Inferência Estatística para Modelos Lineares
com Restrições nos Parâmetros em Condições
Regulares e Não Regulares

Viviana Giampaoli

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. Julio da Motta Singer**

-São Paulo, Novembro de 1999-

Inferência Estatística para Modelos Lineares com Restrições nos Parâmetros em Condições Regulares e Não Regulares

Este exemplar corresponde à redação final devidamente corrigida e defendida por Viviana Giampaoli e aprovada pela comissão julgadora.

São Paulo, 12 de Novembro de 1999.

Banca examinadora:

- Prof. Dr. Júlio da Motta Singer
- Prof. Dr. Nelson Ithiro Tanaka
- Prof. Dr. Josemar Rodrigues
- Prof. Dr. Basilio de Bragança Pereira
- Prof. Dr. Enrico Colosimo

Resumo

Estudamos estatísticas adequadas para testar hipóteses sobre os parâmetros de modelos lineares com restrições no espaço paramétrico. Propomos um teste alternativo para dois casos em que as restrições sobre o espaço paramétrico são impostas pelo objetivo da investigação em condições regulares, (i.e. quando o parâmetro está no interior do espaço paramétrico). Concluimos, através de estudos de simulação, que o teste alternativo é, em geral, mais poderoso que o teste usual. Uma abordagem bayesiana utilizando o fator de Bayes para a análise deste tipo de hipótese também foi considerada. Nesse contexto, mostramos que a metodologia proposta por Irony e Pereira (*Resenhas IME-USP*, (1995) 27-46) para o cálculo do fator de Bayes é mais flexível que proposta por Carlin e Chib (*Journal of the Royal Statistical Society B*, 57, (1995) 473-484) quanto à escolha das distribuições a priori. Também consideramos testes de hipóteses sob condições não regulares (quando a hipótese coloca o parâmetro na fronteira do espaço paramétrico) para as componentes de variância de um modelo de efeitos aleatórios. Aplicamos os resultados dados por Vu e Zhou (*The Annals of Statistics*, 25, (1997) 897-916) sobre a estatística da razão de verossimilhanças, identificando casos específicos em que as condições de regularidade são válidas.

Abstract

Suitable statistics for testing of hypothesis on the parameters of linear models with restrictions on the parametric space were studied. An alternative test for two cases in which the restrictions are imposed by the research objective in regular conditions (i.e. when the parameter is in the interior of the parametric space) was proposed. Using simulation studies, we concluded that, in general, that the alternative test is more powerful than the usual test. A Bayesian approach for the analyses of this type of hypothesis using the Bayes' factor was also considered. In this context, we showed that the methodology proposed by Irony and Pereira (*Resenhas IME-USP*, (1995) 27-46) for calculation of Bayes' factor is more flexible than proposed by Carlin and Chib (*Journal of the Royal Statistical Society B*, 57, (1995) 473-484) with respect to choice of the prior distribution. Tests of hypothesis in nonstandard conditions (when the hypothesis sets the parameter in the boundary of the parametric space) for the components of variance in a random model were also proposed. We applied the results of Vu and Zhou (*The Annals of Statistics*, 25, (1997) 897-916) about the likelihood ratio test to specific cases in which the conditions of regularity are valid.

Índice

1	Introdução	7
1.1	Inferência em espaços paramétricos restritos	7
1.2	Testes de hipóteses sob condições não regulares	13
1.2.1	Proposta deste trabalho	16
2	Comparação de duas populações normais com médias restritas.	18
2.1	Definição da estatística de teste e comparação do poder	18
2.2	Desempenho do teste alternativo baseado em estimadores da média comum	32
2.3	Enfoque bayesiano	40
2.3.1	Cálculo do fator de Bayes segundo a metodologia de Carlin e Chib	42
2.3.2	Cálculo do fator de Bayes segundo a metodologia de Irony e Pereira	47
2.4	Extensão para o caso em que a variância é desconhecida	63
2.4.1	Enfoque Bayesiano	69
2.5	Discussão	76
3	Comparação de parâmetros de regressão restritos num modelo normal linear sem intercepto.	88
3.1	Definição da estatística do teste e comparação do poder	88
3.2	Desempenho do teste alternativo baseado no estimador do parâmetro de regressão comum	100
3.3	Enfoque Bayesiano	106

3.3.1	Cálculo do fator de Bayes segundo a metodologia de Carlin e Chib	106
3.3.2	Calculo do fator de Bayes segundo a metodologia de Irony e Pereira	109
3.4	Extensão para o caso em que a variância é desconhecida	115
3.4.1	Enfoque bayesiano	122
3.5	Discussão	127
4	Testes de hipóteses em condições não regulares	140
4.1	Generalização do Teste da Razão de Verossimilhanças sob Condições Não Regulares	141
4.1.1	Resultados Fundamentais	141
4.1.2	Aplicações	146
4.2	Discussão	178
A	Definição de estimadores para uma média restrita	179
B	Elementos básicos de inferência bayesiana	183
B.1	Teorema de Bayes e Famílias Conjugadas	183
B.2	Cadeias de Markov e amostrador de Gibbs	186
B.3	Testes de hipóteses: enfoque Bayesiano para a escolha de modelos.	189
B.4	Significado da constante c definida na Subseção 2.3.2	192
C	Cálculo da matriz inversa	194
D	Cálculo dos elementos da matriz análoga à matriz de informação	200

Lista de Tabelas

1.1	Pressão diastólica (média de 3 medidas) (mmHg)	9
1.2	Índice de placa bacteriana	10
2.1	Valores da média comum abaixo dos quais o teste alternativo é mais poderoso que o teste usual	33
2.2	Valores da média comum abaixo dos quais o teste alternativo é mais poderoso que o teste usual	34
2.3	Valores do nível de significância empírico	36
2.4	Valores do poder empírico	38
2.5	Estimativas de mínimos quadrados	39
2.6	Estimativas de mínimos quadrados	43
2.7	Diagnóstico de convergência de Geweke	44
2.8	Fator de Bayes com distribuição a priori informativa para a comparação de médias	45
2.9	Valores empíricos a posteriori sob o espaço paramétrico restrito (com distribuições a priori informativas)	45
2.10	fatores de Bayes para diferentes precisões das distribuições a priori dos parâmetros μ , μ_1 e μ_1	46
2.11	Valores empíricos a posteriori sobre o espaço paramétrico irrestrito (com distribuições a priori informativas)	47
2.12	Valores dos hiperparâmetros e os fatores de Bayes correspondentes . . .	56
2.13	Fator de Bayes para o hiperpâmetro $\gamma = 0.349$	57

2.14 Fator de Bayes considerando distribuições a priori normais no espaço paramétrico irrestrito	58
2.15 Fatores de Bayes	61
2.16 Fator de Bayes para distribuições a priori exponenciais	61
2.17 Valores do nível de significância empírico com variância desconhecida . .	68
2.18 Valores do poder empírico com variância desconhecida	70
2.19 Diagnóstico de convergência de Geweke quando σ^2 é desconhecida	71
2.20 Fator de Bayes quando σ^2 é desconhecida	72
2.21 Valores empíricos a posteriori para o espaço paramétrico restrito quando σ^2 é desconhecida	72
2.22 Valores empíricos a posteriori para o espaço paramétrico irrestrito quando σ^2 é desconhecida	73
2.23 Valores dos hiperparâmetros e os fatores de Bayes correspondentes quando σ^2 é desconhecida	75
2.24 Fator de Bayes considerando distribuições a priori normais no espaço paramétrico irrestrito quando σ^2 é desconhecida	76
2.25 quantis amostrais para $n = 5$ e $\sigma = 0.5$	81
2.26 quantis amostrais para $n = 5$ e $\sigma = 1.0$	82
2.27 quantis amostrais para $n = 5$ e $\sigma = 1.5$	83
2.28 quantis amostrais para $n = 5$ e $\sigma = 2.0$	84
2.29 quantis amostrais para $n = 10$ e $\sigma = 0.5$	85
2.30 quantis amostrais para $n = 10$ e $\sigma = 1.0$	86
2.31 quantis amostrais para $n = 10$ e $\sigma = 1.5$	87
3.1 Valores empíricos do nível de significância do teste alternativo para diferentes matrizes da variável explicativa, com variância conhecida	102
3.2 Valores do poder empírico do teste alternativo com variância conhecida .	104
3.3 Valores do poder empírico do teste alternativo com variância conhecida (continuação)	105

3.4	Estimativas de mínimos quadrados	105
3.5	Estimativas de mínimos quadrados dos parâmetros de cada modelo . . .	107
3.6	Fator de Bayes para comparar os parâmetros de regressão, quando a variância é conhecida	107
3.7	Fatores de Bayes para diferentes precisões das distribuições a priori dos parâmetros de regressão	108
3.8	Valores empíricos a posteriori sobre o espaço paramétrico restrito	108
3.9	Valores empíricos a posteriori sobre o espaço paramétrico irrestrito . . .	109
3.10	Fator de Bayes correspondente à distribuição a priori normal truncada .	114
3.11	Valores empíricos do nível de significância com a variância desconhecida considerando $\mathbf{X}$, $\mathbf{X}_2$, $\mathbf{X}_3$ e $\mathbf{X}_4$	119
3.12	Valores do poder empírico, com a variância desconhecida	120
3.13	Valores do poder empírico, com a variância desconhecida (continuação) .	121
3.14	Fator de Bayes para a comparação de parâmetros de regressão, quando a variância é desconhecida	122
3.15	Valores empíricos a posteriori, sob o espaço paramétrico restrito, quando a variância é desconhecida	123
3.16	Fatores de Bayes correspondentes a distribuição a priori normal truncada, para o caso em que a variância é desconhecida, calculados com a metodologia dada por Carlin e Chib	123
3.17	Valores empíricos a posteriori, sob o espaço paramétrico irrestrito, quando a variância é desconhecida	124
3.18	Fatores de Bayes correspondentes à distribuição a priori normal truncada, para o caso em que a variância é desconhecida, calculados com a metodologia dada por Irony e Pereira	126
3.19	quantis amostrais para $\sigma = 0.13$	128
3.20	quantis amostrais para $\sigma = 0.13$ (continuação)	129
3.21	quantis amostrais para $\sigma = 0.13$ (continuação)	130

3.22	quantis amostrais para $\sigma = 0.13$ (continuação)	131
3.23	quantis amostrais para $\sigma = 0.5$	132
3.24	quantis amostrais para $\sigma = 0.5$ (continuação)	133
3.25	quantis amostrais para $\sigma = 0.5$ (continuação)	134
3.26	quantis amostrais para $\sigma = 0.5$ (continuação)	135
3.27	quantis amostrais para $\sigma = 1.0$	136
3.28	quantis amostrais para $\sigma = 1.0$ (continuação)	137
3.29	quantis amostrais para $\sigma = 1.0$ (continuação)	138
3.30	quantis amostrais para $\sigma = 1.0$ (continuação)	139
A.1	vícios para o estimador ξ	181
A.2	Desempenho dos estimadores da média restrita	182
B.1	Regra de decisão	190

Capítulo 1

Introdução

Em muitos problemas de análise de dados deseja-se fazer inferência sobre os elementos de um espaço paramétrico Ω (subconjunto de $\mathbb{R}^p$); por exemplo, há situações em que se pretende saber se o verdadeiro valor de um vetor de parâmetros pertence a um conjunto determinado ou se suas componentes satisfazem uma determinada relação. A essas questões estão associados diferentes tipos de restrições: aquelas impostas pelo espaço paramétrico e aquelas induzidas pelas hipóteses, gerando, em alguns casos, condições não regulares em que a hipótese nula situa o verdadeiro valor do parâmetro na fronteira do espaço paramétrico. Nas seções seguintes apresentaremos detalhes de alguns problemas relacionados com esses tópicos que serão objeto deste trabalho.

1.1 Inferência em espaços paramétricos restritos

Consideramos modelos lineares da forma

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1.1}$$

em que $\mathbf{y}$ é um vetor de observações com dimensão $n \times 1$, $\mathbf{X}$ é uma matriz de constantes de dimensão $n \times p$, conhecida e de posto completo, $\boldsymbol{\beta}$ é o vetor de parâmetros de posição

com dimensão $p \times 1$ e ε é um vetor de erros aleatórios com dimensão $n \times 1$. Assumimos também que $E(\varepsilon) = \mathbf{0}$ e $Var(\varepsilon) = \mathbf{\Gamma} = \sigma^2 \mathbf{Q}$, com $\mathbf{Q}$ denotando uma matriz positiva definida de dimensão $n \times n$ e σ^2 uma constante positiva. Além disso, dados dois vetores $\mathbf{z} = (z_1, \dots, z_p)$ e $\mathbf{a} = (a_1, \dots, a_p)$, utilizamos a notação $\mathbf{z} \geq \mathbf{a}$ para indicar $z_i \geq a_i$, $i = 1, \dots, p$. Nosso foco de atenções está em problemas de inferência relacionados com os seguintes tipos de hipóteses:

- a) $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{c}$ contra $H_1 : \mathbf{C}\boldsymbol{\beta} \neq \mathbf{c}$, em que $\mathbf{C}$ é uma matriz $c \times p$ de constantes conhecidas e $\mathbf{c}$ denota um vetor $c \times 1$ de constantes conhecidas.
- b) $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{c}$ contra $H_1 : \mathbf{C}\boldsymbol{\beta} \geq \mathbf{c}$, com pelo menos uma desigualdade estrita, sob a suposição adicional de que o espaço paramétrico é restrito por desigualdades da forma

$$\mathbf{R}\boldsymbol{\beta} \geq \mathbf{r} \quad (1.2)$$

com $\mathbf{R}$ denotando uma matriz $k \times p$ ($k \leq p$) de posto completo, possivelmente diferente de $\mathbf{C}$ e $\mathbf{r}$ denotando um vetor $k \times 1$ de constantes conhecidas.

Exemplo 1.1: Suponhamos que com base em amostras de tamanho n_1 e n_2 , há interesse em comparar a média da pressão diastólica de indivíduos submetidos a situações de estresse, como a morte de um parente, com aquela de indivíduos em condições normais, isto é, não submetidos a estresse. Um exemplo numérico com $n_1 = n_2 = 11$ é apresentado na Tabela 1.1. Sob a hipótese de normalidade, este é um problema bastante conhecido e é tratado na maioria dos textos de Estatística Básica. Admitamos agora que os indivíduos de ambos grupos são hipertensos, ou seja, que sua pressão diastólica média é maior ou igual a 90 mmHg.

O problema pode então ser definido como de comparação das médias, $(\mu_1, \text{ e } \mu_2)$, de duas populações com distribuição normal, sob a restrição de que ambas são maiores ou iguais a uma constante c . Este problema pode ser expresso no contexto de modelos

Tabela 1.1: Pressão diastólica (média de 3 medidas) (mmHg)

grupo com estresse	grupo controle
89.6	81.7
92.4	89.9
87.1	81.5
92.2	89.4
96.4	94.6
92.2	93.5
92.7	85.5
95.0	95.5
92.2	88.9
109.2	95.4
96.8	97.0

lineares (1.1), com

$$p = 2, \mathbf{X} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} \end{bmatrix}, \Gamma = \sigma^2 \mathbf{I}_n, \boldsymbol{\beta} = (\mu_1, \mu_2)^t \quad (1.3)$$

sob as restrições (1.2) com $\mathbf{1}_{n_i}$ denotando um vetor de dimensão $n_i \times 1$ com todos os elementos iguais a 1, $\mathbf{I}_n$ a matriz identidade de ordem n , $\mathbf{R} = \mathbf{I}_2$, e $\mathbf{r} = [c, c]^t$. Sem perda de generalidade, podemos tomar $c = 0$, e considerar as hipóteses

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 > \mu_2.$$

Exemplo 1.2. : Os dados apresentados na Tabela 1.2 são provenientes de um estudo na área de Odontopediatria em que há interesse em comparar a eficácia de dois tipos de escovas de dentes na remoção da placa bacteriana. Oito crianças utilizam um determinado tipo de escova (A), e outras oito utilizam um outro tipo de escova (B). A quantidade de placa bacteriana é avaliada através de um índice apropriado medido antes e depois da escovação.

Uma análise apropriada para este tipo de experimento, conhecido na literatura como

Tabela 1.2: Índice de placa bacteriana

Escova A		Escova B	
Antes da escovação	Depois da escovação	Antes da escovação	Depois da escovação
0.62	0.57	0.45	0.40
0.70	0.64	1.15	0.65
0.90	0.80	1.27	1.22
0.95	0.92	0.87	0.80
1.20	1.17	0.82	0.77
0.10	0.10	0.97	0.95
0.57	0.55	0.50	0.47
1.35	1.32	0.73	0.35

preteste/pósteste, foi considerada em Singer e Andrade (1997). Essa análise é baseada em modelos lineares da forma (1.1) em que $\mathbf{y} = (y_{11}, \dots, y_{1n_1}, y_{21}, \dots, y_{2n_2})^t$, $\mathbf{X} = \text{diag}\{\mathbf{x}_1, \mathbf{x}_2\}$ com $\mathbf{x}_i = (x_{i1}, \dots, x_{in_i})^t$, x_{1j} (x_{2j}) denotando a resposta preteste, y_{1j} (y_{2j}) a resposta pósteste do j -ésimo sujeito utilizando a escova A (escova B) e β_1 (β_2) a taxa de placa bacteriana residual associada aos indivíduos que utilizam a escova A (escova B), $j = 1, \dots, n_i$. Como esperamos que o índice de placa bacteriana antes da escovação deva ser maior ou igual ao índice depois da escovação, e como os índices devem ser positivos, os parâmetros β_i são tais que $0 \leq \beta_i \leq 1$, ($i = 1, 2$).

Hipóteses do tipo

$$H_0 : \beta_1 = \beta_2,$$

$$H_1 : \beta_1 > \beta_2,$$

podem ser expressas no formato (b) com

$$\mathbf{C} = [1, -1], \quad \mathbf{c} = 0, \quad \mathbf{R} = \mathbf{I}_2 \otimes \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \text{e} \quad \mathbf{r} = \mathbf{1}_2 \otimes \begin{bmatrix} -1 \\ 0 \end{bmatrix}. \quad (1.4)$$

Nas situações em que o espaço paramétrico para β é $\Omega = \mathbb{R}^p$, resultados inferenciais

são bastantes conhecidos. O estudo de testes de hipóteses do tipo (a) é amplamente considerado na teoria clássica de modelos lineares (ver Graybill (1976), por exemplo). Mais especificamente, inferência exata pode ser feita sob a suposição de normalidade dos erros e no caso mais geral, onde essa suposição é relaxada, resultados de inferência assintótica podem ser obtidos sob certas condições de regularidade (ver Sen e Singer (1993), Cap. 5, por exemplo).

Situações em que o espaço paramétrico para β é $[a, b] \times [a, b]$, com a e b números reais, constituirão o objeto deste trabalho.

O caso particular do problema de testar hipóteses do tipo

$$H_0 : \mu = (\mu_1, \mu_2, \dots, \mu_k)^t = 0 \quad (1.5)$$

versus

$$H_1 : \mu \text{ com pelo menos uma componente estritamente positiva} \quad (1.6)$$

em que μ representa o vetor de médias de uma população normal multivariada com matriz de covariância conhecida foi sugerido inicialmente por Bartholomew (1959a,b, 1961). Kudô (1963) mostra que a distribuição nula exata do teste da razão de verossimilhanças é uma mistura de distribuições qui-quadrado, diferentemente do caso usual em que não são assumidas restrições na alternativa (nesse caso a distribuição nula é qui-quadrado central). Resultados semelhantes foram obtidos independentemente por Nüesch (1966), e invalidados parcialmente por um erro apontado em Shorack (1967).

No caso em que a matriz de covariância Γ é completamente desconhecida, Perlman (1969) generaliza este resultado testando

$$H_0 : \mu \in P_1 \text{ versus } H_1 : \mu \in P_2, \quad (1.7)$$

em que P_1 e P_2 denotam conjuntos positivamente homogêneos, isto é, conjuntos para os quais se $x \in P_i$ então $ax \in P_i$ para todo a número real positivo, com $P_1 \subset P_2$; por exemplo,

$P_1 = \{0\}$ e $P_2 = \{\mu : \mu_1 \geq 0, \mu_i \in \mathbb{R} \ i = 2, \dots, p\}$. Ele só obteve os limites superior e inferior para a função distribuição da estatística da razão de verossimilhanças sob a hipótese nula para o caso particular em que $P_1 = \{0\}$ e $P_2 = C$, onde C é um cone, isto é, um conjunto positivamente homogêneo fechado com pelo menos um elemento não nulo. Para o caso particular (1.5) versus (1.6) em que $P_2 = \{\mu : \mu_i \geq 0, i = 2, \dots, p\}$, isto é, o quadrante positivo, Perlman (1969) mostra que a distribuição nula exata da estatística da razão de verossimilhanças é uma mistura de distribuições de cocientes de variáveis χ^2 com pesos que dependem da matriz de covariância que é desconhecida. Quando a matriz de covariância é substituída pela matriz de covariância amostral a distribuição nula exata da estatística da razão de verossimilhanças é uma mistura de distribuições de tipo F . Shapiro (1988) afirma que para o caso mais geral, em que P_2 é um cone, também a distribuição nula exata da estatística da razão de verossimilhanças é uma mistura de distribuições de tipo F .

Uma estatística de teste para o caso em que P_2 é um cone foi recentemente proposta por Tang (1994), e está baseada em semi-espacos de $\mathbb{R}^p$, ou seja, conjuntos do tipo $\{\mathbf{v} \in \mathbb{R}^p | \mathbf{v}^t \mathbf{u} \geq 0\}$ para algum vetor fixo $\mathbf{u}$. Por exemplo, se $p = 2$ e $\mathbf{u} = (1, 1)^t$, temos o semi-espaço $\{\mathbf{v} = (v_1, v_2)^t | v_1 \geq -v_2\}$ que contém o quadrante positivo. Resultados de simulação para tamanhos de amostra maiores que 10 e menores que 62 sugerem que esta estatística é uniformemente mais poderosa que a estatística T^2 de Hotelling. Wang e McDermott (1998) indicam as vantagens limitadas deste teste, dado que seus valores críticos não dependem do espaço paramétrico; esses autores propõem uma estatística da razão de verossimilhanças condicional para testar a hipótese nula (1.5) contra a hipótese alternativa (1.6).

No modelo de regressão normal mais geral, testes de hipóteses do tipo (b), sem as restrições em forma de desigualdade (1.2) sobre o espaço paramétrico, foram estudados por Gouriéroux et al. (1982) e Gouriéroux e Monfort (1995) entre outros. Gouriéroux e Monfort (1995) provam a equivalência assintótica da estatística de multiplicadores de Kuhn-Tucker (ξ_{KT}) com as estatísticas da razão de verossimilhanças (ξ_{RV}) e de Wald

(ξ_W) quando a matriz de covariância dos erros $\Gamma = \sigma^2 \mathbf{Q}$ é conhecida. Gouriéroux et al. (1982) obtêm também a relação $\xi_W \geq \xi_{RV} \geq \xi_{KT}$ entre essas estatísticas no caso em que Γ é desconhecida, porém seus elementos são funções contínuas de um número finito de parâmetros não relacionados com β . Excelentes referências para este tema são Barlow et al. (1972) e Paula (1997).

Nos testes de hipóteses em modelos lineares da forma (1.1) com restrições paramétricas em forma de desigualdades (1.2) aparece o problema da **estimação** dos parâmetros restritos, que tem longa história na literatura estatística. Judge e Takayama (1966) obtêm os estimadores de mínimos quadrados sob a restrição (1.2) com $\Gamma = \sigma^2 \mathbf{I}$ e Liew (1976) estuda as propriedades estatísticas desse estimador. Considerando uma certa função de perda, Toutenburg e Srivastava (1996) propõem estimadores para β quando as componentes do vetor de parâmetros estão contidas num intervalo determinado, que é um caso particular de (1.2).

O caso particular de estimação da média de uma população normal sob a restrição de ela pertencer a um intervalo aberto é tratado por Heiny e Siddiqui (1970), que obtêm o estimador de máxima verossimilhança, Stroud (1974) e Gupta e Rohatgi (1980), que propõem dois outros estimadores. Esses autores apresentam também os vícios e erros quadráticos médios de cada estimador, indicando para que valores do cociente entre a média e o desvio padrão o desempenho correspondente é melhor.

1.2 Testes de hipóteses sob condições não regulares

Consideremos agora estudos longitudinais em que para cada unidade amostral i , ($i = 1, \dots, m$) é observada uma variável resposta ao longo de n_i ocasiões.

Exemplo 1.3: Seja $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^t$ o vetor de respostas da i -ésima unidade amostral. Um modelo linear que associa a resposta da i -ésima unidade amostral a um conjunto de variáveis explicativas é dado por

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{e}_i \quad (1.8)$$

em que $\mathbf{X}_i$ é uma matriz conhecida de dimensão $n_i \times p'$, $\boldsymbol{\alpha}$ é um vetor de dimensão $p' \times 1$ de parâmetros de regressão desconhecidos e $\mathbf{e}_i$ são erros independentes e distribuídos segundo uma distribuição $N(\mathbf{0}, \boldsymbol{\Sigma}_i)$. Assumimos que os elementos de cada matriz $\boldsymbol{\Sigma}_i$ são funções conhecidas de p parâmetros desconhecidos contidos no vetor $\boldsymbol{\theta}$ e que os parâmetros de regressão $\boldsymbol{\alpha}$ variam independentemente dos parâmetros de covariância.

Um caso especial é o modelo de efeitos aleatórios

$$\mathbf{e}_i = \mathbf{Z}_i \mathbf{b}_i + \mathbf{u}_i \quad (1.9)$$

em que $\mathbf{Z}_i$ é uma matriz conhecida com dimensão $n_i \times k'$, $\mathbf{b}_i$ e $\mathbf{u}_i$ são vetores aleatórios independentes e tais que $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$ e $\mathbf{u}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i})$ o que implica,

$$\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \sigma^2 \mathbf{I}_{n_i}. \quad (1.10)$$

Se denotarmos por $d_{i,j}$ o elemento (i, j) da matriz de covariância $\mathbf{D}$, hipóteses nulas da forma $H_0 : d_{i,i} = 0$, colocam $\mathbf{D}$ na fronteira do espaço paramétrico

$$\Omega = \{\mathbf{D} : \mathbf{D} \text{ é definida não negativa (d.n.n.)}\} \quad (1.11)$$

isto é, implicam $\mathbf{D}$ semidefinida positiva. Alguns exemplos concretos são apresentados a seguir.

- i) Se $\mathbf{Z}_i = \mathbf{1}_{n_i}$ e $\mathbf{b}_i$ é um escalar b_i com distribuição $N(0, \sigma_A^2)$, $\mathbf{D} = \sigma_A^2$ e o espaço paramétrico é $\Omega = [0, +\infty)$, um teste da hipótese sobre a necessidade de inclusão do efeito aleatório b_i no modelo (1.9) é um teste de

$$H_0 : \sigma_A^2 = 0 \text{ contra } H_1 : \sigma_A^2 > 0. \quad (1.12)$$

- ii) Suponhamos agora que nosso interesse é testar a necessidade de inclusão de um segundo efeito aleatório no modelo (1.9), admitindo a existência de um primeiro

efeito aleatório. Se

$$\mathbf{D} = \begin{bmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{bmatrix},$$

o espaço paramétrico é dado por

$$\Omega = \left\{ \mathbf{D} : \mathbf{D} = \begin{bmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{bmatrix}, \mathbf{D} \text{ d.n.n, com } \sigma_A^2 > 0 \right\}$$

e a hipótese a testar será

$$H_0 : \mathbf{D} = \begin{bmatrix} \sigma_A^2 & 0 \\ 0 & 0 \end{bmatrix}, \quad (1.13)$$

contra

$$H_1 : \mathbf{D} \text{ d.n.n. com } \sigma_B^2 > 0. \quad (1.14)$$

As hipóteses (1.13) e (1.14) são respectivamente equivalentes a

$$H_0 : \sigma_B^2 = \sigma_{AB} = 0, \quad (1.15)$$

e

$$H_1 : \sigma_B^2 > 0, \sigma_A^2 \sigma_B^2 \geq \sigma_{AB}^2. \quad (1.16)$$

Essencialmente nos problemas apresentados, deseja-se testar a hipótese de que o verdadeiro valor do parâmetro $\boldsymbol{\theta}_0$ está situado num subconjunto de Ω , denotado por Ω_0 , versus a alternativa que $\boldsymbol{\theta}_0$ está situado em Ω_1 , também subconjunto de Ω . Nesses casos a distribuição da estatística da razão de verossimilhanças sob a hipótese nula depende da localização de $\boldsymbol{\theta}_0$ relativamente às fronteiras de Ω_0 , Ω_1 e Ω . Estas questões foram analisadas fundamentalmente por Chernoff (1954) e Self e Liang (1987) no caso em que as observações são procedentes de variáveis aleatórias independentes com a mesma função densidade $f(\mathbf{x}, \boldsymbol{\theta})$ e recentemente por Vu e Zhou (1997) no caso em que as variáveis

aleatórias não são identicamente distribuídas.

Stram e Lee (1994) estudam a distribuição assintótica do teste da razão de verossimilhanças para componentes de variância em modelos de efeitos aleatórios longitudinais utilizando a teoria sobre a distribuição assintótica do teste da razão de verossimilhanças em condições não regulares desenvolvida por Self e Liang (1987) e parecem não levar em conta o fato de que esses resultados só valem para variáveis aleatórias identicamente distribuídas.

1.2.1 Proposta deste trabalho

No Capítulo 2 consideramos o teste de igualdade de médias de duas populações normais, sob a restrição que ambas as médias sejam maiores ou iguais a uma constante c . Se o número de observações nas duas populações é o mesmo, nosso problema é um caso particular de (1.7), considerando uma população normal bivariada com

$$P_1 = \{\boldsymbol{\mu} : \mu_1 = \mu_2, \mu_1 \geq 0, \mu_2 \geq 0\} \text{ e } P_2 = \{\boldsymbol{\mu} : \mu_1 \geq \mu_2, \mu_1 \geq 0, \mu_2 \geq 0\}.$$

Abordamos aqui o caso em que não necessariamente $n_1 = n_2$ e obtemos resultados para amostras pequenas. No caso em que as médias não são restritas, a teoria clássica considera a estatística de teste baseada na diferença das médias amostrais. Por este motivo, trabalhamos com a estatística definida pela projeção desta diferença no primeiro quadrante. Analisamos hipóteses alternativas dos tipos (a) e (b) definidas em relação ao modelo (1.1) quando a variância é conhecida. Também estudamos o caso de variância desconhecida, em que certos cuidados são necessários na escolha do estimador correspondente.

Reanalisamos o mesmo problema sob um enfoque Bayesiano utilizando duas metodologias diferentes para calcular o fator de Bayes. Na primeira, empregamos amostradores de Gibbs (veja por exemplo Gamerman (1997)) cuja desvantagem é a exigência do emprego de distribuições que definam distribuições condicionais completas que possam ser

amostradas, por exemplo, distribuições conjugadas. Utilizando integrais de superfície sobre os espaços definidos pelas hipóteses nula e alternativa, empregamos a técnica proposta por Irony e Pereira (1995) com os mesmos objetivos. Essa técnica permite o uso de qualquer tipo de distribuições a priori.

No Capítulo 3 consideramos o teste de comparação de parâmetros de regressão do modelo linear (1.1) quando estes parâmetros pertencem a um intervalo fechado. Abordamos também este problema sob um o enfoque Bayesiano.

No Capítulo 4 estudamos testes de hipóteses sobre a matriz de covariância num modelo de efeitos aleatórios utilizando a estatística da razão de verossimilhanças proposta por Vu e Zhou (1997) analisando as condições de regularidade necessárias.

Capítulo 2

Comparação de duas populações normais com médias restritas.

