^2 + \lambda \sigma_2^2 & \beta_2 \sigma_{x(2)}^2 & \beta_2 \sigma_{x(2)}^2 \\ \beta_2 \sigma_{x(2)}^2 & \sigma_{x(2)}^2 + \sigma_2^2 & \sigma_{x(2)}^2 \\ \beta_2 \sigma_{x(2)}^2 & \sigma_{x(2)}^2 & \sigma_{x(2)}^2 \end{pmatrix}, \end{aligned}$$

onde $\lambda =$ razão das variâncias dos erros (conhecida), $\nu_1 =$ número de graus de liberdade antes e $\nu_2 =$ número de graus de liberdade após a mudança, além disso $\Sigma_i > 0$, $i = 1, 2$, $\tilde{\vartheta} = (\vartheta_1^T, \vartheta_2^T, k, w_1, w_2, \nu_1, \nu_2)^T$ é o vetor de parâmetros do modelo, onde $\vartheta_i = (\alpha_i, \beta_i, \mu_i, \sigma_{x(i)}^2, \sigma_i^2)^T$,

A distribuição à priori escolhida para $\check{\theta}$ é da forma

$$\pi(\check{\theta}) = p(k) \prod_{i=1}^2 \pi(\check{\theta}_i) \pi(w_i | \nu_i) \pi(\nu_i), \quad (5.2.2)$$

onde $\pi(w_i | \nu_i) \sim IG(\frac{\nu_i}{2}, \frac{\nu_i}{2})$, $i = 1, 2$, e $\pi(\check{\theta}_i)$, $p(k)$ e $\pi(\nu_i)$ coincidem com as distribuições em (5.1.4)

• Função de verossimilhança

A função de verossimilhança baseada nos dados “completos” $D = \{Z_i = (Y_i, X_i, x_i)^\top, i = 1, \dots, n\}$ é a seguinte:

$$\begin{aligned} L(\check{\theta} | D) &\propto |\check{\Sigma}_1 w_1|^{-\frac{k}{2}} |\check{\Sigma}_2 w_2|^{-\frac{(n-k)}{2}} \exp\left\{-\frac{1}{2} \left[(Z^{(1)} - \check{1}_{k \times 1} \otimes \mu_1)^\top \right. \right. \\ &\quad \left. \left. (I_{k \times k} \otimes \check{\Sigma}_1 w_1)^{-1} (Z^{(1)} - \check{1}_{k \times 1} \otimes \mu_1) + (Z^{(2)} - \check{1}_{(n-k) \times 1} \otimes \mu_2)^\top \right. \right. \\ &\quad \left. \left. \cdot (I_{(n-k) \times (n-k)} \otimes \check{\Sigma}_2 w_2)^{-1} (Z^{(2)} - \check{1}_{(n-k) \times 1} \otimes \mu_2) \right] \right\} \\ &\propto |\check{\Sigma}_1 w_1|^{-\frac{k}{2}} |\check{\Sigma}_2 w_2|^{-\frac{(n-k)}{2}} \exp\left\{-\frac{1}{2} \left[\sum_{i=1}^k (Z_i - \mu_1)^\top (\check{\Sigma}_1 w_1)^{-1} (Z_i - \mu_1) \right. \right. \\ &\quad \left. \left. + \sum_{i=k+1}^n (Z_i - \mu_2) (\check{\Sigma}_2 w_2)^{-1} (Z_i - \mu_2) \right] \right\} \\ &\propto L_1(\check{\theta}_1, w_1, \nu_1 | D_1) L_2(\check{\theta}_2, w_2, \nu_2 | D_2) = L_1(\check{\theta}_1 | D_1) L_2(\check{\theta}_2 | D_2), \quad (5.2.3) \end{aligned}$$

onde $D_1 = \{Z_i, i = 1, \dots, k\} = Z^{(1)}$, $D_2 = \{Z_i, i = k+1, \dots, n\} = Z^{(2)}$,

$$\begin{aligned} L_1(\check{\theta}_1 | D_1) &= L_1(\theta_1, w_1, \nu_1 | D_1) \\ &\propto |\Sigma_1 w_1|^{-\frac{k}{2}} \exp\left\{-\frac{1}{2} \sum_{i=1}^k (Z_i - \mu_1)^\top (\Sigma_1 w_1)^{-1} (Z_i - \mu_1)\right\} \\ &= w_1^{-\frac{3k}{2}} (\sigma_{x(1)}^2)^{-\frac{k}{2}} (\sigma_1^2)^{-k} \exp\left\{-\frac{1}{2} \left[\frac{\sum_{i=1}^k (Y_i - \alpha_1 - \beta_1 x_i)^2}{\lambda \sigma_1^2 w_1} \right. \right. \\ &\quad \left. \left. + \frac{\sum_{i=1}^k (X_i - x_i)^2}{\sigma_1^2 w_1} + \frac{\sum_{i=1}^k (x_i - \mu_1)^2}{\sigma_{x(1)}^2 w_1} \right] \right\} \end{aligned}$$

é a verossimilhança para os dados completos, D_1 , e

$$L_2(\check{\theta}_2 | D_2) = L_2(\theta_2, w_2, \nu_2 | D_2), \text{ onde}$$

$$L_2(\theta_2, w_2, \nu_2 | D_2) \propto w_2^{-\frac{3(n-k)}{2}} (\sigma_{x(2)}^2)^{-\frac{(n-k)}{2}} (\sigma_2^2)^{(n-k)} \exp \left\{ -\frac{1}{2} \left[\sum_{i=k+1}^n \frac{(Y_i - \alpha_2 - \beta_2 x_i)^2}{\lambda \sigma_2^2 w_2} + \sum_{i=k+1}^n \frac{(X_i - x_i)^2}{\sigma_1^2 w_2} + \sum_{i=k+1}^n \frac{(x_i - \mu_2)^2}{\sigma_{x(2)}^2 w_2} \right] \right\}$$

é a verossimilhança para o conjunto de dados completos, D_2 , após a mudança.

• **Distribuição à posteriori**

Seja $\mathcal{X} = \{ \underline{x} = (x_1, \dots, x_n)^T : x_i \in \mathbb{R} \}$.

A distribuição à posteriori

$$p(\underline{\theta}^* | D_{\text{obs}}) = \int_{\mathcal{X}} p(\underline{x}, \underline{\theta}^* | D_{\text{obs}}) d\underline{x} \propto \int_{\mathcal{X}} L(\underline{\theta}^* | D) \pi(\underline{\theta}^*) d\underline{x}, \quad (5.2.4)$$

onde $L(\theta^* | D)$ é dada em (5.2.3) acima e $\pi(\underline{\theta}^*)$ é dada em (5.2.2).

5.2.2 Implementação do algoritmo de Gibbs no modelo com mudança em um subconjunto qualquer de parâmetros

Para implementar o algoritmo de Gibbs no modelo com mudança em todos os parâmetros, necessitamos das distribuições condicionais completas à posteriori de x_i (para $i = 1, \dots, n$), e de $\alpha_1, \alpha_2, \beta_1, \beta_2, \mu_1, \mu_2, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma_1^2, \sigma_2^2, \nu_1, \nu_2$ e k , que coincidem com as do modelo t -independente da Seção 5.1.2 quando substituímos w_i por $w_i = \begin{cases} w_1, & \text{se } 1 \leq i \leq k, \\ w_2, & \text{se } k+1 \leq i \leq n. \end{cases}$ As distribuições condicionais completas à posteriori de w_1 e w_2 , nesse caso, são dadas, respectivamente, por (5.2.5) e (5.2.6) a seguir.

$$IG \left(\frac{3k + \nu_1}{2}, \frac{1}{2} \left[\frac{\sum_{i=1}^k (Y_i - \alpha_1 - \beta_1 x_i)^2}{\lambda \sigma_1^2} + \frac{\sum_{i=1}^k (X_i - x_i)^2}{\sigma_1^2} + \frac{\sum_{i=1}^k (x_i - \mu_1)^2}{\sigma_{x(1)}^2} + \nu_1 \right] \right) \quad (5.2.5)$$

e

$$IG \left(\frac{3(n-k) + \nu_2}{2}, \frac{1}{2} \left[\frac{\sum_{i=1}^k (Y_i - \alpha_2 - \beta_2 x_i)^2}{\lambda \sigma_2^2} + \frac{\sum_{i=1}^k (X_i - x_i)^2}{\sigma_2^2} \right] \right)$$

$$\left. + \frac{\sum_{i=1}^k (x_i - \mu_2)^2}{\sigma_{x(2)}^2} + \nu_2 \right\} \quad (5.2.6)$$

No caso do modelo geral que permite mudança apenas em algum subconjunto de parâmetros estritamente contido em $\{\alpha, \beta, \mu, \sigma_x^2, \sigma^2\}$, precisamos ainda conhecer as distribuições condicionais completas de α do modelo com mudança em $\{\beta, \mu, \sigma_x^2, \sigma^2\}$, de β do modelo com mudança em $\{\alpha, \mu, \sigma_x^2, \sigma^2\}$, de μ do modelo com mudança em $\{\alpha, \beta, \sigma_x^2, \sigma^2\}$, de σ_x^2 do modelo com mudança em $\{\alpha, \beta, \mu, \sigma^2\}$ e de σ^2 do modelo com mudança em $\{\alpha, \beta, \mu, \sigma_x^2\}$. Essas distribuições coincidem com as da Seção 5.1.3 substituindo w_i por $w_i = \begin{cases} w_1, & \text{se } 1 \leq i \leq k, \\ w_2, & \text{se } k+1 \leq i \leq n. \end{cases}$ nas fórmulas (5.1.22) a (5.1.26).

Com estas distribuições, e fazendo as devidas restrições, implementamos o algoritmo de Gibbs para o modelo que permite mudança em qualquer subconjunto de parâmetros (como fizemos na Seção 4.3 e na Seção 5.1.3).

5.3 Detecção de mudança e estimação do modelo

Para detectar mudança nos parâmetros do modelo t -independente e t -dependente, utilizamos as mesmas técnicas da seção 4.4 adaptadas a esses modelos.

5.4 Condições para a existência da posteriori no modelo t -dependente sob diferentes escolhas de prioris impróprias

Nessa seção apresentamos teoremas que estabelecem condições para a existência da posteriori sob o modelo t -dependente com graus de liberdade conhecidos e prioris impróprias análogas as usadas na seção 3.5.

5.4.1 Condições no caso do modelo que permite mudança em todos os parâmetros

Nesta subsecção apresentamos alguns teoremas que garantem a existência da posteriori no modelo t -dependente que permite mudança em todos os parâmetros e tem priori imprópria do tipo:

Distribuição à priori

$$\pi(\check{\theta}) = p(k)\pi(\check{\theta}_1)\pi(\check{\theta}_2), \text{ onde } \pi(\check{\theta}_i) = \pi(\theta_i, w_i) = \pi(\theta_i)\pi(w_i), i = 1, 2,$$

onde $w_i \sim IG(\frac{\nu_i}{2}, \frac{\nu_i}{2})$, $i = 1, 2$, para $n \geq 3$, $p(k) = P[K = k]$ é uma distribuição discreta definida num conjunto \mathcal{K} finito e $\pi(\theta_i) \propto \pi(\sigma_{x^{(i)}}^2)\pi(\sigma_i^2)$, $i = 1, 2$, onde $\theta_i = (\alpha_i, \beta_i, \mu_i, \sigma_{x^{(i)}}^2, \sigma_i^2)^T$.

Distribuição à posteriori

No modelo com mudança em todos os parâmetros e priori $\pi(\check{\theta}) = p(k)\pi(\check{\theta}_1)\pi(\check{\theta}_2)$, a distribuição à posteriori $p(\check{\theta}|D_{\text{obs}})$ é proporcional a

$$\int_{\mathcal{X} = \{\mathcal{X} = (x_1, \dots, x_n) : x_i \in \mathbb{R}\}} L(\check{\theta}|D)\pi(\check{\theta})d\mathcal{X}, \text{ onde } L(\check{\theta}|D) \text{ é dado em (5.2.3)}$$

e portanto

$$\begin{aligned} p(\check{\theta}|D_{\text{obs}}) &\propto \int_{\mathcal{X}^{(1)}} L_1(\check{\theta}_1|D_1)\pi(\check{\theta}_1)d\mathcal{X}^{(1)} \int_{\mathcal{X}^{(2)}} L_2(\check{\theta}_2|D_2)\pi(\check{\theta}_2)d\mathcal{X}^{(2)} \cdot p(k), \\ &\propto p_1(\check{\theta}_1|D_{\text{obs}_1}^{(k)})p_2(\check{\theta}_2|D_{\text{obs}_2}^{(n-k)})p(k), \end{aligned}$$

onde $p_1(\check{\theta}_1|D_{\text{obs}_1}^{(k)})$ é a posteriori baseada nos dados até a mudança e $p_2(\check{\theta}_2|D_{\text{obs}_2}^{(n-k)})$ é a posteriori baseada nos dados após a mudança.

Observação:

Para mostrar que $p(\check{\theta}|D_{\text{obs}})$ é própria, devemos mostrar que

$$\sum_{k \in \mathcal{K}} p(k) \int_{W_1} \int_{\Theta_1} p(\check{\theta}_1|D_{\text{obs}_1}^{(k)})d\theta_1dw_1 \int_{W_2} \int_{\Theta_2} p_2(\check{\theta}_2|D_{\text{obs}_2}^{(n-k)})d\theta_2dw_2 < \infty. \quad (5.4.1)$$

onde $W_i = \mathbb{R}_+$, $i = 1, 2$, $\Theta_i = \{\theta_i = (\alpha_i, \beta_i, \mu_i, \sigma_{x(i)}^2, \sigma_i^2)^T : \alpha_i, \beta_i, \mu_i \in \mathbb{R} \text{ e } \sigma_{x(i)}^2, \sigma_i^2 \in \mathbb{R}_+\}$, $i = 1, 2$, e \mathcal{K} é um conjunto finito.

Como \mathcal{K} é finito, então para mostrar a existência da posteriori $p(\check{\theta} | D_{\text{obs}})$, basta mostrar que $p_1(\check{\theta}_1 | D_{\text{obs}1}^{(k)})$ e $p_2(\check{\theta}_2 | D_{\text{obs}2}^{(n-k)})$ são próprias para um k fixado.

Teoremas

Teorema 5.4.1 *No modelo t -dependente com graus de liberdade conhecidos, que permite mudança em todos os parâmetros e tem distribuição à priori dada por*

$$\pi(\check{\theta}) = p(k) \prod_{i=1}^2 \pi(\theta_i^*) = p(k) \prod_{i=1}^2 \pi(\theta_i) \pi(w_i),$$

onde $\pi(w_i) \sim IG(\frac{\nu_i}{2}, \frac{\nu_i}{2})$, $\nu_i \geq 3$, $i = 1, 2$, $\pi(\theta_i) \propto \pi(\sigma_{x(i)}^2) \pi(\sigma_i^2)$ (onde $\pi(\sigma_{x(i)}^2) \sim IG(c_i, d_i)$, $\pi(\sigma_i^2) \sim IG(f_i, g_i)$, $i = 1, 2$) e $p(k)$ é uma distribuição discreta num conjunto finito \mathcal{K} . Então uma condição suficiente para que a distribuição à posteriori $p(\check{\theta} | D_{\text{obs}})$ seja própria é que $k \geq 3$, $(n - k) \geq 3$, $0 < c_1 < k + \frac{(\nu_1 - 3)}{2}$ e $0 < c_2 < (n - k) + \frac{(\nu_2 - 3)}{2}$. Outra condição suficiente é que $k \geq 3$, $n - k \geq 3$, $0 < f_1 < \frac{k + \nu_1 - 1}{2}$ e $0 < f_2 < \frac{n - k + \nu_2 - 1}{2}$.

Demonstração: O Teorema 3.5.2 garante que para k fixado a integral

$$\int_{W_1} \int_{\Theta_1} p_1(\theta_1, w_1 | D_{\text{obs}}^{(k)}) d\theta_1 dw_1$$

converge se $k \geq 3$ e $0 < c_1 < k + \frac{(\nu_1 - 3)}{2}$ (ou então $k \geq 3$ e $0 < f_1 < \frac{k + \nu_1 - 1}{2}$) e a integral

$$\int_{W_2} \int_{\Theta_2} p_2(\theta_2, w_2 | D_{\text{obs}}^{(n-k)}) d\theta_2 dw_2$$

converge se $n - k \geq 3$ e $0 < c_2 < (n - k) + \frac{(\nu_2 - 3)}{2}$ (ou seja $n - k \geq 3$ e $0 < f_2 < \frac{n - k + \nu_2 - 1}{2}$) e daí segue a conclusão. ■

Teorema 5.4.2 *No modelo do Teorema 5.4.1 se a distribuição $\pi(\sigma_{x(i)}^2)$ for substituída por $\pi(\sigma_{x(i)}^2) \propto (\sigma_{x(i)}^2)^{-r_i}$, $r_i > 0$, $i = 1, 2$, então uma condição suficiente*

para a existência da correspondente posteriori do modelo com mudança é que $k \geq 2$, $n - k \geq 2$, $0 < r_i < \frac{1}{2}$ e $f_i < \frac{\nu_i}{2} + 1 - r_i$, $i = 1, 2$.

Demonstração: Como $p_1(\theta_1, w_1 | D_{\text{obs}1}^{(k)})$ e $p_2(\theta_2, w_2 | D_{\text{obs}2}^{(n-k)})$ estão sob as mesmas condições do Teorema 3.5.3 (que envolve o modelo t sem mudança), então para cada k fixado a integral

$$\int_{W_1} \int_{\Theta_1} p_1(\theta_1, w_1 | D_{\text{obs}1}^{(k)}) d\theta_1 dw_1$$

é finita se $k \geq 2$, $0 < r_1 < \frac{1}{2}$ e $0 < f_1 < \frac{\nu_1}{2} + 1 - r_1$ e a integral

$$\int_{W_2} \int_{\Theta_2} p_2(\theta_2, w_2 | D_{\text{obs}2}^{(k)}) d\theta_2 dw_2$$

é finita se $n - k \geq 2$, $0 < r_2 < \frac{1}{2}$ e $0 < f_2 < \frac{\nu_2}{2} + 1 - r_2$ e daí segue a conclusão. ■

Teorema 5.4.3 *No modelo do Teorema 5.4.1, se a distribuição de $\pi(\sigma_i^2)$ for substituída por $\pi(\sigma_i^2) \propto (\sigma_i^2)^{-s_i}$, $i = 1, 2$, e se os vetores de dados $\underline{X}^{(i)}$, $\underline{Y}^{(i)}$ são linearmente independentes para $i = 1, 2$ e tais que as componentes de $X^{(i)}$ (bem como as de $Y^{(i)}$) não são todas iguais, então a distribuição à posteriori do modelo com mudança correspondente é própria se $k \geq 3$, $n - k \geq 3$, $k + \nu_1 > 2s_1 - 1$ e $n - k + \nu_2 > 2s_2 - 1$.*

Demonstração: Analogamente, a demonstração segue usando a relação em (5.4.1) e o Teorema 3.5.4. ■

Teorema 5.4.4 *No modelo do Teorema 5.4.1 se as distribuições de $\pi(\sigma_{x(i)}^2)$ e $\pi(\sigma_i^2)$ forem substituídas respectivamente por $\pi(\sigma_{x(i)}^2) \propto (\sigma_{x(i)}^2)^{-r_i}$ e $\pi(\sigma_i^2) \propto (\sigma_i^2)^{-s_i}$ e os vetores de dados $\underline{X}^{(i)}$ e $\underline{Y}^{(i)}$ são linearmente independentes para $i = 1, 2$ e as componentes de $X^{(i)}$ (bem como as de $Y^{(i)}$) não são todas iguais, então a condição necessária e suficiente para a existência da correspondente posteriori é que $k + r_1 + s_1 > 3$, $(n - k) + r_2 + s_2 > 3$, $0 < r_i < \frac{1}{2}$ e $0 < s_i < \frac{\nu_i}{2} + 2 - r_i$, $i = 1, 2$.*

Demonstração: Usando a relação em (5.4.1) e o Teorema 3.5.5, a demonstração segue. ■

Observação: Nos teoremas 4.6.1 a 4.6.4, se considerarmos que os números de graus de liberdade não são conhecidos, mas são variáveis aleatórias discretas definidas em conjuntos finitos, as demonstrações seriam basicamente as mesmas.