2.1 Definição da estatística de teste e comparação do poder

Sejam $X_1, \dots, X_{n_1}$, n_1 variáveis aleatórias com distribuição $N(\mu_1, \sigma^2)$ e $Y_1, \dots, Y_{n_2}$, n_2 variáveis aleatórias com distribuição $N(\mu_2, \sigma^2)$, com X_j independente de Y_i para todo j e i , $\mu_1 \geq c, \mu_2 \geq c$, com c denotando uma constante não negativa, e admitamos que $\sigma^2 > 0$ é conhecida. Consideramos hipótese do tipo

$$H_0 : \mu_1 = \mu_2, \tag{2.1}$$

contra

$$H_1 : \mu_1 > \mu_2.$$

A função de verossimilhança é dada por

$$f(\mu_1, \mu_2, \sigma^2) = \frac{1}{(2\pi)^{(n_1+n_2)/2} (\sigma^2)^{(n_1+n_2)/2}} \times \exp \left(-\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n_1} (X_i - \mu_1)^2 + \sum_{i=1}^{n_2} (Y_i - \mu_2)^2 \right] \right). \quad (2.2)$$

Escrevendo

$$\sum_{i=1}^{n_1} (X_i - \mu_1)^2 = \sum_{i=1}^{n_1} (X_i - \bar{X})^2 + n_1 (\bar{X} - \mu_1)^2$$

e

$$\sum_{i=1}^{n_2} (Y_i - \mu_2)^2 = \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2 + n_2 (\bar{Y} - \mu_2)^2,$$

obter o estimador de máxima verossimilhança de (μ_1, μ_2) num espaço paramétrico determinado, $\tilde{\Omega}$, é encontrar o ponto onde a função

$$\tilde{f}(\mu_1, \mu_2) = n_1 (\bar{X} - \mu_1)^2 + n_2 (\bar{Y} - \mu_2)^2$$

assume o valor mínimo, que é aquele que minimiza a distância entre $(\bar{X}, \bar{Y})$ e o espaço $\tilde{\Omega}$.

Suponhamos que μ_1 e μ_2 , não estão sujeitos a restrições, ou seja, que $(\mu_1, \mu_2) \in \mathbb{R}^2$. Neste caso, para testar (2.1) podemos utilizar o **teste usual** (ou **irrestrito**) definido a seguir.

Seja D_α a constante que determina o nível de significância α , ou seja,

$$P_{\mu_1=\mu_2=\mu} \left(\frac{\bar{X} - \bar{Y}}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > \frac{D_\alpha}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right) = \alpha, \quad (2.3)$$

que implica

$$1 - \Phi \left(\frac{D_\alpha}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right) = \alpha$$

e conseqüentemente que

$$D_\alpha = \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \Phi^{-1}(1 - \alpha). \quad (2.4)$$

Denotando por $\bar{D} = \bar{X} - \bar{Y}$ o estimador de máxima verossimilhança de $\mu_1 - \mu_2$, rejeitamos H_0 se

$$\bar{D} > D_\alpha,$$

ou seja, se

$$\frac{\bar{D}}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > \Phi^{-1}(1 - \alpha).$$

O poder do teste usual para $\mu_1 - \mu_2 = \delta > 0$ é dado por

$$P_{\mu_1 - \mu_2 = \delta} \left(\frac{\bar{X} - \bar{Y} - \delta}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > \frac{D_\alpha - \delta}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right) = 1 - \Phi \left(\frac{D_\alpha - \delta}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right). \quad (2.5)$$

Propomos a seguir um teste **alternativo** (ou **restrito**), para situações em que $\mu_1 \geq c$ e $\mu_2 \geq c$. Sem perda da generalidade, suponhamos $c = 0$. A região de rejeição da hipótese nula corresponde a $\hat{D} > D_\alpha^*$, onde $\hat{D}$ denota o estimador de máxima verossimilhança de $\mu_1 - \mu_2$ sujeito a $(\mu_1, \mu_2) \in \mathbb{R}^+ \times \mathbb{R}^+$, i.e.,

$$\hat{D} = \hat{\mu}_1 - \hat{\mu}_2$$

em que $\hat{\mu}_1 = \max\{0, \bar{X}\}$ e $\hat{\mu}_2 = \max\{0, \bar{Y}\}$.

Seja D_α^* a constante positiva tal que

$$P_{\mu_1 = \mu_2 = \mu} (\hat{D} > D_\alpha^*) = \alpha, \quad (2.6)$$

ou seja, devemos rejeitar H_0 se $\hat{D} > D_\alpha^*$. O problema é então calcular D_α^* . Nesse sentido, note que

$$\begin{aligned}
P_{\mu_1 - \mu_2 = \delta} (\hat{D} > D_\alpha^*) &= P_{\mu_1 - \mu_2 = \delta} (\bar{X} > D_\alpha^*, \bar{X} \geq 0, \bar{Y} \leq 0) \\
&+ P_{\mu_1 - \mu_2 = \delta} (\bar{X} - \bar{Y} > D_\alpha^*, \bar{X} \geq 0, \bar{Y} \geq 0) \\
&+ P_{\mu_1 - \mu_2 = \delta} (-\bar{Y} > D_\alpha^*, \bar{X} \leq 0, \bar{Y} \geq 0) \\
&+ P_{\mu_1 - \mu_2 = \delta} (0 > D_\alpha^*, \bar{X} \leq 0, \bar{Y} \leq 0). \quad (2.7)
\end{aligned}$$

O terceiro termo do segundo membro de (2.7) é nulo pois se $\bar{Y} \geq 0$ então não é possível que $-\bar{Y} > D_\alpha^* > 0$. O último termo do segundo membro de (2.7) também é nulo, pois como $D_\alpha^* > 0$ segue que $\{0 > D_\alpha^*, \bar{X} \leq 0, \bar{Y} \leq 0\} = \emptyset$. Analisemos os termos restantes. Primeiramente notemos que

$$\begin{aligned}
P_{\mu_1 - \mu_2 = \delta} (\bar{X} > D_\alpha^*, \bar{X} \geq 0, \bar{Y} \leq 0) &= P_{\mu_1 - \mu_2 = \delta} (\bar{X} > D_\alpha^*, \bar{Y} \leq 0) \\
&= \left[1 - \Phi \left(\frac{D_\alpha^* - \mu_1}{\sigma / \sqrt{n_1}} \right) \right] \Phi \left(\frac{-\mu_2}{\sigma / \sqrt{n_2}} \right). \quad (2.8)
\end{aligned}$$

Agora consideremos

$$\begin{aligned}
P_{\mu_1 - \mu_2 = \delta} (\bar{X} - \bar{Y} > D_\alpha^*, \bar{X} \geq 0, \bar{Y} \geq 0) &= P_{\mu_1 - \mu_2 = \delta} (\bar{X} - \bar{Y} > D_\alpha^*, \bar{Y} \geq 0) \\
&= P_{\mu_1 - \mu_2 = \delta} (\bar{X} - \bar{Y} > D_\alpha^*) \\
&\quad - P_{\mu_1 - \mu_2 = \delta} (\bar{X} - \bar{Y} > D_\alpha^*, \bar{Y} \leq 0). \quad (2.9)
\end{aligned}$$

Para o último termo do segundo membro de (2.9), temos

$$P_{\mu_1 - \mu_2 = \delta} (\bar{X} - \bar{Y} > D_\alpha^*, \bar{Y} \leq 0) = P_{\mu_1 - \mu_2 = \delta} \left(\frac{\bar{X} - \mu_1}{\sigma / \sqrt{n_1}} > \frac{D_\alpha^* - \mu_1 + \bar{Y}}{\sigma / \sqrt{n_1}}, \bar{Y} \leq 0 \right). \quad (2.10)$$

Denotando

$$Z_1 = \frac{\bar{X} - \mu_1}{\sigma / \sqrt{n_1}} \quad \text{e} \quad Z_2 = \frac{\bar{Y} - \mu_2}{\sigma / \sqrt{n_2}},$$

temos que (2.10) é igual a

$$\begin{aligned}
& P_{\mu_1 - \mu_2 = \delta} \left(Z_1 > \frac{D_\alpha^* - (\mu_1 - \mu_2)}{\sigma/\sqrt{n_1}} + \frac{\sqrt{n_1}}{\sqrt{n_2}} Z_2, Z_2 \leq \frac{-\mu_2}{\sigma/\sqrt{n_2}} \right) \\
&= \frac{1}{2\pi} \int_{-\infty}^{\frac{-\mu_2}{\sigma/\sqrt{n_2}}} \int_{\frac{D_\alpha^* - \delta}{\sigma/\sqrt{n_1}} + \frac{\sqrt{n_1}}{\sqrt{n_2}} v}^{\infty} \exp\left(\frac{-w^2}{2}\right) dw \exp\left(\frac{-v^2}{2}\right) dv \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{-\mu_2}{\sigma/\sqrt{n_2}}} \left[1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{D_\alpha^* - \delta}{\sigma/\sqrt{n_1}} + \frac{\sqrt{n_1}}{\sqrt{n_2}} v} \exp\left(\frac{-w^2}{2}\right) dw \right] \exp\left(\frac{-v^2}{2}\right) dv \\
&= \Phi\left(\frac{-\mu_2}{\sigma/\sqrt{n_2}}\right) \\
&\quad - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{-\mu_2}{\sigma/\sqrt{n_2}}} \int_{-\infty}^{\frac{D_\alpha^* - \delta}{\sigma/\sqrt{n_1}} + \frac{\sqrt{n_1}}{\sqrt{n_2}} v} \exp\left(\frac{-w^2}{2}\right) dw \exp\left(\frac{-v^2}{2}\right) dv \\
&= \Phi\left(\frac{-\mu_2}{\sigma/\sqrt{n_2}}\right) \\
&\quad - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{-\mu_2}{\sigma/\sqrt{n_2}}} \Phi\left(\frac{D_\alpha^* - \delta}{\sigma/\sqrt{n_1}} + \frac{\sqrt{n_1}}{\sqrt{n_2}} v\right) \exp\left(\frac{-v^2}{2}\right) dv \\
&= \Phi\left(\frac{-\mu_2}{\sigma/\sqrt{n_2}}\right) - h(\mu_2, \delta)
\end{aligned} \tag{2.11}$$

em que

$$h(\mu_2, \delta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{-\mu_2}{\sigma/\sqrt{n_2}}} \Phi\left(\frac{D_\alpha^* - \delta}{\sigma/\sqrt{n_1}} + \frac{\sqrt{n_1}}{\sqrt{n_2}} v\right) \exp\left(\frac{-v^2}{2}\right) dv.$$

De (2.7) - (2.11) obtemos

$$\begin{aligned}
P_{\mu_1 - \mu_2 = \delta} \left(\widehat{D} > D_\alpha^* \right) &= 1 - \Phi\left(\frac{D_\alpha^* - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right) \\
&\quad - \Phi\left(\frac{D_\alpha^* - \mu_1}{\sigma/\sqrt{n_1}}\right) \Phi\left(\frac{-\mu_2}{\sigma/\sqrt{n_2}}\right) + h(\mu_2, \delta) \\
&= 1 - \Phi\left(\frac{D_\alpha^* - \delta}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right) \\
&\quad - \Phi\left(\frac{D_\alpha^* - \delta - \mu_1}{\sigma/\sqrt{n_1}}\right) \Phi\left(\frac{-\mu_2}{\sigma/\sqrt{n_2}}\right) \\
&\quad + h(\mu_2, \delta)
\end{aligned} \tag{2.12}$$

Notemos que sob a hipótese nula, $\mu_1 = \mu_2 = \mu$, temos $\delta = 0$, e então, fixado o nível de significância α , D_α^* deverá ser calculado considerando (2.12) com $\delta = 0$.

Além disso, como $-\Phi\left(\frac{D_\alpha^* - \delta - \mu_1}{\sigma/\sqrt{n_1}}\right) \Phi\left(\frac{-\mu_2}{\sigma/\sqrt{n_2}}\right) + h(\mu_2, \delta)$ é uma função negativa, temos que

$$D_\alpha^* \leq D_\alpha \quad (2.13)$$

indicando que o teste alternativo é menos favorável à hipótese nula que o teste usual quando $\hat{D} = \overline{D}$.

Notemos entretanto, que este cálculo envolve o parâmetro desconhecido μ . Em situações similares, Cox e Hinkley (1974) sugerem testes condicionais, porém tais testes requerem o uso de uma estatística suficiente e completa que não é disponível em nosso caso. Outra alternativa sugerida por esses autores consiste em substituir μ por algum estimador. Nesse caso deve-se procurar um estimador que minimize os desvios do valor nominal do nível de significância, isto é, o estimador, que chamaremos de $\tilde{\mu}$, deve manter controlada a probabilidade de erro de tipo I ou seja,

$$P_{\tilde{\mu}=\mu_1=\mu_2}(\hat{D} > D_\alpha^*) \simeq \alpha.$$

Inicialmente avaliaremos o poder do teste admitindo μ conhecido. Na próxima seção analisaremos o desempenho do teste quando μ é substituído por diferentes estimadores. Lembremos que sob a hipótese alternativa, i.e. para $\delta \neq 0$, (2.12) representará o poder do teste restrito.

Dado que uma comparação analítica dos poderes de ambos testes é intratável devido ao número de parâmetros envolvidos, consideramos uma análise gráfica para tal fim. Tomando $\alpha = 0.05$, construímos os gráficos do poder de ambos os testes como funções de μ_2 , o menor valor entre as médias que estão sendo comparadas, de diferentes valores do tamanho amostral n , da média comum μ sob a hipótese nula e da diferença δ entre as duas médias sob a hipótese alternativa. Calculamos os valores críticos de D_α^* por (2.12), com $\delta = 0$, utilizando o método de Brent e a regra de Simpson (ver, por exemplo, Press

et al., 1994). Os resultados são apresentados nas Figuras 2.1 e 2.2. e indicam que o teste proposto é mais poderoso que o teste usual para certos valores de μ . Para um nível de significância $\alpha = 0.05$, fixados δ e $n = n_1 = n_2$, sem perda de generalidade é possível observar que:

1. Fixado o valor de μ , o poder do teste alternativo se estabiliza quando μ_2 cresce.
2. Fixado o valor de μ_2 , o teste alternativo é mais poderoso que o teste usual para valores de μ menores que um determinado valor, que denotamos por $\mu(\mu_2, \delta, n)$.
3. Denotamos por $\mu_2(\delta, n)$ é o valor a partir do qual todas curvas do poder são praticamente constantes. Se utilizamos valores de $\mu < \mu(\mu_2(\delta, n), \delta, n)$ no cálculo de D_α^* em (2.6) obteremos um teste mais poderoso que o teste usual para detectar a diferença $\mu_1 - \mu_2 = \delta$, para $\mu_2 \geq \mu_2(\delta, n)$.

Fixados os valores de δ , n e μ_2 para determinar $\mu(\mu_2, \delta, n)$, calculamos o poder do teste usual através de (2.5) e o denotamos por $p_u(\alpha, \delta)$. Obtemos $D_{p_u(\alpha, \delta)}^*$ por (2.12), considerando o valor escolhido fixo de μ_2 , ou seja,

$$P_{\mu_1 - \mu_2 = \delta} \left(\widehat{D} > D_{p_u(\alpha, \delta)}^* \right) = p_u(\alpha, \delta)$$

e posteriormente obtemos $\mu(\mu_2, \delta, n)$ como a solução da equação

$$P_{\mu_1 - \mu_2 = 0} \left(\widehat{D} > D_{p_u(\alpha, \delta)}^* \right) = \alpha.$$

Nas Tabelas (2.1) e (2.2) apresentamos os valores de $\mu(\mu_2, \delta, n)$ para diferentes tamanhos de amostras, diferentes valores de δ e $\sigma = 1$. Por exemplo, sejam $n = 5$ e $\delta = 1$. Então, para $\mu_2 = 0.2$ temos $\mu(0.2, 1, 5) = 0.60$; isto significa que para valores de μ menores que 0.60, o poder do teste restrito será maior que o do teste usual; para $\mu = 0.60$ o poder será igual ao do teste usual e para $\mu > 0.60$ esse poder será menor que o do teste usual. Continuando, com $n = 5$ e $\delta = 1$, o último valor que aparece na coluna é $\mu(1.3, 1, 5) = 1.85$;

portanto $\mu_2(1, 5) = 1.3$, ou seja, para valores de $\mu_2 \geq 1.3$, o poder do teste alternativo é igual ao poder do teste usual para valores de μ maiores que $\mu(1.3, 1, 5) = 1.85$, e esse teste é mais poderoso para valores de μ menores que 1.85; essas situações estão representadas com asteriscos nas Tabelas (2.1) e (2.2). Finalmente, tomemos $n = 10$ e $\delta = 0.5$. Para $\mu_2 = 1.0$ temos $\mu(1.0, 0.5, 10) = 1.25$ e $\mu_2(0.5, 10) = 1.0$ ou seja, para valores de μ_2 maiores que 1.0, o teste alternativo tem o mesmo poder ou é mais poderoso que o teste usual, independentemente de qual seja o valor de μ considerado na hipótese nula.

Em resumo

- fixados os valores de δ e n , quando μ_2 cresce $\mu(\mu_2, \delta, n)$ aumenta; logo a região onde o teste alternativo é mais poderoso que o teste usual aumenta de tamanho.
- fixados os valores de δ e μ_2 , quando n aumenta, $\mu(\mu_2, \delta, n)$ aumenta.
- fixado o valor de δ , quando n cresce, $\mu_2(\delta, n)$ diminui; logo a região onde o teste alternativo tem poder igual ou maior que o teste usual aumenta de tamanho.

Como conclusão geral podemos dizer que o poder do teste alternativo aumenta quando a média comum utilizada para definir D_α^* está mais próxima da origem e quando o valor da menor média na hipótese alternativa está mais longe da origem. Na próxima seção vamos estudar o que acontece com o poder do teste alternativo na prática, quando μ é substituído por um estimador.

Figura 2.1: Poder do teste alternativo e do teste usual para $\sigma = 1$ e $n = 5$

Figura 2.2: Poder do teste alternativo e do teste usual para $\sigma = 1$ e $n = 10$

- poder do teste usual Z
- * — poder do teste alternativo para $\mu = 0.0$
- ■ — poder do teste alternativo para $\mu = 0.5$
- + — poder do teste alternativo para $\mu = 1.0$
- × — poder do teste alternativo para $\mu = 1.5$
- o — poder do teste alternativo para $\mu = 2.0$

Figura 2.3: Poder do teste alternativo e do teste usual para $\sigma = 1$ e $n = 15$

- poder do teste usual Z
- * — poder do teste alternativo para $\mu = 0.0$
- ■ — poder do teste alternativo para $\mu = 0.5$
- + — poder do teste alternativo para $\mu = 1.0$
- × — poder do teste alternativo para $\mu = 1.5$
- o — poder do teste alternativo para $\mu = 2.0$

Figura 2.4: Poder do teste alternativo e do teste usual para $\sigma = 2$ e $n = 5$

Figura 2.5: Poder do teste alternativo e do teste usual para $\sigma = 2$ e $n = 10$

- poder do teste usual Z
- * — poder do teste alternativo para $\mu = 0.0$
- ■ — poder do teste alternativo para $\mu = 0.5$
- + — poder do teste alternativo para $\mu = 1.0$
- × — poder do teste alternativo para $\mu = 1.5$
- o — poder do teste alternativo para $\mu = 2.0$

Figura 2.6: Poder do teste alternativo e do teste usual para $\sigma = 2$ e $n = 15$

2.2 Desempenho do teste alternativo baseado em estimadores da média comum

Para estudar o efeito da substituição do valor desconhecido μ na expressão (2.12) utilizada para definir a região crítica, sobre o nível de significância e o poder do teste restrito, consideramos diferentes estimadores. Para tal fim, realizamos um estudo de simulação com 1.000 pares de amostras de tamanho 5 das distribuições $N(\mu_1, \sigma^2)$ e $N(\mu_2, \sigma^2)$.

Para saber se o estimador considerado em cada caso mantém o erro de tipo I próximo do nível de significância, as amostras foram geradas considerando $\mu_1 = \mu_2 = \mu$, para valores de $\mu = 0, 0.1, \dots, 0.5, 1, 1.5$. Em cada amostra foram calculados $\hat{D}$ e D_α^* através de (2.12) utilizando algum estimador de μ . O valor da taxa de rejeição (valor de α empírico), que é a estimativa da probabilidade de erro de tipo I, é calculado como o número de amostras em que $\hat{D} > D_\alpha^*$ dividido pelo número total de amostras.

Estimadores para a média restrita a um intervalo apresentados por Gupta e Rohatgi (1980) estão descritos no Apêndice A, conjuntamente com as expressões de seus vícios e erros quadráticos médios. Com base nesses estimadores, sob a hipótese nula, a média μ pode ser estimada alternativamente por

$$\hat{\mu} = f((\bar{X} + \bar{Y})/2), \quad (2.14)$$

$$\hat{\mu}'_1 = |(\bar{X} + \bar{Y})/2|, \quad (2.15)$$

$$\hat{\mu}'_2 = \frac{1}{2} \left[\sum_{i=1}^n |X_i| + \sum_{i=1}^n |Y_i| \right], \quad (2.16)$$

$$\hat{\mu}_3 = \frac{1}{2} \left[\sum_{i=1}^n f(X_i) + \sum_{i=1}^n f(Y_i) \right]. \quad (2.17)$$

em que $f(x) = \max(0, x)$. Levando em conta o fato de esses estimadores serem viciados, definimos adicionalmente os estimadores $\hat{\mu}_4 = \hat{\mu} - \text{vício}(\hat{\mu})$ e $\hat{\mu}_5 = \hat{\mu}'_1 - \text{vício}(\hat{\mu}'_1)$.

Tabela 2.1: Valores da média comum abaixo dos quais o teste alternativo é mais poderoso que o teste usual

μ_2	delta=0.25			delta=0.5		
	n=5	n=10	n=15	n=5	n=10	n=15
0.0	0.09	0.09	0.09	0.18	0.18	0.19
0.1	0.19	0.19	0.19	0.29	0.29	0.30
0.2	0.29	0.30	0.30	0.39	0.40	0.41
0.3	0.40	0.40	0.40	0.50	0.51	0.52
0.4	0.50	0.51	0.51	0.60	0.62	0.63
0.5	0.60	0.61	0.61	0.71	0.72	0.73
0.6	0.71	0.71	0.72	0.82	0.83	0.83
0.7	0.81	0.81	0.82	0.92	0.93	0.93
0.8	0.91	0.92	0.92	1.02	1.04	0.99
0.9	1.01	1.02	1.03	1.13	1.14	1.01
1.0	1.11	1.13	*	1.23	1.25	*
1.1	1.22	*	*	1.33	*	*
1.2	1.32	*	*	1.44	*	*
1.3	1.43	*	*	1.54	*	*
1.4	1.54	*	*	1.66	*	*
1.5	1.73	*	*	1.92	*	*

μ_2	delta=0.75			delta=1		
	n=5	n=10	n=15	n=5	n=10	n=15
0.0	0.28	0.28	0.29	0.38	0.39	0.41
0.1	0.39	0.40	0.41	0.49	0.51	0.53
0.2	0.49	0.51	0.52	0.60	0.63	0.65
0.3	0.6	0.62	0.64	0.72	0.74	0.76
0.4	0.71	0.73	0.75	0.83	0.86	0.87
0.5	0.82	0.84	0.85	0.94	0.96	0.97
0.6	0.93	0.95	0.97	1.05	1.07	1.05
0.7	1.04	1.05	*	1.15	1.18	1.07
0.8	1.14	1.13	*	1.26	1.29	1.08
0.9	1.25	1.17	*	1.37	1.07	*
1.0	1.35	1.18	*	1.48	*	*
1.1	1.45	1.18	*	1.60	*	*
1.2	1.55	*	*	1.83	*	*
1.3	1.65	*	*	1.85	*	*
1.4	1.74	*	*	*	*	*
1.5	1.81	*	*	*	*	*
1.6	1.85	*	*	*	*	*
1.7	1.86	*	*	*	*	*

Tabela 2.2: Valores da média comum abaixo dos quais o teste alternativo é mais poderoso que o teste usual

μ_2	delta=1.25			delta=1.5		
	n=5	n=10	n=15	n=5	n=10	n=15
0.0	0.48	0.51	0.53	0.6	0.64	0.67
0.1	0.6	0.64	0.66	0.72	0.76	0.79
0.2	0.72	0.75	0.77	0.84	0.88	0.92
0.3	0.83	0.87	0.89	0.95	0.99	*
0.4	0.95	0.98	1.00	1.07	1.08	*
0.5	1.06	1.11	*	1.18	1.13	*
0.6	1.17	*	*	1.29	1.15	*
0.7	1.28	*	*	1.40	*	*
0.8	1.38	*	*	1.50	*	*
0.9	1.49	*	*	1.61	*	*
1.0	1.58	*	*	1.70	*	*
1.1	1.67	*	*	1.79	*	*
1.2	1.73	*	*	1.83	*	*
1.3	1.76	*	*	1.85	*	*
1.4	1.77	*	*	*	*	*

μ_2	delta=1.75			delta=2.0		
	n=5	n=10	n=15	n=5	n=10	n=15
0.0	0.72	0.77	0.80	0.84	0.90	0.79
0.1	0.84	0.89	0.99	0.97	1.02	*
0.2	0.96	1.01	*	1.09	1.10	*
0.3	1.08	1.12	*	1.20	1.15	*
0.4	1.19	1.20	*	1.32	*	*
0.5	1.3	1.25	*	1.44	*	*
0.6	1.41	*	*	1.51	*	*
0.7	1.52	*	*	1.57	*	*
0.8	1.63	*	*	1.61	*	*
0.9	1.73	*	*	1.62	*	*
1.0	1.83	*	*	*	*	*
1.1	1.96	*	*	*	*	*
1.2	2.06	*	*	*	*	*

Gupta e Rohatgi (1980) comparam o desempenho dos estimadores baseando-se nos respectivos erros quadráticos médios; os resultados são apresentados na Tabela A.2. Essa escolha é baseada nos valores de $m = \mu/\sigma$. Se utilizássemos $\hat{\mu}$, que é um bom estimador de μ para valores de μ/σ próximos de zero de acordo com os resultados de Gupta e Rohatgi (1980) (ver Tabela A.2), teríamos $\max\{0, (\bar{X} + \bar{Y})/2\sigma\}$ como estimador de m . Se $(\bar{X} + \bar{Y})/2\sigma > 0$, então $\hat{\mu} = \hat{\mu}'_1 = (\bar{X} + \bar{Y})/2$, que é um bom estimador de μ para $0.25 \leq \mu/\sigma \leq 0.5$ (ver Tabela A.2). Assim para efeitos práticos podemos adaptar a comparação considerando $(\bar{X} + \bar{Y})/2\sigma$ como estimativa de m em cada par de amostras, já que estamos analisando valores de μ próximos de zero.

Levando em conta que as funções vício destes estimadores são positivas e monótonas decrescentes podemos então definir os seguintes estimadores adicionais

$$\hat{\mu}_{GR1} = \begin{cases} \hat{\mu} - \text{vício}(\hat{\mu}) & \text{se } (\bar{X} + \bar{Y})/(2\sigma) < 0.25 \text{ ou } (\bar{X} + \bar{Y})/(2\sigma) \geq 2, \\ \hat{\mu}'_1 - \text{vício}(\hat{\mu}'_1) & \text{se } 0.25 \leq (\bar{X} + \bar{Y})/(2\sigma) \leq 0.5, \\ \hat{\mu}_3 - \text{vício}(\hat{\mu}_3) & \text{se } 0.5 < (\bar{X} + \bar{Y})/(2\sigma) \leq 2, \end{cases}$$

$$\hat{\mu}_{GR2} = \begin{cases} \hat{\mu} - \text{vício}(\hat{\mu}) & \text{se } (\bar{X} + \bar{Y})/(2\sigma) < 0.25 \text{ ou } (\bar{X} + \bar{Y})/(2\sigma) \geq 2, \\ \hat{\mu}'_1 & \text{se } 0.25 \leq (\bar{X} + \bar{Y})/(2\sigma) \leq 0.5, \\ \hat{\mu}_3 - \text{vício}(\hat{\mu}_3) & \text{se } 0.5 < (\bar{X} + \bar{Y})/(2\sigma) \leq 2. \end{cases}$$

e finalmente

$$\hat{\mu}_V = \begin{cases} \hat{\mu} - \text{vício}(\hat{\mu}) & \text{se } [(\bar{X} + \bar{Y})/2]/\sigma < 0.25, \\ \hat{\mu} & \text{se } 0.25 \leq [(\bar{X} + \bar{Y})/2]/\sigma. \end{cases}$$

Na Tabela 2.3 são apresentados os valores de α empírico para o teste alternativo considerando $\hat{\mu}$, $\hat{\mu}_4$ e $\hat{\mu}_5$, $\hat{\mu}_{GR1}$, $\hat{\mu}_{GR2}$ e $\hat{\mu}_V$ como estimadores de μ além dos valores de α empírico para o teste usual. Consideramos $\sigma = 1$ e as mesmas amostras para cada caso.

Tabela 2.3: Valores do nível de significância empírico

μ	teste alternativo					teste usual	
	$\hat{\mu}$	$\hat{\mu}_4$	$\hat{\mu}_5$	$\hat{\mu}_{GR1}$	$\hat{\mu}_{GR2}$	$\hat{\mu}_V$	
0.0	0.026	0.035	0.049	0.038	0.017	0.032	0.048
0.1	0.026	0.032	0.050	0.037	0.025	0.030	0.046
0.2	0.034	0.044	0.059	0.046	0.029	0.042	0.049
0.3	0.037	0.047	0.061	0.050	0.037	0.043	0.046
0.4	0.055	0.062	0.071	0.062	0.048	0.060	0.059
0.5	0.057	0.068	0.076	0.063	0.050	0.057	0.055
1.0	0.056	0.056	0.056	0.053	0.053	0.056	0.053
1.5	0.052	0.052	0.052	0.052	0.052	0.052	0.052
2.5	0.054	0.054	0.054	0.054	0.054	0.054	0.054
5.0	0.049	0.049	0.049	0.049	0.049	0.049	0.049

Observamos que ao considerar o estimador $\hat{\mu}$, a taxa de rejeição é muito menor que a taxa nominal 0.05, para valores de μ próximos de zero. Acreditamos que isto seja devido ao fato de que o vício é uma função monótona decrescente de μ/σ , assumindo valores maiores para valores de μ/σ próximos de zero (ver Tabela A.1 do Apêndice).

Ao considerar como estimadores a $\hat{\mu}_4$, $\hat{\mu}_5$ e $\hat{\mu}_{GR1}$ obtiveram-se taxas de rejeição maiores que 0.05 para valores do parâmetro μ iguais ou maiores que 0.3 e menores que 1.

Com os estimadores $\hat{\mu}_{GR2}$ e $\hat{\mu}_V$ obtivemos taxas de rejeição mais próximas a 0.05, mas $\hat{\mu}_V$ tem a vantagem de ser computacionalmente mais simples. Utilizando o estimador $\hat{\mu}_V$ obtivemos taxas de rejeição menores que 0.05 para valores próximos de zero, ou seja, se o verdadeiro valor de μ estiver perto da fronteira do espaço paramétrico, a probabilidade de rejeitar a hipótese nula quando ela é verdadeira é menor que o nível de significância especificado. Isto implica que o teste baseado nesse estimador tende a ser mais favorável à hipótese nula. Em consequência, pode haver um aumento da probabilidade do erro de tipo II ou uma perda do poder do teste alternativo.

Pelo desempenho obtido para controlar a probabilidade de erro de tipo I, decidimos estudar a influência do uso de $\hat{\mu}_V$ sobre o poder do teste alternativo. O estimador $\hat{\mu}$ foi também considerado para tal fim por ser o estimador de máxima verossimilhança. Com este objetivo realizamos uma nova série de simulações considerando os mesmos pares de

amostras para analisar o desempenho dos dois estimadores. Para cada valor de δ fixo, as amostras foram geradas considerando $\mu_1 = \mu_2 + \delta$, $\sigma = 1$ e $n = 5$, para valores de $\mu_2 = 0, 0.1, \dots, 0.5, 1.0, 1.5$ e $\delta = 1.00, 1.25$ e 1.50 . Em cada amostra obtiveram-se $\hat{D}$ e D_α^* de (2.12) com $\delta = 0$, utilizando ou o estimador $\hat{\mu}$ ou $\hat{\mu}_V$ no lugar de μ . O poder empírico é calculado como o número de amostras em que $\hat{D} > D_\alpha^*$ dividido pelo número total de amostras. Os resultados estão dispostos na Tabela 2.4.

Em ambas situações analisadas o poder empírico do teste alternativo foi maior que o poder do teste usual.¹

A diferença entre o poder empírico do teste alternativo e o poder empírico do teste usual é maior quando a média menor, no caso μ_2 , está mais próxima da origem. Como já foi mencionado, as taxas empíricas obtidas ao utilizar os estimadores $\hat{\mu}$ e $\hat{\mu}_V$ foram menores que a taxa nominal, isto é, a probabilidade de erro de tipo I é menor que a nominal, portanto é esperado um aumento da probabilidade de erro de tipo II ou uma perda do poder do teste alternativo. Logo, ao considerar um estimador de μ , para o qual a taxa empírica fosse mais próxima da taxa nominal, para valores de μ próximos da origem, a diferença em termos do poder do teste alternativo e o teste usual seria maior. Isto é observado na Tabela 2.4 já que a diferença entre os valores do poder empírico do teste alternativo com respeito ao teste usual é maior ao utilizar o estimador $\hat{\mu}_V$, em todos os casos.

O teste alternativo parece ser preferível ao teste usual por ser tão ou mais poderoso que o teste usual na prática, sem alteração da probabilidade de erro de tipo I. Apesar dos resultados resumidos nas Tabelas 2.1 e 2.2 indicarem que existem regiões onde o teste usual é mais poderoso que o teste alternativo, sua implementação com $\hat{\mu}$ ou $\hat{\mu}_V$ substituído μ sugere (ver Tabela 2.4) que independentemente da distância da menor das médias à origem, seu poder é maior que o poder do teste usual. Vale notar que esse ganho no poder não pode ser creditado a um aumento da probabilidade de erro de tipo I,

¹Valores nominais do poder do teste usual para $\delta = 1$ é 0.4746, para $\delta = 1.25$ é 0.6299, para $\delta = 1.50$ é 0.7663.

Tabela 2.4: Valores do poder empírico

$\delta = 1.00$			
	teste alternativo	teste alternativo	teste usual
μ_2	$\hat{\mu}_V$	$\hat{\mu}$	
0.0	0.514	0.485	0.485
0.1	0.510	0.492	0.482
0.2	0.481	0.475	0.461
0.3	0.519	0.518	0.506
0.4	0.471	0.466	0.452
0.5	0.488	0.488	0.470
1.0	0.461	0.461	0.460
1.5	0.471	0.471	0.471
2.5	0.454	0.454	0.454
5.0	0.459	0.459	0.459
$\delta = 1.25$			
	teste alternativo	teste alternativo	teste usual
μ_2	$\hat{\mu}_V$	$\hat{\mu}$	
0.0	0.656	0.646	0.627
0.1	0.671	0.666	0.655
0.2	0.650	0.647	0.628
0.3	0.635	0.635	0.626
0.4	0.613	0.613	0.598
0.5	0.647	0.647	0.635
1.0	0.628	0.628	0.628
1.5	0.625	0.625	0.625
2.5	0.637	0.637	0.637
5.0	0.638	0.638	0.638
$\delta = 1.50$			
	teste alternativo	teste alternativo	teste usual
μ_2	$\hat{\mu}_V$	$\hat{\mu}$	
0.0	0.786	0.783	0.765
0.1	0.781	0.779	0.768
0.2	0.775	0.775	0.766
0.3	0.760	0.759	0.751
0.4	0.765	0.765	0.760
0.5	0.771	0.771	0.770
1.0	0.791	0.791	0.790
1.5	0.769	0.769	0.769
2.5	0.781	0.781	0.781
5.0	0.755	0.755	0.755

Tabela 2.5: Estimativas de mínimos quadrados

Parâmetro	Média	Desvio amostral
μ_1	$\bar{X} = 94.2$	$s_X^2 = 5.71$
μ_2	$\bar{Y} = 90.3$	$s_Y^2 = 5.51$

já que, segundo o que observamos na Tabela 2.3, a utilização desses estimadores mantém controlada essa probabilidade.

As estimativas de mínimos quadrados do Exemplo 1.1, são apresentados na Tabela (2.5).

Considerando $\sigma = s_p = 5.6$, definida em (2.48), para $\alpha = 0.05$ e utilizando $\hat{\mu}_V = \hat{\mu} = (4.16 + 0.26) / 2 = 2.2$, em (2.56) obtemos valor crítico do teste alternativo $D_\alpha^* = 3.6$. Para o teste usual, obtemos valor crítico $D_\alpha = 3.9$. Dado que $\hat{D} = \bar{D} = \bar{X} - \bar{Y} = 3.9$, rejeitamos H_0 unicamente com o teste alternativo

2.3 Enfoque bayesiano

O teste alternativo apresentado requer a estimação de um parâmetro desconhecido na sua implementação. Isto traz como consequência uma perda do poder, diminuindo sua superioridade com respeito ao teste usual, como pode ser observado ao se obterem taxas empíricas menores que o nível de significância nominal na Seção 2.2. Por esta razão, o problema de comparação de médias restritas também foi avaliado sob o ponto de vista Bayesiano, que permite a incorporação das restrições do espaço paramétrico através do uso de distribuições a priori convenientes.

Como nosso objetivo é testar uma hipótese, ou equivalentemente comparar dois modelos o uso do fator de Bayes é apropriado, segundo a análise sugerida por Smith (1991). Se os modelos assumidos são considerados aproximações úteis para um modelo teórico e praticamente indeterminado, e também se tem como objetivo algum tipo de predição, as comparações entre modelos baseadas em medidas de validação cruzada podem ser mais apropriadas (Stone, 1974, 1977; Geisser, 1975).

Entre outras técnicas para comparação de modelos relacionadas podemos citar o pseudo fator de Bayes (Geisser e Eddy, 1979; Gelfand et al, 1992; Gelfand e Day, 1994), o fator de Bayes a posteriori (Aitkin, 1991), ou critérios que minimizem a função de perda a posteriori (Gelfand e Ghosh, 1998).

Também podemos mencionar a teoria de fator de Bayes intrínseco desenvolvida por Berger e Pericchi (1996), os chamados fatores de Bayes fracionais, discutidos em O'Hagan (1995) ou De Santis and Spezzaferri (1997). Estas técnicas utilizam amostras de calibração e podem ser instáveis quando o tamanho correspondente é pequeno como em nosso caso.

Todos estes trabalhos têm por objetivo diminuir a sensibilidade do fator de Bayes à variação das distribuições a priori, e permitem o uso de distribuições a priori impróprias com as quais o fator de Bayes sofre de falta de interpretação (ver O'Hagan, 1995). Neste trabalho só consideraremos distribuições a priori próprias.

Reanalisamos o problema utilizando o fator de Bayes, calculado por duas metodologias

diferentes. Na primeira delas, empregou-se a técnica proposta por Carlin e Chib (1995) que é baseada em amostradores de Gibbs. Tanto a implementação do algoritmo de Gibbs quanto a técnica de Carlin e Chib só podem ser utilizadas se as distribuições a priori pertencem à família de distribuições conjugadas, se existe conjugação condicional completa ou se as densidades condicionais completas a posteriori são densidades factíveis de serem amostradas. No segundo enfoque, o fator de Bayes é obtido através do cálculo direto das distribuições a posteriori (vide Bragança Pereira, 1995, e Irony e Pereira, 1995) e tem a vantagem de poder utilizar qualquer tipo de distribuições a priori, desde que sejam próprias.

Um teste de hipótese do tipo (2.1) sob o modelo linear (1.1), é equivalente à comparação dos modelos

$$M = 1 : y_{ik} = \mu + \varepsilon_{ik}, \varepsilon_{ik} \sim N(0, \sigma^2), \quad (2.18)$$

$$M = 2 : y_{ik} = \mu_i + \varepsilon_{ik}, \text{ com } \mu_1 > \mu_2, \varepsilon_{ik} \sim N(0, \sigma^2), \quad (2.19)$$

com $i = 1, 2$, $k = 1, \dots, n_i$, e $\sigma^2 > 0$. O modelo $M = 1$ é equivalente à hipótese nula e o modelo $M = 2$ é equivalente à hipótese alternativa.

Para decidir a favor ou contra cada modelo podemos utilizar o fator de Bayes, isto é a razão entre as densidades marginais observadas para os dois modelos,

$$B_{21} = \frac{p(\mathbf{y}|M=2)}{p(\mathbf{y}|M=1)} = \frac{p(M=2|\mathbf{y})\pi_1}{p(M=1|\mathbf{y})\pi_2}. \quad (2.20)$$

em que $\pi_j = p(M=j)$ $j = 1, 2$, denotam as probabilidades a priori de opção pelos modelos $M = 1$ e $M = 2$, respectivamente.

Com o objetivo de calcular o fator de Bayes, necessitamos obter as probabilidades a posteriori de cada modelo, $p(M=j|\mathbf{y})$, $j = 1, 2$. Estas são as probabilidades calculadas utilizando as técnicas de Carlin e Chib (1995) ou de Irony e Pereira (1995) como mencionamos acima.

2.3.1 Cálculo do fator de Bayes segundo a metodologia de Carlin e Chib

Nesta seção apresentamos a técnica Monte Carlo para Cadeias de Markov (MCMC) proposta por Carlin e Chib (1995) descrita no Apêndice B para o cálculo do fator de Bayes, usando os dados do Exemplo 1.1. Seja o vetor aleatório $\mathbf{v} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, M)$, cujas componentes são os parâmetros de cada modelo, $\boldsymbol{\theta}_1 = \mu$, $\boldsymbol{\theta}_2 = (\mu_1, \mu_2)$ e M , o parâmetro indexador do modelo. A geração direta de amostras da distribuição marginal $p(M = j|\mathbf{y})$ ou da distribuição conjunta $p(\mathbf{v}|\mathbf{y})$ (ver Apêndice B (B.1)) é complicada, mas as gerações de amostras das densidades condicionais completas a posteriori $p(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, M, \mathbf{y})$, $p(\boldsymbol{\theta}_2|\mu, M, \mathbf{y})$ e $p(M|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \mathbf{y})$ podem ser realizadas sem maiores problemas. Estas densidades condicionais completas a posteriori (descritas no Apêndice B) formam o núcleo de transição da cadeia de Markov que será amostrada através do seguinte algoritmo de Gibbs:

- i) inicializamos o contador das iterações da cadeia ($j = 1$) e escolhemos o valor inicial $\mathbf{v}^{(0)}$.
- ii) obtemos um novo valor $\mathbf{v}^{(j)} = (\boldsymbol{\theta}_1^{(j)}, \boldsymbol{\theta}_2^{(j)}, M^{(j)})$ a partir de $\mathbf{v}^{(j-1)}$ através de sucessivas gerações de valores

$$\begin{aligned}\boldsymbol{\theta}_1^{(j)} &\sim p\left(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2^{(j-1)}, M^{(j-1)}, \mathbf{y}\right), \\ \boldsymbol{\theta}_2^{(j)} &\sim p\left(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1^{(j)}, M^{(j-1)}, \mathbf{y}\right), \\ M^{(j)} &\sim p\left(M|\boldsymbol{\theta}_1^{(j)}, \boldsymbol{\theta}_2^{(j)}, \mathbf{y}\right).\end{aligned}$$

- iii) mudamos o contador j para $j + 1$.
- iv) verificamos se o critério de convergência está satisfeito; em caso afirmativo interrompemos o processo e em caso negativo retornamos para ii).

Tabela 2.6: Estimativas de mínimos quadrados

Modelo	Parâmetro	Média	Desvio Padrão
1	μ	92.2	1.2
2	μ_1	94.2	1.7
	μ_2	90.3	1.7

Após o equilíbrio, o estimador de $p(M = j|\mathbf{y})$ é obtido como o cociente entre o número de iterações em que $M = j$ e o número total de iterações.

Lembramos que, para realizar a amostragem além das distribuições a priori para os parâmetros é preciso especificar as chamadas distribuições “pseudo-priori” $p(\theta_j|M \neq j)$, $j = 1, 2$, que deverão ser próprias.

Como distribuições a priori para as médias μ , μ_i , $i = 1, 2$, consideramos distribuições normais que incorporam a restrição $\mu \geq 90$ para $M = 1$, e $\mu_i \geq 90$, $i = 1, 2$, $\mu_1 > \mu_2$ para $M = 2$, ou seja, valores gerados menores que 90 são excluídos da amostragem. Como hiperparâmetros das distribuições a priori (média e desvio padrão) tomamos inicialmente as estimativas de mínimos quadrados sob os modelos correspondentes, apresentadas na Tabela 2.6.

Posteriormente, para analisar a sensibilidade do fator de Bayes, consideraremos distribuições a priori “menos informativas”.

Apesar de a função de verossimilhança e das distribuições a priori escolhidas serem usuais nas análises Bayesianas, a obtenção das distribuições a posteriori requeridas para o fator de Bayes não é simples. A dificuldade surge em virtude da restrição sobre o espaço paramétrico. Por tal motivo, para realizar a amostragem de Gibbs utilizamos procedimentos numéricos que foram implementados no conjunto de subrotinas BUGS (*Bayesian inference Using Gibbs Sampling*, versão 0.6), desenvolvido por Thomas et al. (1992).

Para obter uma aproximação das “pseudo-prioris”, avaliamos cada modelo separadamente. Considerando, $\pi_1 = 1$ e $\pi_2 = 0$, em (2.20) usamos a distribuição a posteriori resultante como a distribuição “pseudo-priori” para o parâmetro μ no modelo $M = 2$.

Tabela 2.7: Diagnóstico de convergência de Geweke

Modelo	Parâmetro	Estatística Z
1	μ	0.601
2	μ_1	-0.155
	μ_2	0.299

Similarmente, considerando $\pi_1 = 0$ e $\pi_2 = 1$, utilizamos as distribuições a posteriori resultantes como as distribuições “pseudo-prioris ” para os parâmetros μ_1, μ_2 no modelo $M = 1$.

Para determinar a convergência foi utilizado o método de diagnóstico de Geweke (1992) disponível no sistema CODA (*Convergence Diagnosis and Output Analysis Software for Gibbs sampling output*). CODA é um conjunto de funções escritas em S-Plus, desenvolvido inicialmente por Cowles (1994), que se utiliza conjuntamente com as sub-rotinas BUGS. Esse método está descrito no Apêndice B. Os valores da estatística Z ($Z - score$), que é uma diferença padronizada das médias das n_a primeiras e das n_b últimas iterações, (ver B.7), para uma cadeia de tamanho amostral 500, são apresentados na Tabela 2.7; como estes valores pertencem ao intervalo $[-2, 2]$ podemos considerar que existe convergência na cadeia.

Para calcular o fator de Bayes consideramos 500 iterações iniciais, já que verificamos a convergência com este tamanho amostral (ver Tabela 2.7), e as atualizamos com 10000 iterações. A Tabela 2.8 mostra os fatores de Bayes para comparar os modelos 1 e 2 definidos em (2.18) e (2.19), correspondentes às diferentes probabilidades a priori de opção por cada modelo, $\pi_j, j = 1, 2$. Em todos os casos, a convergência das 10000 iterações finais foi verificada através do critério de Geweke.

Uma regra de decisão para determinar com que grau de evidência rejeitamos a hipótese nula é a dada por Kass e Raftery (1995) (ver Tabela B.1 no Apêndice B); essa regra, é baseada em Jeffreys (1961), que afirma: “na prática existe uma pequena diferença se a razão contra a hipótese nula (o fator de Bayes) é de 1 para 10 ou de 1 para 100, mas não existe diferença se esta razão é de 1 para 10^4 ou de 1 para 10^{10} ”. Em todos os casos

Tabela 2.8: Fator de Bayes com distribuição a priori informativa para a comparação de médias

Probabilidades a priori		Fator de Bayes
π_1	π_2	
0.05	0.95	1.49
0.25	0.75	1.64
0.50	0.50	1.64
0.75	0.25	1.69
0.95	0.05	1.70

Tabela 2.9: Valores empíricos a posteriori sob o espaço paramétrico restrito (com distribuições a priori informativas)

Modelo	Parâmetro	Média	Desvio Padrão
1	μ	92.2	0.8
2	μ_1	94.2	1.1
	μ_2	91.0	0.7

avaliados aqui (ver Tabela 2.8), o fator de Bayes mostra que existe apenas uma evidência vaga contra a hipótese nula, ou, equivalentemente contra a afirmação que os dados são mais compatíveis com o modelo $M = 1$.

A Tabela 2.9 apresenta os valores empíricos da média e desvio padrão da distribuição a posteriori para cada parâmetro, no caso $\pi_1 = \pi_2 = 0.5$.

Uma análise de sensibilidade do fator de Bayes à escolha dos hiperparâmetros das distribuições a priori pode ser realizada considerando uma diminuição do valor do parâmetro de precisão (inverso da variância) das distribuições a priori. Isso corresponde a considerar distriuições a priori “menos informativas”. A Tabela (2.10) apresenta os fatores de Bayes calculados para diferentes valores do parâmetro de precisão τ_μ da distribuição a priori do parâmetro μ , e diferentes valores do parâmetros de precisão τ_{μ_1} e τ_{μ_2} das distribuições a priori dos parâmetros μ_1 e μ_2 , respectivamente, para $\pi_1 = \pi_2 = 0.5$.

Tabela 2.10: fatores de Bayes para diferentes precisões das distribuições a priori dos parâmetros μ , μ_1 e μ_2

Parâmetros de precisão			Fator de Bayes
τ_μ	τ_{μ_1}	τ_{μ_2}	
0.65000	0.35000	0.35000	1.65
0.06500	0.03500	0.03500	0.69
0.00650	0.00350	0.00350	0.63
0.00065	0.00035	0.00035	0.07
0.00001	0.00001	0.00001	0.02

Podemos observar que o fator de Bayes diminui quando a precisão diminui ou a variância aumenta, indicando que é sensível à escolha dos hiperparâmetros das distribuições a priori. Ao considerar as distribuições a priori com hiperparâmetros de precisão iguais a 0.00001, ou seja, distribuições a priori não informativas, o fator de Bayes é $B_{21} = 0.02$, que segundo a proposta de Kass e Raftery (1995) sugere a opção pelo modelo correspondente à hipótese nula.

Calculamos também o fator de Bayes no caso em que não impomos restrições sobre as médias, considerando $\pi_1 = \pi_2 = 0.5$; obtivemos $B_{21} = 2.52$, que é maior que o correspondente fator de Bayes associado ao espaço paramétrico restrito ($B_{21} = 1.65$). Isto sugere que a incorporação da restrição diminui a evidência contra a hipótese nula, contrariamente ao esperado. Ao considerar distribuições a priori não informativas com hiperparâmetros de precisão 0.00001 obtivemos $B_{21} = 0.03$, no caso do espaço paramétrico sem restrições, valor maior que aquele obtido sob o espaço paramétrico restrito ($B_{21} = 0.02$).

A Tabela 2.11 apresenta médias e desvios padrões a posteriori sob o espaço paramétrico irrestrito. Notemos que estes valores são similares às médias e desvios padrões a posteriori obtidas sob o espaço paramétrico restrito apresentados na Tabela 2.9, para μ e μ_1 sugerindo que a restrição sobre o espaço paramétrico parece não afetar a estimação a posteriori esses parâmetros. Ou seja, as distribuições condicionais a posteriori de μ e μ_1 não foram influenciadas pelas restrições sobre as distribuições a priori. Os valores de média e desvio padrão a posteriori associados ao parâmetro μ_2 sob o espaço paramétrico sem restrições são diferentes aos obtidos sob o espaço paramétrico irrestrito. Logo

Tabela 2.11: Valores empíricos a posteriori sobre o espaço paramétrico irrestrito (com distribuições a priori informativas)

Modelo	Parâmetro	Média	Desvio Padrão
1	μ	92.2	0.9
2	μ_1	94.2	1.2
	μ_2	90.2	1.2

a distribuição condicional a posteriori de μ_2 foi influenciada pelas restrições sobre as distribuições a priori. A influência das restrições sobre a amostragem da distribuição condicional completa a posteriori de M dados θ_1 e θ_2 , através da qual obtemos a estimativa do fator de Bayes, é evidenciada pelas diferenças entre os fatores de Bayes calculados sob o espaço paramétrico sem restrições e sob o espaço paramétrico restrito.

Os valores das médias a posteriori sob o espaço paramétrico irrestrito são similares as estimativas de mínimos quadrados sob os modelos correspondentes, apresentadas na Tabela 2.6, e que também foram consideradas na análise freqüentista (ver Tabela 2.5) apresentada na Seção 2.2.

2.3.2 Cálculo do fator de Bayes segundo a metodologia de Irony e Pereira

Descreveremos aqui outra metodologia para calcular o fator de Bayes.

Sejam $\{\Omega_1, \Omega_2\}$ uma partição do espaço paramétrico $\Omega (\subset \mathbb{R}^s)$, isto é, $\Omega_1 \cup \Omega_2 = \Omega$, e $\Omega_1 \cap \Omega_2 = \emptyset$ e $\theta \in \Omega$, um parâmetro desconhecido. As hipóteses a serem confrontadas são

$$H_i : \theta \in \Omega_i, \quad (2.21)$$

com $i = 1, 2$, ou equivalentemente, denotando por M o parâmetro indicador do modelo,

$$H_i : M = i \quad (2.22)$$

com $i = 1, 2$. A preferência sobre a veracidade de $H_1 (H_2)$ em prejuízo de $H_2 (H_1)$

é descrita pela probabilidade a priori da hipótese nula (alternativa) $P(M = 1) = \pi_1$ ($P(M = 2) = \pi_2$), onde $\pi_1 + \pi_2 = 1$.

Assim antes de observar os dados, d , a preferência por uma ou por outra hipótese é dada por (π_1, π_2) e após a observação de d , essa preferência é dada por $(\pi_1(d), \pi_2(d))$. No Exemplo 1.1 d é o conjunto de observações da pressão diastólica dos indivíduos do grupo com estresse e do grupo controle.

A razão de probabilidades a posteriori, ou razão de Bayes é

$$R_{12}(d) = \pi_1(d) / \pi_2(d), \quad (2.23)$$

que é equivalente ao fator de Bayes dado em (2.20), ou seja

$$B_{21} = \frac{p(y|M=2)\pi_2}{p(y|M=1)\pi_1} = R_{12}^{-1}.$$

Se $R_{12}(d) \geq c$ não rejeitamos a hipótese nula. A constante c é escolhida levando-se em consideração o grau de importância dos erros. Se $\pi_1 = \pi_2 = 0.5$ e $c = 1$, ambos erros têm a mesma importância. enquanto que, se $c < 1$, o erro de tipo II tem um custo c vezes maior que o erro de tipo I. Para mais detalhes sobre o significado da constante c ver Seção B.4 do Apêndice B.

A densidade a priori de θ sob H_i ($i = 1, 2$) é definida por

$$g(\theta|M=i) = g_i(\theta) = \frac{g(\theta) I_i(\theta)}{\int g(\theta) d\Omega_i} \quad (i = 1, 2) \quad (2.24)$$

com $I_i(\theta)$ denotando a função indicadora de $\theta \in \Omega_i$, isto é,

$$I_i(\theta) = \begin{cases} 1 & \text{se } \theta \in \Omega_i \\ 0 & \text{se } \theta \notin \Omega_i, \end{cases}$$

e $\int g(\theta) d\Omega_i$ representa o volume sob a função densidade de probabilidade a priori $g(\theta)$ na superfície (linha) Ω_i . Se este integral é nulo, como acontece quando a dimensão de Ω_i

é $s_i < s$, consideraremos o integral sobre o conjunto isomorfo a Ω_i em $\mathbb{R}^{s_i}$.

Naturalmente que as densidades a priori (2.24) estão bem definidas se os integrais $\int g(\boldsymbol{\theta}) d\Omega_i < \infty$, ($i = 1, 2$), ou seja se $g(\boldsymbol{\theta})$ é própria.

A densidade a posteriori de $\boldsymbol{\theta}$ sob a hipótese H_i ($i = 1, 2$) é dada por

$$g(\boldsymbol{\theta}|M = i, d) = g_i(\boldsymbol{\theta}|d) = \frac{g(\boldsymbol{\theta}) l(\boldsymbol{\theta}|d) I_i(\boldsymbol{\theta})}{\int g(\boldsymbol{\theta}) l(\boldsymbol{\theta}|d) d\Omega_i}, \quad (i = 1, 2). \quad (2.25)$$

que está bem definida se $\int g(\boldsymbol{\theta}) l(\boldsymbol{\theta}|d) d\Omega_i < \infty$.

Sob a hipótese (2.21), a função preditiva (a função de probabilidade dos dados) $P(d|M = i)$ é definida como

$$f_i(d) = \frac{\int g(\boldsymbol{\theta}) l(\boldsymbol{\theta}|d) d\Omega_i}{\int g(\boldsymbol{\theta}) d\Omega_i}, \quad (2.26)$$

com $i = 1, 2$ em que $l(\boldsymbol{\theta}|d)$ é a função de verossimilhança, i.e., uma função de $\boldsymbol{\theta}$ tal que para cada valor observado d , $l(\boldsymbol{\theta}|d)$ é a função de densidade de probabilidade $f(d|\boldsymbol{\theta})$ avaliada em d . Temos como uma aplicação direta do Teorema de Bayes que a probabilidade a posteriori $\pi_i(d)$ do evento $\{M = i\}$ dado d é

$$\pi_i(d) = P(M = i|d) = \frac{\pi_i f_i(d)}{\pi_1 f_1(d) + \pi_2 f_2(d)},$$

$i = 1, 2$. Logo a razão de probabilidades a posteriori (2.23) é obtida como

$$R_{12} = \frac{\pi_1 f_1(d)}{\pi_2 f_2(d)}. \quad (2.27)$$

Na análise sob esta técnica, especifica-se a densidade conjunta a priori dos parâmetros definida em todo o espaço paramétrico, $g(\boldsymbol{\theta})$ e a partir dessa densidade, definem-se as densidades a priori sob as hipótese nula e alternativa através de (2.24). Sob a técnica proposta por Carlin e Chib (1995), as distribuições a priori sob as hipótese nula e alternativa são especificadas diretamente e independentemente. Isto é evidenciado claramente pelo número de hiperparâmetros envolvidos em cada caso (maior na metodologia

de Carlin e Chib). Nesse caso obtemos amostras das densidades completas a posteriori $p(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, M, \mathbf{y})$, $p(\boldsymbol{\theta}_2|\mu, M, \mathbf{y})$ e $p(M|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \mathbf{y})$, ao passo que no caso da técnica desenvolvida por Irony e Pereira (1995) são obtidas as densidades a posteriori conjuntas (2.25). Embora uma comparação direta entre os resultados obtidos por cada metodologia não seja possível, o desempenho do fator de Bayes com respeito à escolha das distribuições a priori pode ser comparado.

A seguir aplicaremos a técnica de Irony e Pereira (1995) aos dados do Exemplo 1.1 admitindo que a variância dos erros é conhecida. Para exemplificar, consideraremos distribuições a priori normais truncadas e distribuições a priori exponencias para os parâmetros de interesse. Essas distribuições não pertencem à família de distribuições conjugadas. Neste caso,

$$\begin{aligned}\Omega_1 &= \{\boldsymbol{\varpi} = (\mu_1, \mu_2) : \mu_1 \geq 0, \mu_2 \geq 0 \text{ e } \mu_1 = \mu_2\}, \\ \Omega_2 &= \{\boldsymbol{\varpi} = (\mu_1, \mu_2) : \mu_1 \geq 0, \mu_2 \geq 0 \text{ e } \mu_1 > \mu_2\}.\end{aligned}$$

A função de verossimilhança é

$$\begin{aligned}l(\mu_1, \mu_2|\mathbf{x}, \mathbf{y}) &= k\tau^{(n_1+n_2)/2} \\ &\times \exp\left(-\frac{\tau}{2}\left[\sum_{i=1}^{n_1}(X_i - \bar{X})^2 + n_1(\bar{X} - \mu_1)^2\right]\right) \\ &\times \exp\left(-\frac{\tau}{2}\left[\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2 + n_2(\bar{Y} - \mu_2)^2\right]\right),\end{aligned}\quad (2.28)$$

em que $\mathbf{x} = (X_1, \dots, X_{n_1})$, $\mathbf{y} = (Y_1, \dots, Y_{n_2})$, $\tau = \sigma^{-2}$ e $k = (2\pi)^{-(n_1+n_2)/2}$.

Suponhamos que μ_1 e μ_2 têm distribuições a priori normais truncadas com zero, de parâmetros (a_1, γ_1) e (a_2, γ_2) , respectivamente, ou seja, para as quais as funções de densidades são

$$g(\mu_1) = \frac{1}{1 - \Phi(-a_1\sqrt{\gamma_1})} \frac{\sqrt{\gamma_1}}{\sqrt{2\pi}} \exp\left(\frac{-\gamma_1}{2}(\mu_1 - a_1)^2\right) I_{[0, \infty)}, \quad (2.29)$$

$$g(\mu_2) = \frac{1}{1 - \Phi(-a_2\sqrt{\gamma_2})} \frac{\sqrt{\gamma_2}}{\sqrt{2\pi}} \exp\left(\frac{-\gamma_2}{2}(\mu_2 - a_2)^2\right) I_{[0,\infty)}. \quad (2.30)$$

Portanto temos

$$E(\mu_1) = a_1 + \frac{1}{1 - \Phi(-a_1\sqrt{\gamma_1})} \frac{\sqrt{\gamma_1}}{\sqrt{2\pi}} \exp\left(\frac{-\gamma_1}{2}a_1^2\right), \quad (2.31)$$

$$\begin{aligned} Var(\mu_1) &= \frac{1}{\gamma_1} \left\{ 1 - \frac{a_1}{1 - \Phi(-a_1\sqrt{\gamma_1})} \frac{\sqrt{\gamma_1}}{\sqrt{2\pi}} \exp\left(\frac{-\gamma_1}{2}a_1^2\right) \right. \\ &\quad \left. - \left[\frac{1}{1 - \Phi(-a_1\sqrt{\gamma_1})} \frac{\sqrt{\gamma_1}}{\sqrt{2\pi}} \exp\left(\frac{-\gamma_1}{2}a_1^2\right) \right]^2 \right\}. \end{aligned} \quad (2.32)$$

As expressões para $E(\mu_2)$ e $Var(\mu_2)$ são equivalentes, com a substituição de a_1 por a_2 , e de γ_1 por γ_2 . A densidade conjunta a priori é dada por

$$g(\mu_1, \mu_2) = C \frac{\sqrt{\gamma_1\gamma_2}}{2\pi} \exp\left(-\frac{\gamma_1}{2}(\mu_1 - a_1)^2 - \frac{\gamma_2}{2}(\mu_2 - a_2)^2\right) I_{[0,\infty) \times [0,\infty)},$$

onde $C = [(1 - \Phi(-a_1\sqrt{\gamma_1}))(1 - \Phi(-a_2\sqrt{\gamma_2}))]^{-1} = [\Phi(a_1\sqrt{\gamma_1})\Phi(a_2\sqrt{\gamma_2})]^{-1}$.