5.4.2 Condições no caso do modelo com mudança apenas em σ_x^2

Teorema 5.4.5 *No modelo estrutural t -dependente que permite mudança apenas em σ_x^2 , com priori $\pi(\boldsymbol{\theta}) \propto p(k)\pi(w)\pi(\sigma_x^2) \prod_{i=1}^2 (\pi(\sigma_{x(i)}^2))$, onde $p(k) = P[K = k]$, $k \in \mathcal{K}$ (onde \mathcal{K} é finito); $\pi(\sigma_{x(i)}^2) \sim IG(c_i, d_i)$, $i = 1, 2$, $\pi(\sigma^2) \sim IG(f, g)$ e $\pi(w) \sim IG(\frac{\nu}{2}, \frac{\nu}{2})$, então uma condição suficiente para que a posteriori seja própria é que $n \geq 3$.*

Demonstração: O objetivo é mostrar que $\sum_{k \in \mathcal{K}} \int_W \int_{\Theta} p(k, \boldsymbol{\theta}, w | D_{\text{obs}}) d\boldsymbol{\theta} dw < \infty$, ou seja, $\sum_{k \in \mathcal{K}} \int_W \int_{\mathfrak{X}} \int_{\Theta} p(k, w, \boldsymbol{x}, \boldsymbol{\theta} | D_{\text{obs}}) d\boldsymbol{\theta} d\boldsymbol{x} dw < \infty$, onde $\Theta = \{\boldsymbol{\theta} = (\alpha, \beta, \mu, \sigma_{x(1)}^2, \sigma_{x(2)}^2, \sigma^2)^T \in \mathbb{R}^3 \times \mathbb{R}_+^3\}$, $W = \{w \in \mathbb{R}^+\}$, $\mathfrak{X} = \{\boldsymbol{x} = (x_1, \dots, x_n)^T : x_i \in \mathbb{R}\}$ e \mathcal{K} é um conjunto finito, onde

$p(k, w, \boldsymbol{x}, \boldsymbol{\theta} | D_{\text{obs}}) \propto L(k, w, \boldsymbol{\theta} | D) \pi(k, w, \boldsymbol{\theta})$, que é proporcional a

$$\begin{aligned} & w^{-(\frac{3n+\nu}{2}-\frac{1}{2})} (\sigma^2)^{-(n+f+1)} (\sigma_{x(1)}^2)^{-(\frac{k}{2}+c_1+1)} (\sigma_{x(2)}^2)^{-(\frac{n-k}{2}+c_2+1)} \exp \left\{ -\frac{1}{2} \left[\frac{\nu}{w} + \frac{2d_1}{\sigma_{x(1)}^2} \right. \right. \\ & + \frac{2d_2}{\sigma^2} + \frac{2g}{\sigma^2} + \frac{\sum_{i=1}^n (Y - \alpha - \beta x_i)^2}{\lambda \sigma^2 w} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma^2 w} + \frac{\sum_{i=1}^k (x_i - \mu)^2}{\sigma_{x(1)}^2 w} \\ & \left. \left. + \frac{\sum_{i=k+1}^n (x_i - \mu)^2}{\sigma_{x(1)}^2 w} \right] \right\}. \end{aligned} \quad (5.4.2)$$

Seja $f(k, w, \boldsymbol{x}, \boldsymbol{\theta}, D_{\text{obs}})$ a função que coincide com (5.4.2) quando substituimos a

última parcela, $\frac{\sum_{i=k+1}^n (x_i - \mu)^2}{\sigma_{x(1)}^2 w}$ por zero. Então

$$\int_W \int_{\mathfrak{X}} \int_{\Theta} p(k, w, \underline{x}, \theta | D_{\text{obs}}) d\theta d\underline{x} dw \leq \int_W \int_{\mathfrak{X}} \int_{\Theta} f(k, w, \underline{x}, \theta, D_{\text{obs}}) d\theta d\underline{x} dw,$$

onde

$$\begin{aligned} & \int_{\mathfrak{X}} \int_{\Theta} f(k, w, \underline{x}, \theta, D_{\text{obs}}) d\theta d\underline{x} \propto \\ & \int_{\mathfrak{X}} \frac{w^{-(\frac{3n+\nu-1}{2})} \exp\{-\frac{\nu}{2w}\}}{S_{xx}^{\frac{1}{2}} \left[2d + \frac{S_{x_1 x_1}}{w}\right]^{\frac{k+2c_1-1}{2}} \left[2g + \frac{\sum_{i=1}^n (X_i - x_i)^2}{w} + \frac{S_{xx} S_{YY} - S_{xY}^2}{\lambda S_{xx} w}\right]^{n+f-1}} d\underline{x} \\ & \leq \frac{1}{(2d)^{\frac{k+2c_1-1}{2}}} \int_{\mathfrak{X}} \frac{w^{-(\frac{3n+\nu-1}{2})} \exp\{-\frac{\nu}{2w}\}}{S_{xx}^{\frac{1}{2}} \left[2g + \frac{\sum_{i=1}^n (X_i - x_i)^2}{w} + \frac{S_{xx} S_{YY} - S_{xY}^2}{\lambda S_{xx} w}\right]^{n+f-1}} d\underline{x}. \end{aligned}$$

A integral $\int_{\mathfrak{X}} \frac{w^{-(\frac{3n+\nu-1}{2})} \exp\{-\frac{\nu}{2w}\}}{S_{xx}^{\frac{1}{2}} \left[2g + \frac{\sum_{i=1}^n (X_i - x_i)^2}{w} + \frac{S_{xx} S_{YY} - S_{xY}^2}{\lambda S_{xx} w}\right]^{n+f-1}} d\underline{x}$ pode ser analisada facilmente aplicando-se a mesma transformação de variáveis usada na seção 2.8 (que a cada $(\underline{x}, \underline{X}, \underline{Y})$ associa $(\underline{z}, \underline{b}, \underline{c}) = H(\underline{x}, \underline{X}, \underline{Y})$, onde H é a matriz de Helmert).

De fato, após a transformação, essa integral fica igual a

$$\begin{aligned} & \int_{\mathfrak{Z}} \frac{w^{-(\frac{3n+\nu-1}{2})} \exp\{-\frac{\nu}{2w}\}}{\|\underline{z}_x\| \left[2g + \frac{\|\underline{z}_x - \underline{b}_X\|^2}{w} + \frac{(\sqrt{n}\bar{X} - z_n)^2}{w} + \frac{\|\underline{c}_Y\|^2 \text{sen}^2 \theta}{\lambda w}\right]^{n+f-1}} d\underline{z} \\ & \propto \int_{\mathfrak{Z}_x} \frac{w^{-(\frac{3n+\nu-1}{2})} \exp\{-\frac{\nu}{2w}\}}{\|\underline{z}_x\| \left[2g + \frac{\|\underline{z}_x - \underline{b}_X\|^2}{w} + \frac{\|\underline{c}_Y\|^2 \text{sen}^2 \theta}{\lambda w}\right]^{n+f-\frac{3}{2}}} d\underline{z} \\ & \propto \int_{\mathfrak{Z}_x} \frac{\exp\{-\frac{\nu}{2w}\}}{w^{\frac{n+\nu}{2}-f+1} \|\underline{z}_x\| \left[2gw + \|\underline{z}_x - \underline{b}_X\|^2 + \frac{\|\underline{c}_Y\|^2 \text{sen}^2 \theta}{\lambda}\right]^{n+f-\frac{3}{2}}} d\underline{z}_x. \end{aligned}$$

Portanto, a integral da posteriori $\int_W \int_{\mathfrak{X}} \int_{\Theta} p(k, w, \underline{x}, \theta | D_{\text{obs}}) d\theta d\underline{x} dw$ é majorada por uma integral proporcional a

$$\int_W \int_{\mathfrak{Z}_x} \frac{\exp\{-\frac{\nu}{2w}\}}{w^{\frac{n+\nu}{2}-f+1} \|\underline{z}_x\| \left[2gw + \|\underline{z}_x - \underline{b}_X\|^2 + \frac{\|\underline{c}_Y\|^2 \text{sen}^2 \theta}{\lambda}\right]^{n+f-\frac{3}{2}}} d\underline{z}_x dw,$$

que é muito semelhante à integral que aparece na demonstração do Teorema 3.5.3.

A região de integração $W \times \mathfrak{Z}_x$ é particionada em $W \times R_1$ e $W \times R_1^c$, onde $R_1 = \{\underline{z}_x \in \mathbb{R}^{n-1} : \|\underline{z}_x\| < \varepsilon\}$.

AVALIAÇÃO DA INTEGRAL NA REGIÃO $W \times R_1$

Nessa região a integral pode ser majorada por

$$\int_W \int_{\|z_x\| < \varepsilon} \frac{\exp\left\{\frac{-\nu/2}{w}\right\}}{w^{\frac{n+\nu}{2}-f+1} \|z_x\| [2gw]^{n+f-\frac{3}{2}}} dz_x dw,$$

que é proporcional a

$$\int_0^\infty \frac{\exp\left\{\frac{-\nu/2}{w}\right\}}{w^{\frac{3n+\nu-1}{2}}} dw \int_{\|z_x\| < \varepsilon} \frac{1}{\|z_x\|} dz_x,$$

que é finita se e somente se $\frac{3n+\nu-1}{2} > 1$ e $1 < n-1$, isto é, $n \geq 3$.

AVALIAÇÃO DA INTEGRAL NA REGIÃO $W \times R_2$

Nessa região a integral pode ser majorada por

$$\int_W \int_{\|z_x\| < \varepsilon} \frac{\exp\left\{\frac{-\nu/2}{w}\right\}}{w^{\frac{n+\nu}{2}-f+1} \|z_x\| \left[\|z_x - \underline{b}_X\|^2 + \frac{\|\underline{c}_Y\|^2 \text{sen}^2 \theta}{\lambda} \right]^{n+f-\frac{3}{2}}} dz_x dw,$$

que é proporcional ao produto

$$\int_W \frac{\exp\left\{\frac{-\nu/2}{w}\right\}}{w^{\frac{3n+\nu}{2}-f+1}} dw \int_{\|z_x\| < \varepsilon} \frac{1}{\|z_x\| \left[\|z_x - \underline{b}_X\|^2 + \frac{\|\underline{c}_Y\|^2 \text{sen}^2 \theta}{\lambda} \right]^{n+f-\frac{3}{2}}} dz_x,$$

que é finita se e somente se $\frac{3n+\nu-1}{2} > 1$ e $1 + 2(n+f-\frac{3}{2}) > n-1$, isto é, $3n+\nu > 3$ e $n+2f > 1$.

Analisando simultaneamente essas regiões, concluímos que uma condição suficiente para que a posteriori seja própria é que $n \geq 3$. ■

Teorema 5.4.6 *No modelo do Teorema 5.4.5 se trocarmos a distribuição de $\pi(\sigma^2) \propto IG(f, g)$ por $\pi(\sigma^2) \propto (\sigma^2)^{-s}$, $s > 0$, e se \underline{X} e \underline{Y} forem linearmente independentes onde as componentes de \underline{X} e as de \underline{Y} não são todas iguais, então uma condição suficiente para a existência da posteriori é que $n \geq 3$.*

Demonstração: Substituindo f por $s - 1$ e g por zero na demonstração do teorema anterior, temos que a integral da posteriori é majorada por uma integral proporcional

a

$$\int_W \int_{Z_x} \frac{\exp\{-\nu/w\}}{w^{\frac{n+\nu}{2}-f+1} \|z_x\| \left[\|z_x - \tilde{b}_X\|^2 + \frac{\|C_Y\|^2 \sin^2 \theta}{\lambda} \right]^{n+s-\frac{5}{2}}} dz_x dw.$$

Essa integral é semelhante àquela que aparece na demonstração do Teorema 3.5.5. Para analisá-la particionamos a região de integração $W \times Z_x$ da mesma forma que no Teorema 3.5.5 e concluímos que ela é finita se $\frac{n+\nu-2f+2}{2} > 1$, $1 < n - 1$ e $1 + 2(n + s - \frac{5}{2}) > n - 1$, ou seja, $n \geq 3$ e $f < \frac{n+\nu}{2}$. ■

5.5 Distribuição preditiva e avaliação do ajuste

Podemos obter a distribuição preditiva e avaliar o ajuste do modelo t -independente com um ponto de mudança, de forma análoga ao modelo Normal com ponto de mudança da seção 4.5. Além disso, as seções 3.1.3 e 3.1.4 (que tratam do modelo t -independente sem ponto de mudança) são também úteis para ajudar na dedução das fórmulas correspondentes.

Capítulo 6

Considerações Finais

6.0.1 Resultados do estudo sobre ponto de mudança

Nesse trabalho fizemos uma análise bayesiana do modelo de regressão simples com erros de medida e que permite um único ponto de mudança numa seqüência de n observações, sendo que essa “mudança” pode ocorrer em qualquer um dos 2^5 ou 2^6 subconjuntos de parâmetros, dependendo se o modelo é identificável ou não identificável respectivamente, isto é, manteve-se a forma da distribuição das observações (normal ou t -Student) e permitiu-se mudança apenas nos parâmetros dessas distribuições.

Foram estudados basicamente dois modelos. O primeiro, em que o ponto de mudança k pode variar entre 1 e $n - 1$ e o segundo em que o ponto de mudança pode variar entre 1 e n . No primeiro modelo o caso em que não há mudança ocorre apenas quando $\theta_1 = \theta_2$, caso contrário pelo menos um parâmetro muda. No segundo modelo não há mudança quando $k = n$ (ou $\theta_1 = \theta_2$).

Nesse trabalho muitos resultados foram obtidos porém há também várias limitações e vários pontos que precisam ser ampliados ou mesmo completados.

1- Resultados no modelo Normal:

(a)- Algoritmos:

Para obter amostras das distribuições posterioris foram implementados apenas os

algoritmos de Gibbs e M. Collapsed Gibbs (muito embora o algoritmo de Metropolis-Hastings em Gibbs, visto no capítulo 2, também possa ser implementado com base nos dados observados). Observou-se que o algoritmo M.Collapsed Gibbs, baseado no modelo que permite mudança em todos os parâmetros com particular priori própria “pouco informativa” apresentou um melhor desempenho que o algoritmo de Gibbs simples para o particular conjunto de dados simulados utilizado, independentemente se o modelo é ou não é identificável. Além disso aparentemente não foram observadas falhas na convergência desses algoritmos quando os critérios de diagnóstico descritos no Apêndice foram empregados.

(b)- Estimação e métodos para detectar mudança

Para detectar se existe mudança no modelo identificável (ou não identificável), foram derivados três métodos:

- (a) O método 1 que usa a distribuição à posteriori do ponto de mudança k do modelo que permite mudança num subconjunto qualquer de parâmetros, onde $1 \leq k \leq n$. Este método detecta mudança e estima os parâmetros do modelo simultaneamente.
- (b) O método 2 que examina as distribuições marginais à posteriori das diferenças (ou quocientes) entre os parâmetros “antes” e “depois da mudança” no modelo que permite mudança em todos os parâmetros e $1 \leq k \leq n - 1$.
- (c) O método 3 que consiste em usar os critérios de seleção de modelos do Fator de Bayes, Pseudo-Fator de Bayes, Medida L e soma dos resíduos em valor absoluto (descritos na literatura), para comparar o modelo sem mudança com o modelo com mudança em qualquer subconjunto de parâmetros e $1 \leq k \leq n - 1$, usando o mesmo conjunto de dados.

Utilizando um particular conjunto de dados simulados (dados gerados a partir do modelo com mudança em todos os parâmetros após o vigésimo ponto de uma sequência de 50 observações) e escolhendo distribuições à priori “muito pouco

informativas” vimos que os métodos 1 e 3, implementados sob o modelo que permite mudança em todos os parâmetros, detectaram corretamente a existência de mudança independentemente do modelo ser ou não identificável. Entretanto é importante comentar que a medida L baseada no traço da matriz $L_2(\underline{Y}, \delta)$ não serviu para diferenciar entre os modelos com e sem mudança nesse exemplo; além disso, embora seja esperado que os valores das estimativas do Fator de Bayes sejam altos, as estimativas obtidas pelo método de Newton e Raftery (94) nos parecem muito grandes. Nesse caso, o método de Chib (95), descrito no capítulo 4, deveria ser implementado e as estimativas comparadas.

O método 2 não foi suficientemente preciso para detectar qual (quais) parâmetro(s) está(ão) mudando em alguns dos exemplos rodados, principalmente no caso não identificável.

No modelo do método 1, onde $1 \leq k \leq n$, foi feita uma análise descritiva da sensibilidade das estimativas a variações no valor do hiperparâmetro c (onde c é a probabilidade à priori de não haver mudança) com base em um particular conjunto de dados, e usando priori “ muito pouco informativa”. Essa análise sugeriu que as estimativas dos parâmetros da distr. à posteriori, bem como a distribuição marginal à posteriori de k , não são muito sensíveis à mudança no valor desse hiperparâmetro.

(c)- Ajuste

Para avaliar a qualidade do ajuste derivamos as fórmulas da densidade preditiva de validação cruzada (CPO) e dos resíduos bayesianos de validação cruzada para o modelo com mudança onde $1 \leq k \leq n - 1$. Nesse estudo, entretanto, a análise de resíduos está bastante limitada e deve ser mais explorada futuramente.

(d)- Identificabilidade

Para o particular conjunto de dados simulados da seção 4.7, os modelos com e sem restrição de identificabilidade com as correspondentes prioris “pouco informativas”

não diferiram muito com relação a qualidade do ajuste, além disso, embora os comprimentos dos intervalos de credibilidade no modelo identificável são ligeiramente menores, os intervalos de 95% de credibilidade contiveram os valores verdadeiros dos parâmetros em ambos os modelos. Assim, levando em conta também o fato de que na prática dificilmente o valor de $\lambda = \sigma^e / \sigma_u^2$ é conhecido, adotar o modelo não identificável seria uma boa opção nesse particular exemplo.

Seria interessante rodar mais exemplos com dados simulados e ver se este padrão tende a se manter.

(e)- Melhor modelo

Com a metodologia apresentada na seção 4.3 é possível amostrar facilmente de cada uma das 2^5 (ou 2^6 , no caso não identificável) posteriores associadas aos modelos com mudança em todos os possíveis subconjuntos de parâmetros. Desse modo podemos escolher o “melhor modelo”, usando por exemplo alguns dos critérios de seleção de modelos que compõe o método 3 acima.

(f)- Distribuição preditiva

Para gerar uma observação futura no modelo que permite mudança em todos os parâmetros basta simplesmente seguir os passos descritos na seção 4.5.1. O caso de mudança em apenas um subconjunto de parâmetros é totalmente análogo.

(g)- Teoremas

As condições para a existência da posteriori para certas prioris impróprias foram estudadas apenas para o modelo identificável. O estabelecimento das condições no caso do modelo não identificável é totalmente análogo (Na seção 2.8 este estudo é feito para o modelo sem mudança).

(h)- Exemplos

Os exemplos apresentados nesse trabalho basearam-se apenas em dados simulados. Um exemplo com dados reais, como por exemplo os dados de Chang e Huang

(97), complementar esse estudo.

2– Resultados no modelo *t*-Student:

Embora os resultados teóricos (semelhantes aos resultados do modelo normal), sobre os modelos *t*-dependente e *t*-independente, com número de graus de liberdade conhecidos ou desconhecidos, estejam descritos no capítulo 5, falta ainda complementar esse estudo com exemplos numéricos. Os programas computacionais a serem usados no caso *t* são basicamente os mesmos do caso normal, bastando fazer as devidas adaptações.

As condições para a existência da posteriori no modelo *t*, que foram desenvolvidas apenas para o caso *t*-dependente, podem ser estendidas (de forma análoga) para o caso *t*-independente tanto com pontos de mudança como sem pontos de mudança.

6.0.2 Perspectivas para trabalhos futuros

Nesse trabalho analisamos apenas o modelo com erros aditivos normais e *t*-Student. Esse estudo pode ser facilmente estendido para o caso em que os erros têm distribuição elíptica e também para o caso de regressão linear múltipla tanto com erros aditivos como multiplicativos.

A seguir listamos mais alguns tópicos que acreditamos ser de interesse para ampliar e dar continuidade aos resultados obtidos.

- Investigação de métodos bayesianos para analisar a influência local no modelo com erros nas variáveis com ou sem pontos de mudança.
- Análise bayesiana do modelo linear com erros de medida e pontos de mudança onde os erros tem distribuição assimétrica.
- Análise bayesiana do modelo linear de calibração comparativa com erros nas variáveis e pontos de mudança no caso em que os erros são normais e *t*-Student.

- Análise bayesiana do modelo linear generalizado com erros nas variáveis com ponto de mudança.
- Análise bayesiana de modelos não lineares com erros nas variáveis com e sem pontos de mudança.
- Estudo de um método global para detectar simultaneamente a presença de múltiplos pontos de mudança no modelo com erros nas variáveis (baseando-se no modelo PPM, “Partition Product Model”, introduzido por Hartigan(1990) e Barry e Hartigan(1992)).
- Uso de procedimentos seqüenciais para detectar mudança em modelos estruturais com erros nas variáveis com mudança de regime markoviano.

Apêndice

Seção 2.4 Métodos Computacionais Descritos na Literatura

2.4.1 Classificação geral

Na literatura existem vários procedimentos para aproximar quantidades de interesse na estatística bayesiana. Eles podem ser classificados basicamente em três grandes grupos: (1) os procedimentos de aproximação analíticos, (2) os procedimentos de simulação estáticos (ou “não iterativos”) e (3) os procedimentos de simulação dinâmicos (“iterativos”).

Os procedimentos analíticos nos permitem obter aproximações analíticas para a distribuição conjunta e para as distribuições marginais à posteriori, como também para o valor esperado de qualquer função de θ à posteriori (incluindo os momentos por exemplo). O método mais simples usa um resultado assintótico para aproximar a distribuição à posteriori por uma distribuição Normal em torno da moda da posteriori (“Teorema do Limite Central Bayesiano”). A moda frequentemente é obtida por meio de algoritmos de otimização numérica. Como Exemplo desses algoritmos temos o método de Newton, o método “BFGS” (Broyden, Fletcher, Goldfarb, Shanno, que é um método de Quase-Newton) e o algoritmo EM. No caso da distribuição ser multimodal, pode-se aproximar a posteriori localmente, em cada uma das modas, pelas Normais (ou t 's) correspondentes e tomar a distribuição da mistura dessas Normais (ou t 's) como aproximação global da distribuição à posteriori. (Gelman

et al., 1997, páginas 234 e 275). Outro procedimento, também baseado na teoria assintótica, é o método de Laplace (Tanner, 1993), conhecido principalmente por fornecer aproximações analíticas mais precisas (de segunda ordem) para o valor esperado de funções de θ à posteriori (como, por exemplo, os momentos, a própria função geradora de momentos e a densidade preditiva à posteriori), ele pode ser usado também para obter aproximações analíticas de segunda ordem para as distribuições marginais à posteriori (Carlin e Louis, 2000).