A verossimilhança conjunta dos dados e do parâmetro $\varpi = (\mu_1, \mu_2)$ é

$$\begin{aligned} &g(\mu_1, \mu_2) l(\mu_1, \mu_2 | \mathbf{x}, \mathbf{y}) \\ &= C_{KT}^{(n_1+n_2)/2} \frac{\sqrt{\gamma_1\gamma_2}}{2\pi} \\ &\quad \times \exp\left(-\frac{\gamma_1}{2}(\mu_1 - a_1)^2 - \frac{\gamma_2}{2}(\mu_2 - a_2)^2\right) \\ &\quad \times \exp\left(-\frac{\tau}{2} \left[\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + n_1 (\bar{X} - \mu_1)^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2 + n_2 (\bar{Y} - \mu_2)^2 \right] \right). \end{aligned} \quad (2.33)$$

Calcularemos agora os numeradores e denominadores dos cocientes que definem as funções preditivas $f_i(d)$ dadas em (2.26). Definindo $\tilde{g}(\mu) = g(\mu, \mu)$ e $\tilde{\Omega}_1 = \{\mu : \mu \geq 0\}$; o denominador do cociente que define $f_1(d)$ é igual a

$$\begin{aligned}
& \int \tilde{g}(\mu) d\tilde{\Omega}_1 \\
&= C \frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \int_0^\infty \exp\left(-\frac{\gamma_1}{2}(\mu - a_1)^2 - \frac{\gamma_2}{2}(\mu - a_2)^2\right) d\mu \\
&= C \frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \int_0^\infty \exp\left(-\frac{\gamma_1 + \gamma_2}{2} \left(\mu - \frac{\left(\frac{\gamma_1}{2}a_1 + \frac{\gamma_2}{2}a_2\right)}{\frac{\gamma_1 + \gamma_2}{2}}\right)^2 - \frac{\frac{\gamma_1 \gamma_2}{4}(a_1 - a_2)^2}{\frac{\gamma_1 + \gamma_2}{2}}\right) d\mu \\
&= C \exp\left(-\frac{\gamma_1 \gamma_2 (a_1 - a_2)^2}{2(\gamma_1 + \gamma_2)}\right) \frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \int_0^\infty \exp\left(-\frac{\gamma_1 + \gamma_2}{2} \left(\mu - \frac{(\gamma_1 a_1 + \gamma_2 a_2)}{(\gamma_1 + \gamma_2)}\right)^2\right) d\mu \\
&= C \frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \frac{\sqrt{2\pi}}{\sqrt{\gamma_1 + \gamma_2}} \exp\left(-\frac{\gamma_1 \gamma_2 (a_1 - a_2)^2}{2(\gamma_1 + \gamma_2)}\right) \Phi\left(\frac{(\gamma_1 a_1 + \gamma_2 a_2)}{(\gamma_1 + \gamma_2)} \sqrt{\gamma_1 + \gamma_2}\right). \quad (2.34)
\end{aligned}$$

O numerador do cociente que define $f_1(d)$ é

$$\begin{aligned}
& \int \tilde{g}(\mu) l(\mu, \mu | \mathbf{x}, \mathbf{y}) d\tilde{\Omega}_1 \\
&= C k \tau^{(n_1 + n_2)/2} \frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \times \exp\left(-\frac{\tau}{2} \left[\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2 \right]\right) \times \\
& \quad \int_0^\infty \exp\left(-\frac{\gamma_1}{2}(\mu - a_1)^2 - \frac{\gamma_2}{2}(\mu - a_2)^2 - \frac{\tau}{2} [n_1 (\bar{X} - \mu)^2 + n_2 (\bar{Y} - \mu)^2]\right) d\mu \\
&= K \exp\left(-\frac{\gamma_1 \gamma_2 (a_1 - a_2)^2}{2(\gamma_1 + \gamma_2)}\right) \exp\left(-\frac{\tau n_1 n_2 (\bar{X} - \bar{Y})^2}{2(n_1 + n_2)}\right) \\
& \quad \times \int_0^\infty \exp\left(-\frac{\gamma_1 + \gamma_2}{2} \left(\mu - \frac{(\gamma_1 a_1 + \gamma_2 a_2)}{(\gamma_1 + \gamma_2)}\right) - \frac{\tau}{2} (n_1 + n_2) \left(\mu - \frac{n_1 \bar{X} + n_2 \bar{Y}}{(n_1 + n_2)}\right)^2\right) d\mu \\
&= K \exp\left(-\frac{\gamma_1 \gamma_2 (a_1 - a_2)^2}{2(\gamma_1 + \gamma_2)}\right) \exp\left(-\frac{\tau n_1 n_2 (\bar{X} - \bar{Y})^2}{2(n_1 + n_2)}\right) \\
& \quad \exp\left(-\frac{(\gamma_1 + \gamma_2) \tau (n_1 + n_2)}{2((\gamma_1 + \gamma_2) + \tau (n_1 + n_2))} \left(\frac{(\gamma_1 a_1 + \gamma_2 a_2)}{(\gamma_1 + \gamma_2)} - \frac{n_1 \bar{X} + n_2 \bar{Y}}{(n_1 + n_2)}\right)^2\right) \\
& \quad \int_0^\infty \exp\left(-\left(\frac{\gamma_1 + \gamma_2}{2} + \frac{\tau}{2} (n_1 + n_2)\right) \left(\mu - \frac{(\gamma_1 a_1 + \gamma_2 a_2) + \tau (n_1 \bar{X} + n_2 \bar{Y})}{\gamma_1 + \gamma_2 + \tau (n_1 + n_2)}\right)^2\right) d\mu
\end{aligned}$$

$$\begin{aligned}
&= K \exp \left(-\frac{\gamma_1 \gamma_2 (a_1 - a_2)^2}{2 (\gamma_1 + \gamma_2)} \right) \exp \left(-\frac{\tau n_1 n_2 (\bar{X} - \bar{Y})^2}{2 (n_1 + n_2)} \right) \\
&\exp \left(-\frac{(\gamma_1 + \gamma_2) \tau (n_1 + n_2)}{2 ((\gamma_1 + \gamma_2) + \tau (n_1 + n_2))} \left(\frac{(\gamma_1 a_1 + \gamma_2 a_2)}{(\gamma_1 + \gamma_2)} - \frac{n_1 \bar{X} + n_2 \bar{Y}}{(n_1 + n_2)} \right)^2 \right) \\
&\int_0^\infty \exp \left(-\left(\frac{\gamma_1 + \gamma_2}{2} + \frac{\tau}{2} (n_1 + n_2) \right) \left(\mu - \frac{(\gamma_1 a_1 + \gamma_2 a_2) + \tau (n_1 \bar{X} + n_2 \bar{Y})}{\gamma_1 + \gamma_2 + \tau (n_1 + n_2)} \right)^2 \right) d\mu,
\end{aligned}$$

com $K = C k \tau^{(n_1+n_2)/2} \frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \times \exp \left(-\frac{\tau}{2} \left[\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2 \right] \right)$. Então

$$\begin{aligned}
&\int \tilde{g}(\mu) l(\mu, \mu | \mathbf{x}, \mathbf{y}) d\tilde{\Omega}_1 \\
&= K \exp \left(-\frac{\gamma_1 \gamma_2 (a_1 - a_2)^2}{2 (\gamma_1 + \gamma_2)} \right) \exp \left(-\frac{\tau n_1 n_2 (\bar{X} - \bar{Y})^2}{2 (n_1 + n_2)} \right) \\
&\times \exp \left(-\frac{(\gamma_1 + \gamma_2) \tau (n_1 + n_2)}{2 ((\gamma_1 + \gamma_2) + \tau (n_1 + n_2))} \left(\frac{(\gamma_1 a_1 + \gamma_2 a_2)}{(\gamma_1 + \gamma_2)} - \frac{n_1 \bar{X} + n_2 \bar{Y}}{(n_1 + n_2)} \right)^2 \right) \\
&\times \sqrt{\frac{2\pi}{r_1}} \Phi(A\sqrt{r_1}), \tag{2.35}
\end{aligned}$$

com

$$r_1 = (\gamma_1 + \gamma_2 + \tau (n_1 + n_2)) \text{ e } A = \frac{(\gamma_1 a_1 + \gamma_2 a_2) + \tau (n_1 \bar{X} + n_2 \bar{Y})}{\gamma_1 + \gamma_2 + \tau (n_1 + n_2)}.$$

Logo a distribuição preditiva ou verossimilhança modificada sobre a hipótese nula, co-ciente dos integrais (2.34) e (2.35), é

$$\begin{aligned}
f_1(d) &= \frac{K}{C} \sqrt{\frac{2\pi}{r_1}} \left[\frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \frac{\sqrt{2\pi}}{\sqrt{\gamma_1 + \gamma_2}} \right]^{-1} \\
&\times \exp \left(-\frac{\gamma_1 \gamma_2 (a_1 - a_2)^2}{2(\gamma_1 + \gamma_2)} \right) \exp \left(-\frac{\tau n_1 n_2 (\bar{X} - \bar{Y})^2}{2(n_1 + n_2)} \right) \Phi(A\sqrt{r_1}) \\
&\times \exp \left(-\frac{(\gamma_1 + \gamma_2) \tau (n_1 + n_2)}{2((\gamma_1 + \gamma_2) + \tau(n_1 + n_2))} \left(\frac{(\gamma_1 a_1 + \gamma_2 a_2)}{(\gamma_1 + \gamma_2)} - \frac{n_1 \bar{X} + n_2 \bar{Y}}{(n_1 + n_2)} \right)^2 \right) \\
&\times \left[\Phi \left(\frac{(\gamma_1 a_1 + \gamma_2 a_2)}{(\gamma_1 + \gamma_2)} \sqrt{\gamma_1 + \gamma_2} \right) \right]^{-1}. \tag{2.36}
\end{aligned}$$

Analogamente,

$$\begin{aligned}
&\int g(\mu_1, \mu_2) d\Omega_2 \\
&= C \frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \int_0^\infty \int_0^{\mu_1} \exp \left(-\frac{\gamma_1}{2} (\mu_1 - a_1)^2 - \frac{\gamma_2}{2} (\mu_2 - a_2)^2 \right) d\mu_2 d\mu_1 \\
&= C \int_0^\infty \sqrt{\frac{\gamma_1}{2\pi}} \exp \left(\frac{-\gamma_1}{2} [(\mu_1 - a_1)^2] \right) \left\{ \int_0^{\mu_1} \sqrt{\frac{\gamma_2}{2\pi}} \exp \left(\frac{-\gamma_2}{2} (\mu_2 - a_2)^2 \right) d\mu_2 \right\} d\mu_1 \\
&= C \int_0^\infty \sqrt{\frac{\gamma_1}{2\pi}} \exp \left(\frac{-\gamma_1}{2} [(\mu_1 - a_1)^2] \right) [(\Phi(\sqrt{\gamma_2}(\mu_1 - a_2)) - \Phi(-a_2\sqrt{\gamma_2}))] d\mu_1 \\
&= C \int_0^\infty \sqrt{\frac{\gamma_1}{2\pi}} \exp \left(\frac{-\gamma_1}{2} [(\mu_1 - a_1)^2] \right) \Phi(\sqrt{\gamma_2}(\mu_1 - a_2)) d\mu_1 \\
&\quad - C \Phi(-a_2\sqrt{\gamma_2}) \int_0^\infty \sqrt{\frac{\gamma_1}{2\pi}} \exp \left(\frac{-\gamma_1}{2} [(\mu_1 - a_1)^2] \right) d\mu_1 \\
&= C \left[\sqrt{\frac{\gamma_1}{2\pi}} \text{ef}(\gamma_1, a_1, \gamma_2, a_2) - \Phi(-a_2\sqrt{\gamma_2}) \Phi(a_1\sqrt{\gamma_1}) \right], \tag{2.37}
\end{aligned}$$

em que $\text{ef}(b, a, m, n) = \int_0^\infty \exp \left(-\frac{b}{2} (u - a)^2 \right) \Phi(\sqrt{m}(u - n)) du$.

$$\begin{aligned}
& \int g(\mu_1, \mu_2) l(\mu_1, \mu_2 | \mathbf{x}, \mathbf{y}) d\Omega_2 \\
&= K \int_0^\infty \int_0^{\mu_1} g(\mu_1, \mu_2) l(\mu_1, \mu_2) d\mu_2 d\mu_1 \\
&= K \exp\left(-v_1 (\bar{X} - a_1)^2 - v_2 (\bar{Y} - a_2)^2\right) \times \\
&\quad \int_0^\infty \exp(-k_1 (\mu_1 - \bar{\mu}_1)^2) \left\{ \int_0^{\mu_1} \exp(-k_2 (\mu_2 - \bar{\mu}_2)^2) d\mu_2 \right\} d\mu_1, \\
&= K \exp\left(-v_1 (\bar{X} - a_1)^2 - v_2 (\bar{Y} - a_2)^2\right) \times \\
&\quad \sqrt{\frac{2\pi}{2k_2}} \int_0^\infty \exp(-k_1 (\mu_1 - \bar{\mu}_1)^2) \left\{ \Phi\left(\sqrt{2k_2} (\mu_1 - \bar{\mu}_2)\right) - \Phi\left(-\bar{\mu}_2 \sqrt{2k_2}\right) \right\} \\
&= K \exp\left(-v_1 (\bar{X} - a_1)^2 - v_2 (\bar{Y} - a_2)^2\right) \times \\
&\quad \left\{ \sqrt{\pi/k_2} \operatorname{erf}(2k_1, \bar{\mu}_1, 2k_2, \bar{\mu}_2) - \frac{\pi}{\sqrt{k_2 k_1}} \Phi\left(-\bar{\mu}_2 \sqrt{2k_2}\right) \Phi\left(\bar{\mu}_1 \sqrt{2k_1}\right) \right\}, \quad (2.38)
\end{aligned}$$

com

$$\begin{aligned}
\bar{\mu}_1 &= \frac{\gamma_1 a_1 + \tau n_1 \bar{X}}{\gamma_1 + \tau n_1}, v_1 = \frac{\gamma_1 \tau n_1}{2(\gamma_1 + \tau n_1)}, k_1 = \frac{\gamma_1 + \tau n_1}{2}, \\
\bar{\mu}_2 &= \frac{\gamma_2 a_2 + \tau n_2 \bar{Y}}{\gamma_2 + \tau n_2}, v_2 = \frac{\gamma_2 \tau n_2}{2(\gamma_2 + \tau n_2)}, k_2 = \frac{\gamma_2 + \tau n_2}{2}.
\end{aligned}$$

Logo a distribuição preditiva ou verossimilhança modificada sobre a hipótese alternativa, cociente dos integrais (2.37) e (2.38), é

$$\begin{aligned}
f_2(d) &= K \exp\left(-v_1 (\bar{X} - a_1)^2 - v_2 (\bar{Y} - a_2)^2\right) \\
&\quad \times \left\{ \sqrt{\pi/k_2} \operatorname{erf}(2k_1, \bar{\mu}_1, 2k_2, \bar{\mu}_2) - \frac{\pi}{\sqrt{k_2 k_1}} \Phi\left(-\bar{\mu}_2 \sqrt{2k_2}\right) \Phi\left(\bar{\mu}_1 \sqrt{2k_1}\right) \right\} \\
&\quad \div C \left[\sqrt{\frac{\gamma_1}{2\pi}} \operatorname{erf}(\gamma_1, a_1, \gamma_2, a_2) - \Phi\left(-a_2 \sqrt{\gamma_2}\right) \Phi\left(a_1 \sqrt{\gamma_1}\right) \right]. \quad (2.39)
\end{aligned}$$

Escolhendo probabilidades a priori iguais para as hipótese nula e alternativa ($\pi_1 = \pi_2$), o fator de Bayes (2.27) é a razão entre as verossimilhanças (2.36) e (2.39).

Tabela 2.12: Valores dos hiperparâmetros e os fatores de Bayes correspondentes

$\tau_1 = \tau_2$	a_1	γ_1	a_2	γ_2	B_{21}
0.35000	4.1501	0.3298	-0.2207	0.3049	2.81
0.20000	4.1218	0.1652	-0.0864	0.1815	2.45
0.04000	4.0934	0.0283	0.1010	0.0384	1.62
0.03500	4.0946	0.0249	0.1198	0.0336	1.60
0.00350	4.1276	0.0030	0.2172	0.0034	1.60
0.00035	4.1500	0.0003	0.2490	0.0003	1.63

A principal desvantagem do emprego do fator de Bayes para testar hipóteses é que ele poder ser muito sensível à escolha dos hiperparâmetros das distribuições a priori, como observamos na seção anterior. Para avaliar a sensibilidade neste caso, utilizamos os dados do Exemplo 1.1 supondo σ^2 conhecida e aproximadamente igual ao sua estimativa de mínimos quadrados, o que implica $\tau = 0.03$.

Por razões de simplicidade de cálculo e sem perda da generalidade, escolhemos distribuições a priori para os parâmetros μ_1 e μ_2 tais que $E(\mu_1) = \bar{X} - 90 = 4.1636$, e $E(\mu_2) = \bar{Y} - 90 = 0.2636$, e também consideramos diferentes valores de $Var(\mu_1)$ e $Var(\mu_2)$ ou valores de precisão τ_1 e τ_2 (0.35000, 0.20000, 0.04000, 0.03500, 0.00350 e 0.00035). Com tal propósito resolvemos os sistemas de equações (2.31) e (2.32) para μ_1 e μ_2 para obter os valores dos hiperparâmetros a_1 , γ_X , a_2 e γ_Y correspondentes a essas distribuições a priori. Os valores dos hiperparâmetros e os valores do fator de Bayes correspondentes são apresentados na Tabela 2.12.

O fator de Bayes diminui quando a precisão $\tau = \tau_1 = \tau_2$ decresce, tornando-se mais estável (com valores próximos a 1.60) no caso de distribuições a priori menos informativas. Em todos os casos rejeitamos a hipótese nula, porém com menor evidência ao considerar distribuições a priori não informativas.

O fator de Bayes também é sensível à escolha dos hiperparâmetros a_1 e a_2 . Suponhamos fixo o valor de $\gamma_1 = \gamma_2 = \gamma = 0.349$; a Tabela 2.13 apresenta os fatores de Bayes correspondentes a diversos valores de a_1 e a_2 . Podemos observar que o fator de Bayes diminui quando a_1 e/ou a_2 aumentam. Em todos os casos rejeitamos a hipótese nula.

Tabela 2.13: Fator de Bayes para o hiperpâmetro $\gamma = 0.349$

$a_1 \backslash a_2$	0.000	1.000	2.000	3.000	4.000	5.000
0.000	2.44	2.23	2.04	1.88	1.75	1.65
1.000	2.63	2.37	2.15	1.97	1.84	1.73
2.000	2.78	2.50	2.27	2.08	1.94	1.83
3.000	2.87	2.59	2.38	2.22	2.08	1.95
4.000	2.82	2.63	2.48	2.35	2.23	2.09
5.000	2.62	2.56	2.52	2.47	2.38	2.25

Suponhamos agora que não existem restrições no espaço paramétrico a título de comparação, estudemos o comportamento do fator de Bayes nesta situação. O espaço paramétrico irrestrito tem como partição

$$\dot{\Omega}_1 = \{\varpi = (\mu_1, \mu_2) : \mu_1 = \mu_2\}, \quad (2.40)$$

$$\dot{\Omega}_2 = \{\varpi = (\mu_1, \mu_2) : \mu_1 > \mu_2\}. \quad (2.41)$$

Considerando para μ_1 e μ_2 distribuições a priori normais com funções densidades

$$\dot{g}(\mu_1) = \frac{\sqrt{\dot{\gamma}_1}}{\sqrt{2\pi}} \exp\left(\frac{-\dot{\gamma}_1}{2} (\mu_1 - \dot{a}_1)^2\right) \quad (2.42)$$

e

$$\dot{g}(\mu_2) = \frac{\sqrt{\dot{\gamma}_2}}{\sqrt{2\pi}} \exp\left(\frac{-\dot{\gamma}_2}{2} (\mu_2 - \dot{a}_2)^2\right). \quad (2.43)$$

Notemos que os integrais que definem as funções preditivas podem ser calculados de forma similar à dos integrais (2.34), (2.35), (2.37) e (2.38).

Para efeitos de comparação, escolhemos como valores de $E(\mu_1) = \dot{a}_1 = \bar{X} - 90 = 4.1636$, $E(\mu_2) = \dot{a}_2 = \bar{Y} - 90 = 0.2636$ e $(Var(\mu_1))^{-1} = \dot{\gamma}_1 = \dot{\tau}_1 = (Var(\mu_2))^{-1} = \dot{\gamma}_2 = \dot{\tau}_2 = \dot{\tau}$. Na Tabela 2.14, que apresenta os fatores de Bayes correspondentes a diversos valores de $\dot{\tau}$, podemos observar que o fator de Bayes diminui quando $\dot{\tau}$ diminui, sugerindo que

Tabela 2.14: Fator de Bayes considerando distribuições a priori normais no espaço paramétrico irrestrito

$\dot{\tau}$	$B_{21} = R_{12}^{-1}$
0.35000	2.16
0.20000	1.97
0.04000	0.46
0.03500	0.45
0.00350	0.40
0.00035	0.05

quanto menos informativa a distribuição a priori, menores são as evidências contra a hipótese nula.

Comparando os fatores de Bayes das Tabelas (2.12) e (2.14) para os mesmos valores dos hiperparâmetros de dispersão $\tau_1 = \tau_2 = \dot{\tau} = 0.35$, observamos que o fator de Bayes correspondente ao espaço paramétrico restrito ($B_{21} = 2.81$) é maior que o correspondente ao espaço paramétrico irrestrito ($B_{21} = 2.16$), ou seja que a restrição aumenta a evidência contra a hipótese nula. Para o caso de distribuições a priori menos informativas o fator de Bayes correspondente ao espaço paramétrico irrestrito é menor que o correspondente ao espaço paramétrico restrito como esperamos, o que não acontece quando utilizamos a técnica dada por Carlin e Chib (1995).

Para evidenciar a flexibilidade da técnica proposta por Irony e Pereira (1995) suponhamos agora que μ_1 e μ_2 têm distribuições a priori exponenciais, de parâmetros λ_1 e λ_2 , respectivamente, ou seja,

$$\dot{g}(\mu_1) = \lambda_1 \exp(-\lambda_1 \mu_1) I_{[0, \infty)}, \quad (2.44)$$

e

$$\dot{g}(\mu_2) = \lambda_2 \exp(-\lambda_2 \mu_2) I_{[0, \infty)}. \quad (2.45)$$

Então $E(\mu_1) = \lambda_1^{-1}$, $Var(\mu_1) = \lambda_1^{-2}$, $E(\mu_2) = \lambda_2^{-1}$, $Var(\mu_2) = \lambda_2^{-2}$ e a densidade

conjunta a priori é dada por

$$\overset{\circ}{g}(\mu_1, \mu_2) = \lambda_1 \lambda_2 \exp(-\lambda_1 \mu_1 - \lambda_2 \mu_2) I_{[0, \infty) \times [0, \infty)}.$$

A verossimilhança conjunta dos dados e do parâmetro $\varpi = (\mu_1, \mu_2)$ é

$$\begin{aligned} \overset{\circ}{g}(\mu_1, \mu_2) &= \lambda_1 \lambda_2 k \tau^{(n_1+n_2)/2} \times \\ &\exp(-\lambda_1 \mu_1 - \lambda_2 \mu_2) \\ &\times \exp\left(-\frac{\tau}{2} \left[\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + n_1 (\bar{X} - \mu_1)^2 \right]\right) \\ &\times \exp\left(-\frac{\tau}{2} \left[\sum_{i=1}^{n_2} (Y_i - \bar{Y})^2 + n_2 (\bar{Y} - \mu_2)^2 \right]\right) \times I_{[0, \infty) \times [0, \infty)}. \end{aligned}$$

Para obter o fator de Bayes, calcularemos as funções preditivas $f_i(d)$ apresentadas em (2.26). O denominador do cociente que define $f_1(d)$ é igual a

$$\begin{aligned} &\int \tilde{g}(\mu) d\tilde{\Omega}_1 \\ &= \lambda_1 \lambda_2 \int_0^\infty \exp(-(\lambda_1 + \lambda_2)\mu) d\mu \\ &= \frac{\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2)}, \end{aligned} \tag{2.46}$$

em que $\tilde{g}(\mu) = \overset{\circ}{g}(\mu, \mu)$ e $\tilde{\Omega}_1 = \{\mu : \mu > 0\}$. O numerador do cociente que define $f_0(d)$ é

$$\begin{aligned} &\int \tilde{g}(\mu) l(\mu, \mu | \mathbf{x}, \mathbf{y}) d\tilde{\Omega}_1 \\ &= K_1 \int_0^\infty \exp\left(-(\lambda_1 + \lambda_2)\mu - \frac{\tau}{2} \left[n_1 (\bar{X} - \mu)^2 + n_2 (\bar{Y} - \mu)^2 \right]\right) d\mu, \end{aligned} \tag{2.47}$$

com $K_1 = \lambda_1 \lambda_2 k \tau^{(n_1+n_2)/2} \times \exp\left(-\frac{\tau}{2} \left[\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2 \right]\right)$.

Analogamente, o denominador do cociente que define $f_2(d)$ é igual a

$$\begin{aligned}
& \int \overset{\circ}{g}(\mu_1, \mu_2) d\Omega_2 \\
&= \lambda_1 \lambda_2 \int_0^\infty \exp(-\lambda_1 \mu_1) \int_0^{\mu_1} \exp(-\lambda_2 \mu_2) d\mu_2 d\mu_1 \\
&= \lambda_1 \lambda_2 \int_0^\infty \exp(-\lambda_1 \mu_1) \left[\frac{1}{\lambda_2} - \frac{1}{\lambda_2} \exp(-\lambda_2 \mu_1) \right] d\mu_1 \\
&= \lambda_1 \lambda_2 \left[\frac{1}{\lambda_1 \lambda_2} - \frac{1}{\lambda_2} \int_0^\infty \exp(-(\lambda_1 + \lambda_2) \mu_1) d\mu_1 \right] \\
&= \lambda_1 \left[\frac{1}{\lambda_1} - \frac{1}{(\lambda_1 + \lambda_2)} \right] \\
&= 1 - \frac{\lambda_1}{(\lambda_1 + \lambda_2)}.
\end{aligned}$$

O numerador do cociente que define $f_2(d)$ é

$$\begin{aligned}
& \int \overset{\circ}{g}(\mu_1, \mu_2) l(\mu_1, \mu_2 | \mathbf{x}, \mathbf{y}) d\Omega_2 \\
&= K_1 \int_0^\infty \int_0^{\mu_1} \exp\left(-\lambda_1 \mu_1 - \lambda_2 \mu_2 - \frac{\tau}{2} \left[n_1 (\bar{X} - \mu_1)^2 + n_2 (\bar{Y} - \mu_2)^2 \right] \right) d\mu_2 d\mu_1.
\end{aligned}$$

Considerando as distribuições a priori com valores esperados iguais às médias amostrais temos $\lambda_1 = 0.24 \simeq 1/4.1636$ e $\lambda_2 = 3.79 \simeq 1/0.2636$, implicando que o fator de Bayes é igual a 5.22 e portanto que a hipótese nula deve ser rejeitada, com maior evidência que ao considerar distribuições a priori normais truncadas.

Para analisar a sensibilidade do fator de Bayes à escolha dos hiperparâmetros λ_1 e λ_2 , apresentamos na Tabela 2.15 o fator de Bayes para diferentes valores de λ_1 e λ_2 .

Podemos observar que, fixado λ_1 , o fator de Bayes aumenta quando λ_2 aumenta. O contrario acontece se fixamos λ_2 . Em todos, os casos o fator de Bayes é maior que 1 levando à rejeição da hipótese nula. Para efeito de comparação com os fatores de Bayes obtidos sob distribuições a priori normais truncadas, consideramos distribuições a priori tais que $\lambda_1^2 = \lambda_2^2 = \tau$. Os resultados são apresentados na Tabela 2.16.

Tabela 2.15: Fatores de Bayes

$\lambda_2 \backslash \lambda_1$	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0.5	2.87	2.57	2.16	1.88	1.70	1.57	1.48	1.41	1.36	1.32
1.0	3.42	2.81	2.26	1.93	1.72	1.58	1.49	1.42	1.36	1.32
1.5	3.89	3.00	2.34	1.97	1.74	1.59	1.49	1.42	1.37	1.33
2.0	4.26	3.14	2.40	1.99	1.75	1.60	1.50	1.43	1.37	1.33
2.5	4.56	3.25	2.45	2.02	1.77	1.61	1.50	1.43	1.37	1.33
3.0	4.79	3.34	2.48	2.03	1.78	1.62	1.51	1.43	1.37	1.33
3.5	4.98	3.42	2.51	2.05	1.78	1.62	1.51	1.43	1.38	1.33
4.0	5.14	3.47	2.54	2.06	1.79	1.62	1.51	1.43	1.38	1.33
4.5	5.26	3.52	2.56	2.07	1.80	1.63	1.52	1.44	1.38	1.33
5.0	5.37	3.56	2.58	2.08	1.80	1.63	1.52	1.44	1.38	1.33

Tabela 2.16: Fator de Bayes para distribuições a priori exponenciais

τ	$B_{21} = R_{12}^{-1}$
0.35000	2.96
0.20000	2.77
0.04000	1.76
0.03500	1.68
0.00350	0.62
0.00035	0.21

Para o caso em que $\lambda_1^2 = \lambda_2^2 = 0.35$, (valor amostral da precisão), rejeitamos a hipótese nula. Os valores dos fatores de Bayes da Tabela 2.16 são maiores que os valores dos fatores de Bayes correspondentes da Tabela 2.12 para o caso de distribuições a priori informativas e menores para os casos de distribuições a priori não informativas. Notemos que o fator de Bayes é mais sensível à escolha dos hiperparâmetros ao considerar distribuições a priori exponenciais que distribuições a priori normais truncadas.

Uma comparação direta dos resultados obtidos pelas duas técnicas não é possível pois, com a primeira técnica são especificadas distribuições a priori diferentes e independentes sobre cada partição do espaço paramétrico, e com segunda técnica estas distribuições a priori surgem a partir da especificação da distribuição a priori sobre todo o espaço paramétrico. Apesar disto podemos observar que em todos os casos, inclusive ao considerar distribuições a priori não informativas, os fatores de Bayes calculados com a técnica dada por Carlin e Chib (1995) são menores que aqueles obtidos com a técnica dada por Irony e Pereira (1995). Também podemos observar como o fator de Bayes é menos sensível à escolha do hiperparâmetro de dispersão ao considerar as distribuições a priori normais truncadas sob a segunda técnica. A diferença, apesar de utilizar distribuições a priori que possam, sob certos critérios, ser consideradas similares, radica em como são obtidas as probabilidades a posteriori $P(M = j|d)$, $j = 1, 2$. Na técnica de Carlin e Chib (1995), são obtidas através de amostragem de Gibbs das distribuições condicionais completas e na técnica de Irony e Pereira (1995) são calculadas diretamente através das funções preditivas (2.26). A amostragem de Gibbs não pode ser implementada com distribuições a priori exponenciais como foi feito com a segunda metodologia, porque as distribuições condicionais completas não permitem a amostragem.

2.4 Extensão para o caso em que a variância é desconhecida

Nesta seção estendemos os resultados das seções anteriores para o caso em que a variância σ^2 é desconhecida. Analisamos a distribuição empírica do teste alternativo para diferentes estimadores dessa variância. Reanalisamos o Exemplo 1.1, sob o enfoque bayesiano, propondo também uma distribuição a priori para o parâmetro σ^2 .

Um estimador não viesado de σ^2 no modelo definido na Seção 2.1 é dado por

$$s_p^2 = \frac{(n_1 - 1) s_X^2 + (n_2 - 1) s_Y^2}{n_1 + n_2 - 2}, \quad (2.48)$$

em que

$$s_X^2 = \frac{\sum_{i=1}^{n_1} (X_i - \bar{X})^2}{n_1 - 1}, \quad s_Y^2 = \frac{\sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}{n_2 - 1}.$$

Pode-se mostrar (ver Searle, 1971, pag. 99) que

$$\frac{(n_1 + n_2 - 2) s_p^2}{\sigma^2} \sim \chi_{n_1 + n_2 - 2}^2,$$

e que sob a hipótese nula, $\bar{D}/s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ tem distribuição t-Student com $m = (n_1 + n_2 - 2)$ graus de liberdade quando as médias não são restritas.

Seja d_α a constante que determina o nível de significância α , ou seja,

$$\begin{aligned} P_{\mu_1=\mu_2=\mu} (\bar{D} > d_\alpha) &= P_{\mu_1=\mu_2=\mu} \left(\frac{\bar{D}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > \frac{d_\alpha}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right) \\ &= \alpha, \end{aligned}$$

o que implica

$$1 - \mathcal{T}_m \left(\frac{d_\alpha}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right) = \alpha,$$

em que $\mathcal{T}_m$ denota a função de distribuição t-Student com m graus de liberdade.

Conseqüentemente,

$$d_\alpha = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \mathcal{T}_m^{-1}(1 - \alpha). \quad (2.49)$$

Logo o teste de H_0 contra H_1 , em (2.1), corresponde a rejeitar a hipótese nula se

$$\overline{D} > d_\alpha, \quad (2.50)$$

ou seja, se

$$\frac{\overline{D}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > \mathcal{T}_m^{-1}(1 - \alpha).$$

O poder desse teste em $\mu_1 - \mu_2 = \delta > 0$ é

$$P_{\mu_1 - \mu_2 = \delta} \left(\frac{\overline{X} - \overline{Y} - \delta}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > \frac{d_\alpha - \delta}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right) = 1 - \mathcal{T}_m \left(\frac{d_\alpha - \delta}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right). \quad (2.51)$$

Analogamente ao caso em que a variância é conhecida, propomos um teste **alternativo** (ou **restrito**), para situações em que $\mu_1 \geq c$ e $\mu_2 \geq c$. Sem perda de generalidade, suponhamos $c = 0$. Sejam d_α^* a constante positiva que determina o nível de significância α , obtida de

$$P_{\mu_1 - \mu_2 = 0} (\widehat{D} > d_\alpha^*) = \alpha. \quad (2.52)$$

Por razões similares àquelas utilizadas para calcular (2.7), temos

$$\begin{aligned} P_{\mu_1 - \mu_2 = \delta} (\widehat{D} > d_\alpha^*) &= P_{\mu_1 - \mu_2 = \delta} (\overline{X} > d_\alpha^*, \overline{X} \geq 0, \overline{Y} \leq 0) \\ &\quad + P_{\mu_1 - \mu_2 = \delta} (\overline{X} - \overline{Y} > d_\alpha^*, \overline{X} \geq 0, \overline{Y} \geq 0). \end{aligned} \quad (2.53)$$

Calculemos cada termo do segundo membro de (2.53).

$$\begin{aligned} \text{i) } P_{\mu_1 - \mu_2 = \delta} (\overline{X} > d_\alpha^*, \overline{X} \geq 0, \overline{Y} \leq 0) &= P_{\mu_1 - \mu_2 = \delta} (\overline{X} > d_\alpha^*, \overline{Y} \leq 0) \\ &= \left[1 - \mathcal{T}_m \left(\frac{d_\alpha^* - \delta - \mu_2}{s_p / \sqrt{n_1}} \right) \right] \mathcal{T}_m \left(\frac{-\mu_2}{s_p / \sqrt{n_2}} \right). \end{aligned} \quad (2.54)$$

$$\begin{aligned}
\text{ii) } P_{\mu_1 - \mu_2 = \delta} (\bar{X} - \bar{Y} > d_\alpha^*, \bar{X} \geq 0, \bar{Y} \geq 0) &= P_{\mu_1 - \mu_2 = \delta} (\bar{X} - \bar{Y} > d_\alpha^*, \bar{Y} \geq 0) \\
&= P_{\mu_1 - \mu_2 = \delta} (\bar{X} - \bar{Y} > d_\alpha^*) \quad (2.55) \\
&\quad - P_{\mu_1 - \mu_2 = \delta} (\bar{X} - \bar{Y} > d_\alpha^*, \bar{Y} \leq 0).
\end{aligned}$$

Sejam

$$T_1 = \frac{\bar{X} - \mu_1}{s_p / \sqrt{n_1}}, \quad T_2 = \frac{\bar{Y} - \mu_2}{s_p / \sqrt{n_2}},$$

então $T_i \sim \mathcal{T}_m$ (ver por exemplo, Searle, 1971, pag. 107).

O último termo do segundo membro de (2.55) pode ser calculado como

$$\begin{aligned}
&P_{\mu_1 - \mu_2 = \delta} (\bar{X} - \bar{Y} > d_\alpha^*, \bar{Y} \leq 0) \\
&= P_{\mu_1 - \mu_2 = \delta} \left(\frac{\bar{X} - \mu_1}{s_p / \sqrt{n_1}} > \frac{d_\alpha^* - \mu_1 + \bar{Y}}{s_p / \sqrt{n_1}}, \bar{Y} \leq 0 \right) \\
&= P_{\mu_1 - \mu_2 = \delta} \left(T_1 > \frac{d_\alpha^* - (\mu_1 - \mu_2)}{s_p / \sqrt{n_1}} + \frac{\sqrt{n_1}}{\sqrt{n_2}} T_2, T_2 \leq \frac{-\mu_2}{s_p / \sqrt{n_2}} \right) \\
&= \int_{-\infty}^{\frac{-\mu_2}{s_p / \sqrt{n_2}}} \int_{\frac{d_\alpha^* - \delta}{s_p \sqrt{n_1}} + \frac{\sqrt{n_1}}{\sqrt{n_2}} t_2}^{\infty} f_T(t_1) f_T(t_2) dt_1 dt_2 \\
&= \int_{-\infty}^{\frac{-\mu_2}{s_p / \sqrt{n_2}}} \left(1 - \int_{-\infty}^{\frac{d_\alpha^* - \delta}{s_p \sqrt{n_1}} + \frac{\sqrt{n_1}}{\sqrt{n_2}} t_2} f_T(t_1) dt_1 \right) f_T(t_2) dt_2 \\
&= \mathcal{T}_m \left(\frac{-\mu_2}{s_p / \sqrt{n_2}} \right) - \int_{-\infty}^{\frac{-\mu_2}{s_p / \sqrt{n_2}}} \int_{-\infty}^{\frac{d_\alpha^* - \delta}{s_p \sqrt{n_1}} + \frac{\sqrt{n_1}}{\sqrt{n_2}} t_2} f_T(t_1) f_T(t_2) dt_1 dt_2 \\
&= \mathcal{T}_m \left(\frac{-\mu_2}{s_p / \sqrt{n_2}} \right) - \tilde{h}(\mu_2, \delta),
\end{aligned}$$

em que $f_T(\cdot)$ denota a função de densidade de probabilidades de uma função de distribuição t-Student com m graus de liberdade e

$$\begin{aligned}
\tilde{h}(\mu_2, \delta) &= \frac{1}{\sqrt{m} B\left(\frac{1}{2}, \frac{m}{2}\right)} \times \\
&\quad \int_{-\infty}^{\frac{-\mu_2}{s_p / \sqrt{n_2}}} \mathcal{T}_m \left(\frac{d_\alpha^* - \delta}{s_p / \sqrt{n_1}} + \frac{\sqrt{n_1}}{\sqrt{n_2}} t \right) \left(1 + \frac{t^2}{m+1} \right)^{-(m+1)/2} dt.
\end{aligned}$$

Logo, de (2.53)-(2.55), temos

$$\begin{aligned}
P_{\mu_1 - \mu_2 = \delta} \left(\widehat{D} > d_\alpha^* \right) &= 1 - \mathcal{T}_m \left(\frac{d_\alpha^* - \delta}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right) \\
&\quad - \mathcal{T}_m \left(\frac{d_\alpha^* - \delta - \mu_2}{s_p / \sqrt{n_1}} \right) \mathcal{T}_m \left(\frac{-\mu_2}{s_p / \sqrt{n_2}} \right) \\
&\quad + \tilde{h}(\mu_2, \delta). \tag{2.56}
\end{aligned}$$

Fixado o nível de significância α , segue de (2.52) que o valor d_α^* deverá ser calculado utilizando (2.56) considerando $\delta = 0$, isto é, assumindo verdadeira a hipótese nula.

Nesses resultados, o estimador s_p^2 não leva em conta o fato de o espaço paramétrico ser restrito. Como alternativa podemos substituí-lo pelo estimador de máxima verossimilhança obtido sob restrição, $\widehat{\sigma}_R^2$. Para obtê-lo, consideremos a função de verossimilhança (2.2) avaliada nos estimadores de máxima verossimilhança $\widehat{\mu}_1$ e $\widehat{\mu}_2$ de μ_1 e μ_2 respectivamente, no espaço paramétrico restrito, ou seja,

$$\begin{aligned}
f(\sigma^2 | \widehat{\mu}_1, \widehat{\mu}_2) &= \frac{1}{(2\pi)^{n_1+n_2/2} (\sigma^2)^{n_1+n_2/2}} \\
&\quad \times \exp \left(-\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n_1} (X_i - \widehat{\mu}_1)^2 + \sum_{i=1}^{n_2} (Y_i - \widehat{\mu}_2)^2 \right] \right), \tag{2.57}
\end{aligned}$$

cuja derivada é

$$\begin{aligned}
\frac{\partial f}{\partial \sigma^2} &= \frac{1}{(2\pi)^{n_1+n_2/2} \sigma^{n_1+n_2+1}} \times \exp \left(-\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n_1} (X_i - \widehat{\mu}_1)^2 + \sum_{i=1}^{n_2} (Y_i - \widehat{\mu}_2)^2 \right] \right) \\
&\quad \times \left[-(n_1 + n_2) + \frac{1}{\sigma^2} \left[\sum_{i=1}^{n_1} (X_i - \widehat{\mu}_1)^2 + \sum_{i=1}^{n_2} (Y_i - \widehat{\mu}_2)^2 \right] \right]. \tag{2.58}
\end{aligned}$$

A solução da equação obtida ao igualar (2.58) a zero é o estimador de máxima verossimilhança de σ^2 sob a restrição $\{\mu_1 \geq 0, \mu_2 \geq 0\}$ dado por

$$\widehat{\sigma}_R^2 = \frac{\sum_{i=1}^{n_1} (X_i - \widehat{\mu}_1)^2 + \sum_{i=1}^{n_2} (Y_i - \widehat{\mu}_2)^2}{n_1 + n_2}. \tag{2.59}$$

$\hat{\mu}_1 = \max \{0, \bar{X}\}$ e $\hat{\mu}_2 = \max \{0, \bar{Y}\}$ definindo

$$\begin{aligned} s_{Xr}^2 &= \begin{cases} n_1^{-1} S_X^2 & \text{se } \bar{X} \geq 0, \\ n_1^{-1} \sum_{i=1}^{n_1} X_i^2 & \text{se } \bar{X} < 0, \end{cases} \\ &= \begin{cases} n_1^{-1} S_X^2 & \text{se } \bar{X} \geq 0, \\ n_1^{-1} S_X^2 + \bar{X}^2 & \text{se } \bar{X} < 0, \end{cases} \end{aligned}$$

e

$$s_{Yr}^2 = \begin{cases} n_2^{-1} S_Y^2 & \text{se } \bar{Y} \geq 0, \\ n_2^{-1} S_Y^2 + \bar{Y}^2 & \text{se } \bar{Y} < 0, \end{cases}$$

com $S_X^2 = \sum_{i=1}^{n_1} (X_i - \bar{X})^2 = (n_1 - 1)^{-1} s_X^2$ e $S_Y^2 = \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2 = (n_2 - 1)^{-1} s_Y^2$ temos que

$$\hat{\sigma}_R^2 = \frac{n_1 s_{Xr}^2 + n_2 s_{Yr}^2}{n_1 + n_2}$$

Um estimador alternativo, que pondera os desvios com os mesmos pesos que s_p^2 é dado por

$$s_{pr}^2 = \frac{n_1 s_{Xr}^2 + n_2 s_{Yr}^2}{n_1 + n_2 - 2} \quad (2.60)$$

Os valores esperados de s_{Xr}^2 ou s_{Yr}^2 são obtidos por Heiny e Siddiqui (1970). Ambos os estimadores mencionados acima são viciados e os vícios dependem de μ .

Com a substituição de s_p^2 , por $\hat{\sigma}_R^2$ ou s_{pr}^2 , a distribuição t-Student de $\bar{D}/s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ considerada em (2.56) não pode mais ser utilizada. Como alternativa estudamos as distribuições empíricas da estatística de teste proposta em (2.56) e daquelas obtidas com a substituição de s_p^2 por $\hat{\sigma}_R^2$ ou s_{pr}^2 , através de um estudo de simulação.

Geramos $r = 10000$ pares de amostras aleatórias de tamanho n provenientes de uma distribuição normal com média μ e variância σ^2 . Para cada par de amostras calculamos

$$\frac{\hat{D}}{s_p \sqrt{\frac{2}{n}}}, \frac{\hat{D}}{s_{pr} \sqrt{\frac{2}{n}}} \text{ e } \frac{\hat{D}}{\hat{\sigma}_R \sqrt{\frac{2}{n}}},$$

e obtivemos os respectivos quantis amostrais de ordem p ($=0.10, 0.20, 0.25, 0.50, 0.75$,

Tabela 2.17: Valores do nível de significância empírico com variância desconhecida

	teste alternativo	teste alternativo	teste usual
μ_2	$\hat{\mu}_V$	$\hat{\mu}$	
0.0	0.042	0.030	0.048
0.1	0.052	0.041	0.055
0.2	0.056	0.040	0.052
0.3	0.056	0.043	0.044
0.4	0.057	0.048	0.048
0.5	0.063	0.059	0.052
1.0	0.054	0.053	0.047
1.5	0.045	0.045	0.045

0.90, 0.95, 0.975, 0.995, 0.9975, 0.999, 0.9995) para as diferentes combinações de $n = 5, 10$, $\mu = 0, 0.1, 0.2, 0.5$ e $\sigma^2 = 0.5, 1.0, 1.5, 2.0$. Estes quantis são apresentados nas Tabelas 2.25 a 2.28. O uso do estimador $\hat{\sigma}_R$ favorece a hipótese nula, já que em todos os casos, os quantis baseados no desvio padrão amostral s_p são menores ou iguais àqueles baseados no estimador de máxima verossimilhança $\hat{\sigma}_R$. Por outro lado, sob a hipótese nula há um menor número de rejeições com o teste baseado em s_{pr} do que com o teste baseado em s_p , pois em todas as situações, os quantis baseados no desvio padrão amostral s_p são maiores ou iguais àqueles baseados no estimador alternativo s_{pr} . Assim acreditamos que o teste proposto baseado em s_p , seja uma alternativa razoável, já que não favorece nem a hipótese nula, como a estatística baseada em $\hat{\sigma}_R$, nem a hipótese alternativa, como a estatística baseada em s_{pr} , além das vantagens distribucionais.

De maneira similar ao que foi considerado na Seção 2.2, estudamos o efeito da substituição do valor desconhecido μ na expressão (2.56) pelos estimadores $\hat{\mu}$ e $\hat{\mu}_V$, sobre o nível de significância e o poder do teste restrito. Com este intuito, realizamos um estudo de simulação com 1.000 pares de amostras de tamanho 5 das distribuições $N(\mu_1, 1)$ e $N(\mu_2, 1)$. Na Tabela 2.17 são apresentados os valores de α empírico para o teste alternativo além os valores de α empírico para o teste usual.

Observamos um desempenho melhor do estimador $\hat{\mu}_V$, para valores de μ_2 menores

que 0.2. Para valores de μ_2 iguais a 0.2, 0.3, 0.4 e 0.5 as taxas empíricas obtidas com o estimador do estimador $\hat{\mu}_V$ foram levemente maiores que o nível nominal.

Também estudamos a influência da substituição de μ por $\hat{\mu}$ e $\hat{\mu}_V$ sobre o poder do teste alternativo. Com este objetivo realizamos uma nova série de simulações. Para cada valor de δ ($= 1.00, 1.25, 1.50$) fixo, as amostras foram geradas considerando $\mu_1 = \mu_2 + \delta$, $\sigma = 1$ e $n = 5$, para valores de $\mu_2 = 0, 0.1, \dots, 0.5, 1.0, 1.5$. Em cada amostra obtiveram-se $\hat{D}$ e d_α^* de (2.56) com $\delta = 0$, utilizando $\hat{\mu}$ ou $\hat{\mu}_V$ no lugar de μ . O poder empírico é calculado como o número de amostras em que $\hat{D} > d_\alpha^*$ dividido pelo número total de amostras. Os resultados estão dispostos na Tabela 2.18.²

Na Tabela 2.18, podemos observar que o poder empírico do teste alternativo é maior ou igual que o poder empírico do teste usual em todos os casos. Além disso, o poder empírico do teste alternativo baseado em $\hat{\mu}_V$ é maior que o poder empírico do teste alternativo baseado $\hat{\mu}$ quando os valores de μ_2 estão próximos de zero. Como, sob esta alternativa a probabilidade empírica de erro de tipo I, é mais próxima ao nível nominal recomendamos o uso do estimador $\hat{\mu}_V$.

Para Exemplo 1.1, as estimativas são $\bar{X} = 94.2$, $s_X^2 = 32.63$, $\bar{Y} = 90.3$, $s_Y^2 = 30.40$, $s_p = 5.6$. Para $\alpha = 0.05$ e usando $\hat{\mu}_V = 92.2$, em (2.56), obtemos $d_\alpha^* = 3.6$. Para o teste usual, temos $d_\alpha = 4.1$. Dado que $\hat{D} = \bar{D} = \bar{X} - \bar{Y} = 3.9$, rejeitamos H_0 unicamente com o teste alternativo.

2.4.1 Enfoque Bayesiano

Nesta seção aplicaremos as duas metodologias apresentadas na Seção 2.3 para calcular o fator de Bayes correspondentes aos dados do Exemplo 1.1.

Inicialmente aplicamos a metodologia dada por Carlin e Chib (1995). As distribuições a priori para o parâmetro μ no caso $M = 1$, e μ_1 e μ_2 no caso $M = 2$, foram as mesmas que consideramos para o caso de variância conhecida. Para o parâmetro σ^2 uma distribuição gama inversa foi considerada como distribuição a priori, ou equivalentemente

²Valor nominal do teste usual para $\delta = 1.00$ é 0.394, para $\delta = 1.25$ é 0.545, para $\delta = 1.50$ é 0.689.

Tabela 2.18: Valores do poder empírico com variância desconhecida

$\delta = 1.00$			
	teste alternativo	teste alternativo	teste usual
μ_2	$\hat{\mu}_V$	$\hat{\mu}$	
0.0	0.467	0.452	0.432
0.1	0.465	0.454	0.433
0.2	0.450	0.442	0.397
0.3	0.466	0.465	0.428
0.4	0.431	0.428	0.400
0.5	0.451	0.451	0.421
1.0	0.415	0.415	0.410
1.5	0.423	0.423	0.423
$\delta = 1.25$			
	teste alternativo	teste alternativo	teste usual
μ_2	$\hat{\mu}_V$	$\hat{\mu}$	
0.0	0.605	0.595	0.553
0.1	0.590	0.574	0.546
0.2	0.634	0.630	0.591
0.3	0.577	0.573	0.537
0.4	0.569	0.569	0.541
0.5	0.596	0.596	0.565
1.0	0.577	0.577	0.573
1.5	0.568	0.568	0.563
$\delta = 1.50$			
	teste alternativo	teste alternativo	teste usual
μ_2	$\hat{\mu}_V$	$\hat{\mu}$	
0.0	0.728	0.726	0.684
0.1	0.723	0.719	0.679
0.2	0.736	0.736	0.706
0.3	0.699	0.699	0.671
0.4	0.746	0.746	0.709
0.5	0.721	0.721	0.704
1.0	0.717	0.717	0.714
1.5	0.662	0.662	0.662

Tabela 2.19: Diagnóstico de convergência de Geweke quando σ^2 é desconhecida

parâmetro	<i>Z - score</i>	
	cadeia 1	cadeia 2
μ	-0.214	1.180
μ_1	-0.780	-0.323
μ_2	-0.951	-1.110
$\tau = \sigma^{-1}$	-0.926	0.645

uma distribuição a priori gama com hiperparâmetros ν_1 e ν_2 para o parâmetro de precisão $\tau = 1/\sigma^2$. Consideramos dois casos: (i) uma distribuição gama com parâmetros $\nu_1 = \nu_2 = 0.03$ (correspondentes à estimativa de mínimos quadrados de τ) (ii) uma distribuição gama não informativa, com média igual à estimativa de mínimos quadrados de τ , ou seja, com $\nu_1 = 1$, e $\nu_2 = 33.3$.

Não consideramos uma distribuição “pseudo-priori” para σ^2 porque este é um parâmetro de ruído comum aos dois modelos; como os dados são informativos sobre σ^2 independentemente do valor de M , as distribuições a posteriori resultantes serão as mesmas (ver Carlin e Chib, 1995)

Como foi feito no caso da variância conhecida, para obter uma aproximação das “pseudo-prioris”, avaliamos cada modelo separadamente e usamos as distribuições a posteriori resultantes como distribuições “pseudo-prioris”.

Os valores do *Z - score* do critério de convergência de Geweke são apresentados na Tabela 2.19; como estes valores pertencem ao intervalo $[-2, 2]$ podemos considerar que existe convergência na cadeia 1 de tamanho amostral 500, associada à primeira distribuição a priori e na cadeia 2 de tamanho amostral 1000, associada à segunda distribuição a priori. Inicialmente foi considerada uma cadeia de tamanho 500 também para a segunda distribuição a priori, mas como não se verificou a convergência, consideramos uma cadeia de tamanho 1000.

A Tabela 2.20 mostra os fatores de Bayes para comparar os modelos 1 e 2 definidos em (2.18) e (2.19), correspondentes às duas diferentes distribuições a priori consideradas para cada modelo, $(\pi_j, j = 1, 2)$; em cada caso consideramos 500 iterações iniciais para

Tabela 2.20: Fator de Bayes quando σ^2 é desconhecida

Probabilidades a priori		Fatores de Bayes	
		Distribuição a Priori	Distribuição a Priori
π_1	π_2	Informativa	Não Informativa
0.05	0.95	1.70	1.26
0.25	0.75	1.75	1.76
0.50	0.50	1.72	1.69
0.75	0.25	1.67	1.66
0.95	0.05	1.79	1.74

Tabela 2.21: Valores empíricos a posteriori para o espaço paramétrico restrito quando σ^2 é desconhecida

Parâmetro	Distribuição a priori			
	Informativa		Não Informativa	
	Média	Desvio Padrão	Média	Desvio Padrão
μ	92.230	0.847	92.230	0.852
μ_1	94.190	1.138	94.190	1.147
μ_2	91.040	0.694	91.030	0.692
τ	0.003	0.001	0.032	0.009

a primeira distribuição a priori e 1000 iterações iniciais para a segunda distribuição a priori, e as atualizamos com 10000 iterações..

Em todos os casos o fator de Bayes sugere que existe evidência vaga contra a hipótese nula, ou, equivalentemente, que o modelo $M = 2$ é preferível (ver Tabela B.1 no Apêndice B).

A Tabela 2.21 apresenta os valores empíricos a posteriori da média e desvio padrão para cada parâmetro, quando consideramos $\pi_1 = \pi_2 = 0.5$.

Calculamos também o fator de Bayes no caso em que não impomos restrições sobre o espaço paramétrico, e obtivemos $B_{21} = 2.72$ e $B_{21} = 2.65$, para a primeira e segunda distribuições a priori consideradas, respectivamente; esses valores são maiores que os correspondes fatores de Bayes associados ao espaço paramétrico restrito ($B_{21} = 1.72$ e $B_{21} = 1.69$). Aqui também a incorporação da restrição diminui a evidência contra a hipótese nula.

Tabela 2.22: Valores empíricos a posteriori para o espaço paramétrico irrestrito quando σ^2 é desconhecida

Parâmetro	Distribuição a priori			
	Informativa		Não Informativa	
	Média	Desvio Padrão	Média	Desvio Padrão
μ	92.190	0.886	92.220	0.857
μ_1	94.160	1.157	94.160	1.168
μ_2	90.220	1.148	90.210	1.155
τ	0.003	0.001	0.032	0.009

A Tabela 2.22 apresenta médias e desvios padrões a posteriori sob o espaço paramétrico irrestrito. Notemos que estes valores para μ e μ_1 são similares aos apresentados na Tabela 2.21, que apresenta média e desvio padrão a posteriori sobre o espaço paramétrico restrito, ou seja, as amostragens das distribuições condicionais a posteriori de μ e μ_1 não foram influenciadas pelas restrições sobre as distribuições a priori. A distribuição condicional a posteriori de μ_2 não foi influenciada pelas restrições sobre as distribuições a priori, já que as médias a posteriori sob o espaço paramétrico sem restrições é diferente da média a posteriori sob o espaço paramétrico restrito. Isto também foi observado para o caso em que a variância é suposta conhecida (ver Tabela 2.9). As médias e desvios padrão a priori e posteriori de τ iguais aos obtidos sob o espaço paramétrico restrito para ambas distribuições a priori consideradas.

Aplicando a metodologia de Irony e Pereira (1995) para calcular o fator de Bayes, suponhamos que $\tau = \sigma^{-2}$ tem distribuição a priori gama de parâmetros ν_1 e ν_2 , ou seja, que sua função densidade é

$$f(\tau) = \frac{\nu_2^{\nu_1}}{\Gamma(\nu_1)} \tau^{\nu_1-1} \exp(-\tau\nu_2), \quad \tau > 0. \quad (2.61)$$

Se μ_1 e μ_2 têm, como anteriormente, distribuições a priori normais truncadas (2.29) e (2.30), a densidade conjunta a priori é

$$g'(\mu_1, \mu_2, \tau) = C_1 \exp\left(-\frac{\gamma_1}{2}(\mu - a_1)^2 - \frac{\gamma_2}{2}(\mu - a_2)^2\right) \tau^{\nu_1-1} \exp(-\tau\nu_2), \quad \tau > 0,$$

com $C_1 = C\nu_2^{\nu_1} \sqrt{\gamma_1 \gamma_2} (\Gamma(\nu_1) 2\pi)^{-1}$.

Agora as componentes da partição do espaço paramétrico são

$$\begin{aligned}\Omega_1 &= \{\varpi = (\mu_1, \mu_2, \tau) : \mu_1 \geq 0, \mu_2 \geq 0 \text{ e } \mu_1 = \mu_2, \tau > 0\}, \\ \Omega_2 &= \{\varpi = (\mu_1, \mu_2, \tau) : \mu_1 \geq 0, \mu_2 \geq 0 \text{ e } \mu_1 > \mu_2, \tau > 0\}.\end{aligned}$$

Analogamente a (2.34) e (2.37), considerando $\Omega'_1 = \{(\mu, \tau) : \mu \geq 0, \tau > 0\}$, temos

$$\begin{aligned}\int g'(\mu, \mu, \tau) d\Omega'_1 &= C \frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \frac{\sqrt{2\pi}}{\sqrt{\gamma_1 + \gamma_2}} \exp\left(-\frac{\gamma_1 \gamma_2 (a_1 - a_2)^2}{2(\gamma_1 + \gamma_2)}\right) \\ &\quad \times \Phi\left(\frac{(\gamma_1 a_1 + \gamma_2 a_2)}{(\gamma_1 + \gamma_2)} \sqrt{\gamma_1 + \gamma_2}\right) \\ \int g(\mu_1, \mu_2, \tau) d\Omega_2 &= C \left[\sqrt{\frac{\gamma_1}{2\pi}} \text{ef}(\gamma_1, a_1, \gamma_2, a_2) - \Phi(-a_2 \sqrt{\gamma_2}) \Phi(a_1 \sqrt{\gamma_1}) \right],\end{aligned}$$

$$\begin{aligned}&\int g'(\mu, \mu, \tau) l(\mu, \mu, \tau) d\Omega'_1 \\ &= C_1 k \int_0^\infty \exp\left(-\frac{\gamma_1}{2}(\mu - a_1)^2 - \frac{\gamma_2}{2}(\mu - a_2)^2\right) \\ &\quad \times \left\{ \int_0^\infty \tau^{\alpha_1-1} \times \exp\left(-\tau \left[\frac{1}{2} \left\{ \sum_{i=1}^{n_1} (X_i - \mu)^2 + \sum_{i=1}^{n_2} (Y_i - \mu)^2 \right\} + \nu_2 \right] \right) d\tau \right\} d\mu, \\ &= C_1 k \Gamma(\alpha_1) \int_0^\infty \exp\left(-\frac{\gamma_1}{2}(\mu - a_1)^2 - \frac{\gamma_2}{2}(\mu - a_2)^2\right) \\ &\quad \left[\frac{1}{2} \left\{ \sum_{i=1}^{n_1} (X_i - \mu)^2 + \sum_{i=1}^{n_2} (Y_i - \mu)^2 \right\} + \nu_2 \right]^{-\alpha_1} d\mu, \tag{2.62}\end{aligned}$$

com $\alpha_1 = \left(\frac{n_1+n_2+2\nu_1}{2}\right)$.

Tabela 2.23: Valores dos hiperparâmetros e os fatores de Bayes correspondentes quando σ^2 é desconhecida

					Fatores de Bayes	
$\tau_1 = \tau_2$	a_1	γ_1	a_2	γ_2	$\nu_1 = \nu_2 = 0.03$	$\nu_1 = 1$ e $\nu_2 = 33.3$
0.35000	4.1501	0.3298	-0.2207	0.3049	0.49	0.49
0.20000	4.1218	0.1652	-0.0864	0.1815	0.78	0.77
0.04000	4.0934	0.0283	0.1010	0.0384	1.21	1.20
0.03500	4.0946	0.0249	0.1198	0.0336	1.22	1.21
0.00350	4.1276	0.0030	0.2172	0.0034	1.61	1.60
0.00035	4.1500	0.0003	0.2490	0.0003	1.71	1.69

$$\begin{aligned}
& \int g(\mu_1, \mu_2, \tau) l(\mu_1, \mu_2, \tau) d\Omega_2 \\
&= C_1 k \int_0^\infty \int_0^{\mu_2} \exp\left(-\frac{\gamma_1}{2}(\mu - a_1)^2 - \frac{\gamma_2}{2}(\mu - a_2)^2\right) \\
&\quad \times \left\{ \int_0^\infty \tau^{\alpha_1-1} \exp\left[-\tau \left(\frac{1}{2} \left\{ \sum_{i=1}^{n_1} (X_i - \mu_1)^2 + \sum_{i=1}^{n_2} (Y_i - \mu_2)^2 \right\} + \nu_2\right)\right] d\tau \right\} d\mu_2 d\mu_1 \\
&= C_1 k \Gamma(\alpha_1) \\
&\quad \times \int_0^\infty \exp\left(-\frac{\gamma_1}{2}(\mu - a_1)^2 - \frac{\gamma_2}{2}(\mu - a_2)^2\right) \\
&\quad \left[\frac{1}{2} \left\{ \sum_{i=1}^{n_1} (X_i - \mu_1)^2 + \sum_{i=1}^{n_2} (Y_i - \mu_2)^2 \right\} + \nu_2 \right]^{-\alpha_1} d\mu_1 d\mu_2. \tag{2.63}
\end{aligned}$$

Se consideramos $\nu_1 = \nu_2 = 0.03$, e $\nu_1 = 1$ e $\nu_2 = 33.3$, e os valores dos hiperparâmetros apresentados na Tabela 2.12, os valores de Bayes correspondentes são apresentados na Tabela 2.23.