O principal problema com os métodos de aproximação pela Normal e de Laplace é que o número de parâmetros do modelo precisa ser pequeno para que esses métodos possam fornecer aproximações razoáveis (assim eles podem ser usados, por exemplo, para estudar subconjuntos de parâmetros à posteriori ou distribuições condicionais à posteriori). Esses métodos requerem ainda que o tamanho da amostra seja “grande” (e não sabemos quão grande é “grande”), requerem também que certas condições de regularidade estejam satisfeitas, além disso dependem também da parametrização (o método de Laplace, por exemplo, requer ainda que a posteriori seja unimodal).

Outro procedimento analítico alternativo para o cálculo de integrais na estatística bayesiana é utilizar integração numérica. Os métodos mais conhecidos de integração numérica são os da “Quadratura Gaussiana”, o “método de Romberg” (Lange, 1999) e o método “Quadratura de Gauss-Hermite” (Smith, 1987). Esses métodos, entretanto, são usados apenas quando o número de parâmetros do problema é pequeno.

Outra maneira de obter aproximações para as quantidades de interesse na inferência estatística bayesiana é usar simulação. Existem dois tipos de técnicas de simulação: as estáticas (ou “não iterativas”) e as dinâmicas (ou “iterativas”).

Nas técnicas estáticas geramos “amostras aleatórias” diretamente da distribuição à posteriori ou indiretamente por meio de uma distribuição conhecida que serve como intermediária e portanto a qualidade das aproximações pode ser melhorada sempre

que aumentarmos o número de elementos gerados.

Os métodos de simulação direta mais conhecidos são: o método da “grade de pontos” e o método da “aproximação trapezoidal” (Gelman et al., 2000). No método da “grade de pontos” a idéia é aproximar uma distribuição à posteriori contínua por uma discreta, calculando-se o valor da posteriori (ou de seu núcleo) nos pontos da grade e depois gerando uma amostra a partir da distribuição discreta correspondente.

Quando os valores da posteriori calculados nos pontos da grade são ligados por segmentos de reta no caso unidimensional, seções de um plano no caso bidimensional, etc. e uma amostra é gerada a partir dessa aproximação, obtemos o chamado “método trapezoidal”. O problema com essas técnicas é que se o espaço paramétrico não for bem pequeno elas são inviáveis para um estudo conjunto dos parâmetros.

Nos métodos de simulação indiretos, as amostras não são geradas diretamente da posteriori, mas sim de uma distribuição $g(\theta)$ conhecida (a função $g(\theta)$ é também chamada de “função importance sampling”).

Na literatura os métodos mais conhecidos de simulação indireta são: o “método de rejeição” e o método “weighted bootstrap”, que é bastante semelhante ao “sampling-importance-resampling” (Tanner, 1993; Geweke, 1989; Smith and Gelfand, 1992).

Quando o objetivo é apenas obter aproximações para o *valor esperado de funções de θ à posteriori* (e além disso não sabemos gerar amostras diretamente da posteriori mas sim de uma distribuição $g(\theta)$ conhecida), então podemos usar um método denominado “importance sampling” (Geweke, 1989). Esse procedimento, para essa finalidade é mais eficiente do que gerar uma amostra (usando qualquer um dos métodos mencionados acima) e depois obter as estimativas das quantidades desejadas (por exemplo, por meio do método simples de Monte Carlo para o cálculo de integrais).

Os três métodos acima apresentam limitações importantes nas aplicações, pois eles são eficientes apenas quando a função $g(\theta)$ (da qual geramos as amostras) for aproximadamente uniformemente proporcional ao núcleo da posteriori para todo θ do domínio da posteriori. O problema de achar uma distribuição g com essas características, e que ao mesmo tempo seja simples de ser amostrada, pode ser uma tarefa praticamente impossível se a dimensão do espaço de parâmetros for grande.

Como pudemos ver, todos os métodos de aproximação analítica como os de simulação não iterativa têm limitações nas aplicações. Gelman et al. (1997) sugerem o uso combinado desses procedimentos para se obter melhores aproximações. Um exemplo (descrito na página 312 de seu livro) considera primeiramente uma aproximação da distribuição à posteriori por uma mistura de Normais (ou t 's) a partir dessa distribuição aproximada, gera-se uma amostra de valores de θ e a seguir utiliza-se o “algoritmo de rejeição” ou o algoritmo “weighted bootstrap” para “filtrar” essa amostra. Note que aqui a distribuição da mistura está fazendo o mesmo papel da distribuição “importance sampling” $g(\theta)$ descrita anteriormente, outro método semelhante é combinar o “método trapezoidal” com o algoritmo de rejeição (esse método é descrito na página 305 de Gelman et al., 1997).

Segundo Gelman et al. (1997), mesmo utilizando essas combinações (de métodos), ainda assim as simulações obtidas a partir dessas aproximações podem não ser adequadas, dependendo do modelo que se está estudando. Em tais casos eles sugerem que a amostra gerada por esses métodos seja utilizada para se obter “bons” valores iniciais, que por sua vez são usados em procedimentos iterativos (ou dinâmicos) de simulação, como os métodos MCMC (“Markov Chain Monte Carlo”).

Um método de simulação MCMC tem por objetivo (poder) gerar uma cadeia de Markov cuja distribuição estacionária coincida com a distribuição alvo $\pi(\theta)$ da qual queremos amostrar (por exemplo, a posteriori). Usando as observações geradas por

essa cadeia podemos explorar aspectos de interesse sobre $\pi(\theta)$ como, por exemplo, os momentos, os quantis e as densidades marginais, como também podemos gerar amostra da distribuição preditiva à posteriori (com a qual podemos explorar certos aspectos dessa distribuição).

Existe na literatura uma quantidade muito grande de algoritmos do tipo MCMC (ver os livros de Chen, 2000, e Robert e Casella, 1999). Os algoritmos mais populares são o “amostrador de Gibbs” (Geman e Geman, 1984; Gelfand e Smith, 1990) e o algoritmo de Metropolis-Hastings (Metropolis et al., 1993; Hastings, 1970).

Neste trabalho analisamos o modelo com erros nas variáveis Normal e t -Student, com e sem mudança nos parâmetros, usando métodos MCMC como técnica principal (embora algumas das técnicas mencionadas anteriormente também sejam utilizadas para a obtenção de bons valores iniciais) . A razão dessa escolha (métodos MCMC) é que o número de “parâmetros” do modelo pode ser bem grande, especialmente quando se trabalha com a posteriori baseada nos dados completos, que envolve $n + p$ “parâmetros” onde n é o tamanho da amostra. (O valor de p no caso por exemplo do modelo de regressão linear simples com restrição de identificabilidade e com ponto de mudança é onze).

Embora possamos trabalhar com a “posteriori baseada nos dados observados”, que no caso da regressão simples com um ponto de mudança e restrição de identificabilidade envolve apenas 11 parâmetros (o que é relativamente pequeno), a idéia nesse trabalho é propor um método geral que possa também vir a ser usado em em regressão múltipla.

Os métodos MCMC são muito poderosos, não tendo restrições quanto ao número de parâmetros envolvidos e ao mesmo tempo são relativamente simples de serem implementados.

2.4.2 Métodos MCMC: idéias básicas

Quando uma distribuição $\pi(\theta)$, $\theta \in E \subset \mathbb{R}^p$ (por exemplo, a posteriori) é difícil de ser amostrada (ou porque p é muito grande, ou porque a própria distribuição é intratável) então uma solução para o problema seria obtida se pudéssemos construir uma cadeia de Markov homogênea e aperiódica com “espaço de estados” E e distribuição estacionária igual a $\pi(\theta)$, e daí então amostrar dessa cadeia. Isso porque se uma cadeia de Markov é irredutível e aperiódica a distribuição estacionária (ou distribuição invariante) é única e coincide com a distribuição limite (ou distribuição de equilíbrio), portanto, para n “suficientemente grande” a observação gerada na n -ésima iteração (da cadeia) “essencialmente” é uma observação da distribuição de $\pi(\theta)$ (assim como todas as observações geradas a partir da $(n + 1)$ -ésima iteração).

Para implementar a idéia acima temos na literatura os algoritmos denominados MCMC (“Markov-Chain-Monte Carlo”). Num algoritmo MCMC constrói-se um núcleo de transição de tal modo que a cadeia de Markov gerada por ele tenha distribuição estacionária igual a $\pi(\theta)$. Entretanto, para que a distribuição limite dessa cadeia exista e coincida com $\pi(\theta)$ (onde $\pi(\theta)$ é a única distribuição estacionária da cadeia), é necessário que ela seja irredutível e aperiódica. Na literatura muitos autores têm estabelecido condições para a convergência de vários desses algoritmos MCMC. Por exemplo, Roberts e Smith (1994) estabelecem condições suficientes relativamente simples sobre $\pi(\theta)$ de modo que as cadeias induzidas pelos algoritmos de Gibbs e Metropolis-Hastings (que tem $\pi(\theta)$ como distribuição estacionária) sejam irredutíveis e aperiódicas (e portanto converjam em distribuição para uma variável com distribuição $\pi(\theta)$). Outra referência é Roberts e Tweedie (1996) que mostra a ergodicidade geométrica de uma classe bem ampla de algoritmos do tipo Metropolis-Hastings. O problema é que mesmo quando condições teóricas estão “satisfeitas” pode não ocorrer convergência dessas cadeias. Por exemplo, uma cadeia pode ser tec-

nicamente irredutível mas na prática se comportar como se fosse redutível quando a probabilidade de transição de um determinado estado para outro é próxima de zero. Isso ocorre, por exemplo, quando a densidade $\pi(\theta)$ tem várias modas (máximos locais) e a massa da densidade entre essas modas (embora positiva) é bem pequena (nesse caso se o ponto inicial da cadeia pertence à região em torno de uma das modas e essa região tem uma massa substancial da densidade, então a cadeia tenderá a não sair dessa região).

Nesse trabalho não vamos explorar resultados teóricos sobre condições de convergência (como os de Roberts e Tweedie, 1996; Meyn e Tweedie, 1993; Roberts e Smith, 1994), mas vamos simplesmente fazer uma análise exploratória (a partir dos dados simulados) para verificar se houve falha na convergência desses algoritmos, utilizando diagnósticos de convergência. Na literatura existem muitas técnicas de diagnóstico com essa finalidade (ver Best, Cowler e Venes, 1996, Cowles e Carlson, 1996, e Mangelsen et al., 1999). É muito importante ressaltar entretanto que nenhuma dessas técnicas fornece uma “prova” de convergência (a começar pelo simples fato de que apenas um número finito de observações é analisado). Na subseção 2.4.6 apresentamos algumas das técnicas de diagnóstico mais conhecidas na literatura.

Os algoritmos MCMC mais conhecidos são os algoritmos de Gibbs e o Metropolis-Hastings. Entretanto, na literatura existe uma grande quantidade de métodos, incluindo variações dos algoritmos de Gibbs e Metropolis-Hastings, como também combinações desses algoritmos, como os métodos híbridos (ver Chen, 2000, capítulo 2, e Robert e Casella, 1999). Entretanto, nessa seção descrevemos apenas os métodos básicos de Gibbs e Metropolis-Hastings. Nas subseções 2.4.3 e 2.4.4 descrevemos duas variações do método de Gibbs, destinadas a acelerar a convergência do algoritmo original. o “grouped Gibbs” e o “collapsed Gibbs”, e na subseção 2.4.5 descrevemos o algoritmo híbrido de Metropolis-Hastings em Gibbs.

O algoritmo de Metrópolis-Hastings

Seja $\pi(\theta)$, $\theta \in A \subset \mathbb{R}^p$ a distribuição alvo da qual desejamos amostrar (por exemplo, a posteriori $p(\theta|D_{\text{obs}})$).

No algoritmo de Metropolis-Hastings, primeiramente escolhe-se arbitrariamente uma densidade (ou núcleo) de transição (auxiliar) em uma etapa $q(\theta_b|\theta_a)$, $\forall \theta_a, \theta_b$, de uma cadeia com espaço de estados igual ao domínio de $\pi(\theta)$ (o conjunto A). Esta densidade deve ser tal que na $(i+1)$ -ésima iteração do algoritmo possamos gerar $\check{\theta}$ a partir de $q(\cdot|\theta^{(i)} = \theta)$, $\forall \theta \in A$. Além disso, quanto mais próxima $q(\check{\theta}|\theta^{(i)} = \theta)$ estiver de $\pi(\check{\theta})$, mais eficiente será o algoritmo, ou seja, $q(\cdot|\theta^{(i)})$ faz um papel mais ou menos semelhante à função “importance sampling” citada na subseção 2.4.1.

Assim, em cada iteração do algoritmo de Metropolis-Hastings amostramos $\check{\theta}$ de $q(\cdot|\theta^{(i)})$ e tomamos

$$\theta^{(i+1)} = \begin{cases} \check{\theta}, & \text{com probabilidade } \alpha(\check{\theta}|\theta^{(i)}) = \min \left\{ \frac{\pi(\check{\theta})q(\theta^{(i)}|\check{\theta})}{\pi(\theta^{(i)})q(\check{\theta}|\theta^{(i)})}; 1 \right\}, \\ \theta^{(i)}, & \text{com probabilidade } 1 - \alpha(\check{\theta}|\theta^{(i)}). \end{cases} \quad (6.0.1)$$

Na prática, o algoritmo de M-H (Metropolis-Hastings) pode ser implementado da seguinte maneira:

Passo 0: Escolher um ponto inicial $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})^T$ e fazer $i = 0$.

Passo 1: Gerar um ponto $\check{\theta} = (\check{\theta}_1, \dots, \check{\theta}_p)^T$ da densidade de transição auxiliar $q(\cdot|\theta^{(i)})$ e gerar U de uma $U(0, 1)$.

Passo 2: Fazer $\alpha(\check{\theta}|\theta^{(i)}) = \min \left\{ \frac{\pi(\check{\theta})q(\theta^{(i)}|\check{\theta})}{\pi(\theta^{(i)})q(\check{\theta}|\theta^{(i)})}; 1 \right\}$.

Se $U < \alpha(\check{\theta}|\theta^{(i)}) \Rightarrow \theta^{(i+1)} = \check{\theta}$, caso contrário, $\theta^{(i+1)} = \theta^{(i)}$.

Passo 3: Fazer $i = i + 1$ e voltar ao passo 1.

Observações

1. $\alpha(\check{\theta}|\theta^{(i)})$ é a probabilidade de aceitar o valor $\check{\theta}$ na $(i + 1)$ -ésima iteração do algoritmo de M-H.
2. É bem simples ver que o algoritmo de M-H acima define uma cadeia de Markov com núcleo de transição (ou probabilidade de transição no caso discreto) do estado θ_a na i -ésima iteração para o estado θ_b na $(i + 1)$ -ésima iteração dada por

$$p(\theta_b|\theta_a) = \begin{cases} q(\theta_b|\theta_a)\alpha(\theta_b|\theta_a), & \text{se } \theta_b \neq \theta_a, \forall \theta_a, \theta_b \in A \subset \mathbb{R}^n, \\ 1 - \sum_{\theta \neq \theta_a} q(\theta|\theta_a)\alpha(\theta|\theta_a), & \text{se } \theta_a = \theta_b \end{cases}$$

3. Usando o núcleo de transição $p(\theta_b|\theta_a)$ acima, é simples também verificar que $\pi(\theta)$ satisfaz a condição de reversibilidade $\pi(\theta_a)p(\theta_b|\theta_a) = \pi(\theta_b)p(\theta_a|\theta_b)$, $\forall \theta_a, \theta_b$, o que implica que $\pi(\theta)$ é uma distribuição estacionária da cadeia definida por esse núcleo (ou seja, o algoritmo proposto induz uma CM com distribuição estacionária igual a $\pi(\theta)$).
4. Não é necessário conhecer a constante de normalização da distribuição alvo $\pi(\theta)$, já que ela se cancela na expressão entre chaves dada em (6.0.1).
5. Se $\pi(\theta_b) = q(\theta_b|\theta_a)$, $\forall \theta_b, \theta_a$, então $\alpha(\theta_b|\theta_a) = \min \left\{ \frac{\pi(\theta_b)q(\theta_a|\theta_b)}{\pi(\theta_a)q(\theta_b|\theta_a)}, 1 \right\} = 1$. Nesse caso, o valor $\check{\theta}$ gerado na $(i + 1)$ -ésima iteração pela densidade de transição auxiliar $q(\cdot|\theta^{(i)})$ é aceito com probabilidade $\alpha(\check{\theta}|\theta^{(i)}) = 1$ (isso quer dizer que se a densidade auxiliar coincide com a densidade alvo aceitamos o candidato $\check{\theta}$ com probabilidade 1).
6. No caso particular em que $q(\theta_b|\theta_a) = q(\theta_a|\theta_b)$, $\forall \theta_a, \theta_b$, caímos no algoritmo de Metropolis (Metropolis et al., 1953) e nesse caso $\alpha(\theta_b|\theta_a) = \min \left\{ \frac{\pi(\theta_b)}{\pi(\theta_a)}, 1 \right\}$ (e portanto $\alpha(\check{\theta}|\theta^{(i)}) = \min \left\{ \frac{\pi(\check{\theta})}{\pi(\theta^{(i)})}, 1 \right\}$).
7. No caso particular em que $q(\theta_b|\theta_a)$ não é função de θ_a , $\forall \theta_a, \theta_b$, ($q(\theta_b|\theta_a) =$

$h(\theta_b)$), caímos no algoritmo denominado “Independence Chain Metropolis-Hastings” (Tierney, 1994) e nesse caso

$$\alpha(\theta_b|\theta_a) = \min \left\{ \frac{\pi(\theta_b) h(\theta_a)}{\pi(\theta_a) h(\theta_b)}, 1 \right\}$$

(portanto, na $(i + 1)$ ésima iteração $\alpha(\check{\theta}|\theta^{(i)}) = \min \left\{ \frac{\pi(\check{\theta}) h(\theta^{(i)})}{\pi(\theta^{(i)}) h(\check{\theta})}, 1 \right\}$).

8. No caso particular em que $\pi(\theta)$ pode ser escrita como $\pi(\theta) \propto \psi(\theta).h(\theta)$, onde $h(\theta)$ é uma densidade que pode ser facilmente amostrada e $\psi(\theta)$ é uma função uniformemente limitada, então tomando-se a função de transição $q(\theta_b|\theta_a) = h(\theta_b), \forall \theta_a, \theta_b$, então

$$\alpha(\theta_b|\theta_a) = \min \left\{ \frac{\psi(\theta_b)h(\theta_b) h(\theta_a)}{\psi(\theta_a)h(\theta_a) h(\theta_b)}, 1 \right\} = \min \left\{ \frac{\psi(\theta_b)}{\psi(\theta_a)}, 1 \right\}$$

(e portanto na $(i + 1)$ -ésima iteração $\alpha(\check{\theta}|\theta^{(i)}) = \min \left\{ \frac{\psi(\check{\theta})}{\psi(\theta^{(i)})}, 1 \right\}$ (ver Chib e Greenberg, 1995)).

9. Outras escolhas para a função de transição auxiliar q são descritas em Chib e Greenberg (1995).
10. Quando o espaço (amostral) é contínuo e $\pi(\theta)$ é unimodal com caudas não muito pesadas, uma escolha conveniente de $q(\cdot|\theta^{(i)})$ é a distribuição Normal (pela simetria e pela simplicidade para gerar amostras).

- Assim, quando $q(\check{\theta}|\theta^{(i)})$ é $N(\check{\theta}|\theta^{(i)}, \hat{\Sigma})$ (onde $N(\check{\theta}|\theta^{(i)}, \hat{\Sigma})$ significa que $\check{\theta} \sim N(\theta^{(i)}, \hat{\Sigma})$), temos que

$$q(\check{\theta}|\theta^{(i)}) = \frac{1}{\sqrt{2\pi}} |\hat{\Sigma}| \exp \left\{ -\frac{1}{2} (\check{\theta} - \theta^{(i)})^\top \hat{\Sigma} (\check{\theta} - \theta^{(i)}) \right\}$$

e portanto

$$q(\theta^{(i)}|\check{\theta}) = \frac{1}{\sqrt{2\pi}} |\hat{\Sigma}|^{-1} \exp \left\{ -\frac{1}{2} (\theta^{(i)} - \check{\theta})^\top \hat{\Sigma}^{-1} (\theta^{(i)} - \check{\theta}) \right\}$$

(ou seja, $q(\underline{\theta}^{(i)}|\underline{\theta}^*)$ é $N(\underline{\theta}^{(i)}|\underline{\theta}^*, \hat{\Sigma})$); nesse caso, $\alpha(\theta^*|\theta^{(i)}) = \min\left\{\frac{\pi(\underline{\theta}^*)}{\pi(\underline{\theta}^{(i)})}, 1\right\}$, onde $\underline{\theta}^*$ é gerado de $q(\cdot|\theta^{(i)})$ que é $N(\underline{\theta}^{(i)}, \hat{\Sigma})$ e temos o algoritmo de Metropolis.