O fator de Bayes aumenta quando a precisão $\tau = \tau_1 = \tau_2$ decresce, o contrario acontece quando a variância é conhecida (ver Tabela 2.12). Rejeitamos a hipótese nula ao considerar distribuições a priori não informativas. Comparando os fatores de Bayes das Tabelas 2.12 e 2.23 para os mesmos valores dos hiperparâmetros observamos que os valores de Bayes são menores quando a variância é desconhecida.

Estudemos o comportamento do fator de Bayes supondo agora que não existem res-

Tabela 2.24: Fator de Bayes considerando distribuições a priori normais no espaço paramétrico irrestrito quando σ^2 é desconhecida

Fatores de Bayes		
$\dot{\tau}$	$\nu_1 = \nu_2 = 0.03$	$\nu_1 = 1$ e $\nu_2 = 33.3$
0.35000	0.35	0.35
0.20000	0.67	0.67
0.04000	1.30	1.26
0.03500	1.31	1.31
0.00350	1.63	1.62
0.00035	1.72	1.71

trições no espaço paramétrico. O espaço paramétrico irrestrito tem como partição os conjuntos (2.40) e (2.41) Considerando para μ_1 e μ_2 distribuições a priori normais com funções densidades (2.42) e (2.43), respectivamente.

Para efeitos de comparação, escolhemos como valores de $E(\mu_1) = \dot{a}_1 = \bar{X} - 90 = 4.1636$, $E(\mu_2) = \dot{a}_2 = \bar{Y} - 90 = 0.2636$ e $(Var(\mu_1))^{-1} = \dot{\gamma}_1 = \dot{\tau}_1 = (Var(\mu_2))^{-1} = \dot{\gamma}_2 = \dot{\tau}_2 = \dot{\tau}$. Na Tabela 2.24, que apresenta os fatores de Bayes correspondentes a diversos valores de $\dot{\tau}$, podemos observar que o fator de Bayes aumenta quando $\dot{\tau}$ diminui, sugerindo que quanto menos informativa a distribuição a priori, maiores são as evidências contra a hipótese nula, o contrario acontece quando a variância é considerada conhecida (ver Tabela 2.14).

Comparando os fatores de Bayes das Tabelas (2.23) e (2.14) para os mesmos valores dos hiperparâmetros de dispersão, os fatores de Bayes são menores quando o espaço paramétrico é irrestrito para $\dot{\tau} = 0.35$ e $\dot{\tau} = 0.20$, e próximos a 1.71 para $\dot{\tau} = 0.00035$.

2.5 Discussão

Embora exista uma vasta literatura sobre a controvérsia entre procedimentos bayesianos e não-bayesianos para testes de hipóteses, neste texto nós não pretendemos tratar disto em detalhes. Kass e Raftery (1995) descrevem vários aspectos da comparação entre o fator de Bayes e os testes freqüentistas discutidos na literatura. Entre eles podemos

destacar:

1. O fator de Bayes satisfaz o princípio da verossimilhança (ver, por exemplo, Berger e Wolpert, 1984).
2. Os testes freqüentistas tendem a rejeitar a hipótese nula quase sistematicamente para grandes amostras, enquanto o fator de Bayes não (ver Raftery, 1986, 1995).
3. Erradamente se compara o fator de Bayes com o o nível descritivo, pois de maneira equivocada espera-se que o nível descritivo seja similar à probabilidade a posteriori de que a hipótese nula seja correta. (ver Berger e Mortera, 1991 e Edwards et al, 1963).
4. O fator de Bayes é mais fácil de ser aplicado em modelos não encaixados (ver Cox 1961, 1962).

Existem outras técnicas para o cálculo do fator de Bayes ou mais especificamente das probabilidades a posteriori, se o tamanho da amostra é moderado ou grande o que geralmente torna os métodos numéricos disponíveis sejam ineficientes para calcular integrais intratáveis analiticamente. Estas técnicas utilizam aproximações assintóticas baseadas em o método de Laplace com suas variantes, através do critério de Schwarz (ver Schwarz, 1978, e Kass e Raftery, 1995). Outra solução que aparece na literatura é através de estimações amostrais da densidade a posteriori (ver DiCiccio, Kass, Raftery e Wasserman, 1997). No nosso caso lidamos com amostras de tamanho pequeno ou moderado em que as metodologias apresentadas puderam ser aplicadas.

O problema prático mais sério do fator de Bayes é a escolha das distribuições a priori. A teoria de inferência bayesiana subjetivista sugere que estas distribuições a priori sejam escolhidas representando alguma opinião sobre o parâmetro. O mais usual, sob o ponto de vista dos bayesianos objetivistas é a escolha de distribuições a priori “não informativas” em algum sentido, uma revisão de distribuições a priori não informativas é dada em Kass e Wasserman (1996). O problema é que com freqüência distribuições

a priori não informativas são impróprias causando a falta de interpretação do fator de Bayes (ver O'Hagan, 1995). As distribuições a priori consideradas neste trabalho foram todas próprias.

Existem na literatura propostas de fatores de Bayes modificados para solucionar este problema, como mencionamos na apresentação do enfoque bayesiano, mas podem ser instáveis para o caso de pequenas amostras (ver Berger e Pericchi, 1996).

Outra solução alternativa que aparece na literatura é considerar distribuições a priori realistas e informativas e calcular os limites superior e inferior do fator de Bayes nessa classe de distribuições a priori. Iwata (1994) propõe várias classes de distribuições a priori informativas e calcula os limites inferiores do fator de Bayes para cada uma delas para um teste de hipótese num modelo de regressão normal linear. Esses limites não podem ser calculados neste estudo porque não temos uma expressão analítica do fator de Bayes.

O teste alternativo (freqüentista) apresentado é útil para o caso de pequenas amostras, e mesmo que o ganho do poder não seja grande, sua importância radica no fato de poder obter conclusões diferentes daquelas obtidas pelo teste usual (que não leva em conta a restrição sobre espaço paramétrico). O teste alternativo tem a vantagem de não requerer a especificação de distribuições a priori que influenciam a decisão de rejeitar a hipótese nula, como foi observado nas análises de sensibilidade do fator de Bayes. Também observamos que, ao calcular o fator de Bayes com a metodologia dada por Carlin e Chib (1995) e com a metodologia dada por Irony e Pereira (1995) com determinadas distribuições a priori supondo o espaço paramétrico irrestrito, obtivemos valores maiores que sob o espaço paramétrico restrito, ou seja, que a ausência de restrição aumenta a evidência contra a hipótese nula, o que contradiz a intuição. Um estudo sobre o erro numérico associado a cada uma das metodologias empregadas poderia ser realizado para esclarecer este ponto. Com o mesmo objetivo se poderia considerar um maior número de distribuições a priori para tentar definir os limites superior e inferior do fator de Bayes. Esta última opção é mais trabalhosa de ser implementado com a metodologia proposta por Carlin e Chib (1995). Os resultados obtidos com as duas técnicas são

difíceis de serem comparados, pois mesmo com distribuições a priori similares, segundo algum critério especificado, a comparação das distribuições a posteriori marginais é difícil. Nas Figuras 2.7 e 2.8 apresentamos os gráficos das densidades a priori consideradas sob a hipótese nula e alternativa respectivamente, para cada uma das metodologias. Para a metodologia de Carlin e Chib (1995) os hiperparâmetros foram $\tau_\mu = 0.65$, $\tau_{\mu_1} = 0.35$, $\tau_{\mu_2} = 0.35$, e o fator de Bayes obtido foi de 1.65. Para a metodologia de Irony e Pereira (1995) os hiperparâmetros foram $\tau_1 = \tau_2 = 0.35$, $a_1 = 4.1501$, $\gamma_1 = 0.3298$, $a_2 = -0.2207$, $\gamma_2 = 0.3049$, e o fator de Bayes obtido foi de 2.81. Podemos observar que as distribuições a priori são diferentes apesar de que ambas provêm de distribuições a priori normais truncadas com as mesmas médias e desvios padrão. Na primeira metodologia são obtidas amostras das distribuições marginais e portanto a densidade a posteriori marginal estimada vai depender do critério de estimação escolhido (ver Silverman, 1986). Estas densidades a posteriori marginais estimadas podem ser comparadas com as obtidas com a segunda metodologia a partir da densidade a posteriori conjunta.

O teste alternativo e a segunda metodologia requerem sua implementação através de programas de cálculo específicos, que têm o mesmo grau de dificuldade.

Figura 2.7. Densidades a priori sob a hipótese nula: a) metodologia de Carlin e Chib, b) metodologia de Irony e Pereira

Figura 2.8. Densidades a priori sob a hipótese alternativa: a) metodologia de Carlin e Chib, b) metodologia de Irony e Pereira

Tabela 2.25: quantis amostrais para $n = 5$ e $\sigma = 0.5$

quantil(%)	$\mu = 0.0$			$\mu = 0.1$		
	teste baseado em			teste baseado em		
	s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
10	-1.760	-1.858	-1.967	-2.084	-2.185	-2.330
20	-1.442	-1.545	-1.612	-1.772	-1.849	-1.982
25	-1.362	-1.446	-1.523	-1.648	-1.717	-1.843
50	0.000	0.000	0.000	0.000	0.000	0.000
75	0.282	0.292	0.315	0.475	0.489	0.531
90	0.754	0.794	0.843	1.008	1.030	1.126
95	1.061	1.127	1.187	1.383	1.437	1.547
97.5	1.349	1.405	1.508	1.756	1.813	1.964
99	1.729	1.843	1.933	2.239	2.321	2.503
99.5	2.034	2.130	2.274	2.476	2.561	2.768
99.75	2.300	2.467	2.571	2.787	2.944	3.116
99.9	2.753	2.937	3.078	3.251	3.323	3.634
99.95	3.215	3.300	3.595	3.329	3.763	3.722
quantil(%)	$\mu = 0.2$			$\mu = 0.5$		
	teste baseado em			teste baseado em		
	s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
10	-2.462	-2.498	-2.752	-2.802	-2.809	-3.133
20	-2.100	-2.152	-2.348	-2.409	-2.410	-2.694
25	-1.989	-2.026	-2.224	-2.252	-2.252	-2.518
50	0.000	0.000	0.000	0.007	0.007	0.008
75	0.579	0.585	0.647	0.695	0.695	0.777
90	1.177	1.192	1.316	1.396	1.397	1.561
95	1.549	1.581	1.732	1.849	1.854	2.067
97.5	1.923	1.963	2.149	2.319	2.323	2.592
99	2.360	2.439	2.638	2.840	2.840	3.176
99.5	2.723	2.793	3.045	3.309	3.310	3.699
99.75	3.017	3.097	3.373	3.620	3.649	4.048
99.9	3.364	3.481	3.761	4.244	4.244	4.745
99.95	4.007	4.203	4.480	4.422	4.422	4.944

Tabela 2.26: quantis amostrais para $n = 5$ e $\sigma = 1.0$

quantil(%)	$\mu = 0.0$			$\mu = 0.1$		
	teste baseado em			teste baseado em		
	s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
10	-1.759	-1.853	-1.966	-1.912	-2.009	-2.138
20	-1.480	-1.557	-1.655	-1.569	-1.656	-1.754
25	-1.378	-1.461	-1.540	-1.483	-1.552	-1.658
50	0.000	0.000	0.000	0.000	0.000	0.000
75	0.277	0.290	0.310	0.371	0.382	0.415
90	0.748	0.781	0.836	0.869	0.900	0.972
95	1.048	1.106	1.172	1.200	1.245	1.342
97.5	1.380	1.447	1.542	1.502	1.565	1.679
99	1.752	1.837	1.959	1.867	1.955	2.088
99.5	2.030	2.151	2.270	2.152	2.280	2.406
99.75	2.219	2.504	2.481	2.369	2.532	2.649
99.9	2.595	2.899	2.902	2.957	3.017	3.306
99.95	2.808	3.374	3.139	3.097	3.246	3.463
quantil(%)	$\mu = 0.2$			$\mu = 0.5$		
	teste baseado em			teste baseado em		
	s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
10	-2.072	-2.156	-2.317	-2.521	-2.592	-2.818
20	-1.758	-1.833	-1.965	-2.084	-2.117	-2.330
25	-1.651	-1.730	-1.846	-1.964	-1.999	-2.196
50	0.000	0.000	0.000	0.000	0.000	0.000
75	0.463	0.474	0.517	0.647	0.649	0.723
90	0.998	1.024	1.116	1.258	1.267	1.406
95	1.319	1.365	1.475	1.664	1.684	1.861
97.5	1.702	1.755	1.903	2.070	2.108	2.314
99	2.127	2.210	2.378	2.433	2.536	2.721
99.5	2.407	2.488	2.691	2.747	2.798	3.071
99.75	2.689	2.780	3.006	3.036	3.155	3.394
99.9	3.110	3.190	3.477	3.504	3.504	3.918
99.95	3.517	3.533	3.932	4.340	4.790	4.852

Tabela 2.27: quantis amostrais para $n = 5$ e $\sigma = 1.5$

quantil(%)	$\mu = 0.0$			$\mu = 0.1$		
	teste baseado em			teste baseado em		
	s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
10	-1.699	-1.822	-1.899	-1.901	-1.999	-2.125
20	-1.430	-1.530	-1.599	-1.594	-1.703	-1.782
25	-1.350	-1.417	-1.510	-1.496	-1.596	-1.673
50	0.000	0.000	0.000	0.000	0.000	0.000
75	0.274	0.284	0.306	0.342	0.356	0.383
90	0.753	0.789	0.842	0.851	0.890	0.952
95	1.062	1.131	1.188	1.179	1.238	1.318
97.5	1.367	1.439	1.528	1.457	1.548	1.629
99	1.794	1.883	2.006	1.847	1.943	2.065
99.5	1.988	2.146	2.222	2.181	2.263	2.439
99.75	2.261	2.457	2.528	2.448	2.555	2.737
99.9	2.654	2.857	2.967	2.916	2.985	3.260
99.95	2.858	3.054	3.195	3.008	3.151	3.363
quantil(%)	$\mu = 0.2$			$\mu = 0.5$		
	teste baseado em			teste baseado em		
	s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
10	-2.025	-2.104	-2.264	-2.256	-2.322	-2.523
20	-1.642	-1.711	-1.835	-1.887	-1.926	-2.110
25	-1.535	-1.607	-1.717	-1.776	-1.832	-1.985
50	0.000	0.000	0.000	0.000	0.000	0.000
75	0.410	0.423	0.459	0.570	0.580	0.637
90	0.915	0.945	1.023	1.128	1.147	1.261
95	1.238	1.294	1.384	1.483	1.516	1.658
97.5	1.557	1.633	1.741	1.864	1.902	2.084
99	1.970	2.094	2.202	2.287	2.346	2.557
99.5	2.339	2.428	2.615	2.615	2.723	2.924
99.75	2.684	2.860	3.001	3.002	3.131	3.356
99.9	3.059	3.172	3.420	3.416	3.563	3.819
99.95	3.564	3.568	3.985	3.722	3.977	4.161

Tabela 2.28: quantis amostrais para $n = 5$ e $\sigma = 2.0$

quantil(%)	$\mu = 0.0$			$\mu = 0.1$		
	teste baseado em			teste baseado em		
	s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
10	-1.627	-1.750	-1.819	-1.948	-2.039	-2.178
20	-1.389	-1.486	-1.553	-1.577	-1.694	-1.763
25	-1.309	-1.394	-1.463	-1.489	-1.569	-1.664
50	0.000	0.000	0.000	0.000	0.000	0.000
75	0.278	0.290	0.311	0.315	0.328	0.352
90	0.751	0.790	0.840	0.821	0.850	0.918
95	1.077	1.131	1.204	1.118	1.168	1.249
97.5	1.355	1.449	1.514	1.429	1.482	1.598
99	1.765	1.884	1.973	1.823	1.895	2.039
99.5	2.026	2.181	2.265	2.132	2.208	2.383
99.75	2.304	2.409	2.576	2.396	2.543	2.679
99.9	2.753	2.808	3.078	2.735	2.910	3.058
99.95	3.049	3.268	3.409	3.013	3.013	3.369
quantil(%)	$\mu = 0.2$			$\mu = 0.5$		
	teste baseado em			teste baseado em		
	s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
10	-1.915	-2.046	-2.141	-2.166	-2.219	-2.422
20	-1.620	-1.718	-1.811	-1.859	-1.916	-2.078
25	-1.531	-1.605	-1.712	-1.739	-1.810	-1.944
50	0.000	0.000	0.000	0.000	0.000	0.000
75	0.373	0.387	0.417	0.497	0.503	0.555
90	0.888	0.923	0.993	1.018	1.044	1.138
95	1.209	1.259	1.352	1.391	1.430	1.556
97.5	1.532	1.571	1.712	1.704	1.757	1.905
99	1.932	2.039	2.160	2.084	2.174	2.330
99.5	2.207	2.362	2.467	2.442	2.472	2.730
99.75	2.604	2.673	2.911	2.674	2.768	2.989
99.9	3.036	3.226	3.395	3.288	3.600	3.676
99.95	3.391	3.552	3.791	3.566	3.708	3.987

Tabela 2.29: quantis amostrais para $n = 10$ e $\sigma = 0.5$

quantil(%)	$\mu = 0.0$			$\mu = 0.1$		
	teste baseado em			teste baseado em		
	s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
10	-1.631	-1.679	-1.720	-1.963	-1.981	-2.069
20	-1.395	-1.430	-1.471	-1.707	-1.722	-1.800
25	-1.303	-1.336	-1.373	-1.626	-1.649	-1.714
50	0.000	0.000	0.000	0.000	0.000	0.000
75	0.295	0.299	0.311	0.528	0.531	0.556
90	0.763	0.775	0.804	1.040	1.053	1.097
95	1.064	1.089	1.121	1.362	1.375	1.435
97.5	1.309	1.339	1.380	1.667	1.679	1.757
99	1.625	1.658	1.713	1.996	2.046	2.104
99.5	1.921	1.964	2.025	2.363	2.376	2.491
99.75	2.156	2.195	2.273	2.553	2.622	2.691
99.9	2.452	2.575	2.585	2.807	2.843	2.959
99.95	2.751	2.824	2.900	2.941	2.970	3.100
quantil(%)	$\mu = 0.2$			$\mu = 0.5$		
	teste baseado em			teste baseado em		
	s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
10	-2.236	-2.261	-2.357	-2.539	-2.539	-2.677
20	-1.979	-1.997	-2.086	-2.204	-2.204	-2.324
25	-1.906	-1.914	-2.009	-2.092	-2.092	-2.205
50	-0.003	-0.003	-0.003	0.004	0.004	0.005
75	0.632	0.634	0.666	0.701	0.701	0.739
90	1.207	1.210	1.272	1.338	1.338	1.410
95	1.555	1.561	1.639	1.746	1.746	1.841
97.5	1.888	1.896	1.990	2.091	2.091	2.205
99	2.273	2.285	2.396	2.500	2.500	2.635
99.5	2.522	2.528	2.658	2.852	2.852	3.006
99.75	2.765	2.766	2.915	3.178	3.178	3.350
99.9	2.992	3.053	3.154	3.474	3.474	3.662
99.95	3.388	3.474	3.571	3.597	3.597	3.791

Tabela 2.30: quantis amostrais para $n = 10$ e $\sigma = 1.0$

quantil(%)	$\mu = 0.0$			$\mu = 0.1$		
	teste baseado em			teste baseado em		
	s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
10	-1.580	-1.643	-1.665	-1.807	-1.858	-1.905
20	-1.358	-1.397	-1.432	-1.575	-1.599	-1.660
25	-1.284	-1.318	-1.353	-1.486	-1.516	-1.567
50	0.000	0.000	0.000	0.000	0.000	0.000
75	0.291	0.296	0.307	0.415	0.419	0.437
90	0.760	0.776	0.802	0.910	0.919	0.960
95	1.045	1.076	1.101	1.221	1.244	1.287
97.5	1.309	1.346	1.380	1.462	1.483	1.541
99	1.609	1.645	1.696	1.809	1.871	1.907
99.5	1.813	1.862	1.911	2.078	2.105	2.191
99.75	2.111	2.183	2.225	2.339	2.356	2.466
99.9	2.351	2.402	2.478	2.566	2.665	2.705
99.95	2.490	2.614	2.625	2.917	2.969	3.075
quantil(%)	$\mu = 0.2$			$\mu = 0.5$		
	teste baseado em			teste baseado em		
	s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
10	-2.000	-2.012	-2.108	-2.434	-2.443	-2.565
20	-1.725	-1.753	-1.818	-2.087	-2.094	-2.200
25	-1.643	-1.657	-1.732	-2.008	-2.015	-2.116
50	0.000	0.000	0.000	-0.008	-0.008	-0.008
75	0.524	0.528	0.553	0.651	0.651	0.687
90	1.054	1.060	1.111	1.264	1.266	1.332
95	1.371	1.387	1.446	1.627	1.627	1.715
97.5	1.662	1.680	1.752	1.943	1.945	2.049
99	1.983	2.017	2.090	2.344	2.351	2.471
99.5	2.255	2.277	2.377	2.574	2.579	2.713
99.75	2.438	2.493	2.570	2.904	2.904	3.061
99.9	2.804	2.804	2.956	3.346	3.346	3.527
99.95	2.996	3.030	3.158	3.680	3.681	3.879

Tabela 2.31: quantis amostrais para $n = 10$ e $\sigma = 1.5$

quantil(%)	$\mu = 0.0$			$\mu = 0.1$		
	teste baseado em			teste baseado em		
	s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
10	-1.613	-1.638	-1.701	-1.829	-1.865	-1.928
20	-1.361	-1.399	-1.435	-1.580	-1.610	-1.665
25	-1.279	-1.302	-1.349	-1.482	-1.508	-1.562
50	0.000	0.000	0.000	0.000	0.000	0.000
75	0.285	0.291	0.300	0.371	0.377	0.391
90	0.776	0.787	0.817	0.866	0.880	0.913
95	1.067	1.092	1.125	1.164	1.179	1.227
97.5	1.315	1.338	1.386	1.426	1.449	1.503
99	1.627	1.664	1.715	1.769	1.799	1.865
99.5	1.833	1.895	1.932	2.056	2.081	2.167
99.75	2.067	2.082	2.179	2.312	2.357	2.438
99.9	2.351	2.385	2.478	2.549	2.646	2.687
99.95	2.635	2.847	2.778	2.754	2.758	2.903
quantil(%)	$\mu = 0.2$			$\mu = 0.5$		
	teste baseado em			teste baseado em		
	s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
10	-1.872	-1.923	-1.973	-2.076	-2.100	-2.189
20	-1.625	-1.657	-1.713	-1.859	-1.870	-1.960
25	-1.545	-1.582	-1.629	-1.760	-1.772	-1.855
50	0.000	0.000	0.000	0.000	0.000	0.000
75	0.446	0.451	0.470	0.598	0.600	0.630
90	0.955	0.965	1.007	1.144	1.147	1.206
95	1.260	1.287	1.328	1.477	1.484	1.556
97.5	1.515	1.534	1.597	1.772	1.785	1.868
99	1.889	1.924	1.991	2.132	2.146	2.248
99.5	2.128	2.152	2.243	2.352	2.355	2.479
99.75	2.412	2.427	2.542	2.592	2.620	2.733
99.9	2.729	2.765	2.876	2.803	2.858	2.954
99.95	2.949	2.950	3.109	3.002	3.002	3.165

Capítulo 3

Comparação de parâmetros de regressão restritos num modelo normal linear sem intercepto.

Neste capítulo abordamos o problema de comparação de parâmetros de regressão num modelo normal linear sem intercepto, quando estão restritos a um intervalo fechado. A comparação de médias de uma população normal pode ser encarada como um caso particular. Trataremos o caso em que o espaço paramétrico é um intervalo fechado não homogêneo, e que portanto não é um cone. Na literatura não são apresentados testes de hipóteses com este tipo de restrições em modelos lineares.

3.1 Definição da estatística do teste e comparação do poder

Os modelos de regressão com parâmetros restritos são apropriados para vários problemas práticos, como aquele descrito no Exemplo 1.2. Consideremos então o modelo linear (1.1)

com a restrição $0 \leq \beta_i \leq 1, i = 1, 2$; estamos interessados em testar a hipótese

$$H_0 : \beta_1 = \beta_2 = \beta, \quad (3.1)$$

contra a alternativa

$$H_1 : \beta_1 > \beta_2. \quad (3.2)$$

Assumimos inicialmente que σ^2 é conhecida. Seja $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2)$ o estimador de mínimos quadrados de β sob o modelo de regressão linear (1.1), i.e.

$$\widehat{\beta}_i = \left(\sum_{k=1}^{n_i} x_{ik}^2 \right)^{-1} \sum_{k=1}^{n_i} x_{ik} y_{ik}. \quad (3.3)$$

Da teoria clássica de modelos lineares sabemos que $\widehat{\beta}_i$ tem distribuição normal com média β_i e variância

$$v_i = \sigma^2 \left(\sum_{k=1}^{n_i} x_{ik}^2 \right)^{-1}, \quad (3.4)$$

$i = 1, 2$, e que portanto $\widehat{\beta}_1 - \widehat{\beta}_2$, segue uma distribuição normal de média $\beta_1 - \beta_2$ e variância $v_1 + v_2$. Considerando o problema de testar (3.1) quando não existem restrições sobre o espaço paramétrico, i.e., quando $\beta_i \in \mathbb{R}, i = 1, 2$, a região de rejeição do teste **usual** (ou **irrestrito**) a um nível de significância α , é dada por $\overline{D} = \widehat{\beta}_1 - \widehat{\beta}_2 > D_\alpha$ em que

$$D_\alpha = \sqrt{v_1 + v_2} \Phi^{-1}(1 - \alpha).$$

O poder do teste usual é dado por

$$P_{\beta_1 - \beta_2 = \delta} \left(\frac{\overline{D} - \delta}{\sqrt{v_1 + v_2}} \geq \frac{D_\alpha - \delta}{\sqrt{v_1 + v_2}} \right) = 1 - \Phi \left(\frac{D_\alpha - \delta}{\sqrt{v_1 + v_2}} \right).$$

Seja $\widehat{\beta}_{ir}$ o estimador de máxima verossimilhança de β_i , sob o espaço paramétrico restrito,

i.e.,

$$\widehat{\beta}_{ir} = \begin{cases} \widehat{\beta}_i & \text{se } 0 \leq \widehat{\beta}_i \leq 1 \\ 0 & \text{se } \widehat{\beta}_i < 0 \\ 1 & \text{se } \widehat{\beta}_i > 1. \end{cases} \quad (3.5)$$

Assim como no Capítulo 2, propomos um teste **alternativo** (ou **restrito**) para o qual a região de rejeição de H_0 corresponde a $\widehat{D} > D_\alpha^*$, em que $\widehat{D} = \widehat{\beta}_{1r} - \widehat{\beta}_{2r}$ e D_α^* é uma constante positiva tal que

$$P_{\beta_1=\beta_2=\beta}(\widehat{D} \geq D_\alpha^*) = \alpha. \quad (3.6)$$

Como $|\widehat{D}| \leq 1$, então $0 < D_\alpha^* \leq 1$.