- Quando $q(\underline{\theta}^*|\theta^{(i)})$ é $N(\underline{\theta}^*|\hat{\theta}, \hat{\Sigma})$ então $q(\theta^{(i)}|\underline{\theta}^*)$ é $N(\theta^{(i)}|\hat{\theta}, \hat{\Sigma})$ e

$$\frac{q(\underline{\theta}^*|\theta^{(i)})}{q(\theta^{(i)}|\underline{\theta}^*)} = \frac{\exp\left\{-\frac{1}{2}(\underline{\theta}^* - \hat{\theta})^\top \hat{\Sigma}^{-1}(\underline{\theta}^* - \hat{\theta})\right\}}{\exp\left\{-\frac{1}{2}(\theta^{(i)} - \hat{\theta})^\top \hat{\Sigma}^{-1}(\theta^{(i)} - \hat{\theta})\right\}},$$

onde $\underline{\theta}^*$ é gerado de $N(\hat{\theta}, \hat{\Sigma})$. Nesse caso temos o algoritmo “Independence Metropolis-Hastings”.

Uma possível escolha para $\hat{\theta}$ é a moda de $\pi(\theta)$ (que pode ser obtida por algum método de otimização como o BFGS (Gill et al., 1981) ou o algoritmo EM). Uma possível escolha para $\hat{\Sigma}$ é $-\left[\frac{\partial^2 \log \pi(\theta)}{\partial \theta \partial \theta^\top}\right]^{-1}$ calculada no ponto $\hat{\theta}$.

- No caso em que $q(\underline{\theta}^*|\theta^{(i)})$ é $N(\underline{\theta}^*|\theta^{(i)}, \Sigma^{(i)})$ então $q(\theta^{(i)}|\underline{\theta}^*)$ é $N(\theta^{(i)}|\underline{\theta}^*, \Sigma^*)$ e

$$\frac{q(\underline{\theta}^*|\theta^{(i)})}{q(\theta^{(i)}|\underline{\theta}^*)} = \frac{|\Sigma^{(i)}| \exp\left\{-\frac{1}{2}(\underline{\theta}^* - \theta^{(i)})^\top \Sigma^{(i)-1}(\underline{\theta}^* - \theta^{(i)})\right\}}{|\Sigma^*|^{-1} \exp\left\{-\frac{1}{2}(\theta^{(i)} - \underline{\theta}^*)^\top \Sigma^{*-1}(\theta^{(i)} - \underline{\theta}^*)\right\}},$$

onde $\underline{\theta}^*$ é gerado de $q(\cdot|\theta^{(i)})$ que é $N(\theta^{(i)}, \Sigma^{(i)})$.

Uma possível escolha para $\Sigma^{(i)}$ e Σ^* é calcular $\left[-\frac{\partial^2 \log \pi(\theta)}{\partial \theta \partial \theta^\top}\right]^{-1}$ nos pontos $\theta^{(i)}$ e $\underline{\theta}^*$, respectivamente.

Quando as caudas de $\pi(\theta)$ forem “pesadas” outra alternativa conveniente para $q(\cdot|\theta^{(i)})$ é a distribuição t -Student.

11. Para a eficiência do algoritmo de M-H é importante escolher bem os parâmetros de dispersão da distribuição de transição auxiliar $q(\cdot|\theta^{(i)})$. No caso do algoritmo de Metropolis, Roberts et al. (1996) mostraram que se $\pi(\theta)$ e $q(\cdot|\theta^{(i)})$ são normais univariadas com parâmetros de dispersão Σ e $\hat{\Sigma}$, respectivamente, então, $\hat{\Sigma} = (2,4)^2 \Sigma$ é a escolha “ótima”, ou seja, que minimiza a

auto-correlação de primeira ordem da cadeia, e que conseqüentemente leva a uma taxa de aceitação de aproximadamente 44% (no caso multivariado os autores mostraram que essa escolha leva a uma taxa de aceitação de aproximadamente 23% para $p > 5$). Baseados nesse resultado, Gelman et al. (1997) sugerem que no caso não Normal iniciemos o processo usando $\hat{\Sigma} = (2, 4)^2 \left[\frac{\partial^2 \log \pi(\theta)}{\partial \theta \partial \theta} \Big|_{\theta = \hat{\theta}} \right]^{-1}$, onde $\hat{\theta}$ é a moda de $\pi(\theta)$, e após um certo número de iterações atualizemos $\hat{\Sigma}$ da seguinte forma: $\hat{\Sigma} = c\tilde{\Sigma}$, onde $\tilde{\Sigma}$ é a matriz de covariância à posteriori obtida a partir dos dados simulados e c é uma constante a ser ajustada de tal modo que a taxa de aceitação da cadeia $\alpha(\tilde{\theta} | \theta^{(i)})$ seja de aproximadamente 44% no caso unidimensional e 23% no caso multidimensional.

12. Nos casos mais complexos, escolher a densidade auxiliar de transição como uma mistura de normais (ou t 's) pode ser uma solução (por exemplo, se $\pi(\theta)$ é multimodal).
13. Para melhorar a estabilidade do algoritmo é melhor trabalhar com a escala logaritmica e ao invés de $\frac{\pi(\tilde{\theta})q(\theta^{(i)}|\tilde{\theta})}{\pi(\theta^{(i)})q(\tilde{\theta}|\theta^{(i)})}$ calculamos

$$\exp\{\ln[\pi(\tilde{\theta})q(\theta^{(i)}|\tilde{\theta})] - \ln[\pi(\theta^{(i)})q(\tilde{\theta}|\theta^{(i)})]\}.$$

O algoritmo original de Gibbs

Seja $\pi(\theta)$, $\theta \in A \subset \mathbb{R}^p$, a distribuição de interesse (por exemplo, a posteriori) e suponhamos que amostrar dessa distribuição é extremamente complicado, mas que as distribuições condicionais completas $\pi(\theta_i | \theta_{(-i)})$, $i = 1, \dots, p$ (ou $\pi(\theta_i | \theta_{(-i)}, D_{\text{obs}})$, no caso em que $\pi(\theta)$ é uma distribuição à posteriori), são distribuições totalmente

conhecidas e podem ser “facilmente” amostradas por algum método, então, nesse caso, o algoritmo de Gibbs é uma alternativa. A idéia de gerar dados das distribuições condicionais completas foi discutida pela primeira vez por Geman e Geman (1984).

O Algoritmo de Gibbs na sua forma mais simples é descrito a seguir.

Passo 0: Escolha um valor inicial $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_p^{(0)})^T$ e faça $i = 0$ (onde i é o contador das iterações);

Passo 1: obtenha um novo valor $\varrho^{(i+1)}$ a partir de $\varrho^{(i)}$ da seguinte maneira:

- gere $\theta_1^{(i+1)}$ de $\pi(\theta_1|\theta_2^{(i)}, \theta_3^{(i)}, \theta_4^{(i)}, \dots, \theta_p^{(i)})$,
- gere $\theta_2^{(i+1)}$ de $\pi(\theta_2|\theta_1^{(i+1)}, \theta_3^{(i)}, \theta_4^{(i)}, \dots, \theta_p^{(i)})$,
- gere $\theta_3^{(i+1)}$ de $\pi(\theta_3|\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_4^{(i)}, \dots, \theta_p^{(i)})$,
- \vdots
- gere $\theta_p^{(i+1)}$ de $\pi(\theta_p|\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{(i+1)}, \dots, \theta_{p-1}^{(i+1)})$,

Passo 2: faça $i = i + 1$ e volte ao passo 1 (até a convergência).

Observações

1. O passo 1 constitui um ciclo ou uma iteração do algoritmo de Gibbs.
2. As componentes θ_i podem ser escalares ou vetores.
3. $\pi(\theta_i|\theta_{(-i)})$ pode ser uma distribuição desconhecida mas deve ser facilmente amostrada (por exemplo, pelo algoritmo de rejeição adaptativo (Gilks, 1995), pelo algoritmo de M-H, ou por qualquer outro método).
4. O algoritmo original de Gibbs descrito acima define uma CM com núcleo de transição (ou probabilidade de transição) do estado $\varrho^{(i)}$ para o estado $\varrho^{(i+1)}$

dado por

$$p(\boldsymbol{\theta}^{(i+1)}|\boldsymbol{\theta}^{(i)}) = \prod_{j=1}^p \pi(\theta_j^{(i+1)}|\theta_{k<i}^{(i+1)}, \theta_{k>i}^{(i)}).$$

No caso particular de espaço de estados discreto e $p = 2$, escolhida uma ordem para o movimento, por exemplo atualizar primeiro a primeira componente θ_1 , depois a segunda componente θ_2 , é fácil ver que

$$\begin{aligned} p[\boldsymbol{\theta}^{(i+1)} = (c, d)^T | \boldsymbol{\theta}^{(i)} = (a, b)^T] &= p[(a, b) \rightarrow (c, b) \rightarrow (c, d)] \\ &= p[(c, d)|(c, b)] \cdot p[(c, b)|(a, b)] = p(\theta_2^{(i+1)} = d | \theta_1^{(i+1)} = c) \cdot p(\theta_1^{(i+1)} = c | \theta_2^{(i)} = b). \end{aligned}$$

5. Embora o núcleo de transição $p(\cdot|\cdot)$ acima geralmente não satisfaz a condição de reversibilidade, ele define uma cadeia que tem $\pi(\boldsymbol{\theta})$ como uma distribuição estacionária. Sem perda de generalidade, mostramos esse fato apenas para o caso $p = 2$:

Demonstração:

$$\begin{aligned} &\int_{\Theta_1} p(\boldsymbol{\theta}^{(i+1)}|\boldsymbol{\theta}^{(i)})\pi(\boldsymbol{\theta}^{(i)})d\boldsymbol{\theta}^{(i)} \\ &= \int_{\Theta_1} \int_{\Theta_2} \pi(\theta_2^{(i+1)}|\theta_1^{(i+1)})\pi(\theta_1^{(i+1)}|\theta_2^{(i)})\pi(\theta_1^{(i)}, \theta_2^{(i)})d\theta_1^{(i)}d\theta_2^{(i)} \\ &= \pi(\theta_2^{(i+1)}|\theta_1^{(i+1)}) \int_{\Theta_1} \pi(\theta_1^{(i+1)}|\theta_2^{(i)})\pi(\theta_2^{(i)})d\theta_2^{(i)} \\ &= \pi(\theta_2^{(i+1)}|\theta_1^{(i+1)}) \int_{\Theta_1} \pi(\theta_1^{(i+1)}, \theta_2^{(i)})d\theta_2^{(i)} = \pi(\theta_2^{(i+1)}|\theta_1^{(i+1)})\pi(\theta_1^{(i+1)}) \\ &= \pi(\theta_1^{(i+1)}, \theta_2^{(i+1)}) = \pi(\boldsymbol{\theta}^{(i+1)}). \end{aligned}$$

6. Existe na literatura, vários algoritmos que são variações do algoritmo de Gibbs (ver Chen, 2000, Carlin e Louis, 2000, e Robert e Casella, 1999).

2.4.3 O algoritmo de Gibbs em Blocos (“Grouped Gibbs”)

Quando o algoritmo original de Gibbs converge lentamente devido as fortes correlações entre os parâmetros do modelo ou devido as autocorrelações entre as it-

erações do algoritmo para cada parâmetro, pode-se tomar certas medidas para melhorar o desempenho desse algoritmo como, por exemplo, reparametrizar o modelo ou usar variáveis auxiliares, como também utilizar variações do algoritmo original como, por exemplo, os métodos “grouped Gibbs” (Liu, 1994, e Roberts e Sahu, 1997), “collapsed Gibbs” (Liu, 1994) e “covariance-adjusted MCMC” (Liu et al., 1994), dentre outros.

No algoritmo original de Gibbs, amostramos cada componente θ_i (do vetor de parâmetros $\boldsymbol{\theta}$), de sua respectiva distribuição condicional completa. No “Gibbs em blocos” particionamos as componentes de $\boldsymbol{\theta}$ em grupos (de um ou mais componentes); e os grupos que tiverem mais de uma componente são amostrados conjuntamente (de suas respectivas distribuições condicionais completas conjuntas). Segundo Chen (2000), o procedimento de amostrar em blocos traz benefícios e segundo Gemermann (1996), quando amostramos os parâmetros conjuntamente deixa-se de fazer a cadeia percorrer ao longo dos eixos para fazê-la percorrer trajetórias ditadas pela distribuição condicional completa do bloco. Segundo Roberts e Sahu (1997), nem sempre o uso de blocos traz benefícios e eles mostram alguns exemplos em que o uso desse método é ruim. A seguir ilustramos o método por meio de um exemplo.

Seja $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)^T$ e a distribuição alvo $\pi(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|D_{\text{obs}})$ (onde $p(\boldsymbol{\theta}|D_{\text{obs}})$ é a posteriori de $\boldsymbol{\theta}$) e suponha que sabemos amostrar $(\theta_1, \theta_2, \theta_3)$ conjuntamente de $\pi(\theta_1, \theta_2, \theta_3|\theta_4, \theta_5, \theta_6, D_{\text{obs}})$ como também (θ_4, θ_5) de $\pi(\theta_4, \theta_5|\theta_1, \theta_2, \theta_3, \theta_6, D_{\text{obs}})$ e θ_6 de $\pi(\theta_6|\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, D_{\text{obs}})$, então teríamos o seguinte algoritmo:

Passo 0: escolha $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}, \theta_4^{(0)}, \theta_5^{(0)}, \theta_6^{(0)})^T$ e faça $i = 0$.

Passo 1: gere $(\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{(i+1)})$ de $\pi(\theta_1, \theta_2, \theta_3|\theta_4^{(i)}, \theta_5^{(i)}, \theta_6^{(i)}, D_{\text{obs}})$.

Passo 2: gere $(\theta_4^{(i+1)}, \theta_5^{(i+1)})$ de $\pi(\theta_4, \theta_5|\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{(i+1)}, \theta_6^{(i)}, D_{\text{obs}})$.

Passo 3: gere $\theta_6^{(i+1)}$ de $\pi(\theta_6|\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{(i+1)}, \theta_4^{(i+1)}, \theta_5^{(i+1)}, \theta_6^{(i)}, D_{\text{obs}})$.

Passo 4: faça $i = i + 1$ e volte ao passo 1, até a convergência.

Observação: É fácil ver que a CM induzida pelo algoritmo de Gibbs em blocos tem $\pi(\theta)$ como distribuição invariante. A demonstração para o caso particular em que $\theta = (\theta_1, \theta_2, \theta_3)^T$, onde primeiramente (θ_1, θ_2) é amostrado de $\pi(\theta_1, \theta_2 | \theta_3, D_{\text{obs}})$ e depois θ_3 é amostrado de $\pi(\theta_3 | \theta_1, \theta_2, D_{\text{obs}})$ é dado abaixo:

demonstração;

O núcleo de transição de $\theta^{(i)}$ para $\theta^{(i+1)}$ é dado por

$$p(\theta^{(i+1)} | \theta^{(i)}) = \pi(\theta_1^{(i+1)}, \theta_2^{(i+1)} | \theta_3^{(i)}, D_{\text{obs}}) \pi(\theta_3^{(i+1)} | \theta_1^{(i+1)}, \theta_2^{(i+1)}, D_{\text{obs}}).$$

Nesse caso,

$$\begin{aligned} & \int_{\Theta_1} \int_{\Theta_2} \int_{\Theta_3} p(\theta^{(i+1)} | \theta^{(i)}) \pi(\theta^{(i)} | D_{\text{obs}}) d\theta^{(i)} \\ &= \int_{\Theta_1} \int_{\Theta_2} \int_{\Theta_3} \pi(\theta_1^{(i+1)}, \theta_2^{(i+1)} | \theta_3^{(i)}, D_{\text{obs}}) \pi(\theta_3^{(i+1)} | \theta_1^{(i+1)}, \theta_2^{(i+1)}, D_{\text{obs}}) \\ & \quad \cdot \pi(\theta_1^{(i)}, \theta_2^{(i)}, \theta_3^{(i)} | D_{\text{obs}}) d\theta_1^{(i)} d\theta_2^{(i)} d\theta_3^{(i)} \\ &= \pi(\theta_3^{(i+1)} | \theta_1^{(i+1)}, \theta_2^{(i+1)}, D_{\text{obs}}) \int_{\Theta_3} \pi(\theta_1^{(i+1)}, \theta_2^{(i+1)} | \theta_3^{(i)}, D_{\text{obs}}) \pi(\theta_3^{(i)} | D_{\text{obs}}) d\theta_3 \\ &= \pi(\theta_3^{(i+1)} | \theta_1^{(i+1)}, \theta_2^{(i+1)}, D_{\text{obs}}) \int_{\Theta_3} \pi(\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{(i)} | D_{\text{obs}}) d\theta_3 \\ &= \pi(\theta_3^{(i+1)} | \theta_1^{(i+1)}, \theta_2^{(i+1)}, D_{\text{obs}}) \pi(\theta_1^{(i+1)}, \theta_2^{(i+1)} | D_{\text{obs}}) = \pi(\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{(i+1)} | D_{\text{obs}}) \\ &= \pi(\theta^{(i+1)} | D_{\text{obs}}). \end{aligned}$$

2.4.4 O algoritmo “Collapsed Gibbs”

O algoritmo “collapsed Gibbs” foi proposto por Liu (1994) com o objetivo de acelerar a convergência e o “mixing” do amostrador de Gibbs. Este método é uma variação do algoritmo original de Gibbs e para ilustrar como ele funciona usamos o seguinte exemplo:

Seja $\theta = (\theta_1, \theta_2, \theta_3)^T$ e a distribuição alvo $\pi(\theta) = \pi(\theta | D_{\text{obs}})$ (a posteriori de θ).

Se, por exemplo, pudermos eliminar por integração a componente θ_3 de $\pi(\theta_1, \theta_2, \theta_3 | D_{\text{obs}})$ (ou seja, obter a marginal à posteriori $\pi(\theta_1, \theta_2 | D_{\text{obs}})$), então os passos do algoritmo são os seguintes:

Passo 0: faça $i = 0$, escolha $\varrho^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)})^T$;

Passo 1: gerar $(\theta_1^{(i+1)}, \theta_2^{(i+1)})$ de $\pi(\theta_1, \theta_2 | D_{\text{obs}})$;

Passo 2: gerar $\theta_3^{(i+1)}$ de $\pi(\theta_3 | \theta_1^{(i+1)}, \theta_2^{(i+1)}, D_{\text{obs}})$;

Passo 3: fazer $i = i + 1$ e voltar para o passo 1.

No caso em que não sabemos amostrar diretamente (θ_1, θ_2) simultaneamente, podemos usar uma versão desse algoritmo denominada “modified collapsed Gibbs sampler” e, nesse caso, no passo 1 podemos usar um mini-Gibbs para amostrar $(\theta_1^{(i+1)}, \theta_2^{(i+1)})$ de $\pi(\theta_1, \theta_2 | D_{\text{obs}})$, Liu, 1994, menciona que nesse caso a cadeia induzida pelo mini-Gibbs deve convergir antes de prosseguir para o passo 2 e o algoritmo seria o seguinte, em cada iteração:

Passo 1: gerar (θ_1, θ_2) de $\pi(\theta_1, \theta_2 | D_{\text{obs}})$ da seguinte maneira:

- gerar θ_2 de $\pi(\theta_2 | \theta_1, D_{\text{obs}})$,
- gerar θ_1 de $\pi(\theta_1 | \theta_2, D_{\text{obs}})$;

Passo 2: gerar θ_3 de $\pi(\theta_3 | \theta_1, \theta_2, D_{\text{obs}})$.

Mais precisamente, teríamos os seguintes passos:

Passo 0: faça $i = 0$, faça $j = 0$ e escolha $\theta_{1j}^{(0)}$.

Passo 1: gere $(\theta_{1,j+1}^{(i+1)}, \theta_{2,j+1}^{(i+1)})$ da seguinte maneira:

- gere $\theta_{2,j+1}^{(i+1)}$ de $\pi(\theta_2 | \theta_{1,j}^{(i)}, D_{\text{obs}})$,
- gere $\theta_{1,j+1}^{(i+1)}$ de $\pi(\theta_1 | \theta_{2,j+1}^{(i+1)}, D_{\text{obs}})$;

Passo 2: faça $j = j + 1$, se $j < N$ volte ao passo 1 (onde N é fixado e suficientemente grande);

Passo 3: gere $\theta_3^{(i+1)}$ de $\pi(\theta_3 | \theta_{1,N}^{(i+1)}, \theta_{2,N}^{(i+1)})$;

Passo 5: fazer $i = i + 1$ e volte ao passo 1, até a convergência.

Observações

1. O algoritmo “collapsed Gibbs” induz uma CM onde $\pi(\underline{\theta})$ é uma distribuição invariante (ou estacionária). É muito simples demonstrar isso no exemplo acima

demonstração

Temos nesse caso que o núcleo de transição $p(\underline{\theta}^{(i+1)} | \underline{\theta}^{(i)})$ do estado $\underline{\theta}^{(i)}$ para o estado $\underline{\theta}^{(i+1)}$ é dado por

$$p(\underline{\theta}^{(i+1)} | \underline{\theta}^{(i)}) = \pi(\theta_1^{(i+1)}, \theta_2^{(i+1)} | D_{\text{obs}}) \pi(\theta_3^{(i+1)} | \theta_1^{(i+1)}, \theta_2^{(i+1)}, D_{\text{obs}})$$

e portanto

$$\begin{aligned} & \int_{\Theta_1} \int_{\Theta_2} \int_{\Theta_3} p(\underline{\theta}^{(i+1)} | \underline{\theta}^{(i)}) \pi(\underline{\theta}^{(i)} | D_{\text{obs}}) d\underline{\theta}^{(i)} \\ &= \int_{\Theta_1} \int_{\Theta_2} \int_{\Theta_3} \pi(\theta_1^{(i+1)} \theta_2^{(i+1)} | D_{\text{obs}}) \pi(\theta_3^{(i+1)} | \theta_1^{(i+1)}, \theta_2^{(i+1)}, D_{\text{obs}}) \\ & \quad \cdot \pi(\theta_1^{(i)}, \theta_2^{(i)}, \theta_3^{(i)} | D_{\text{obs}}) d\theta_1^{(i)} d\theta_2^{(i)} d\theta_3^{(i)} \\ &= \pi(\theta_1^{(i+1)} \theta_2^{(i+1)} | D_{\text{obs}}) \pi(\theta_3^{(i+1)} | \theta_1^{(i+1)}, \theta_2^{(i+1)}, D_{\text{obs}}) \\ &= \pi(\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{(i+1)} | D_{\text{obs}}) = \pi(\underline{\theta}^{(i+1)} | D_{\text{obs}}) \end{aligned}$$

2. O algoritmo modificado também induz uma CM com uma distribuição estacionária igual a $\pi(\underline{\theta})$, a demonstração desse fato para o mesmo exemplo acima segue abaixo.