Para calcular D_α^* , notemos que

$$\begin{aligned} P_{\beta_1-\beta_2=\delta}(\widehat{D} \geq D_\alpha^*) &= P_{\beta_1-\beta_2=\delta}(\widehat{\beta}_1 \geq D_\alpha^*, 0 \leq \widehat{\beta}_1 \leq 1, \widehat{\beta}_2 < 0) \\ &\quad + P_{\beta_1-\beta_2=\delta}(\widehat{\beta}_1 - 1 \geq D_\alpha^*, 0 \leq \widehat{\beta}_1 \leq 1, \widehat{\beta}_2 > 1) \\ &\quad + P_{\beta_1-\beta_2=\delta}(-\widehat{\beta}_2 \geq D_\alpha^*, 0 \leq \widehat{\beta}_2 \leq 1, \widehat{\beta}_1 < a) \\ &\quad + P_{\beta_1-\beta_2=\delta}(1 - \widehat{\beta}_2 \geq D_\alpha^*, 0 \leq \widehat{\beta}_2 \leq 1, \widehat{\beta}_1 > 1) \\ &\quad + P_{\beta_1-\beta_2=\delta}(1 \geq D_\alpha^*, \widehat{\beta}_2 < 0, \widehat{\beta}_1 > 1) \\ &\quad + P_{\beta_1-\beta_2=\delta}(\widehat{\beta}_1 - \widehat{\beta}_2 \geq D_\alpha^*, 0 \leq \widehat{\beta}_1, \widehat{\beta}_2 \leq 1). \end{aligned} \quad (3.7)$$

O segundo termo do segundo membro de (3.7) é nulo pois não é possível que $\widehat{\beta}_1 \geq D_\alpha^* + 1$, e $\widehat{\beta}_1 \leq 1$. Também é nulo o terceiro termo pois como $D_\alpha^* > 0$, não é possível $-D_\alpha^* \geq \widehat{\beta}_2 \geq 0$. Logo (3.7) fica reduzida a

$$\begin{aligned} P_{\beta_1-\beta_2=\delta}(\widehat{D} \geq D_\alpha^*) &= P_{\beta_1-\beta_2=\delta}(\widehat{\beta}_1 \geq D_\alpha^*, 0 \leq \widehat{\beta}_1 \leq 1, \widehat{\beta}_2 < 0) \\ &\quad + P_{\beta_1-\beta_2=\delta}(1 - \widehat{\beta}_2 \geq D_\alpha^*, 0 \leq \widehat{\beta}_2 \leq 1, \widehat{\beta}_1 > 1) \\ &\quad + P_{\beta_1-\beta_2=\delta}(1 \geq D_\alpha^*, \widehat{\beta}_2 < 0, \widehat{\beta}_1 > 1) \\ &\quad + P_{\beta_1-\beta_2=\delta}(\widehat{\beta}_1 - \widehat{\beta}_2 \geq D_\alpha^*, 0 \leq \widehat{\beta}_1, \widehat{\beta}_2 \leq 1). \end{aligned} \quad (3.8)$$

Avaliemos cada termo do segundo membro de (3.8); o primeiro termo é dado por

$$\begin{aligned}
& P_{\beta_1 - \beta_2 = \delta} \left(\widehat{\beta}_1 \geq D_\alpha^*, 0 \leq \widehat{\beta}_1 \leq 1, \widehat{\beta}_2 < 0 \right) \\
&= P_{\beta_1 - \beta_2 = \delta} \left(1 \geq \widehat{\beta}_1 \geq D_\alpha^*, \widehat{\beta}_2 < 0 \right) \\
&= P_{\beta_1 - \beta_2 = \delta} \left(1 \geq \widehat{\beta}_1 \geq D_\alpha^* \right) P_{\beta_1 - \beta_2 = \delta} \left(\widehat{\beta}_2 < 0 \right) \\
&= P_{\beta_1 - \beta_2 = \delta} \left(1 \geq \widehat{\beta}_1 \geq D_\alpha^* \right) P_{\beta_1 - \beta_2 = \delta} \left(\widehat{\beta}_2 < 0 \right) \\
&= P_{\beta_1 - \beta_2 = \delta} \left(\frac{1 - \beta_1}{\sqrt{v_1}} \geq Z_1 \geq \frac{D_\alpha^* - \beta_1}{\sqrt{v_1}} \right) P_{\beta_1 - \beta_2 = \delta} \left(Z_2 < \frac{-\beta_2}{\sqrt{v_2}} \right) \\
&= \left(\Phi \left(\frac{1 - \beta_1}{\sqrt{v_1}} \right) - \Phi \left(\frac{D_\alpha^* - \beta_1}{\sqrt{v_1}} \right) \right) \Phi \left(\frac{-\beta_2}{\sqrt{v_2}} \right). \tag{3.9}
\end{aligned}$$

O segundo termo é

$$\begin{aligned}
& P_{\beta_1 - \beta_2 = \delta} \left(1 - \widehat{\beta}_2 \geq D_\alpha^*, 0 \leq \widehat{\beta}_2 \leq 1, \widehat{\beta}_1 > 1 \right) \tag{3.10} \\
&= P_{\beta_1 - \beta_2 = \delta} \left(0 \leq \widehat{\beta}_2 \leq 1 - D_\alpha^*, \widehat{\beta}_1 > 1 \right) \\
&= \left(\Phi \left(\frac{1 - D_\alpha^* - \beta_2}{\sqrt{v_2}} \right) - \Phi \left(\frac{-\beta_2}{\sqrt{v_2}} \right) \right) \left(1 - \Phi \left(\frac{1 - \beta_1}{\sqrt{v_1}} \right) \right).
\end{aligned}$$

Para o terceiro termo temos,

$$\begin{aligned}
P_{\beta_1 - \beta_2 = \delta} \left(1 \geq D_\alpha^*, \widehat{\beta}_2 < 0, \widehat{\beta}_1 > 1 \right) &= P_{\beta_1 - \beta_2 = \delta} \left(\widehat{\beta}_2 < 0 \right) P_{\beta_1 - \beta_2 = \delta} \left(\widehat{\beta}_1 > 1 \right) \\
&= \Phi \left(\frac{-\beta_2}{\sqrt{v_2}} \right) \left(1 - \Phi \left(\frac{1 - \beta_1}{\sqrt{v_1}} \right) \right). \tag{3.11}
\end{aligned}$$

Finalmente, o último termo do segundo membro de (3.8) é dado por

$$\begin{aligned}
P_{\beta_1 - \beta_2 = \delta} \left(\widehat{\beta}_1 - \widehat{\beta}_2 \geq D_\alpha^*, 0 \leq \widehat{\beta}_1, \widehat{\beta}_2 \leq 1 \right) &= P_{\beta_1 - \beta_2 = \delta} \left(\widehat{\beta}_1 - \widehat{\beta}_2 \geq D_\alpha^* \right) \tag{3.12} \\
&- P_{\beta_1 - \beta_2 = \delta} \left(\widehat{\beta}_1 - \widehat{\beta}_2 \geq D_\alpha^*, \widehat{\beta}_2 \leq 0 \right) \\
&- P_{\beta_1 - \beta_2 = \delta} \left(\widehat{\beta}_1 - \widehat{\beta}_2 \geq D_\alpha^*, \widehat{\beta}_2 \geq 1 - D_\alpha^* \right) \\
&- P_{\beta_1 - \beta_2 = \delta} \left(0 \leq \widehat{\beta}_2 \leq 1 - D_\alpha^*, \widehat{\beta}_1 \geq 1 \right)
\end{aligned}$$

Analisemos os termos do segundo membro de (3.12). Para o segundo termo podemos escrever

$$\begin{aligned}
& P_{\beta_1 - \beta_2 = \delta} \left(\widehat{\beta}_1 - \widehat{\beta}_2 \geq D_\alpha^*, \widehat{\beta}_2 \leq 0 \right) \\
&= \frac{1}{2\pi} \int_{-\infty}^{\frac{-\beta_2}{\sqrt{v_2}}} \int_{\frac{D_\alpha^* - \delta + z_2 \sqrt{v_2}}{\sqrt{v_1}}}^{\infty} \exp\left(-\frac{z_1^2}{2}\right) \exp\left(-\frac{z_2^2}{2}\right) dz_1 dz_2 \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{-\beta_2}{\sqrt{v_2}}} \left[1 - \Phi\left(\frac{D_\alpha^* - \delta + z_2 \sqrt{v_2}}{\sqrt{v_1}}\right) \right] \exp\left(-\frac{z_2^2}{2}\right) dz_2 \\
&= \Phi\left(\frac{-\beta_2}{\sqrt{v_2}}\right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{-\beta_2}{\sqrt{v_2}}} \Phi\left(\frac{D_\alpha^* - \delta + z_2 \sqrt{v_2}}{\sqrt{v_1}}\right) \exp\left(-\frac{z_2^2}{2}\right) dz_2.
\end{aligned}$$

O terceiro termo do segundo membro de (3.12) pode ser calculado como

$$\begin{aligned}
& P_{\beta_1 - \beta_2 = \delta} \left(\widehat{\beta}_1 - \widehat{\beta}_2 \geq D_\alpha^*, \widehat{\beta}_2 \geq 1 - D_\alpha^* \right) \\
&= \frac{1}{2\pi} \int_{\frac{1 - D_\alpha^* - \beta_2}{\sqrt{v_2}}}^{\infty} \int_{\frac{D_\alpha^* - \delta + z_2 \sqrt{v_2}}{\sqrt{v_1}}}^{\infty} \exp\left(-\frac{z_1^2}{2}\right) \exp\left(-\frac{z_2^2}{2}\right) dz_1 dz_2 \\
&= \frac{1}{\sqrt{2\pi}} \int_{\frac{1 - D_\alpha^* - \beta_2}{\sqrt{v_2}}}^{\infty} \left[1 - \Phi\left(\frac{D_\alpha^* - \delta + z_2 \sqrt{v_2}}{\sqrt{v_1}}\right) \right] \exp\left(-\frac{z_2^2}{2}\right) dz_2 \\
&= \left[1 - \Phi\left(\frac{1 - D_\alpha^* - \beta_2}{\sqrt{v_2}}\right) \right] - \frac{1}{\sqrt{2\pi}} \int_{\frac{1 - D_\alpha^* - \beta_2}{\sqrt{v_2}}}^{\infty} \Phi\left(\frac{D_\alpha^* - \delta + z_2 \sqrt{v_2}}{\sqrt{v_1}}\right) \exp\left(-\frac{z_2^2}{2}\right) dz_2.
\end{aligned}$$

Também temos que o último termo do segundo membro de (3.12) é dado por

$$\begin{aligned}
P_{\beta_1 - \beta_2 = \delta} \left(0 \leq \widehat{\beta}_2 \leq 1 - D_\alpha^*, \widehat{\beta}_1 \geq 1 \right) &= \left[\Phi\left(\frac{1 - D_\alpha^* - \beta_2}{\sqrt{v_2}}\right) - \Phi\left(\frac{-\beta_2}{\sqrt{v_2}}\right) \right] \\
&\quad \times \left[1 - \Phi\left(\frac{1 - \beta_1}{\sqrt{v_1}}\right) \right].
\end{aligned}$$

Logo, de (3.12), segue que:

$$\begin{aligned}
& P_{\beta_1 - \beta_2 = \delta} \left(\widehat{\beta}_1 - \widehat{\beta}_2 \geq D_\alpha^*, 0 \leq \widehat{\beta}_1, \widehat{\beta}_2 \leq 1 \right) \\
&= P_{\beta_1 - \beta_2 = \delta} \left(\widehat{\beta}_1 - \widehat{\beta}_2 \geq D_\alpha^* \right) + \Phi \left(\frac{1 - \beta_1}{\sqrt{v_1}} \right) \left[\Phi \left(\frac{1 - D_\alpha^* - \beta_1}{\sqrt{v_2}} \right) - \Phi \left(\frac{-\beta_1}{\sqrt{v_2}} \right) \right] \\
&\quad + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{-\beta_2}{\sqrt{v_2}}} \Phi \left(\frac{D_\alpha^* - \delta + z_2 \sqrt{v_2}}{\sqrt{v_1}} \right) \exp \left(-\frac{z_2^2}{2} \right) dz_2 \\
&\quad + \frac{1}{\sqrt{2\pi}} \int_{\frac{1 - D_\alpha^* - \beta_2}{\sqrt{v_2}}}^{\infty} \Phi \left(\frac{D_\alpha^* - \delta + z_2 \sqrt{v_2}}{\sqrt{v_1}} \right) \exp \left(-\frac{z_2^2}{2} \right) dz_2 - 1.
\end{aligned}$$

De (3.7) a (3.12), obtemos

$$\begin{aligned}
& P_{\beta_1 - \beta_2 = \delta} \left(\widehat{D} \geq D_\alpha^* \right) \\
&= \left(\Phi \left(\frac{1 - \beta_1}{\sqrt{v_1}} \right) - \Phi \left(\frac{D_\alpha^* - \beta_1}{\sqrt{v_1}} \right) \right) \Phi \left(\frac{-\beta_2}{\sqrt{v_2}} \right) \\
&\quad + \left(\Phi \left(\frac{1 - D_\alpha^* - \beta_2}{\sqrt{v_2}} \right) - \Phi \left(\frac{-\beta_2}{\sqrt{v_2}} \right) \right) \left(1 - \Phi \left(\frac{1 - \beta_1}{\sqrt{v_1}} \right) \right) \\
&\quad + P_{\beta_1 - \beta_2 = \delta} \left(\widehat{\beta}_1 - \widehat{\beta}_2 \geq D_\alpha^* \right) + \Phi \left(\frac{1 - \beta_1}{\sqrt{v_1}} \right) \left[\Phi \left(\frac{1 - D_\alpha^* - \beta_2}{\sqrt{v_2}} \right) - \Phi \left(\frac{-\beta_2}{\sqrt{v_2}} \right) \right] \\
&\quad + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{-\beta_2}{\sqrt{v_2}}} \Phi \left(\frac{D_\alpha^* - \delta + z_2 \sqrt{v_2}}{\sqrt{v_1}} \right) \exp \left(-\frac{z_2^2}{2} \right) dz_2 \\
&\quad + \frac{1}{\sqrt{2\pi}} \int_{\frac{1 - D_\alpha^* - \beta_2}{\sqrt{v_2}}}^{\infty} \Phi \left(\frac{D_\alpha^* - \delta + z_2 \sqrt{v_2}}{\sqrt{v_1}} \right) \exp \left(-\frac{z_2^2}{2} \right) dz_2 - 1 \\
&\quad + \Phi \left(\frac{-\beta_2}{\sqrt{v_2}} \right) \left(1 - \Phi \left(\frac{1 - \beta_1}{\sqrt{v_1}} \right) \right) \\
&= \Phi \left(\frac{1 - D_\alpha^* - \beta_2}{\sqrt{v_2}} \right) - \Phi \left(\frac{D_\alpha^* - \beta_1}{\sqrt{v_1}} \right) \Phi \left(\frac{-\beta_2}{\sqrt{v_2}} \right) \\
&\quad + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{-\beta_2}{\sqrt{v_2}}} \Phi \left(\frac{D_\alpha^* - \delta + z_2 \sqrt{v_2}}{\sqrt{v_1}} \right) \exp \left(-\frac{z_2^2}{2} \right) dz_2 \\
&\quad + \frac{1}{\sqrt{2\pi}} \int_{\frac{1 - D_\alpha^* - \beta_2}{\sqrt{v_2}}}^{\infty} \Phi \left(\frac{D_\alpha^* - \delta + z_2 \sqrt{v_2}}{\sqrt{v_1}} \right) \exp \left(-\frac{z_2^2}{2} \right) dz_2 \\
&\quad + P_{\beta_1 - \beta_2 = \delta} \left(\widehat{\beta}_1 - \widehat{\beta}_2 \geq D_\alpha^* \right) - 1.
\end{aligned}$$

Assim, o poder associado ao teste alternativo é dado por

$$\begin{aligned}
P_{\beta_1=\beta_2=\delta} \left(\widehat{D} \geq D_\alpha^* \right) &= \Phi \left(\frac{1 - D_\alpha^* - \beta_2}{\sqrt{v_2}} \right) - \Phi \left(\frac{D_\alpha^* - \beta_1}{\sqrt{v_1}} \right) \Phi \left(\frac{-\beta_2}{\sqrt{v_2}} \right) \\
&+ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{-\beta_2}{\sqrt{v_2}}} \Phi \left(\frac{D_\alpha^* - \delta + z\sqrt{v_2}}{\sqrt{v_1}} \right) \exp \left(-\frac{z^2}{2} \right) dz \\
&+ \frac{1}{\sqrt{2\pi}} \int_{\frac{1-D_\alpha^*-\beta_2}{\sqrt{v_2}}}^{\infty} \Phi \left(\frac{D_\alpha^* - \delta + z_2\sqrt{v_2}}{\sqrt{v_1}} \right) \exp \left(-\frac{z_2^2}{2} \right) dz_2 \\
&- \Phi \left(\frac{D_\alpha^* - \delta}{\sqrt{v_1 + v_2}} \right). \tag{3.13}
\end{aligned}$$

Para calcular D_α^* , devemos resolver a equação (3.6), i.e., ou equivalentemente a equação obtida ao se igualar (3.13) a α com $\delta = 0$. Na prática o valor do parâmetro $\beta = \beta_1 = \beta_2$, sob a hipótese nula é desconhecido. Como já foi mencionado no Capítulo 2, Cox and Hinkley (1974) sugerem testes condicionais, mas não dispomos, neste caso, de uma estatística completa e suficiente requerida para isso. Seguindo outra alternativa sugerida por esses autores, propomos substituir β por um estimador. Neste caso um estimador apropriado é

$$\widehat{\beta}_r = \begin{cases} \widehat{\beta}_0 & \text{se } 0 \leq \widehat{\beta}_0 \leq 1 \\ 0 & \text{se } \widehat{\beta}_0 \leq 0 \\ 1 & \text{se } \widehat{\beta}_0 \geq 1, \end{cases} \tag{3.14}$$

onde $\widehat{\beta}_0 = \left(\sum_{i=1}^2 \sum_{k=1}^{n_i} x_{ik}^2 \right)^{-1} \sum_{i=1}^2 \sum_{k=1}^{n_i} x_{ik} y_{ik}$.

Na próxima seção, avaliaremos se $\widehat{\beta}_r$ mantém a probabilidade de erro de tipo I controlada, ou seja, se

$$P_{\beta_1=\beta_2=\widehat{\beta}_r} \left(\widehat{D} \geq D_\alpha^* \right) \simeq \alpha, \tag{3.15}$$

e compararemos os poderes de ambos os testes.

Com o intuito de comparar os poderes do teste usual e do teste alternativo consideramos uma análise gráfica semelhante àquela realizada no Capítulo 2, pois o número de parâmetros envolvidos é grande dificulta fazer uma comparação analítica.

Tomamos $\alpha = 0.05$ sem perda de generalidade, e construímos gráficos do poder de ambos testes para diferentes valores i) do parâmetro comum β sob a hipótese nula, ii) da diferença δ entre os dois parâmetros sob a hipótese alternativa iii) o menor destes parâmetros (β_2). Para calcular D_α^* em (3.6) utilizamos o método de Brent e a regra de Simpson (ver, por exemplo, Press et al, 1994). Os resultados são apresentados nas Figuras 3.1 and 3.2. A escolha de v_1 e v_2 foi baseada nos dados do Exemplo 1.2 apresentados na Tabela 1.2. A análise sugere que o poder do teste alternativo aumenta se: i) o valor comum do parâmetro de regressão sob a hipótese nula (β) utilizado para calcular o valor crítico (D_α^*) está perto da fronteira; ii) a distância (δ) entre os parâmetros de regressão sob a hipótese alternativa aumenta. Assim podemos concluir que, quando a constante β utilizada para calcular o valor crítico (D_α^*) está situada próxima dos extremos do intervalo que define o espaço paramétrico, o teste alternativo é mais poderoso que o teste usual; de outra forma os poderes de ambos testes são iguais, exceto quando o menor dos parâmetros está perto da fronteira.

Figura 3.1. Poder do teste restrito (alternativo) e do teste irrestrito (usual) para

- poder do teste usual Z
- + — poder do teste alternativo para $\beta = 0.010$
- x — poder do teste alternativo para $\beta = 0.025$
- * — poder do teste alternativo para $\beta = 0.050$
- □ — poder do teste alternativo para $\beta = 0.075$
- o — poder do teste alternativo para $\beta = 0.100$

Figura 3.1. Poder do teste restrito (alternativo) e do teste irrestrito (usual) para

$\nu_1 = \nu_2 = 0.001$ e $\beta \leq 0.1$ (continuação)

- poder do teste usual Z
- + - poder do teste alternativo para $\beta = 0.010$
- x - poder do teste alternativo para $\beta = 0.025$
- * - poder do teste alternativo para $\beta = 0.050$
- □ - poder do teste alternativo para $\beta = 0.075$
- o - poder do teste alternativo para $\beta = 0.100$

Figura 3.2. Poder do teste restrito (alternativo) e do teste irrestrito (usual) para

- poder do teste usual Z
- + - poder do teste alternativo para $\beta = 1.000$
- × - poder do teste alternativo para $\beta = 0.990$
- * - poder do teste alternativo para $\beta = 0.975$
- □ - poder do teste alternativo para $\beta = 0.950$
- + · poder do teste alternativo para $\beta = 0.925$
- o - poder do teste alternativo para $\beta = 0.900$

Figura 3.2. Poder do teste restrito (alternativo) e do teste irrestrito (usual) para

$\nu_1 = \nu_2 = 0.001$ e $\beta \geq 0.900$ (continuação)

- poder do teste usual Z
- + - poder do teste alternativo para $\beta = 1.000$
- × - poder do teste alternativo para $\beta = 0.990$
- * - poder do teste alternativo para $\beta = 0.975$
- □ - poder do teste alternativo para $\beta = 0.950$
- + · poder do teste alternativo para $\beta = 0.925$
- o - poder do teste alternativo para $\beta = 0.900$

3.2 Desempenho do teste alternativo baseado no estimador do parâmetro de regressão comum

Com o propósito de estudar o efeito da substituição de β por um estimador na expressão (3.13), utilizada para definir a região crítica, sobre o nível de significância e o poder do teste alternativo, consideramos o estimador $\hat{\beta}_r$ (3.14).

Notemos que neste caso, o estimador de β depende da matriz de variáveis explicativas; por esta razão, no estudo de simulação que realizamos foram consideradas 4 matrizes diferentes cujos elementos são possíveis valores da variável explicativa. As quatro matrizes de especificação tem a forma $\mathbf{X}_k = \begin{bmatrix} \mathbf{x}_{k1}^t & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{k1}^t \end{bmatrix}$ em que $\mathbf{x}_{k1}$ e $\mathbf{x}_{k1}$ são vetores de dimensão 8, para $k = 1, \dots, 4$. Os valores utilizados foram

$$\begin{aligned} \mathbf{x}_{11}^t &= (0.620, 0.700, 0.900, 0.950, 1.200, 0.100, 0.570, 1.350)^t \\ \mathbf{x}_{12}^t &= (0.450, 1.150, 1.270, 0.870, 0.820, 0.970, 0.500, 0.730)^t, \end{aligned} \quad (3.16)$$

$$\begin{aligned} \mathbf{x}_{21}^t &= (0.794, 1.005, 0.824, 0.516, 0.234, 0.088, 0.452, 1.231)^t \\ \mathbf{x}_{22}^t &= (0.864, 1.158, 0.748, 1.451, 0.789, 0.366, 1.137, 1.034)^t, \end{aligned} \quad (3.17)$$

$$\begin{aligned} \mathbf{x}_{31}^t &= (1.298, 1.401, 1.211, 0.807, 1.078, 0.983, 1.371, 1.392)^t, \\ \mathbf{x}_{32}^t &= (1.362, 0.859, 1.055, 1.335, 1.015, 0.909, 1.628, 1.278)^t, \end{aligned} \quad (3.18)$$

$$\begin{aligned} \mathbf{x}_{41}^t &= (0.444, 0, 0.304, 0.761, 0.286, 0.550, 0.220, 0.774)^t, \\ \mathbf{x}_{42}^t &= (0.421, 0.021, 0.287, 0.545, 0.669, 0.250, 0.761, 0.467)^t. \end{aligned} \quad (3.19)$$

Os elementos dos vetores $\mathbf{x}_{11}$ e $\mathbf{x}_{12}$ correspondem aos valores observados da variável

explicativa indicados na Tabela 1.2. Os elementos dos demais vetores correspondem a amostras aleatórias de tamanho 8 de distribuições normais. Os elementos do vetor $\mathbf{x}_{21}$ correspondem a amostras aleatórias de uma distribuição normal com média $\bar{x}_{11} = \frac{1}{8} \sum_{i=1}^8 x_{11i} = 0.799$ e variância $s_{11}^2 = \frac{1}{7} \sum_{i=1}^8 (x_{11i} - \bar{x}_{11})^2 = (0.393)^2$. Por outro lado, os elementos do vetor $\mathbf{x}_{22}$ correspondem a amostras aleatórias de uma distribuição normal com média $\bar{x}_{12} = \frac{1}{8} \sum_{i=1}^8 x_{12i} = 0.845$ e variância $s_{12}^2 = \frac{1}{7} \sum_{i=1}^8 (x_{12i} - \bar{x}_{12})^2 = (0.288)^2$. Os elementos do vetor $\mathbf{x}_{31}$ correspondem a amostras aleatórias de uma distribuição normal com média $\bar{x}_{11} + s_{11} = 0.799 + 0.393 = 1.192$ e variância s_{11}^2 ; e os elementos do vetor $\mathbf{x}_{32}$ correspondem a amostras aleatórias de uma distribuição normal com média $\bar{x}_{12} + s_{12} = 0.845 + 0.288 = 1.133$ e variância s_{12}^2 . Por último, os elementos do vetor $\mathbf{x}_{41}$ correspondem a amostras aleatórias de uma distribuição normal com média $\bar{x}_{11} - s_{11} = 0.406$ e variância s_{11}^2 ; e os elementos do vetor $\mathbf{x}_{42}$ correspondem a amostras aleatórias de uma distribuição normal com média $\bar{x}_{12} - s_{12} = 0.557$ e variância s_{12}^2 .

Para cada vetor de variáveis explicativas foram geradas 1000 amostras de valores da variável resposta com distribuição normal multivariada com vetor de médias $\mathbf{x}_{i,j}\beta$ e matriz de covariância $\sigma^2 \mathbf{I}_8$, com $\sigma^2 = (0.13)^2 = 0.0169$ (0.13 é o valor médio das estimativas de mínimos quadrados de σ sob a hipótese nula e sob a hipótese alternativa) para $\beta = 0.70, 0.80, 0.85, 0.90, 0.925, 0.950, 0.975, 0.99$, e 1. Em cada par de amostras foram calculados $\hat{D}$ e D_α^* através da fórmula (3.13) utilizando $\hat{\beta}_r$ definido em (3.14). O valor da taxa de rejeição (valor de α empírico), que é a estimativa da probabilidade de erro de tipo I, é calculado como o número de amostras em que $\hat{D} > D_\alpha^*$ dividido pelo número total de amostras. Os valores obtidos estão apresentados na Tabela 3.1.

Em todos os casos considerados podemos observar que para valores próximos ao extremo do intervalo, o valor de α empírico para o teste alternativo é menor que o nível de significância, o que leva a uma diminuição do seu poder relativamente ao poder nominal.

Também estudamos a influência da substituição de β pelo estimador $\hat{\beta}_r$ (3.14) na expressão (3.13) sobre o poder do teste alternativo. Com este objetivo realizamos uma nova série de simulações. Para cada valor de δ fixo, as amostras foram geradas considerando

Tabela 3.1: Valores empíricos do nível de significância do teste alternativo para diferentes matrizes da variável explicativa, com variância conhecida

β	$\mathbf{X}_1$		$\mathbf{X}_2$	
	teste alternativo	teste usual	teste alternativo	teste usual
0.700	0.051	0.051	0.054	0.054
0.800	0.042	0.042	0.045	0.044
0.850	0.054	0.053	0.056	0.052
0.900	0.052	0.053	0.058	0.054
0.925	0.043	0.055	0.051	0.056
0.950	0.047	0.058	0.044	0.053
0.975	0.023	0.066	0.028	0.036
0.990	0.018	0.052	0.022	0.056
1.000	0.007	0.051	0.012	0.043

β	$\mathbf{X}_3$		$\mathbf{X}_4$	
	teste alternativo	teste usual	teste alternativo	teste usual
0.700	0.053	0.053	0.047	0.047
0.800	0.050	0.050	0.049	0.044
0.850	0.041	0.041	0.054	0.045
0.900	0.049	0.044	0.054	0.055
0.925	0.058	0.052	0.049	0.054
0.950	0.045	0.045	0.046	0.061
0.975	0.028	0.033	0.028	0.056
0.990	0.024	0.041	0.029	0.047
1.000	0.028	0.059	0.022	0.050

$\beta_2 = \beta_1 - \delta$ e $\sigma = 0.13$ para $\beta_1 = 0.90, 0.925, 0.950, 0.975, 0.990$ e 1 . Se obtiveram, em cada amostra $\hat{D}$ e D_α^* de (3.13) com $\delta = 0$, utilizando o estimador $\hat{\beta}_r$. O poder empírico é calculado como o número de amostras em que $\hat{D} > D_\alpha^*$ dividido pelo número total de amostras. Os resultados estão dispostos nas Tabelas 3.2 e 3.3

Podemos observar que o poder do teste alternativo é maior ou igual ao poder do teste usual, à exceção dos valores indicados com asterisco, onde o valor de β_1 está próximo ao extremo do intervalo; isto já era esperado, dado que a probabilidade empírica de erro de tipo I se mostrou menor que o nível nominal para estes valores.

As estimativas de mínimos quadrados do Exemplo 1.2, são apresentados na Tabela (3.4).

Assim considerando em (3.13), $v_1 = v_2 = 0.0027$, com v_1 e v_2 definidas em (3.4), obtemos $D_\alpha = 0.121$ e $D_\alpha^* = 0.119$, como $\hat{D} = \hat{\beta}_1 - \hat{\beta}_2 = 0.126$, não existem evidências para rejeitar a hipótese nula, em ambos testes. Mas se considerássemos três casas decimais nas estimativas de mínimos quadrados apresentadas na Tabela (3.4) e também em $v_1 = v_2 = 0.003$, obtemos $D_\alpha = 0.127$ e $D_\alpha^* = 0.125$, como $\hat{D} = \hat{\beta}_1 - \hat{\beta}_2 = 0.126$, rejeitaríamos só com o teste alternativo.

As mesmas conclusões obtidas no Capítulo 2 podem ser estendidas ao problema de comparação de parâmetros de regressão. O uso do estimador $\hat{\beta}_r$ definido em (3.14) traz como consequência, em alguns casos, uma diminuição do poder do teste alternativo para valores de β_1 próximos ao extremo do intervalo. O desempenho de outros estimadores, como aqueles propostos por Toutenburg e Srivastava (1996), pode ser estudado.

Tabela 3.2: Valores do poder empírico do teste alternativo com variância conhecida

	X_1		X_2	
$\delta = 0.1$				
β	teste alternativo	teste usual	teste alternativo	teste usual
0.900	0.399	0.396	0.357	0.354
0.925	0.403	0.395	0.405	0.387
0.950	0.418	0.403	0.383	0.367
0.975	0.421	0.403	0.382	0.362
0.990	0.397	0.395	0.357	0.363*
1.000	0.386	0.400*	0.331	0.341*
$\delta = 0.125$				
β	teste alternativo	teste usual	teste alternativo	teste usual
0.900	0.499	0.498	0.485	0.466
0.925	0.511	0.504	0.454	0.445
0.950	0.504	0.489	0.455	0.440
0.975	0.525	0.504	0.482	0.461
0.990	0.496	0.486	0.474	0.460
1.000	0.479	0.483*	0.487	0.487
$\delta = 0.15$				
β	teste alternativo	teste usual	teste alternativo	teste usual
0.900	0.648	0.648	0.622	0.609
0.925	0.650	0.644	0.612	0.604
0.950	0.663	0.651	0.648	0.633
0.975	0.671	0.655	0.615	0.604
0.990	0.672	0.655	0.644	0.619
1.000	0.693	0.684	0.636	0.612

Tabela 3.3: Valores do poder empírico do teste alternativo com variância conhecida (continuação)

	X_3		X_4	
$\delta = 0.1$				
β	teste alternativo	teste usual	teste alternativo	teste usual
0.900	0.572	0.571	0.207	0.190
0.925	0.578	0.573	0.193	0.182
0.950	0.602	0.596	0.195	0.194
0.975	0.614	0.594	0.168	0.172
0.990	0.589	0.585	0.146	0.172*
1.000	0.595	0.590	0.163	0.196*
$\delta = 0.125$				
β	teste alternativo	teste usual	teste alternativo	teste usual
0.900	0.718	0.715	0.240	0.223
0.925	0.737	0.734	0.233	0.219
0.950	0.709	0.704	0.229	0.219
0.975	0.734	0.716	0.238	0.235
0.990	0.726	0.711	0.205	0.226*
1.000	0.723	0.702	0.180	0.217*
$\delta = 0.15$				
β	teste alternativo	teste usual	teste alternativo	teste usual
0.900	0.872	0.872	0.325	0.302
0.925	0.872	0.870	0.313	0.287
0.950	0.861	0.859	0.279	0.264
0.975	0.880	0.876	0.297	0.317
0.990	0.860	0.849	0.293	0.322*
1.000	0.870	0.860	0.241	0.276*

Tabela 3.4: Estimativas de mínimos quadrados

Parâmetro	Valor
β	$\hat{\beta}_0 = 0.8915$
β_1	$\hat{\beta}_1 = 0.9548$
β_2	$\hat{\beta}_2 = 0.8292$

3.3 Enfoque Bayesiano

3.3.1 Cálculo do fator de Bayes segundo a metodologia de Carlin e Chib

Nesta seção calculamos o fator de Bayes para os dados do Exemplo 1.2 utilizando a metodologia baseada em métodos MCMC dada por Carlin and Chib (1995), descrita no Apêndice B. Supondo inicialmente que σ^2 é conhecida e igual a 0.0169. Para começar, notemos que testar a hipótese (3.1) contra a alternativa (3.2) é equivalente a escolher entre dois modelos não encaixados para os dados observados. Seja M o parâmetro (com valores inteiros) que indexa o modelo. Então podemos explicitar

$$M = 1 : y_{ik} = \beta x_{ik} + \varepsilon_{ik}, \quad \varepsilon_{ik} \stackrel{IID}{\sim} N(0, \sigma^2), \quad (3.20)$$

$$M = 2 : y_{ik} = \beta_i x_{ik} + \varepsilon_{ik}, \quad \varepsilon_{ik} \stackrel{IID}{\sim} N(0, \sigma^2), \quad (3.21)$$

com $i = 1, 2, k = 1, \dots, n_i$, $\theta_1 = \beta$, $\theta_2 = (\beta_1, \beta_2)^t$ e $\sigma^2 > 0$.

A função de verossimilhança sob o modelo $M = 2$ é

$$l(\theta_2) = \frac{1}{\sigma^{n_1+n_2}} \exp \left(-\frac{1}{2\sigma^2} \left(\sum_{k=1}^{n_1} (y_{1k} - \beta_1 x_{1k})^2 + \sum_{k=1}^{n_2} (y_{2k} - \beta_2 x_{2k})^2 \right) \right), \quad (3.22)$$

com a restrição $0 \leq \beta_i \leq 1$, $i = 1, 2$ e $\beta_1 > \beta_2$. A verossimilhança sob o modelo M_1 é similar a (3.22) com β_1 e β_2 substituídos por β . Como distribuições a priori para os parâmetros de regressão, consideramos distribuições normais com a restrição $0 \leq \beta \leq 1$ em $M = 1$, e $0 \leq \beta_i \leq 1$, $i = 1, 2$ e $\beta_1 > \beta_2$ em $M = 2$. Como parâmetros das distribuições a priori tomamos inicialmente as estimativas de mínimos quadrados sob os modelos correspondentes, apresentadas na Tabela 3.5.