Demonstração: Seja

$$p(\underline{\theta}^{(i+1)}|\underline{\theta}^{(i)}) = \pi(\theta_1^{(i+1)}|\theta_2^{(i)}, D_{\text{obs}})\pi(\theta_2^{(i+1)}|\theta_1^{(i+1)}, D_{\text{obs}})\pi(\theta_3^{(i+1)}|\theta_1^{(i+1)}, \theta_2^{(i+1)}, D_{\text{obs}})$$

o núcleo de transição de $\underline{\theta}^{(i)}$ para $\underline{\theta}^{(i+1)}$. Então,

$$\begin{aligned} & \int_{\Theta_1} \int_{\Theta_2} \int_{\Theta_3} p(\underline{\theta}^{(i+1)}|\underline{\theta}^{(i)})\pi(\underline{\theta}^{(i)}|D_{\text{obs}})d\underline{\theta}^{(i)} \\ &= \int_{\Theta_1} \int_{\Theta_2} \int_{\Theta_3} p(\underline{\theta}^{(i+1)}|\underline{\theta}^{(i)})\pi(\underline{\theta}^{(i)}|D_{\text{obs}})d\theta_1^{(i)}d\theta_2^{(i)}d\theta_3^{(i)} \\ &= \int_{\Theta_1} \int_{\Theta_2} \int_{\Theta_3} \pi(\theta_1^{(i+1)}|\theta_2^{(i)}, D_{\text{obs}})\pi(\theta_2^{(i+1)}|\theta_1^{(i+1)}, D_{\text{obs}})\pi(\theta_3^{(i+1)}|\theta_1^{(i+1)}, \theta_2^{(i+1)}, D_{\text{obs}}) \\ & \quad \cdot \pi(\theta_1^{(i)}, \theta_2^{(i)}, \theta_3^{(i)}|D_{\text{obs}})d\theta_1^{(i)}d\theta_2^{(i)}d\theta_3^{(i)} \\ &= \int_{\Theta_1} \int_{\Theta_2} \pi(\theta_1^{(i+1)}|\theta_2^{(i)}, D_{\text{obs}})\pi(\theta_2^{(i+1)}|\theta_1^{(i+1)}, D_{\text{obs}})\pi(\theta_3^{(i+1)}|\theta_1^{(i+1)}, \theta_2^{(i+1)}, D_{\text{obs}}) \\ & \quad \cdot \pi(\theta_1^{(i)}, \theta_2^{(i)}|D_{\text{obs}})d\theta_1^{(i)}d\theta_2^{(i)} \\ &= \int_{\Theta_2} \pi(\theta_1^{(i+1)}|\theta_2^{(i)}, D_{\text{obs}})\pi(\theta_2^{(i+1)}|\theta_1^{(i+1)}, D_{\text{obs}})\pi(\theta_3^{(i+1)}|\theta_1^{(i+1)}, \theta_2^{(i+1)}, D_{\text{obs}})\pi(\theta_2^{(i)}|D_{\text{obs}})d\theta_2^{(i)} \\ &= \pi(\theta_3^{(i+1)}|\theta_1^{(i+1)}, \theta_2^{(i+1)}, D_{\text{obs}})\pi(\theta_2^{(i+1)}|\theta_1^{(i+1)}, D_{\text{obs}}) \int_{\Theta_2} \pi(\theta_1^{(i+1)}|\theta_2^{(i)}, D_{\text{obs}})\pi(\theta_2^{(i)}|D_{\text{obs}})d\theta_2^{(i)} \\ &= \pi(\theta_3^{(i+1)}|\theta_1^{(i+1)}, \theta_2^{(i+1)}, D_{\text{obs}})\pi(\theta_2^{(i+1)}|\theta_1^{(i+1)}, D_{\text{obs}}) \int_{\Theta_2} \pi(\theta_1^{(i+1)}, \theta_2^{(i)}|D_{\text{obs}})d\theta_2^{(i)} \\ &= \pi(\theta_3^{(i+1)}|\theta_1^{(i+1)}, \theta_2^{(i+1)}, D_{\text{obs}})\pi(\theta_2^{(i+1)}|\theta_1^{(i+1)}, D_{\text{obs}})\pi(\theta_1^{(i+1)}|D_{\text{obs}}) \\ &= \pi(\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{(i+1)}|D_{\text{obs}}) = \pi(\underline{\theta}^{(i+1)}|D_{\text{obs}}) \end{aligned}$$

3. Segundo Liu (1994), uma vantagem de eliminar componentes por integração (“collapsed down components”) é que em geral as auto-correlações entre as iterações são reduzidas. No exemplo acima vemos que $\theta_3^{(i+1)}$ depende de $\theta_1^{(i+1)}$ e $\theta_2^{(i+1)}$, mas $\theta_1^{(i+1)}$ e $\theta_2^{(i+1)}$ não dependem de $\theta_3^{(i)}$.
4. Segundo Chen (2000), e de acordo com Liu (1994), o collapsed Gibbs “funciona melhor” que o “grouped Gibbs”, enquanto que o “grouped” é melhor que o Gibbs original. Também espera-se que o “collapsed Gibbs” seja melhor do que o “modified collapsed Gibbs” e que este por sua vez seja melhor do que o Gibbs original. Porém, entre o “grouped” e o “modified collapsed Gibbs”, é

difícil dizer qual é o melhor e vai depender das correlações dentro dos grupos (ou blocos) (se elas forem “altas” espera-se que o “grouped” funcione melhor).

2.4.5 O algoritmo Híbrido de Metropolis-Hastings em Gibbs

Seja $\pi(\theta)$ a distribuição da qual queremos amostrar e suponhamos que nem todas as distribuições condicionais completas, $\pi(\theta_i|\theta_{(-i)})$, possam ser amostradas facilmente (por exemplo, quando apenas os seus núcleos são conhecidos). Em princípio podemos utilizar qualquer um dos métodos de simulação estática descritos na seção 2.4.1 para amostrar dessas distribuições mais complicadas (de serem amostradas) e assim completar o ciclo do amostrador de Gibbs. O problema entretanto é que esses métodos são numericamente intensivos e, ao serem re-utilizados em cada uma das iterações do algoritmo principal de Gibbs, leva a um “custo” computacional muito grande no final (sem contar o fato que em cada iteração do algoritmo de Gibbs amostramos um grande número de observações para salvar apenas uma observação, e também há o problema do “re-envelopamento” em cada iteração quando usamos o método de rejeição ou o método “weighted bootstrap”).

Uma solução elegante para amostrar das condicionais completas complicadas (sendo ou não log-côncavas), especialmente quando o número dessas condicionais é relativamente grande, é utilizar o algoritmo de Metropolis-Hastings (M-H). Nesse caso, geramos sub-cadeias de M-H dentro do ciclo amostrador de Gibbs. Por exemplo, se $\pi(\theta_j|\theta_{(-j)})$ é difícil de ser amostrada, então na $(i + 1)$ -ésima iteração do algoritmo principal de Gibbs geramos $\theta_j^{(i+1)}$ de $\pi(\theta_j|\theta_1^{(i)}, \dots, \theta_{j-1}^{(i)}, \theta_{j+1}^{(i-1)}, \dots, \theta_p^{(i-1)})$ por meio do algoritmo de M-H (onde $\theta_j^{(i+1)}$ é o resultado da última iteração de M-H). Esse procedimento é conhecido como “Metropolis-Hastings em Gibbs”, ou também como “passos de Metropolis-Hastings” ou “Localized Metropolis-Hastings”.

Observações

1. Esse método híbrido é tal que a cadeia de Markov induzida por ele tem $\pi(\underline{\theta})$ (a distribuição alvo) como uma distribuição estacionária ou invariante (e portanto para haver convergência da cadeia para essa distribuição é necessário que a cadeia seja irredutível e aperiódica).
2. O número de iterações T do sub-algoritmo de M-H é arbitrário. Entretanto, a escolha de um valor muito grande para T (digamos 10.000) é puro desperdício, especialmente nas primeiras iterações do algoritmo principal de Gibbs, pois estaríamos ainda longe da distribuição estacionária. Na prática é muito comum usar $T = 1$, entretanto, não existe uma recomendação de um valor de T ideal. Intuitivamente, concordamos que nas primeiras iterações do algoritmo principal (“burn in period”), T deveria ser pequeno, mas à medida que o número de iterações aumenta, T deveria também aumentar.

A seguir, por meio de um exemplo, apresentamos os passos do algoritmo de Metropolis-Hastings em Gibbs na sua forma mais simples (isto é, no caso em que o número de iterações, T , dos sub-algoritmos de M-H para cada iteração do algoritmo principal de Gibbs é $T = 1$).

Seja $\pi(\underline{\theta}) = p(\underline{\theta}|D_{\text{obs}})$ a distribuição à posteriori de $\underline{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)^T$, onde $\pi(\theta_1|\theta_2, \theta_3, \theta_4, D_{\text{obs}})$, $\pi(\theta_2|\theta_1, \theta_3, \theta_4, D_{\text{obs}})$ e $\pi(\theta_4|\theta_1, \theta_2, \theta_3, D_{\text{obs}})$ podem ser amostradas facilmente, mas $\pi(\theta_3|\theta_1, \theta_2, \theta_4, D_{\text{obs}})$ é difícil de ser amostrada. Então temos os seguintes passos:

Passo 0: Escolher um vetor inicial $\underline{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}, \theta_4^{(0)})^T$ e fazer $i = 0$;

Passo 1: gerar $\underline{\theta}^{(i+1)} = (\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{(i+1)}, \theta_4^{(i+1)})$ da seguinte forma:

- $\theta_1^{(i+1)} \sim \pi(\theta_1|\theta_2^{(i)}, \theta_3^{(i)}, \theta_4^{(i)}, D_{\text{obs}})$,
- gerar $\theta_2^{(i+1)} \sim \pi(\theta_2|\theta_1^{(i+1)}, \theta_3^{(i)}, \theta_4^{(i)}, D_{\text{obs}})$,
- gerar $\theta_3^{(i+1)} \sim \pi(\theta_3|\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_4^{(i)}, D_{\text{obs}})$ da seguinte forma:

sub-passo 1: gere o candidato $\theta_3^{*(i+1)}$ de $q(\cdot|\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{(i)}, \theta_4^{(i)})$ e gere $U \sim N(0, 1)$,

sub-passo 2: faça $\alpha(\theta_3^{*(i+1)}|\theta_3^{(i)}) = \min \left\{ \frac{\pi(\theta_3^{*(i+1)}|\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_4^{(i)}, D_{\text{obs}}) \cdot q(\theta_3^{(i)}|\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{*(i+1)}, \theta_4^{(i)})}{\pi(\theta_3^{(i)}|\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_4^{(i)}, D_{\text{obs}}) \cdot q(\theta_3^{*(i+1)}|\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{(i)}, \theta_4^{(i)})}; 1 \right\}$

sub-passo 3: se $U < \alpha(\theta_3^{*(i+1)}|\theta_3^{(i)})$ então faça $\theta_3^{(i+1)} = \theta_3^{*(i+1)}$, caso contrário faça $\theta_3^{(i+1)} = \theta_3^{(i)}$.

- gerar $\theta_4^{(i+1)} \sim \pi(\theta_4|\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{(i+1)}, D_{\text{obs}})$;

Passo 2: faça $i = i + 1$ e volte ao passo 1.

Observações

1. Se $q(\theta_3^{*(i+1)}|\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{(i)}, \theta_4^{(i)}) = q(\theta_3^{(i)}|\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{*(i+1)}, \theta_4^{(i)})$, temos o algoritmo de Metropolis.
2. Se $q(\theta_3^{*(i+1)}|\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{(i)}, \theta_4^{(i)})$ não depende de $\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{(i)}$ e $\theta_4^{(i)}$, temos o algoritmo “Independence M-H”.
3. Se $\pi(\theta_3|\theta_1, \theta_2, \theta_4)$ puder ser escrita como o produto $\psi(\theta_3|\theta_1, \theta_2, \theta_4) \cdot h(\theta_3|\theta_1, \theta_2, \theta_4)$, então

$$\alpha(\theta_3^{*(i+1)}|\theta_3^{(i)}) = \min \left\{ \frac{\psi(\theta_3^{*(i+1)}|\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_4^{(i)})}{\psi(\theta_3^{(i)}|\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_4^{(i)})}; 1 \right\}.$$

2.4.6 Diagnósticos de Convergência

Conceitualmente dizemos que um algoritmo MCMC converge no tempo T_0 quando os dados gerados após T_0 podem ser considerados “com segurança” como uma amostra da distribuição alvo e portanto os dados simulados antes de T_0 (no período pré-convergência, ou “burn-in period”) não devem fazer parte nas inferências.

Essa definição, embora não seja muito precisa (pois o termo “com segurança” não está bem definido) serve como um guia. Na prática, gostaríamos de saber não somente se o algoritmo converge, mas o valor de T_0 e também o número de iterações que devem ser consideradas após esse tempo T_0 para que a amostra seja representativa.

Na literatura existem vários métodos para diagnosticar possíveis falhas na convergência dos algoritmos MCMC e nessa subseção apresentamos alguns deles. Entretanto, é importante ressaltar que não existe método que possa “provar” convergência (pelo simples fato de que são utilizadas apenas um número finito de observações simuladas). Por esse motivo, vários autores recomendam que seja utilizada uma combinação de várias técnicas de diagnóstico para analisar um mesmo problema. Desse modo espera-se ter um “certo grau de confiança” na análise da convergência. Entretanto, é importante ter em mente que mesmo utilizando um grande número de técnicas de diagnóstico, não podemos afirmar com certeza que os dados simulados são representativos da distribuição que queremos amostrar.

A seguir consideramos alguns aspectos ligados aos problemas de simulação iterativa.

NÚMERO DE CADEIAS GERADAS

Um dos tópicos de grande debate é se é melhor analisar uma única cadeia MCMC ou então múltiplas cadeias partindo-se de pontos iniciais distintos no espaço paramétrico.

Uma única cadeia pode aparentemente exibir um comportamento aceitável. Entretanto, ela pode estar “confinada” a apenas uma certa região do espaço paramétrico, por exemplo em torno de um máximo local (e nesse caso a amostra final não seria representativa da distribuição alvo, já que a cadeia não “varreu” ou explorou

todo o domínio dessa distribuição).

Para não correr esse risco, podemos considerar o uso de múltiplas cadeias partindo-se de pontos iniciais bem dispersos no espaço paramétrico. Esse procedimento, embora tenha um melhor poder exploratório, depende totalmente da escolha dos valores iniciais (ou seja, da habilidade em escolher esses pontos “bem dispersos”). Outro problema é que o número de simulações descartadas (referente ao período “burn-in”) é multiplicado pelo número de cadeias, o que pode ser um grande desperdício.

ESCOLHA DOS VALORES INICIAIS

Para a escolha dos valores iniciais, Carlin e Louis (2000) recomendam fazer uma grade de pontos na região do espaço paramétrico (centrado por exemplo na média da priori e cobrindo mais ou menos três desvios-padrão a partir dessa média) e depois tomar sistematicamente todo ponto da grade como um valor inicial. Gelman et al. (1997) recomendam um procedimento aleatório para a escolha dos valores iniciais partindo do pressuposto de que a distribuição alvo é multimodal. O procedimento é o seguinte:

- Achar as modas da posteriori (por exemplo pelo algoritmo EM ou algum outro procedimento de otimização numérica);
- Aproximar a posteriori por uma mistura de k Normais multivariadas $N(\theta|\hat{\theta}_i, \widehat{V}_{\theta_i})$, $i = 1, \dots, k$, onde k é o número de modas encontradas no passo anterior, $\hat{\theta}_i$ é a i -ésima moda e $\widehat{V}_{\theta_i} = - \left[\frac{\partial \log p(\theta|D_{obs})}{\partial \theta \partial \theta'} \Big|_{\theta=\hat{\theta}_i} \right]^{-1}$. Ou então por uma mistura de t 's multivariadas com quatro graus de liberdade, $t_\nu(\mu_i, \Sigma_i)$, onde $\nu = 4$, $\mu_i = \hat{\theta}_i$ e $\Sigma_i = \widehat{V}_{\theta_i}$.

Os pesos w_i de cada distribuição são dados por $w_i = q(\hat{\theta}_i|D_{obs}) \cdot |\widehat{V}_{\theta_i}|^{\frac{1}{2}}$, onde

$q(\hat{\theta}_i|D_{obs})$ é a posteriori não normalizada calculada no ponto $\hat{\theta}_i$.

- Amostram 1000 valores da distribuição de mistura e depois fazem uma reamostragem ponderada (“weighted bootstrap”) para “filtrar” uma amostra por exemplo de dez pontos, que serão usados como pontos iniciais do processo iterativo (ver página 312 de Gelman et al., 1997).

MONITORAMENTO DA CONVERGÊNCIA

Os diagnósticos podem ser qualitativos ou quantitativos. Os qualitativos utilizam dispositivos gráficos para descrever os dados simulados e os quantitativos fornecem medidas resumo desses dados. Essas técnicas podem ser univariadas ou multivariadas. As univariadas descrevem cada parâmetro individualmente e as multivariadas, a posteriori conjunta. A maioria das técnicas de diagnóstico exploram a idéia de viés (ou “distância” entre a quantidade de interesse numa particular iteração e seu verdadeiro valor) outras exploram a idéia de variância (ou precisão das estimativas).

(a) MÉTODOS QUANTITATIVOS

Um dos diagnósticos mais conhecidos é a medida de Gelman e Rubin (1992), que descrevemos a seguir.

Medida de Gelman e Rubin (1992)

A idéia desse método é utilizar m cadeias paralelas ($m \geq 2$) partindo de pontos iniciais bem dispersos dentro do espaço paramétrico, e para cada quantidade de interesse ψ , comparar a variabilidade “entre” as cadeias com a variabilidade “dentro” das cadeias, calculadas com base nas últimas $n = T - T_0$ iterações (onde T é o número total de iterações).

Segundo Gelman et al. (1997), essa medida foi originalmente calculada como

$$\sqrt{\hat{R}} = \sqrt{\frac{n-1}{n} + \frac{B}{nW}}, \text{ onde } W = \frac{1}{m} \sum_{j=1}^m S_j^2 \text{ e } B = n \frac{1}{m-1} \sum_{j=1}^m (\bar{\psi}_{\cdot j} - \bar{\bar{\psi}})^2,$$

onde $S_j^2 = \frac{1}{n-1} \sum_{i=1}^n (\psi_{ij} - \bar{\psi}_{\cdot j})^2$, $\bar{\psi}_{\cdot j} = \frac{1}{n} \sum_{i=1}^n \psi_{ij}$, e $\bar{\bar{\psi}} = \frac{1}{m} \sum_{i=1}^m \bar{\psi}_{\cdot j} = \frac{1}{mn} \sum_{i=1}^n \sum_{j=1}^m \psi_{ij}$, ou seja, W é a média das m variâncias dentro das cadeias e $\frac{B}{n}$ é a variância entre as médias das m cadeias paralelas ($\sqrt{\hat{R}}$ é conhecido como “Potential Scale Reduction Factors(PSRF)”).

Interpretação: No caso em que há convergência, espera-se que a variabilidade entre as cadeias seja pequena quando comparada com a variabilidade total (e portanto pequena quando comparada com a variabilidade dentro) e nesse caso \hat{R} deve estar próximo de 1 (ou então que o quantil 0.975 da distribuição amostral é ≤ 1.2).

Limitações do método:

- (a) O sucesso desse método depende *totalmente* da habilidade do usuário em encontrar valores iniciais bem espalhados no domínio da distribuição à posteriori (sem o conhecimento da mesma);
- (b) É uma técnica univariada.

Observação 1: O programa BOA (“Bayesian Output Analysis”)(Smith, 2003) calcula a medida acima (PSRF) fazendo a correção proposta por Brooks e Gelman (1998), dada pela seguinte fórmula

$$\sqrt{\left(\frac{n+1}{n} + \frac{B}{nW} \frac{m+1}{m}\right) \frac{df+3}{df+1}},$$

onde df é o número de graus de liberdade de uma densidade t que aproxima a distribuição à posteriori. (Esta medida é também conhecida como “Corrected Scale Reduction Factor (CSRFB)”). O programa calcula também a mediana e o quantil 0.975 da distribuição amostral para CSRFB.

O programa BOA calcula também uma medida denominada “Multivariate Potential Scale Reduction Factor (MPSRF)”, proposta por Brooks e Gelman(1998), que é uma extensão multivariada da medida PSRF.

Observação 2 Os diagnósticos de Raftery e Lewis (1992), Geweke (1992) e Heidelberg e Welch (1983) são implementados pelo programa BOA.

(b) MÉTODOS GRÁFICOS

Os diagnósticos mais comuns consistem em contruir, para cada parâmetro, os seguintes gráficos:

- Gráfico da série temporal (seqüência simulada).
- Gráfico da densidade (ou do núcleo da densidade) estimada.

No caso em que a densidade condicional completa da componente θ_i é conhecida, a densidade estimada de θ_i à posteriori é dada por

$$\begin{aligned} & \hat{\pi}(\theta_i|\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p, D_{\text{obs}}) \\ &= \frac{1}{m(T - T_0)} \sum_{i=1}^m \sum_{g=T_0+1}^T \pi(\theta_i|\theta_1^{(g)}, \dots, \theta_{i-1}^{(g)}, \theta_{i+1}^{(g)}, \dots, \theta_p^{(g)}, D_{\text{obs}}), \end{aligned}$$

onde m é o número de cadeias paralelas e T é o número total de iterações de cada cadeia.

- Gráfico da função de autocorrelação amostral.

Observação: O programa BOA (Smith, 2003) fornece também outras opções como

- Gráfico das médias móveis (“running mean”).
- Gráfico da medida de diagnóstico de Gelman e Rubin.

- Gráfico da medida de Brooks e Gelman (que é uma extensão multivariada da medida de Gelman e Rubin).
- Gráfico da medida de diagnóstico de Geweke.

GUIA PRÁTICO

Uma possível estratégia para diagnosticar falha na convergência dos algoritmos:

- (1) Gerar m cadeias paralelas e simultaneamente uma cadeia mais longa (digamos com 50000 iterações), onde $3 \leq m \leq 10$.
- (2) Para cada componente do vetor de parâmetros, fazer os seguintes gráficos:
 - (a) Série temporal usando cadeia mais longa.
 - (b) Autocorrelação usando a cadeia mais longa.
 - (c) Fazer o gráfico das médias móveis para cada parâmetro
 - (1) usando a cadeia mais longa ou
 - (2) usando as m cadeias paralelas
 - (i) plotando as m cadeias num mesmo sistema de coordenadas ou
 - (ii) plotando todos os parâmetros num mesmo sistema de coordenadas para cada cadeia.
 - (d) Fazer um gráfico com as $m + 1$ densidades estimadas (ou núcleos das densidades estimadas) num mesmo sistema de coordenadas.
 - (e) Fazer o gráfico com a medida de Gelman e Rubin (que requer $m \geq 2$ cadeias). Aumentar o número de simulações (iterações do algoritmo) caso o gráfico não se estabilize em torno de 1 (ou o quantil $0,975 \leq 1, 2$).
 - (f) Fazer o gráfico de Geweke (implementado no programa BOA).

- (3) Fazer o gráfico com a medida multivariada de Brooks e Gelman (que utiliza as m cadeias e está implementado no programa BOA).
- (4) Avaliar as correlações cruzadas entre os parâmetros.

Observação sobre o período pré-convergência:

Os autores Gelman et al. (1997) recomendam descartar sempre a primeira metade de cada seqüência, outros autores, apenas 1%. Uma outra alternativa é utilizar por exemplo os dispositivos 2a a 2f para auxiliar nesta questão.

Seção 2.5.1

A seguir apresentamos os cálculos para a obtenção das distribuições condicionais completas da posteriori $p(\alpha, \beta, \mu, \sigma_x^2, \sigma^2 | D_{\text{obs}})$ dada em (2.3.3).

• Fórmula (2.5.1)

$$\begin{aligned}
 \pi(\alpha | \beta, \mu, \sigma_x^2, \sigma^2, D_{\text{obs}}) &\propto \exp \left\{ -\frac{1}{2|\Sigma|} \left[-2\beta\sigma_x^2 \sum_{i=1}^n (Y_i - \alpha - \beta\mu)(X_i - \mu) \right. \right. \\
 &\quad \left. \left. + (\sigma_x + \sigma^2) \sum_{i=1}^n (Y_i - \alpha - \beta\mu)^2 \right] \right\} \exp \left\{ -\frac{1}{2} \frac{(\alpha - a)^2}{\sigma_\alpha^2} \right\} \\
 &\propto \exp \left\{ -\frac{1}{2} \left[\frac{2\beta\sigma_x^2 \alpha \sum_{i=1}^n (X_i - \mu) + (\sigma_x^2 + \sigma^2) (n\alpha^2 - 2\alpha \sum_{i=1}^n (Y_i - \beta\mu))}{|\Sigma|} \right. \right. \\
 &\quad \left. \left. + \frac{\alpha^2 - 2\alpha a}{\sigma_\alpha^2} \right] \right\} \\
 &\propto \exp \left\{ -\frac{1}{2} \left[\frac{2\alpha n \beta \sigma_x^2 (\bar{X} - \mu) + n\alpha^2 (\sigma_x^2 + \sigma^2) - 2\alpha n (\sigma_x^2 + \sigma^2) (\bar{Y} - \beta\mu)}{|\Sigma|} \right. \right. \\
 &\quad \left. \left. + \frac{\alpha^2 - 2\alpha a}{\sigma_\alpha^2} \right] \right\} \\
 &\propto \exp \left\{ -\frac{1}{2} \left[\alpha^2 \left(\frac{n(\sigma_x^2 + \sigma^2)}{|\Sigma|} + \frac{1}{\sigma_\alpha^2} \right) \right. \right. \\
 &\quad \left. \left. - 2\alpha \left(\frac{-n\beta\sigma_x^2 (\bar{X} - \mu) + n(\sigma_x^2 + \sigma^2) (\bar{Y} - \beta\mu)}{|\Sigma|} + \frac{a}{\sigma_\alpha^2} \right) \right] \right\}.
 \end{aligned}$$

Fazendo

$$\begin{aligned}
 A &= \frac{n(\sigma_x^2 + \sigma^2)}{|\Sigma|} + \frac{1}{\sigma_\alpha^2} \quad \text{e} \\
 B &= \frac{-n\beta\sigma_x^2 (\bar{X} - \mu) + n(\sigma_x^2 + \sigma^2) (\bar{Y} - \beta\mu)}{|\Sigma|} + \frac{a}{\sigma_\alpha^2} \\
 &= \frac{n\sigma_x^2 (\bar{Y} - \beta\bar{X}) + n\sigma^2 (\bar{Y} - \beta\mu)}{|\Sigma|} + \frac{a}{\sigma_\alpha^2},
 \end{aligned}$$

a expressão acima é proporcional a

$$\begin{aligned}
 \exp \left\{ -\frac{1}{2} [\alpha^2 A - 2\alpha B] \right\} &\propto \exp \left\{ -\frac{A}{2} \left[\alpha^2 - 2\alpha \frac{B}{A} \right] \right\} \\
 &\propto \exp \left\{ -\frac{1}{2 \cdot \frac{1}{A}} \left[\alpha - \frac{B}{A} \right]^2 \right\}.
 \end{aligned}$$

ou seja, é uma $N\left(\frac{B}{A}, \frac{1}{A}\right)$ e segue a fórmula (2.5.1).

• **Fórmula (2.5.3)**

$$\begin{aligned}
\pi(\mu|\alpha, \beta, \sigma_x^2, \sigma^2, D_{\text{obs}}) &\propto \exp\left\{-\frac{1}{2}\left[\left((\beta^2\sigma_x^2 + \lambda\sigma^2)(n\mu^2 - 2n\bar{X}\mu) \right. \right. \right. \\
&\quad \left. \left. - 2\beta\sigma_x^2\left(\sum_{i=1}^n(Y_i - \alpha)(X_i - \mu) - \beta\mu\sum_{i=1}^n(X_i - \mu)\right) + (\sigma_x^2 + \sigma^2)(n\mu^2\beta^2 \right. \right. \\
&\quad \left. \left. - 2\mu\beta\sum_{i=1}^n(Y_i - \alpha)\right)\right] / \left[|\Sigma| + \frac{(\mu - m)^2}{\sigma_\mu^2}\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\left((\beta^2\sigma_x^2 + \lambda\sigma^2)n\mu^2 - 2n\bar{X}(\beta^2\sigma_x^2 + \lambda\sigma^2)\mu + 2\beta\sigma_x^2n(\bar{Y} - \alpha)\mu \right. \right. \right. \\
&\quad \left. \left. + 2\beta^2\sigma_x^2n(\bar{X} - \mu)\mu + (\sigma_x^2 + \sigma^2)n\beta^2\mu^2 - 2n(\sigma_x^2 + \sigma^2)\beta(\bar{Y} - \alpha)\mu\right)\right] / \\
&\quad \left. / \left[|\Sigma| + \frac{\mu^2 - 2m\mu}{\sigma_\mu^2}\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\left(\frac{n(\beta^2\sigma_x^2 + \lambda\sigma^2) - 2n\beta^2\sigma_x^2 + n\beta^2(\sigma_x^2 + \sigma^2)}{|\Sigma|} + \frac{1}{\sigma_\mu^2}\right)\mu^2 \right. \right. \\
&\quad \left. \left. - 2n\mu\left(\bar{X}(\beta^2\sigma_x^2 + \lambda\sigma^2) - \beta\sigma_x^2(\bar{Y} - \alpha) - \beta^2\sigma_x^2\bar{X} + (\sigma_x^2 + \sigma^2)(\bar{Y} - \alpha)\beta\right)\right] / \right. \\
&\quad \left. / \left[|\Sigma| + \frac{m}{\sigma_\mu^2}\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\left(\frac{n(\lambda\sigma^2 + \beta^2\sigma^2)\sigma_\mu^2 + |\Sigma|}{|\Sigma|\sigma_\mu^2}\right)\mu^2 \right. \right. \\
&\quad \left. \left. - 2n\mu\left(\sigma_\mu^2(\bar{X}\lambda\sigma^2 - \beta\sigma_x^2(\bar{Y} - \alpha) + \beta(\sigma_x^2 + \sigma^2)(\bar{Y} - \alpha)) + m|\Sigma|\right)\right]\right\} \\
&\propto \exp\left\{\left[\left(\frac{n\sigma_\mu^2\sigma^2(\lambda + \beta^2) + |\Sigma|}{|\Sigma|\sigma_\mu^2}\right)\mu^2 - 2\mu\left(\frac{n\sigma_\mu^2(\lambda\bar{X}\sigma^2 + \beta\sigma^2(\bar{Y} - \alpha)) + m|\Sigma|}{|\Sigma|\sigma_\mu^2}\right)\right]\right\}.
\end{aligned}$$

Portanto,

$$\mu|\alpha, \beta, \sigma_x, \sigma, D_{\text{obs}} \sim N\left(\frac{n\sigma_\mu^2(\lambda\bar{X}\sigma^2 + \beta\sigma^2(\bar{Y} - \alpha)) + m|\Sigma|}{n\sigma_\mu^2\sigma^2(\lambda + \beta^2) + |\Sigma|}; \frac{\sigma_\mu^2|\Sigma|}{n\sigma_\mu^2\sigma^2(\lambda + \beta^2) + |\Sigma|}\right)$$

• **Fórmula (2.5.5)**

$$\pi(\mu|\beta, \sigma_x^2, \sigma^2, D_{\text{obs}}) \propto \pi(\mu, \beta, \sigma_x^2, \sigma^2|D_{\text{obs}}) = \int_{\mathbb{R}} \pi(\alpha, \mu, \beta, \sigma_x^2, \sigma^2|D_{\text{obs}}) d\alpha$$

(onde o integrando é olhado como função de μ e α), onde

$$\begin{aligned}
& \int_{\mathbb{R}} \pi(\alpha, \beta, \mu, \sigma_x^2, \sigma^2 | D_{\text{obs}}) d\alpha \propto \int_{\mathbb{R}} \exp \left\{ -\frac{1}{2} \left[\left((\beta^2 \sigma_x^2 + \lambda \sigma^2) \sum_{i=1}^n (X_i - \mu)^2 \right. \right. \right. \\
& \quad \left. \left. - 2\beta \sigma_x^2 \sum_{i=1}^n (Y_i - \alpha - \beta \mu)(X_i - \mu) + (\sigma_x^2 + \sigma^2) \sum_{i=1}^n (Y_i - \alpha - \beta \mu)^2 \right) / |\Sigma| \right. \\
& \quad \left. \left. + \frac{(\alpha - a)^2}{\sigma_\alpha^2} + \frac{(\mu - m)^2}{\sigma_\mu^2} \right] \right\} d\alpha, \text{ onde } |\Sigma| \text{ é dada em (2.3.3),} \\
& \propto \exp \left\{ -\frac{1}{2} \left[\left((\beta^2 \sigma_x^2 + \lambda \sigma^2)(n\mu^2 - 2n\bar{X}\mu) - 2\beta \sigma_x^2 \sum_{i=1}^n (Y_i - \beta \mu)(X_i - \mu) \right. \right. \right. \\
& \quad \left. \left. + (\sigma_x^2 + \sigma^2) \sum_{i=1}^n (Y_i - \beta \mu)^2 \right) / |\Sigma| + \frac{(\mu - m)^2}{\sigma_\mu^2} \right] \right\} \\
& \cdot \int_{\mathbb{R}} \exp \left\{ -\frac{1}{2} \left[\left(2\beta \sigma_x^2 \alpha \sum_{i=1}^n (X_i - \mu) + (\sigma_x^2 + \sigma^2) \left(n\alpha^2 - 2\alpha \sum_{i=1}^n (Y_i - \beta \mu) \right) \right) / \right. \right. \\
& \quad \left. \left. / |\Sigma| + \frac{(\alpha - a)^2}{\sigma_\alpha^2} \right] \right\} d\alpha \\
& \propto \exp \left\{ -\frac{1}{2} \left[\left((\beta^2 \sigma_x^2 + \lambda \sigma^2)(n\mu^2 - 2n\bar{X}\mu) - 2\beta \sigma_x^2 (-n\mu\bar{Y} - n\beta\mu\bar{X} + n\beta\mu^2) \right. \right. \right. \\
& \quad \left. \left. + (\sigma_x^2 + \sigma^2)(n\beta^2\mu^2 - 2\beta\mu n\bar{Y}) \right) / \right. \\
& \quad \left. / |\Sigma| + \frac{\mu^2 - 2m\mu}{\sigma_\mu^2} \right] \right\} \cdot \int_{\mathbb{R}} \exp \left\{ -\frac{1}{2} \left[\left(2n\beta \sigma_x^2 (\bar{X} - \mu)\alpha + n(\sigma_x^2 + \sigma^2)\alpha^2 \right. \right. \right. \\
& \quad \left. \left. - 2n\alpha(\bar{Y} - \beta\mu)(\sigma_x^2 + \sigma^2) \right) / |\Sigma| + \frac{\alpha^2 - 2a\alpha}{\sigma_\alpha^2} \right] \right\} \\
& \propto \exp \left\{ -\frac{1}{2} \left[\mu^2 \left((n(\beta^2 \sigma_x^2 + \lambda \sigma^2) - 2n\beta^2 \sigma_x^2 + (\sigma_x^2 + \sigma^2)n\beta^2) / |\Sigma| + \frac{1}{\sigma_\mu^2} \right) \right. \right. \\
& \quad \left. \left. - 2\mu \left((n\bar{X}(\beta^2 \sigma_x^2 + \lambda \sigma^2) - n\bar{Y}\beta \sigma_x^2 - n\beta^2 \sigma_x^2 \bar{X} + \beta n\bar{Y}(\sigma_x^2 + \sigma^2)) / |\Sigma| + \frac{m}{\sigma_\mu^2} \right) \right] \right\} \\
& \cdot \int_{\mathbb{R}} \exp \left\{ -\frac{1}{2} \left[\alpha^2 \left(\frac{n(\sigma_x^2 + \sigma^2)}{|\Sigma|} + \frac{1}{\sigma_\alpha^2} \right) \right. \right. \\
& \quad \left. \left. - 2\alpha \left(\frac{-n\beta \sigma_x^2 (\bar{X} - \mu) + n(\bar{Y} - \beta\mu)(\sigma_x^2 + \sigma^2)}{|\Sigma|} + \frac{a}{\sigma_\alpha^2} \right) \right] \right\} d\alpha \\
& = \exp \left\{ -\frac{1}{2} \mu^2 \left(\frac{n\lambda \sigma^2 + n\sigma^2 \beta^2}{|\Sigma|} + \frac{1}{\sigma_\mu^2} \right) - 2\mu \left(\frac{n\bar{X}\lambda \sigma^2 + n\beta\bar{Y}\sigma^2}{|\Sigma|} + \frac{n}{\sigma_\mu^2} \right) \right. \\
& \quad \left. \cdot \int_{\mathbb{R}} \exp \left\{ -\frac{1}{2} \left[\left(\frac{n\sigma_\alpha^2(\sigma_x^2 + \sigma^2) + |\Sigma|}{\sigma_\alpha^2 |\Sigma|} \right) \alpha^2 \right. \right. \right.
\end{aligned}$$

$$-2\alpha \left[\left(\frac{-n\beta\sigma_x^2\bar{X} + n\bar{Y}(\sigma_x^2 + \sigma^2) - n\beta\sigma^2\mu}{|\Sigma|} + \frac{a}{\sigma_\alpha^2} \right) \right] d\alpha .$$

Chamando

$$\begin{aligned} A &= \frac{n\lambda\sigma^2 + n\sigma^2\beta^2}{|\Sigma|} + \frac{1}{\sigma_\mu^2}; \\ B &= \frac{n\bar{X}\lambda\sigma^2 + n\beta\bar{Y}\sigma^2}{|\Sigma|} + \frac{m}{\sigma_\mu^2} = \frac{n\sigma^2(\lambda\bar{X} + \beta\bar{Y})}{|\Sigma|} + \frac{m}{\sigma_\mu^2}; \\ C &= \frac{n\sigma_\alpha^2(\sigma_x^2 + \sigma^2) + |\Sigma|}{\sigma_\alpha^2|\Sigma|} \quad e \\ D &= \frac{-n\beta\sigma_x^2\bar{X} + n\bar{Y}(\sigma_x^2 + \sigma^2) - n\beta\sigma^2\mu}{|\Sigma|} + \frac{a}{\sigma_\alpha^2} \quad (\text{que é função de } \mu), \end{aligned}$$

temos que a expressão acima é igual a

$$\begin{aligned} &\exp \left\{ -\frac{1}{2}[\mu^2 A - 2\mu B] \right\} \int_{\mathbb{R}} \exp \left\{ -\frac{1}{2}[C\alpha^2 - 2\alpha D] \right\} d\alpha \\ &= \exp \left\{ -\frac{1}{2}[A\mu^2 - 2B\mu] \right\} \int_{\mathbb{R}} \exp \left\{ -\frac{C}{2} \left[\alpha^2 - 2\alpha \frac{D}{C} + \left(\frac{D}{C} \right)^2 - \left(\frac{D}{C} \right)^2 \right] \right\} d\alpha \\ &= \exp \left\{ -\frac{1}{2}[A\mu^2 - 2B\mu] \right\} \exp \left\{ \frac{D^2}{2C} \right\} \int_{\mathbb{R}} \exp \left\{ -\frac{1}{2 \cdot \frac{C}{2}} \left(\alpha - \frac{D}{C} \right)^2 \right\} d\alpha \\ &\propto \exp \left\{ -\frac{1}{2}(A\mu^2 - 2B\mu) \right\} \exp \left\{ \frac{D^2}{2C} \right\}, \end{aligned}$$

onde

$$D = \frac{\sigma_\alpha^2(-n\beta\sigma_x^2\bar{X} + n\bar{Y}(\sigma_x^2 + \sigma^2)) + a|\Sigma|}{|\Sigma|\sigma_\alpha^2} - \frac{n\beta\sigma^2\sigma_\alpha^2}{|\Sigma|\sigma_\alpha^2}\mu .$$

Chamando

$$E = \frac{\sigma_\alpha^2(-n\beta\sigma_x^2\bar{X} + n\bar{Y}(\sigma_x^2 + \sigma^2)) + a|\Sigma|}{|\Sigma|\sigma_\alpha^2} \quad e \quad F = \frac{n\beta\sigma^2}{|\Sigma|},$$

então temos que a expressão acima é proporcional a

$$\begin{aligned} &\exp \left\{ -\frac{1}{2}[A\mu^2 - 2B\mu] \right\} \exp \left\{ \frac{1}{2C}[E - F\mu]^2 \right\} \\ &= \exp \left\{ -\frac{1}{2} \left[A\mu^2 - 2B\mu - \frac{(E - F\mu)^2}{C} \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[A\mu^2 - 2B\mu - \frac{F^2\mu^2}{C} + \frac{2FE}{C}\mu \right] \right\} \\ &= \exp \left\{ -\frac{1}{2} \left[\left(A - \frac{F^2}{C} \right) \mu^2 - 2\mu \left(B - \frac{FE}{C} \right) \right] \right\}. \end{aligned}$$

Chamando

$$G = A - \frac{F^2}{C} \quad \text{e} \quad H = B - \frac{FE}{C},$$

temos que

$$\pi(\mu|\beta, \sigma_x^2, \sigma^2, D_{\text{obs}}) \propto \exp\left\{-\frac{1}{2}[G\mu^2 - 2H\mu]\right\} \propto \exp\left\{-\frac{1}{2\frac{1}{G}}\left(\mu - \frac{H}{G}\right)^2\right\},$$

ou seja, é uma $N\left(\frac{H}{G}, \frac{1}{G}\right)$, onde

$$\begin{aligned} G &= \frac{n\sigma^2(\lambda + \beta^2)}{|\Sigma|} + \frac{1}{\sigma_\mu^2} - \frac{n^2\beta^2(\sigma^2)^2\sigma_\alpha^2}{|\Sigma|(n\sigma_\alpha^2(\sigma_x^2 + \sigma^2) + |\Sigma|)} \\ H &= \frac{n\sigma^2(\lambda\bar{X} + \beta\bar{Y})}{|\Sigma|} + \frac{m}{\sigma_\mu^2} - \frac{n\beta\sigma^2\{n\sigma_\alpha^2(\bar{Y}(\sigma_x^2 + \sigma^2) - \beta\sigma_x^2\bar{X}) + a|\Sigma|\}}{|\Sigma|[n\sigma_\alpha^2(\sigma_x^2 + \sigma^2) + |\Sigma|]} \\ |\Sigma| &= \beta^2\sigma^2\sigma_x^2 + \lambda\sigma^2\sigma_x^2 + \lambda(\sigma^2)^2. \end{aligned}$$

• Fórmula (2.5.6)

$$\begin{aligned} \pi(\beta|\alpha, \mu, \sigma_x^2, \sigma^2, D_{\text{obs}}) &\propto |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\left[\left((\beta^2\sigma_x^2 + \lambda\sigma^2)\sum_{i=1}^n (X_i - \mu)^2\right.\right.\right. \\ &\quad \left.\left.\left.- 2\beta\sigma_x^2\sum_{i=1}^n (Y_i - \alpha - \beta\mu)(X_i - \mu) + (\sigma_x^2 + \sigma^2)\sum_{i=1}^n (Y_i - \alpha - \beta\mu)^2\right)\right] \right. \\ &\quad \left. / \left|\Sigma| + \frac{(\beta - b)^2}{\sigma_\beta^2}\right]\right\} \\ &\propto |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\left[\left(\beta^2\sigma_x^2\sum_{i=1}^n (X_i - \mu)^2 - 2\beta\sigma_x^2\left(\sum_{i=1}^n (Y_i - \alpha)(X_i - \mu)\right.\right.\right.\right. \\ &\quad \left.\left.\left.- \beta\mu\sum_{i=1}^n (X_i - \mu)\right) + (\sigma_x^2 + \sigma^2)\left(n\beta^2\mu^2 - 2\beta\mu\sum_{i=1}^n (Y_i - \alpha)\right)\right]\right] / \left|\Sigma| + \frac{\beta^2 - 2\beta b}{\sigma_\beta^2}\right]\right\} \\ &\propto |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\left[\left(\beta^2\sigma_x^2\sum_{i=1}^n (X_i - \mu)^2 - 2\beta\sigma_x^2\sum_{i=1}^n (Y_i - \alpha)(X_i - \mu)\right.\right.\right. \\ &\quad \left.\left.\left.+ 2n\beta^2\sigma_x^2\mu(\bar{X} - \mu) + n\beta^2\mu^2(\sigma_x^2 + \sigma^2) - 2n\beta\mu(\bar{Y} - \alpha)(\sigma_x^2 + \sigma^2)\right)\right] \right. \\ &\quad \left. / \left|\Sigma| + \frac{\beta^2 - 2\beta b}{\sigma_\beta^2}\right]\right\} \\ &\propto |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\left[\beta^2\left(\frac{\sigma_x^2\sum_{i=1}^n (X_i - \mu)^2 + 2n\sigma_x^2\mu(\bar{X} - \mu) + n\mu^2(\sigma_x^2 + \sigma^2)}{|\Sigma|} + \frac{1}{\sigma_\beta^2}\right)\right]\right\} \end{aligned}$$

$$\begin{aligned}
& -2\beta \left(\frac{\sigma_x^2 \sum_{i=1}^n (Y_i - \alpha)(X_i - \mu) + n\mu(\bar{Y} - \alpha)(\sigma_x^2 + \sigma^2)}{|\Sigma|} + \frac{b}{\sigma_\beta^2} \right) \Bigg\} \\
& \propto |\Sigma|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2|\Sigma|\sigma_\beta^2} \left[\beta^2 \left\{ \sigma_x^2 \sum_{i=1}^n (X_i - \mu)^2 + 2n\sigma_x^2\mu(\bar{X} - \mu) \right. \right. \right. \\
& \quad \left. \left. \left. + n\mu^2(\sigma_x^2 + \sigma^2) \right\} + |\Sigma| \right] - 2\beta \left\{ \sigma_\beta^2 \left[\sigma_x^2 \left(\sum_{i=1}^n X_i Y_i - n\alpha\bar{X} \right) \right. \right. \right. \\
& \quad \left. \left. \left. + n\mu\sigma^2(\bar{Y} - \alpha) \right] + b|\Sigma| \right\} \right\}
\end{aligned}$$

• Fórmula (2.5.7)

$$\begin{aligned}
\pi(\sigma_x^2 | \alpha, \beta, \mu, \sigma^2, D_{\text{obs}}) & \propto |\Sigma|^{-\frac{n}{2}} (\sigma_x^2)^{-(c+1)} \exp \left\{ -\frac{1}{2} \left[\left(\beta^2 \sigma_x^2 \sum_{i=1}^n (X_i - \mu)^2 \right. \right. \right. \\
& \quad \left. \left. \left. - 2\beta \sigma_x^2 \sum_{i=1}^n (Y_i - \alpha - \beta\mu)(X_i - \mu) + \sigma_x^2 \sum_{i=1}^n (Y_i - \alpha - \beta\mu)^2 \right) / \left(|\Sigma| + \frac{2d}{\sigma_x^2} \right) \right] \right\} \\
& \propto |\Sigma|^{-\frac{n}{2}} (\sigma_x^2)^{-(c+1)} \exp \left\{ -\frac{1}{2} \left[\frac{\sigma_x^2 \sum_{i=1}^n (Y_i - \alpha - \beta X_i)^2}{|\Sigma|} + \frac{2d}{\sigma_x^2} \right] \right\}
\end{aligned}$$

• Fórmula (2.5.8)

$$\begin{aligned}
& \pi(\sigma^2 | \alpha, \beta, \mu, \sigma_x^2, D_{\text{obs}}) \propto \\
& |\Sigma|^{-\frac{n}{2}} (\sigma^2)^{-(f+1)} \exp \left\{ -\frac{1}{2} \left[\frac{\lambda \sigma^2 \sum_{i=1}^n (X_i - \mu)^2 + \sigma^2 \sum_{i=1}^n (Y_i - \alpha - \beta\mu)^2}{|\Sigma|} + \frac{2g}{\sigma^2} \right] \right\} \\
& = |\Sigma|^{-\frac{n}{2}} (\sigma^2)^{-(f+1)} \exp \left\{ -\frac{1}{2} \left[\frac{\sigma^2}{|\Sigma|} \left(\sum_{i=1}^n (Y_i - \alpha - \beta\mu)^2 + \lambda \sum_{i=1}^n (X_i - \mu)^2 \right) + \frac{2g}{\sigma^2} \right] \right\}
\end{aligned}$$

Seção 2.8

- Derivação da fórmula (2.8.2)

$$\int_{-\infty}^{\infty} p(\theta, \underline{x} | D_{\text{obs}}) d\alpha \propto \sigma_x^{2^{-(\frac{n}{2}+c+1)}} (\sigma^2)^{-(n+f+1)} \exp \left\{ -\frac{1}{2} \left[\frac{2d}{\sigma_x^2} + \frac{2g}{\sigma^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma^2} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_x^2} \right] \right\} \int_{-\infty}^{\infty} \exp \left(-\frac{1}{2\lambda\sigma^2} \sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2 \right) d\alpha$$

onde

$$\begin{aligned} & \int_{-\infty}^{\infty} \exp \left(-\frac{1}{2\lambda\sigma^2} \sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2 \right) d\alpha \\ &= \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\lambda\sigma^2} \left[\sum_{i=1}^n (Y_i - \beta x_i)^2 + n\alpha^2 - 2\alpha \sum_{i=1}^n (Y_i - \beta x_i) \right] \right\} d\alpha \\ &= \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\lambda\sigma^2} \left[n\alpha^2 - 2n\alpha(\bar{Y} - \beta\bar{x}) + \sum_{i=1}^n (Y_i - \beta x_i)^2 \right] \right\} d\alpha \\ &= \int_{-\infty}^{\infty} \exp \left(-\frac{1}{2\lambda\sigma^2} \sum_{i=1}^n (Y_i - \beta x_i)^2 \right) \exp \left\{ -\frac{n}{2\lambda\sigma^2} [\alpha^2 - 2\alpha(\bar{Y} - \beta\bar{x})] \right\} d\alpha \\ &= \exp \left\{ -\frac{1}{2\lambda\sigma^2} \left[\sum_{i=1}^n (Y_i - \beta x_i)^2 - n(\bar{Y} - \beta\bar{x})^2 \right] \right\} \int_{-\infty}^{\infty} \exp \left\{ -\frac{n}{2\lambda\sigma^2} [\alpha - (\bar{Y} - \beta\bar{x})]^2 \right\} d\alpha \\ &\propto \exp \left\{ -\frac{1}{2\lambda\sigma^2} \left[\sum_{i=1}^n (Y_i - \beta x_i)^2 - n(\bar{Y} - \beta\bar{x})^2 \right] \right\} (\sigma^2)^{\frac{1}{2}}. \end{aligned}$$

Portanto,

$$\int_{-\infty}^{\infty} p(\theta, \underline{x} | D_{\text{obs}}) d\alpha \propto \sigma_x^{2^{-(\frac{n}{2}+c+1)}} (\sigma^2)^{-(n+f+\frac{1}{2})} \exp \left\{ -\frac{1}{2} \left[\frac{2d}{\sigma_x^2} + \frac{2g}{\sigma^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma^2} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_x^2} + \frac{\sum_{i=1}^n (Y_i - \beta x_i)^2 - n(\bar{Y} - \beta\bar{x})^2}{\lambda\sigma^2} \right] \right\}.$$

- Derivação da fórmula (2.8.3)

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\theta, \underline{x} | D_{\text{obs}}) d\alpha d\mu \propto \sigma_x^{2^{-(\frac{n}{2}+c+\frac{3}{2})}} (\sigma^2)^{-(n+f+1)} \exp \left\{ -\frac{1}{2} \left[\frac{2d}{\sigma_x^2} + \frac{2g}{\sigma^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma^2} + \frac{\sum_{i=1}^n (Y_i - \beta x_i)^2 - n(\bar{Y} - \beta\bar{x})^2}{\lambda\sigma^2} \right] \right\} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\sigma_x^2} \sum_{i=1}^n (x_i - \mu)^2 \right\} d\mu,$$

onde

$$\int_{-\infty}^{\infty} \exp \left[-\frac{1}{2\sigma_x^2} \sum_{i=1}^n (x_i - \mu)^2 \right] d\mu$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\sigma_x^2} (\sum x_i^2 - 2n\bar{x}\mu + n\mu^2) \right\} d\mu \\
&= \int_{-\infty}^{\infty} \exp \left\{ -\frac{n}{2\sigma_x} (\mu^2 - 2\mu\bar{x}) \right\} \exp \left(-\frac{\sum_{i=1}^n x_i^2}{2\sigma_x^2} \right) d\mu \\
&= \exp \left(-\frac{\sum_{i=1}^n x_i^2}{2\sigma_x^2} \right) \exp \left(\frac{n\bar{x}^2}{2\sigma_x^2} \right) \int_{-\infty}^{\infty} \exp \left\{ -\frac{n}{2\sigma_x^2} (\mu^2 - 2\mu\bar{x} + \bar{x}^2) \right\} d\mu \\
&\propto \exp \left\{ -\frac{1}{2\sigma_x^2} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \right\} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\frac{\sigma_x^2}{n}} (\mu - \bar{x})^2 \right\} d\mu \\
&\propto (\sigma_x^2)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma_x^2} \sum_{i=1}^n (x_i - \bar{x})^2 \right\}.
\end{aligned}$$

Portanto,

$$\begin{aligned}
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\theta, \tilde{x} | D_{\text{obs}}) d\alpha d\mu &\propto (\sigma_x^2)^{-(\frac{n}{2}+c+\frac{1}{2})} (\sigma^2)^{-(n+f+\frac{1}{2})} \exp \left\{ -\frac{1}{2} \left[\frac{2d}{\sigma_x^2} + \frac{2g}{\sigma^2} \right. \right. \\
&\quad \left. \left. + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma^2} + \frac{\sum_{i=1}^n (Y_i - \beta x_i)^2 - n(\bar{Y} - \beta\bar{x})^2}{\lambda\sigma^2} + \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_x^2} \right] \right\}
\end{aligned}$$

• Derivação da fórmula (2.8.4)

$$\begin{aligned}
&\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\theta, \tilde{x} | D_{\text{obs}}) d\alpha d\mu d\beta \\
&\propto (\sigma_x^2)^{-(\frac{n}{2}+c+1)} (\sigma^2)^{-(n+f+\frac{1}{2})} \exp \left\{ -\frac{1}{2} \left[\frac{2d}{\sigma_x^2} + \frac{2g}{\sigma^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma^2} + \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_x^2} \right] \right\} \\
&\int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\lambda\sigma^2} \left(\sum_{i=1}^n (Y_i - \beta x_i)^2 - n(\bar{Y} - \beta\bar{x})^2 \right) \right\} d\beta,
\end{aligned}$$

onde

$$\begin{aligned}
&\int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\lambda\sigma^2} \left(\sum_{i=1}^n (Y_i - \beta x_i)^2 - n(\bar{Y} - \beta\bar{x})^2 \right) \right\} d\beta \\
&\propto \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\lambda\sigma^2} \left[\sum_{i=1}^n Y_i^2 - 2\beta \sum_{i=1}^n Y_i x_i + \beta^2 \sum_{i=1}^n x_i^2 - n\bar{Y}^2 + 2n\bar{Y}\beta\bar{x} - n\beta^2\bar{x}^2 \right] \right\} d\beta.
\end{aligned}$$

Sejam $S_{YY} = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2$, $S_{xx} = \sum_{i=1}^n x_i^2 - n\bar{x}^2$, $S_{Yx} = \sum_{i=1}^n Y_i x_i - n\bar{x}\bar{Y}$.

Portanto, a última integral é igual a

$$\int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\lambda\sigma^2} [S_{YY} - 2\beta S_{Yx} + \beta^2 S_{xx}] \right\} d\beta$$

$$\begin{aligned}
&= \exp \left\{ -\frac{S_{YY}}{2\lambda\sigma^2} \right\} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\lambda\sigma^2} (\beta^2 S_{xx} - 2\beta S_{Yx}) \right\} \\
&= \exp \left\{ -\frac{S_{YY}}{2\lambda\sigma^2} \right\} \int_{-\infty}^{\infty} \left\{ -\frac{S_{xx}}{2\lambda\sigma^2} \left[\beta^2 - 2\beta \frac{S_{Yx}}{S_{xx}} \right] \right\} d\beta \\
&= \exp \left\{ -\frac{S_{YY}}{2\lambda\sigma^2} \right\} \exp \left\{ \frac{S_{Yx}^2}{2\lambda\sigma^2 S_{xx}} \right\} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\lambda \frac{\sigma^2}{S_{xx}}} \left(\beta - \frac{S_{Yx}}{S_{xx}} \right)^2 \right\} d\beta \\
&\propto \exp \left\{ -\frac{1}{2\lambda\sigma^2} \left(S_{YY} - \frac{S_{xY}^2}{S_{xx}} \right) \right\} \sqrt{\frac{\lambda\sigma^2}{S_{xx}}} \\
&\propto \frac{(\sigma^2)^{\frac{1}{2}}}{(S_{xx})^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2\lambda\sigma^2} \frac{[S_{YY}S_{xx} - S_{xY}^2]}{S_{xx}} \right\}
\end{aligned}$$

e, portanto,

$$\begin{aligned}
&\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\theta, \mathbf{x} | D_{\text{obs}}) d\alpha d\mu d\beta \propto (\sigma_x^2)^{-(\frac{n}{2}+c+\frac{1}{2})} \frac{(\sigma^2)^{-(n+f)}}{(S_{xx})^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \left[\frac{2d}{\sigma_x^2} \right. \right. \\
&\left. \left. + \frac{2g}{\sigma^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma^2} + \frac{S_{xx}}{\sigma_x^2} + \frac{S_{YY}S_{xx} - S_{xY}^2}{\lambda\sigma^2 S_{xx}} \right] \right\}
\end{aligned}$$

• Derivação da fórmula (2.8.5)

$$\begin{aligned}
&\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\theta, \mathbf{x} | D_{\text{obs}}) d\alpha d\mu d\beta d\sigma_x^2 \propto \frac{(\sigma^2)^{-(n+f)}}{(S_{xx})^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \left[\frac{2g}{\sigma^2} \right. \right. \\
&\left. \left. + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma^2} + \frac{(S_{YY}S_{xx} - S_{xY}^2)}{\lambda\sigma^2 S_{xx}} \right] \right\} \int_0^{\infty} (\sigma_x^2)^{-(\frac{n}{2}+c+\frac{1}{2})} \exp \left\{ -\frac{1}{2} \left(\frac{2d}{\sigma_x^2} + \frac{S_{xx}}{\sigma_x^2} \right) \right\} d\sigma_x^2,
\end{aligned}$$

onde

$$\begin{aligned}
&\int_0^{\infty} (\sigma_x^2)^{-(\frac{n+2c-1}{2}+1)} \exp \left\{ -\frac{1}{2\sigma_x^2} (2d + S_{xx}) \right\} d\sigma_x^2 \\
&= \int_0^{\infty} (s_x^2)^{-(\frac{n+2c-1}{2}+1)} \exp \left\{ \frac{-\frac{1}{2}[2d + S_{xx}]}{\sigma_x^2} \right\} d\sigma_x^2,
\end{aligned}$$

onde o integrando é o núcleo de uma Gama Invertida com parâmetros $\frac{n+2c-1}{2}$ e $\frac{2d+S_{xx}}{2}$ e portanto a última integral é igual a

$$\frac{\Gamma \left(\frac{n+2c-1}{2} \right)}{\frac{1}{2}(2d + S_{xx})^{\frac{n+2c-1}{2}}} \propto \frac{1}{(2d + S_{xx})^{\frac{n+2c-1}{2}}}.$$

Portanto,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\theta, \mathbf{x} | D_{\text{obs}}) d\alpha d\mu d\sigma_x^2 \propto \frac{(\sigma^2)^{-(n+f)}}{(S_{xx})^{\frac{1}{2}} (2d + S_{xx})^{\frac{n+2c-1}{2}}} \exp \left\{ -\frac{1}{2} \left[\frac{2g}{\sigma^2} \right. \right.$$

$$\left. + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma^2} + \frac{(S_{YY}S_{xx} - S_{xY}^2)}{\lambda\sigma^2 S_{xx}} \right] \Bigg\}.$$

- Derivação da fórmula (2.8.6)

$$\int_{\Theta} p(\theta, \underline{x} | D_{\text{obs}}) d\theta \propto \frac{1}{(S_{xx})^{\frac{1}{2}} (2d + S_{xx})^{\frac{n+2c-1}{2}}} \cdot \int_0^{\infty} (\sigma^2)^{-(n+f)} \exp \left\{ \frac{-\frac{1}{2} \left[\sum_{i=1}^n (X_i - x_i)^2 + \frac{S_{YY}S_{xx} - S_{xY}^2}{\lambda S_{xx}} + 2g \right]}{\sigma^2} \right\} d\sigma_x^2,$$

onde

$$\begin{aligned} & \int_0^{\infty} (\sigma^2)^{-(n+f)} \exp \left\{ \frac{-\frac{1}{2} \left[\sum_{i=1}^n (X_i - x_i)^2 + \frac{S_{YY}S_{xx} - S_{xY}^2}{\lambda S_{xx}} + 2g \right]}{\sigma^2} \right\} d\sigma_x^2 \\ &= \int_0^{\infty} (\sigma^2)^{-(n+f-1+1)} \exp \left\{ \frac{-\frac{1}{2} \left[\sum_{i=1}^n (X_i - x_i)^2 + \frac{S_{YY}S_{xx} - S_{xY}^2}{\lambda S_{xx}} + 2g \right]}{\sigma^2} \right\} d\sigma_x^2 \\ &= \frac{\Gamma(n+f-1)}{\frac{1}{2} \left[\sum_{i=1}^n (X_i - x_i)^2 + \frac{S_{YY}S_{xx} - S_{xY}^2}{\lambda S_{xx}} + 2g \right]^{n+f-1}} \\ &\propto \frac{1}{\left[\sum_{i=1}^n (X_i - x_i)^2 + \frac{S_{YY}S_{xx} - S_{xY}^2}{\lambda S_{xx}} + 2g \right]^{n+f-1}} \\ &= \frac{1}{\left[\sum_{i=1}^n (X_i - x_i)^2 + \frac{S_{YY}S_{xx} - S_{xY}^2}{\lambda S_{xx}} + 2g \right]^{n+f-1}}. \end{aligned}$$

Portanto,

$$\int_{\Theta} p(\theta, \underline{x} | D_{\text{obs}}) d\theta \propto \frac{1}{(S_{xx})^{\frac{1}{2}} (2d + S_{xx})^{\frac{n+2c-1}{2}} \left(\sum_{i=1}^n (X_i - x_i)^2 + \frac{S_{YY}S_{xx} - S_{xY}^2}{\lambda S_{xx}} + 2g \right)^{n+f-1}}.$$

Seção 4.4.3

(a) O critério do Fator de Bayes

Suponha que estamos interessados em comparar dois modelos: o modelo M_1 (com vetor de parâmetros ϱ_{M_1} e priori $\pi(\varrho_{M_1})$) e o modelo M_2 (com vetor de parâmetros ϱ_{M_2} e priori $\pi(\varrho_{M_2})$) para o mesmo conjunto de dados \underline{Y} .

Então o Fator de Bayes (BF) para o modelo 1 versus o modelo 2 é dado por

$$BF_{12} \frac{\pi(M_1|\underline{Y})/\pi(M_2|\underline{Y})}{\pi(M_1)/\pi(M_2)} = \frac{\pi(\underline{Y}|M_1)}{\pi(\underline{Y}|M_2)},$$

onde $\pi(\underline{Y}|M_i) = \int_{\Theta_{M_i}} \pi(\underline{Y}|\varrho_{M_i})\pi(\varrho_{M_i})d\varrho_{M_i}$, $i = 1, 2$, isto é, $\pi(\underline{Y}|M_i)$ é a distribuição marginal de \underline{Y} sob o modelo M_i (ou distribuição priori preditiva sob o modelo M_i) e portanto $\pi(\underline{Y}|M_i)$ é essencialmente a constante de proporcionalidade (ou de normalização) da posteriori sob o modelo M_i , isto é, $\pi(\varrho|\underline{Y}, M_i)$.

Interpretação: Valores altos de BF_{12} favorecem o modelo 1.

Observação: Se $\pi(\varrho_{M_1})$ ou $\pi(\varrho_{M_2})$ for imprópria, o fator de Bayes não está bem definido pois envolve uma constante arbitrária (e portanto não é interpretável). Além desse problema, essa medida tem sido criticada pois fere princípios bayesianos.

Jeffreys (1961) e Kass e Raftery (1995) sugerem as seguintes tabelas de referência para o Fator de Bayes (BF_{12}):

	BF_{12}	Evidência a favor do modelo M_1
Jeffrey	$1 \leq BF \leq 3.2$	Fraca
	$3.2 < BF \leq 10$	Moderada
	$10 < BF \leq 100$	Forte
	$BF > 100$	Decisiva
Kass e Raftery	$1 \leq BF \leq 3$	Fraca
	$3 < BF \leq 20$	Moderada
	$20 < BF \leq 150$	Forte
	$BF > 150$	Muito forte

Na literatura existem vários métodos para estimar essa razão de constantes de proporcionalidade (ver Chen et al., 2000). Um método é estimar a verossimilhança marginal de cada modelo $\hat{p}(\underline{Y}|M_i)$ e tomar o quociente.

Newton e Raftery (1994) propõe um método direto de estimar $p(\underline{Y}|M_i)$:

$$\hat{p}(\underline{Y}|M_i) = \left[\frac{1}{G} \sum_{g=1}^G \frac{1}{f(\underline{Y}|\underline{\theta}^{(g)})} \right]^{-1},$$

onde $(\underline{\theta}^{(1)}, \dots, \underline{\theta}^{(G)})$ é uma amostra da posteriori $p(\underline{\theta}|\underline{Y}, M_i)$ e $f(\underline{Y}|\underline{\theta}_{M_i})$ é a função densidade das observações. Esse método, embora altamente eficiente, pode ficar instável, já que o denominador pode ficar muito próximo de zero para algumas parcelas dessa soma.

Um método relativamente fácil de ser implementado e também seguro é o método proposto por Chib (1995) que se baseia também na verossimilhança marginal dos modelos sendo comparados (isto é, se o objetivo é comparar dois modelos M_1 e M_2 , então o fator de Bayes $\text{BF}_{1,2}$ é estimado por $\widehat{\text{BF}}_{1,2} = \frac{\hat{p}(\underline{Y}|M_1)}{\hat{p}(\underline{Y}|M_2)}$). Esse método é usado somente no caso em que todas as distribuições condicionais completas são conhecidas (isto é, no caso em que o algoritmo de Gibbs possa ser utilizado).

Descrição do método de Chib

Pela regra de Bayes, temos que a posteriori $p(\underline{\theta}|\underline{Y}) = \frac{p(\underline{Y}|\underline{\theta})p(\underline{\theta})}{p(\underline{Y})}$, e portanto $p(\underline{Y}) = \frac{p(\underline{Y}|\underline{\theta})p(\underline{\theta})}{p(\underline{\theta}|\underline{Y})}$. Essa igualdade vale para todo $\underline{\theta}$, e em particular para um $\underline{\theta} = \underline{\theta}'$ qualquer (nesse método é usual tomar $\underline{\theta}'$ como a moda ou a média da posteriori para garantir estabilidade numérica). Portanto, $p(\underline{Y}) = \frac{p(\underline{Y}|\underline{\theta}')p(\underline{\theta}')}{p(\underline{\theta}'|\underline{Y})}$, onde $p(\underline{\theta}'|\underline{Y})$ deve ser estimado, já que freqüentemente conhecemos apenas o núcleo da posteriori $p(\underline{\theta}|\underline{Y})$.

A seguir descrevemos o método para o caso particular em que $\underline{\theta}$ é composto de três blocos, $\underline{\theta} = (\underline{\theta}_1, \underline{\theta}_2, \underline{\theta}_3)$, e as condicionais completas $p(\underline{\theta}_3|\underline{\theta}_1, \underline{\theta}_2, \underline{Y})$, $p(\underline{\theta}_2|\underline{\theta}_1, \underline{\theta}_3, \underline{Y})$

e $p(\theta_1|\theta_2, \theta_3, \underline{Y})$ são conhecidas explicitamente. Nesse caso, $p(\theta'|Y) =$

$p(\theta_3|\theta_1, \theta_2, \underline{Y}) \cdot p(\theta_2|\theta_1, \underline{Y}) \cdot p(\theta_1|Y)$ e devemos estimar apenas $p(\theta_2|\theta_1, \underline{Y})$ e $p(\theta_1|Y)$, já que $p(\theta'|\theta_1, \theta_2, \underline{Y})$ é conhecido.

Para estimar $p(\theta_1|Y)$ podemos usar o estimador de Monte Carlo $\hat{p}(\theta_1|Y) = \frac{1}{G} \sum_{g=1}^G p(\theta_1|\theta_2^{(g)}, \theta_3^{(g)}, \underline{Y})$, usando uma amostra $(\theta_2^{(1)}, \theta_3^{(1)}, \dots, \theta_2^{(G)}, \theta_3^{(G)})$ obtida pela algoritmo de Gibbs quando amostramos da posteriori $p(\theta_1, \theta_2, \theta_3|Y)$, onde $\theta_1 = \theta_1'$.

Para estimar $p(\theta_2|\theta_1, \underline{Y})$, devemos notar primeiramente que

$$p(\theta_2|\theta_1, \underline{Y}) = \int_{\Theta_3} p(\theta_2, \theta_3|\theta_1, \underline{Y}) d\theta_3 = \int_{\Theta_3} p(\theta_2|\theta_1, \theta_3, \underline{Y}) p(\theta_3|\theta_1, \underline{Y}) d\theta_3,$$

e portanto se pudermos gerar $(\theta_3^{(1)}, \dots, \theta_3^{(G)})$ de $p(\theta_3|\theta_1, \underline{Y})$, um estimador de Monte Carlo de $p(\theta_2|\theta_1, \underline{Y})$ é dado por

$$\hat{p}(\theta_2|\theta_1, \underline{Y}) = \frac{1}{G} \sum_{g=1}^G p(\theta_2|\theta_1, \theta_3^{(g)}, \underline{Y}).$$

O problema então é como gerar essa amostra.

A seguir descrevemos o algoritmo de Gibbs para amostrar de $p(\theta_3|\theta_1, \underline{Y})$.

Passo 1: dar um valor inicial $\theta_3 = \theta_3^{(0)}$ e fazer $i = 0$,

Passo 2: gerar $\theta_2^{(i+1)}$ de $p(\theta_2|\theta_1, \theta_3^{(i)}, \underline{Y}) = p(\theta_2|\theta_1 = \theta_1', \theta_3 = \theta_3^{(i)}, \underline{Y})$,
gerar $\theta_3^{(i+1)}$ de $p(\theta_3|\theta_1, \theta_2^{(i+1)}, \underline{Y}) = p(\theta_3|\theta_1 = \theta_1', \theta_2 = \theta_2^{(i+1)}, \underline{Y})$,

Passo 3: fazer $i = i + 1$ e voltar ao passo 2, até a convergência.

Finalmente, tendo obtido $\hat{p}(\theta_1|Y)$ e $\hat{p}(\theta_2|\theta_1, \underline{Y})$, temos que

$$\hat{p}(Y) = \frac{p(\underline{Y}|\theta')p(\theta')}{\hat{p}(\theta'|Y)},$$

ou seja, $\log \hat{p}(Y) = \log p(\underline{Y}|\theta') + \log p(\theta') - \log \hat{p}(\theta'|Y)$, onde

$$\log \hat{p}(\theta'|Y) = \log p(\theta_3|\theta_1, \theta_2, \underline{Y}) + \log \hat{p}(\theta_2|\theta_1, \underline{Y}) + \log \hat{p}(\theta_1|Y).$$

(b) O critério do Pseudo Fator de Bayes – formulação geral

Vimos que o Fator de Bayes não está definido quando a priori é imprópria. Uma medida alternativa robusta nesse caso é o Pseudo Fator de Bayes, definido por Geisser e Eddy (1979).

O Pseudo Fator de Bayes (PBF) para dois modelos competitivos M_1 e M_2 é definido por

$$\text{PBF}_{12} = \frac{\prod_{i=1}^n \pi_1(Y_i|Y_{(-i)})}{\prod_{i=1}^n \pi_2(Y_i|Y_{(-i)})} = \frac{\prod_{i=1}^n \text{CPO}_i^{(M_1)}}{\prod_{i=1}^n \text{CPO}_i^{(M_2)}},$$

onde $\text{CPO}_i^{(M_l)} = \pi_l(Y_i|Y_{(-i)})$, que é a densidade de validação cruzada no modelo M_l , $l = 1, 2$.

Interpretação: Um valor alto de PBF na fórmula acima favorece o modelo M_1 .

(c) O critério “BIC” (“Bayesian Information Criterion”)

Esta é uma medida proposta por Schwarz (1978) que pode ser usada para comparar modelos e que não depende de conhecimento à priori, mas apenas da verossimilhança.

O “BIC” do modelo M_i é definido por

$$\text{BIC}_{M_i} = -2 \log \left(\sup_{M_i} f(Y|\vartheta) \right) + p_i \log n = -2 \log(f_{M_i}(Y|\hat{\vartheta}_i)) + p_i \log n,$$

onde $\hat{\vartheta}_i$ é o ponto de máximo de $f_{M_i}(Y|\vartheta_i)$, ou seja, $\hat{\vartheta}_i$ é o estimador de máxima verossimilhança de ϑ_i e $p_i \log n$ é a penalidade (onde n é o tamanho da amostra e p_i é a dimensão do espaço paramétrico do modelo M_i).

Interpretação: O modelo com o menor BIC é o “melhor”.

Schwarz (1978) mostrou que para n grande e sob certas condições, $\Delta\text{BIC} \approx -2 \log \text{BF}$, ou seja, $\text{BF} \approx \exp\{-\frac{\Delta\text{BIC}}{2}\}$, onde BF é o Fator de Bayes e $\Delta\text{BIC} = \text{BIC}_{M_1} - \text{BIC}_{M_2}$. Ver também Gelfand e Dey (1994).

Referências bibliográficas

- Arellano-Valle, R. B. (1994). *Distribuições elípticas: propriedades, inferência e aplicações a modelos de regressão*. 262p. Tese (Doutorado). Instituto de Matemática e Estatística. Universidade de São Paulo.
- Berger, J. O. (1980). *Statistical Decision Theory – Foundations, Concepts and Methods*. New York, Heidelberg, Berlin: Springer-Verlag.
- Berger, J. O. (2000). Bayesian analysis: A look at today and thoughts of tomorrow. *Journal of the American Statistical Association*, **95**(452).
- Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference (with discussion). *Journal of The Royal Statistical Society, B*, **41**, 113–147.
- Best, N. G.; Cowles, M. K. and Vines, K. (1995). *CODA: Convergence diagnosis and output analysis software for Gibbs sampling output, version 0.30*. Technical Report, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University.
- Box, G. E. P. and Tiao, G. (1973). *Bayesian Inference in Statistical Analysis*. Londo: Addison-Wesley.
- Brooks, S. and Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computation and Graphical Statistics*, **7**(4), 434–455.
- Brodsky, B. E.; Darkhovsky, B. S. (1993). *Nonparametric Methods in Change-Point Problems*. Kluwer Academic Publishers.
- Broemeling, L. D.; Chin Choy, J. H. (1981). Detecting structural change in linear models. *Communications in Statistics*, **4**, 33–47.
- Carlin, B. P.; Gelfand, A.; Smith, A. F. (1992). Hierarchical bayesian analysis of changepoint problems. *Applied Statistics*, **41**(2), 389–405.

- Carlin, B. P. and Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall/CRC.
- Carroll, R. J.; Ruppert, D. and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*. Chapman & Hall.
- Casella, G. and George, E. L. (1992). Explaining the Gibbs sampler. *The American Statistician*, **46**, 167–174.
- Caussinus, H. e Lyazrhi, F. (1997). *Choosing a linear model with a random number of change-points and outliers*.
- Chang, M. P.; Huang, W. T. (1997). Inferences for the linear errors-in-variables with changepoint models. *J. Amer. Statist. Association*, **92**(437), 171–178.
- Chong, T-L.T. (1999). Estimation of and testing for structural break in the errors-in-variables model. *Journal of Applied Statistical Science*, **9**(2), 147–158.
- Chen, J. e Gupta, A. K. (1997). Testing and locating variance change-points with applications to stock prices. *Journal of the American Statistical Association*, **92**, 739–747.
- Chen, M. H.; Shao, Q. M. and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer-Verlag.
- Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, **49**, 327–335.
- Chin Choy, J. H. e Broemeling, L. D. (1980). Some bayesian inference for a changing linear model. *Technometrics*, **22**(1), 71–78.
- Chung, Y. e Dey, D. K. (1996). Bayes factor for change point problem with conjugate prior. *Journal of the Korean Statistical Society*, **25**(4).
- Cowles, M. K. and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *J. Amer. Statist. Association*, **91**, 883–904.
- Dey, D. K. (1997). Bayesian approach to change point problems. *Commun. Statist. - Theory Meth.*, **26**(8), 2035–2047.
- Dey, D. K.: Maiti, T. (2002). *An Appraisal from Bayesian Perspective*. Technical Report.

- Ferreira, P. E. (1975). A bayesian analysis of a switching regression model: known number of regimes. *Journal of the American Statistical Association*, **70**, 370–374.
- Florens, J. P.; Mouchart, M. e Richards, J. F. (1974). Bayesian inference in errors-in-variables models. *Journal of Multivariate Analysis*, **4**, 419–456.
- Fuller, W. A. (1980). *Measurement Error Models*. New York: Erickson, T.
- Gamermann, D. (1996). *Simulação Estocástica via Cadeias de Markov*. 12° SINA-PE-Caxambu.
- Gbur, E. E. e Dahm, P. F. (1985). Estimation of the linear-linear segmented regression model in the presence of measurement error. *Communication in Statistics, A*, **14**(4), 809–826.
- Geisser, S. (1982). Predictive inference in duscussion of G. E. Box. *Journal of the Royal Statistical Society, Series A*, **143**, 416–417.
- Geisser, S. and Eddy, W. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, **74**, 153–160.
- Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: asymptotics and exact calculations. *J. R. Statist. Soc., B*, **56**, 501–514.
- Gelfand, A. E.; Dey, D.K. and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods (with discussion). In *Bayesian Statistics*, **4**, J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds., Oxford: Oxford University Press, 147–167.
- Gelfand, A. E., and Sahu, S. K. (1999). Identifiability improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association*, **94**, 247–253.
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398–409.
- Gelman, A.; Carlin, J.; Stern, H., and Rubin, D. B. (1997). *Bayesian Data Analysis*. London: Chapman and Hall/CRC.

- Gelman, A. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–511.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.
- Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, **57**, 1317–1339.
- Geweke, J. (1993). Bayesian treatment of the independent Student-*t* linear model. *Journal of Applied Econometrics*, **8**, 519–540.
- Gilks, W. R.; Best, N., and Tan, K. K. C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics*, **44**, 455–472.
- Gill, P. E.; Murray, W; Wrigert, M. H. (1981). *Practical Optimization*. Academic Press. London, New York, Toronto, Sydney and S. Francisco.
- Grimshaw, S. D. (1992). Estimation of the linear-Plateau segmented regression model in the presence of measurement error. *Communication in Statistics, A*, **21**(8), 2399–2413.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.
- Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika*, **57**, 1–16.
- Holbert, D. (1982). A bayesian analysis of a switching linear model. *Journal of Econometrics*, **19**, 77–87.
- Ibrahim, J. G., Chen, M. H., Sinha (2001). Criterion based methods for bayesian model assessment. *Statistica Sinica*, **11**, 419–443.
- Jeffrey, H. (1961). *Theory of Probability*. 3.ed. Oxford: University Press.
- Johnson, R. A.; Wichern, D. W. (1998). *Applied Multivariate Statistical Analysis*. Prentice Hall, W. J. 07458.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *J. Amer. Statist. Assoc.*, **90**, 773–795.

- Kass, R. E. and Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of The American Statistical Association*, **91**, 1343–1370.
- Kim, D. (1991). A bayesian significance test of the stationarity of regression parameters. *Biometrika*, **78**(3), 667–75.
- Land, M. e Broemeling, L. (1983). Bayesian forecasting with changing linear models. *Communications in Statistics, A*, **12**(12), 1421–1430.
- Lange, K. (1999). *Numerical Analysis for Statisticians*. New York: Springer-Verlag, Inc.
- Leiva, R. A. (1994). Bayesian prediction of the number of changepoints in the piecewise linear model. *Journal of the Italian Statistical Society*, **3**(2), 271–289.
- Liu, J. S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, **89**, 958–966.
- Liu, J. S., Wong, W. H., and Kong, A. (1995). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentations. *Biometrika*, **81**, 27–40.
- Lyazrhi, F. (1997). Bayesian criteria for discriminating among regression models with one possible change point. *Journal of Statistical Planning and Inference*, **59**, 317–353.
- Mengersen, K. L.; Robert, C. P. and Guihenec-Jouyavx, C. (1999). MCMC convergence diagnostics: a review (with discussion). In *Bayesian Statistics 6*, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds., Oxford; Oxford University Press, pp. 415–440.
- Meyn, S. P. and Tweedie, R. I. (1993). *Markov Chains and Stochastic Stability*. London: Springer-Verlag.
- Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. W.; Teller, A. H. and Teller, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Physics*, **21**, 1087–1092.
- Moen, D. H.; Salazar, D. e Broemeling, L. D. Structural changes in multivariate regression models. *Communications in Statistics, A*, **14**, 1757–1768.

- Newton, M. A. and Raftery, A. E. (1994). Aproximate Bayesian inference by the weighted likelihood bootstrap (with discussion). *J. Roy. Statist. Soc., Ser. B*, **56**, 1–48.
- Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeys a two separate regimes. *Journal of the American Statistical Association*, **53**, 873–880.
- Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association*, **55**, 324–330.
- Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer-Verlag.
- Roberts, G. O. and Sahu, S. K. (1997). Updating schemes, correlation structure, blocking and parametrization for the Gibbs sampler. *Journal of the Royal Statistical Society, Series B*, **59**, 291–317.
- Roberts, G. O. and Smith, A. F. M. (1993). Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithm. *Stochastic Processes and their Applications*, **49**, 207–216.
- Roberts, G. O. and Tweedie, R. I. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, **83**, 95–110.
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.*, **6**, 461–464.
- Searle, S. R. (1971). *Linear Models*. New York, London, Sydney, Toronto: John Wiley & Sons, Inc.
- Smith, A. F. M. (1975). A bayesian approach to inference about change point in sequence of random variables. *Biometrika*, **62**, 407–416.
- Smith, A. F. M. and Gelfand, A. E. (1992). Bayesian statistics without tears. A sampling-resampling perspective. *The American Statistician*, **46**, 84–88.
- Smith, A. F. M.; Skeene, A. A.; Shaw, J. E. H., and Naylor (1987). Progress with numerical and graphical methods for practical Bayesian statistics. *The Statistician*, **36**, 75–82.

- Smith, B. J. (2003). Bayesian Output Analysis Program (BOA) Version 1.0 User's Manual. Disponível em <http://www.public-health.uiowa.edu>.
- Spiegelhalter, D. J.; Best, N. G.; Carlin, B. P. e Van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, **64**, 583–639.
- Stefanski, L. A. (2000). Measurement error models. *Journal of The American Statistical Association*, **95**(452), 1353–1358.
- Stephens, D. A. and Dellaportas, P. (1992). Bayesian analysis of generalized linear models with covariate measurement error. *Bayesian Statistics*, **4**, 813–820. Oxford University Press.
- Tanner, M. A. (1993). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. 2.ed. New York: Springer-Verlag.
- Teeter, R. (1979). Effects of measurement error in piecewise regression models. American Statistical Association, 1979. *Proceedings of the Social Statistics Section*, 422–426.
- Teeter, R. (1985). The application of linear piecewise regression to basal body temperature data. *Biom. J.*, **27**(7), 750–773.
- Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics*, **22**, 1701–1762.
- Wang, C. H. e Lee, C. B. (1993). Bayesian analysis for a change in the intercept of simple linear regression. *Communication in Statistics, A*, **22**(11), 3031–3057.
- Worsley, K. J. (1983). The power of likelihood ratio and cumulative sum tests for a change in a binomial probability. *Biometrika*, **70**, 455–465.
- Zacks, S. (1983). Survey of classical and bayesian approaches to the change-point problem: fixed sample and sequential procedures of testing and estimation. *Recent Advances in Statistics*, Academic Press, 245–269.