Como não é possível calcular diretamente as distribuições a posteriori pelo fato de o espaço paramétrico ser restrito, utilizamos o conjunto de subrotinas BUGS 0.6 (Thomas et al., 1992). Em cada situação analisada, consideramos cada modelo separadamente para

Tabela 3.5: Estimativas de mínimos quadrados dos parâmetros de cada modelo

Modelo	Parâmetro	Média	Desvio Padrão
1	β	0.8915	0.0387
2	β_1	0.9548	0.0517
	β_2	0.8292	0.0512

Tabela 3.6: Fator de Bayes para comparar os parâmetros de regressão, quando a variância é conhecida

Probabilidades a priori		Fator de Bayes
π_1	π_2	
0.05	0.95	0.89
0.25	0.75	0.96
0.50	0.50	1.01
0.75	0.25	0.99
0.95	0.05	0.90

obter os hiperparâmetros das distribuições “pseudo-priori ” e avaliamos a convergência utilizando o método de Geweke (1992) baseado numa cadeia de tamanho amostral 1000.

Calculamos o fator de Bayes em cada caso, considerando 1000 iterações iniciais e atualizando o processo com 10000 iterações. A Tabela 3.6 apresenta os fatores de Bayes para comparar os modelos 1 e 2 definidos em (3.20) e (3.21) com diferentes probabilidades a priori de opção de cada modelo, π_j , $j = 1, 2$.

Com exceção do caso $\pi_1 = \pi_2 = 0.5$, em todos os casos o fator de Bayes não indica evidências favoráveis à hipótese alternativa. Como o fator de Bayes pode ser sensível à escolha das distribuições a priori, consideramos diferentes valores dos parâmetros de precisão τ_β , τ_{β_1} , τ_{β_2} das distribuições a priori de β , β_1 e β_2 respectivamente, com probabilidades a priori de opção de cada modelo, $\pi_1 = \pi_2 = 0.5$, e calculamos os fatores de Bayes correspondentes, que são apresentados na Tabela 3.7.

Esses resultados confirmam a sensibilidade do fator de Bayes à escolha dos hiperparâmetros de precisão τ_β , τ_{β_1} , τ_{β_2} , em particular, esse fator decresce quando a precisão diminui. Em todos os casos o fator de Bayes não evidencia a rejeição da hipótese nula, a exceção do caso em que consideramos como valores dos parâmetros suas estimativas de

Tabela 3.7: Fatores de Bayes para diferentes precisões das distribuições a priori dos parâmetros de regressão

Parâmetros de precisão			Fator de Bayes
τ_{β}	τ_{β_1}	τ_{β_2}	
667.69000	374.130000	381.470000	1.01
66.769000	37.413000	38.147000	0.52
6.676900	3.741300	3.814700	0.18
0.066769	0.037413	0.038147	0.02

Tabela 3.8: Valores empíricos a posteriori sobre o espaço paramétrico restrito

Modelo	Parâmetro	Média	Desvio padrão
1	β	0.890	0.035
2	β_1	0.941	0.033
	β_2	0.824	0.045

mínimos quadrados donde existe uma vaga evidencia contra a hipótese nula.

A Tabela 3.8 apresenta os valores empíricos a posteriori da média e do desvio padrão de cada parâmetro no caso $\pi_1 = \pi_2 = 0.5$.

O fator de Bayes para a comparação dos modelos 1 e 2 sem restrições no espaço paramétrico com probabilidades a priori de opção de cada modelo $\pi_1 = \pi_2 = 0.5$, é $B_{21} = 1.13$. Como observamos no Capítulo 2, esse fator é maior que o corresponde fator de Bayes para o caso em que o espaço paramétrico é restrito ($B_{21} = 1.01$), o que contraria a intuição.

A Tabela 3.9 apresenta os valores empíricos a posteriori para os parâmetros de cada modelo no caso em que o espaço paramétrico é irrestrito; eles são próximos dos valores empíricos a posteriori obtidos sob o espaço paramétrico restrito dados na Tabela 3.8, as estimativas de β_1 são os que apresentam uma diferença maior. Portanto a restrição sobre o espaço paramétrico quase não modifica a estimação dos parâmetros a posteriori, logo as distribuições condicionais a posteriori dos parâmetros não foram influenciadas pelas restrições sobre as distribuições a priori.

Tabela 3.9: Valores empíricos a posteriori sobre o espaço paramétrico irrestrito

Modelo	Parâmetro	Média	Desvio padrão
1	β	0.891	0.036
2	β_1	0.957	0.045
	β_2	0.825	0.046

3.3.2 Cálculo do fator de Bayes segundo a metodologia de Irony e Pereira

Consideremos agora o problema de comparar os parâmetros de regressão no caso em que a variância dos erros é conhecida, aplicando a metodologia descrita na Seção 2.3.2. As hipóteses (3.1) e (3.2), definem que espaço paramétrico Ω seja formado pela união dos conjuntos

$$\begin{aligned}\Omega_1 &= \{\boldsymbol{\beta} = (\beta_1, \beta_2) : \beta_1 = \beta_2, 0 \leq \beta_i \leq 1, i = 1, 2\}, \\ \Omega_2 &= \{\boldsymbol{\beta} = (\beta_1, \beta_2) : \beta_1 > \beta_2, 0 \leq \beta_i \leq 1, i = 1, 2\}.\end{aligned}$$

Por razões de simplicidade de cálculo, escrevemos a função de verossimilhança (3.22) como

$$l(\beta_1, \beta_2) = \frac{\tau^{(n_1+n_2)/2}}{2\pi} \exp\left(\frac{-\tau}{2} \left(\sum_{i=1}^2 (s_{i1} - \beta_i s_{i2} + \beta_i^2 s_{i3})\right)\right)$$

com

$$s_{i1} = \sum_{k=1}^{n_i} y_{ik}^2, \quad (3.23)$$

$$s_{i2} = \sum_{k=1}^{n_i} 2y_{ik}x_{ik}, \quad (3.24)$$

$$s_{i3} = \sum_{k=1}^{n_i} x_{ik}^2 \quad (3.25)$$

para $i = 1, 2$.

Como exemplo, calculamos o fator de Bayes supondo que os β_i ($i = 1, 2$) têm dis-

tribuições que não pertencem a família de distribuições conjugadas.

Inicialmente suponhamos que β_1 e β_2 têm distribuições a priori Beta, ou seja, distribuições com funções densidades dadas por

$$g(\beta_i) = \frac{1}{B(a_i, b_i)} \beta_i^{a_i-1} (1 - \beta_i)^{b_i-1}, \quad (3.26)$$

em que $B(a_i, b_i)$ a função beta de parâmetros a_i, b_i . Logo

$$\begin{aligned} E(\beta_i) &= a_i / (a_i + b_i), \\ Var(\beta_i) &= a_i b_i / ((a_i + b_i)^2 (a_i + b_i + 1)). \end{aligned}$$

A densidade conjunta a priori é dada por

$$g(\beta_1, \beta_2) = C \prod_{i=1}^2 \beta_i^{a_i-1} (1 - \beta_i)^{b_i-1},$$

em que $C = [B(a_1, b_1) B(a_2, b_2)]^{-1}$.

A verossimilhança conjunta dos dados e do parâmetro (β_1, β_2) é o produto

$$g(\beta_1, \beta_2) l(\beta_1, \beta_2).$$

Para calcular o fator de Bayes, obtemos os integrais que definem as funções preditivas $f_i(d)$ definidas em (2.26). O denominador do cociente que define $f_1(d)$ é igual a

$$\begin{aligned} & \int g(\beta, \beta) d\bar{\Omega}_1 \\ &= C \int_0^1 \beta^{(a_1+a_2-1)-1} (1 - \beta)^{(b_1+b_2-1)-1} d\beta \\ &= CB(a_1 + a_2 - 1, b_1 + b_2 - 1), \end{aligned} \quad (3.27)$$

em que $\bar{\Omega}_1 = \{\beta : 0 \leq \beta \leq 1\}$ e o numerador do cociente é

$$\begin{aligned} & \int g(\beta, \beta) l(\beta, \beta) d\bar{\Omega}_1 \\ &= C \frac{\tau^{(n_1+n_2)/2}}{2\pi} \exp\left(\frac{-\tau}{2} (s_{11} + s_{21})\right) \\ & \int_0^1 \beta^{A-1} (1-\beta)^{B-1} \exp\left(\frac{-\tau}{2} (\beta^2 (s_{13} + s_{23}) - \beta (s_{12} + s_{22}))\right) d\beta, \end{aligned} \quad (3.28)$$

com $A = (a_1 + a_2 - 1)$, $B = (b_1 + b_2 - 1)$, com s_{i1} , para $i = 1, 2$, definidos em (3.23).

Uma forma fechada para (3.28) pode ser obtida utilizando a expansão em série da função $\exp\left(\frac{-\tau}{2} (\beta^2 s_3 - \beta s_2)\right)$, mas para fins práticos (3.28) pode ser aproximada numericamente uma vez escolhidos os valores de a_i , b_i para $i = 1, 2$, isto é escolhidas as distribuições a priori dos parâmetros.

O denominador do cociente que define $f_2(d)$ é

$$\begin{aligned} & \int g(\beta_1, \beta_2) d\Omega_2 \\ &= C \int_0^1 \beta_1^{a_1-1} (1-\beta_1)^{b_1-1} \left(\int_0^{\beta_1} \beta_2^{a_2-1} (1-\beta_2)^{b_2-1} d\beta_2 \right) d\beta_1, \end{aligned} \quad (3.29)$$

e o numerador do cociente que define $f_2(d)$ é dado por

$$\begin{aligned} & \int g(\beta_1, \beta_2) l(\beta_1, \beta_2) d\Omega_2 \\ &= C \frac{\tau^{(n_1+n_2)/2}}{2\pi} \exp\left(\frac{-\tau}{2} (s_{11} + s_{21})\right) \\ &= C \int_0^1 \beta_1^{a_1-1} (1-\beta_1)^{b_1-1} \exp\left(\frac{-\tau}{2} (\beta_1^2 s_{13} - \beta_1 s_{12})\right) \\ & \left(\int_0^{\beta_1} \beta_2^{a_2-1} (1-\beta_2)^{b_2-1} \exp\left(\frac{-\tau}{2} (\beta_2^2 s_{23} - \beta_2 s_{22})\right) d\beta_2 \right) d\beta_1. \end{aligned} \quad (3.30)$$

Para o Exemplo 1.2 consideramos $\tau = (0.13)^{-2} = 59.17$ correspondente à precisão amostral, obtivemos $B_{21} = 1.22$ para $a_1 = b_1 = 1$, e $a_2 = b_2 = 1$, ou seja, para distribuições a priori uniformes $(0,1)$, sugerindo rejeição da hipótese nula.

Suponhamos agora que β_1 e β_2 têm distribuições normais truncadas em $[0,1]$ com parâmetros (ξ_1, γ_1) e (ξ_2, γ_2) , respectivamente, ou seja, para as quais as funções densidades são

$$g'(\beta_i) = \frac{1}{\Phi((1-\xi_i)\sqrt{\gamma_i}) - \Phi(-\xi_i\sqrt{\gamma_i})} \frac{\sqrt{\gamma_i}}{\sqrt{2\pi}} \exp\left(\frac{-\gamma_i}{2}(\beta_i - \xi_i)^2\right), \quad (3.31)$$

para $i = 1, 2$. Fazendo

$$J(\xi_i, \gamma_i) = \frac{\frac{1}{\sqrt{2\pi}} [\exp(\frac{-\gamma_i}{2}\xi_i^2) - \exp(\frac{-\gamma_i}{2}(1-\xi_i)^2)]}{\Phi((1-\xi_i)\sqrt{\gamma_i}) - \Phi(-\xi_i\sqrt{\gamma_i})},$$

temos que

$$E(\beta_i) = \xi_i + \gamma_i^{-1/2} J(\xi_i, \gamma_i), \quad (3.32)$$

$$\begin{aligned} Var(\beta_i) = & \frac{1}{\gamma_i} \left\{ 1 + \frac{\sqrt{\gamma_i}}{\sqrt{2\pi}} \frac{\exp(\frac{-\gamma_i}{2}(1-\xi_i)^2)}{\Phi((1-\xi_i)\sqrt{\gamma_i}) - \Phi(-\xi_i\sqrt{\gamma_i})} \right. \\ & \left. - \xi_i\sqrt{\gamma_i} [J(\xi_i, \gamma_i)]^2 \right\}. \end{aligned} \quad (3.33)$$

A densidade conjunta a priori é dada por

$$g'(\beta_1, \beta_2) = C_1 \prod_{i=1}^2 \frac{\sqrt{\gamma_i}}{\sqrt{2\pi}} \exp\left(\frac{-\gamma_i}{2}(\beta_i - \xi_i)^2\right),$$

em que $C_1 = [\prod_{i=1}^2 \Phi((1-\xi_i)\sqrt{\gamma_i}) - \Phi(-\xi_i\sqrt{\gamma_i})]^{-1}$.

Obtemos agora os integrais que definem as funções preditivas $f_i(d)$ apresentadas em (2.26). O denominador do cociente que define $f_1(d)$ é dado por

$$\begin{aligned} & \int g'(\beta, \beta) d\bar{\Omega}_1 \\ = & C_1 \frac{\sqrt{\gamma_1\gamma_2}}{2\pi} \int_0^1 \exp\left(-\left(\frac{\gamma_1}{2}(\beta - \xi_1)^2 + \frac{\gamma_2}{2}(\beta - \xi_2)^2\right)\right) d\beta, \end{aligned} \quad (3.34)$$

e o numerador é

$$\begin{aligned}
& \int g'(\beta, \beta) l(\beta, \beta) d\bar{\Omega}_1 \\
&= C_1 \frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \frac{\tau^{(n_1+n_2)/2}}{2\pi} \exp\left(\frac{-\tau}{2}(s_{11} + s_{21})\right) \\
& \int_0^1 \exp\left(-\left(\frac{\gamma_1}{2}(\beta - \xi_1)^2 + \frac{\gamma_2}{2}(\beta - \xi_2)^2\right)\right) k(\beta) d\beta,
\end{aligned} \tag{3.35}$$

em que $k(\beta) = \exp\left(\frac{-\tau}{2}(\beta^2(s_{13} + s_{23}) - \beta(s_{12} + s_{22}))\right)$. O denominador de $f_2(d)$ é dado por

$$\int g(\beta_1, \beta_2) d\Omega_2 = C_1 \frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \tag{3.36}$$

$$\times \int_0^1 \exp\left(\frac{-\gamma_1}{2}(\beta_1 - \xi_1)^2\right) \left(\int_0^{\beta_1} \exp\left(\frac{-\gamma_2}{2}(\beta_2 - \xi_2)^2\right) d\beta_2\right) d\beta_1, \tag{3.37}$$

e o numerador por

$$\begin{aligned}
& \int g(\beta_1, \beta_2) l(\beta_1, \beta_2) d\Omega_2 \\
&= C_1 \frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \frac{\tau^{(n_1+n_2)/2}}{2\pi} \exp\left(\frac{-\tau}{2}(s_{11} + s_{21})\right) \\
&= C_1 \frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \int_0^1 \exp\left(\frac{-\gamma_1}{2}(\beta_1 - \xi_1)^2\right) \exp\left(\frac{-\tau}{2}(\beta_1^2 s_{13} - \beta_1 s_{12})\right) \\
& \left(\int_0^{\beta_1} \exp\left(\frac{-\gamma_2}{2}(\beta_2 - \xi_2)^2\right) \exp\left(\frac{-\tau}{2}(\beta_2^2 s_{23} - \beta_2 s_{22})\right) d\beta_2\right) d\beta_1.
\end{aligned} \tag{3.38}$$

Para analisar a sensibilidade do fator de Bayes com respeito à escolha dos hiperparâmetros das distribuições a priori, calculamos $\xi_i, \gamma_i, i = 1, 2$, através do método dos momentos, ou seja, resolvendo $E(\beta_i) = \hat{\beta}_i$ onde $E(\beta_i)$ está definido em (3.32) e $\hat{\beta}_i$ em (3.3) conjuntamente com $(Var(\beta_i))^{-1} = \tau_{\beta_i}$ definida em (3.33), para diferentes valores de τ_{β_i} , e obtivemos os fatores de Bayes correspondentes. Os resultados são apresentados na Tabela 3.10.

Contrariamente ao que foi observado na Tabela 3.7, o fator de Bayes não é sensível

Tabela 3.10: Fator de Bayes correspondente à distribuição a priori normal truncada

Parâmetros							Fator de Bayes
τ_{β}	τ_{β_1}	τ_{β_2}	ξ_1	γ_1	ξ_2	γ_2	B_{21}
667.69000	374.130000	381.470000	0.114	337.933	0.116	349.484	1.001
66.769000	37.413000	38.147000	0.289	29.701	0.307	31.750	1.002
6.676900	3.741300	3.814700	0.264	4.173	0.347	4.722	1.005
0.066769	0.037413	0.038147	0.992	0.076	0.856	0.077	1.002

à escolha dos hiperparâmetros de precisão, e apresenta apenas evidências vagas para a rejeição da hipótese nula.

Se agora consideramos que o espaço paramétrico irrestrito é formado pela união dos conjuntos

$$\dot{\Omega}_1 = \{\beta = (\beta_1, \beta_2) : \beta_1 = \beta_2\} \text{ e } \dot{\Omega}_2 = \{\beta = (\beta_1, \beta_2) : \beta_1 > \beta_2\},$$

e que β_1 e β_2 têm distribuições normais de parâmetros $\left(\dot{\xi}_1, \dot{\gamma}_1\right)$ e $\left(\dot{\xi}_2, \dot{\gamma}_2\right)$ respectivamente, ou seja, para as quais as funções densidades são

$$\dot{g}(\beta_i) = \frac{\sqrt{\dot{\gamma}_i}}{\sqrt{2\pi}} \exp\left(-\frac{\dot{\gamma}_i}{2} \left(\beta_i - \dot{\xi}_i\right)^2\right), \quad (3.39)$$

para $i = 1, 2$. Os integrais que definem as funções preditivas podem ser calculados de forma similar à dos integrais (3.34), (3.35), (3.37) e (3.38). Para efeitos de comparação escolhemos como estimativas de $E(\beta_1) = \dot{\xi}_1$, $E(\beta_2) = \dot{\xi}_2$, $(Var(\beta_1))^{-1} = \dot{\gamma}_1$ e $(Var(\beta_2))^{-1} = \dot{\gamma}_2$ a $\hat{\beta}_1 = 0.9548$, $\hat{\beta}_2 = 0.8292$, $\hat{\tau}_{\beta_1} = 374.13$, $\hat{\tau}_{\beta_2} = 381.47$, respectivamente. O fator de Bayes neste caso é $B_{21} = 1$, praticamente igual ao fator de Bayes no caso do espaço paramétrico com restrição ($B_{21} = 1.001$).

Podemos concluir que para os dados do Exemplo 1.2, para o caso em que a variância é conhecida, analisados tanto com a metodologia frequentista através do teste alternativo, quanto com as metodologias bayesianas só existem evidências muito vagas contra a hipótese nula.

3.4 Extensão para o caso em que a variância é desconhecida

Se no modelo (1.1) σ^2 for desconhecida, o valor crítico D_α no teste usual pode ser substituído por

$$d_\alpha = \sqrt{\widehat{v}_1 + \widehat{v}_2} T_{(n_1 + n_2 - 2)}^{-1} (1 - \alpha) \quad (3.40)$$

com

$$\widehat{v}_i = s_p^2 \left(\sum_{k=1}^{n_i} x_{ik}^2 \right)^{-1},$$

em que

$$s_p^2 = \frac{\sum_{i=1}^2 \sum_{k=1}^{n_i} y_{ik}^2 - \sum_{i=1}^2 \widehat{\beta}_i \sum_{k=1}^{n_i} x_{ik} y_{ik}}{n_1 + n_2 - 2}. \quad (3.41)$$

e $\widehat{\beta}_i$ é definido em (3.3).

O poder do teste usual (para o espaço paramétrico irrestrito) é dado por

$$P_{\beta_1 - \beta_2 = \delta} \left(\frac{\overline{D} - \delta}{\sqrt{\widehat{v}_1 + \widehat{v}_2}} > \frac{d_\alpha - \delta}{\sqrt{\widehat{v}_1 + \widehat{v}_2}} \right) = 1 - T_{(n_1 + n_2 - 2)} \left(\frac{d_\alpha - \delta}{\sqrt{\widehat{v}_1 + \widehat{v}_2}} \right). \quad (3.42)$$

Similarmente ao caso em que a variância é conhecida, propomos um teste alternativo para o qual a região de rejeição da hipótese nula corresponde a $\widehat{D} > d_\alpha^*$ onde d_α^* é a solução de

$$P_{\beta_1 - \beta_2 = 0} \left(\widehat{D} > d_\alpha^* \right) = \alpha. \quad (3.43)$$

Por razões similares àquelas utilizadas ao calcular (3.13), temos

$$\begin{aligned}
P_{\beta_1 - \beta_2 = \delta} \left(\widehat{D} > d_\alpha^* \right) &= T_k \left(\frac{1 - d_\alpha^* - \beta_2}{\sqrt{\widehat{v}_2}} \right) - T_k \left(\frac{d_\alpha^* - \beta_1}{\sqrt{\widehat{v}_1}} \right) T_k \left(\frac{-\beta_2}{\sqrt{\widehat{v}_2}} \right) \\
&+ \frac{1}{\sqrt{k} B \left(\frac{1}{2}, \frac{k}{2} \right)} \int_{-\infty}^{\frac{-\beta_2}{\sqrt{\widehat{v}_2}}} T_k \left(\frac{d_\alpha^* - \delta + t\sqrt{\widehat{v}_2}}{\sqrt{\widehat{v}_1}} \right) \left(1 + \frac{t^2}{k} \right)^{-(k+1)/2} dt \\
&+ \frac{1}{\sqrt{k} B \left(\frac{1}{2}, \frac{k}{2} \right)} \int_{\frac{1 - d_\alpha^* - \beta_2}{\sqrt{\widehat{v}_2}}}^{\infty} T_k \left(\frac{d_\alpha^* - \delta + t\sqrt{\widehat{v}_2}}{\sqrt{\widehat{v}_1}} \right) \left(1 + \frac{t^2}{k} \right)^{-(k+1)/2} dt \\
&- T_k \left(\frac{d_\alpha^* - \delta}{\sqrt{\widehat{v}_1 + \widehat{v}_2}} \right), \tag{3.44}
\end{aligned}$$

em que $k = n_1 + n_2 - 2$. Sob a hipótese alternativa, $\beta_1 - \beta_2 = \delta > 0$ e (3.44) é o poder do teste alternativo.

Para calcular d_α^* em (3.43) usamos o método de Brent e a regra de Simpson (ver, por exemplo, Press et al., 1994) e utilizando argumentos similares ao caso em que σ^2 é conhecida, podemos avaliar situações em que o teste alternativo é mais poderoso que o teste usual.

No Exemplo 1.2, as estimativas requeridas são apresentadas na Tabela 3.5 em que $\widehat{v}_1 = (s_{\beta_1})^2 = (0.0517)^2$ e $\widehat{v}_2 = (s_{\beta_2})^2 = (0.0512)^2$.

Para $\alpha = 0.05$, de (3.42) e (3.43) obtivemos $d_\alpha = 0.128$ e $d_\alpha^* = 0.125$. Como $\widehat{D} = \overline{D} = \widehat{\beta}_1 - \widehat{\beta}_2 = 0.126$, só rejeitamos a hipótese nula com o teste alternativo. A generalização para o caso em que os limites do intervalo são diferentes de 0 e 1 é direta.

O estimador s_p^2 não leva em conta a restrição no espaço paramétrico. Como alternativa podemos então considerar o estimador de máxima verossimilhança

$$\widehat{\sigma}_R^2 = \frac{\sum_{i=1}^2 \sum_{k=1}^{n_i} y_{ik}^2 - \sum_{i=1}^2 \widehat{\beta}_{ir} \sum_{k=1}^{n_i} x_{ik} y_{ik}}{n_1 + n_2}, \tag{3.45}$$

e o estimador alternativo dado por

$$s_{pr}^2 = \frac{\sum_{i=1}^2 \sum_{k=1}^{n_i} y_{ik}^2 - \sum_{i=1}^2 \widehat{\beta}_{ir} \sum_{k=1}^{n_i} x_{ik} y_{ik}}{n_1 + n_2 - 2}. \tag{3.46}$$

A distribuição t-Student considerada em (3.44) não pode mais ser utilizada com a substituição de s_p^2 por $\hat{\sigma}_R^2$ ou s_{pr}^2 . Alguns artigos relacionados com teste de hipóteses para parâmetros restritos em modelos de regressão utilizam o estimador usual de σ^2 para a estatística da razão de verossimilhanças (ver, por exemplo, Wolak (1987)). Consideramos um estudo de simulação em que comparamos o desempenho do teste alternativo baseado nos três estimadores da variância s_p^2 , $\hat{\sigma}_R^2$ ou s_{pr}^2 , definidos em (3.41), (3.45) e (3.46), respectivamente.

Para cada vetor de variáveis explicativas, $\mathbf{x}_{1k}$ e $\mathbf{x}_{2k}$ definidos em (3.16), (3.17), (3.18) e (3.19), $k = 1, \dots, 4$, respectivamente, foram geradas 1000 amostras de valores da variável resposta com distribuição normal multivariada com vetor de médias $\mathbf{x}_{i,k}\beta$, com $i = 1, 2$, e matriz de covariância $\sigma^2 \mathbf{I}_8$, com $\sigma^2 = (0.13)^2 = 0.0169$. Para cada par de amostras calculamos

$$\frac{\hat{D}}{s_{pr} \sqrt{(\mathbf{x}_{1k}^t \mathbf{x}_{1k})^{-1} + (\mathbf{x}_{2k}^t \mathbf{x}_{2k})^{-1}}}, \frac{\hat{D}}{s_p \sqrt{(\mathbf{x}_{1k}^t \mathbf{x}_{1k})^{-1} + (\mathbf{x}_{2k}^t \mathbf{x}_{2k})^{-1}}} \text{ e } \frac{\hat{D}}{\hat{\sigma}_R \sqrt{(\mathbf{x}_{1k}^t \mathbf{x}_{1k})^{-1} + (\mathbf{x}_{2k}^t \mathbf{x}_{2k})^{-1}}},$$

com $k = 1, \dots, 4$.

Obtivemos os respectivos quantis amostrais de ordem

$p(= 0.10, 0.20, 0.25, 0.50, 0.75, 0.90, 0.95, 0.975, 0.995, 0.9975, 0.999, 0.9995)$ calculados para as diferentes combinações de $\beta = 0.8, 0.90, 0.95, 0.99$ e 1 e $\sigma^2 = 0.13, 0.5$ e 1.0 . Estes quantis são apresentados nas Tabelas 3.19 a 3.30. Em todos os casos, os quantis baseados no desvio padrão amostral s_p são maiores do que ou iguais àqueles baseados no estimador s_{pr} . Assim, sob a hipótese nula há um menor número de rejeições com o teste baseado em s_{pr} do que com o teste baseado em s_p . Não podemos estabelecer um padrão de comportamento geral para comparar os quantis baseados em s_p e $\hat{\sigma}_R$, dado que as relações entre eles dependem tanto da posição de β , com respeito ao extremo do intervalo quanto da matriz de valores da variável explicativa. Portanto, o teste alternativo baseado em s_p constitui uma opção razoável, para comparar os parâmetros de regressão.

Estudamos o efeito da substituição de β , por $\hat{\beta}_r$ em (3.44) sobre o nível de significância

e sobre o poder do teste alternativo. Consideramos um novo estudo de simulação em que para cada vetor de variáveis explicativas, $\mathbf{x}_{1k}$ e $\mathbf{x}_{2k}$ definidos em (3.16), (3.17), (3.18) e (3.19), $k = 1, \dots, 4$, respectivamente, foram geradas 1000 amostras de valores da variável resposta com distribuição normal multivariada com vetor de médias $\mathbf{x}_{i,k}\beta_i$, com $i = 1, 2$, e matriz de covariância $\sigma^2\mathbf{I}_8$, com $\sigma^2 = (0.13)^2 = 0.0169$.

Para saber se o estimador $\hat{\beta}_r$ definido em (3.14) mantém a probabilidade de erro de tipo I próxima do nível de significância, as amostras foram geradas considerando $\beta_1 = \beta_2 = \beta$, para valores de $\beta = 0.70, 0.80, 0.85, 0.90, 0.925, 0.950, 0.975, 0.99$, e 1. Em cada amostra foram calculados $\hat{D}$, e d_α^* através de (3.44) utilizando o estimador $\hat{\beta}_r$. Os valores das taxas de rejeição obtidos estão apresentados na Tabela 3.11.

Em todos os casos considerados, podemos observar que para valores próximos ao extremo do intervalo, o valor de α empírico para o teste alternativo é menor que o nível de significância o que leva a uma diminuição do poder.

Também estudamos a influência do uso de $\hat{\beta}_r$ sobre o poder do teste alternativo. Com este intuito, realizamos uma nova série de simulações. Para cada valor de δ fixo, as amostras foram geradas considerando $\beta_2 = \beta_1 - \delta$ e $\sigma = 0.13$ para valores de $\beta_1 = 0.90, 0.925, 0.950, 0.975, 0.990$ e 1. Obtiveram-se, em cada amostra $\hat{D}$ e d_α^* de (3.44) com $\delta = 0$, utilizando o estimador $\hat{\beta}_r$ no lugar de β . O poder empírico é calculado como o número de amostras em que $\hat{D} > d_\alpha^*$ dividido pelo número total de amostras. Os resultados estão dispostos nas Tabelas 3.12 e 3.13.

Podemos observar que o poder do teste alternativo é maior do que ou igual ao poder do teste usual, à exceção dos valores indicados com asterisco onde o valor de β_1 está próximo ao extremo do intervalo, como era esperado, já que o valor da taxa de rejeição empírica foi menor que o nível de significância. O estudo do desempenho de outros estimadores de β em relação ao poder do teste alternativo é um tópico para futuras pesquisas.

Tabela 3.11: Valores empíricos do nível de significância com a variância desconhecida considerando $\mathbf{X}$, $\mathbf{X}_2$, $\mathbf{X}_3$ e $\mathbf{X}_4$

β	$\mathbf{X}_1$		$\mathbf{X}_2$	
	teste alternativo	teste usual	teste alternativo	teste usual
0.700	0.053	0.053	0.049	0.049
0.800	0.042	0.043	0.058	0.057
0.850	0.052	0.050	0.053	0.050
0.900	0.049	0.042	0.051	0.048
0.925	0.052	0.048	0.050	0.048
0.950	0.046	0.043	0.050	0.052
0.975	0.029	0.036	0.027	0.053
0.990	0.040	0.054	0.020	0.056
1.000	0.017	0.033	0.017	0.052

β	$\mathbf{X}_3$		$\mathbf{X}_4$	
	teste alternativo	teste usual	teste alternativo	teste usual
0.700	0.051	0.051	0.049	0.048
0.800	0.045	0.045	0.058	0.053
0.850	0.050	0.050	0.046	0.038
0.900	0.053	0.050	0.054	0.052
0.925	0.047	0.041	0.040	0.042
0.950	0.048	0.049	0.043	0.055
0.975	0.044	0.051	0.027	0.042
0.990	0.041	0.057	0.028	0.052
1.000	0.027	0.054	0.021	0.049

Tabela 3.12: Valores do poder empírico, com a variância desconhecida

	$\mathbf{X}_1$		$\mathbf{X}_2$	
$\delta = 0.1$				
β	teste alternativo	teste usual	teste alternativo	teste usual
0.900	0.381	0.372	0.343	0.331
0.925	0.365	0.349	0.356	0.338
0.950	0.386	0.362	0.356	0.332
0.975	0.392	0.377	0.373	0.350
0.990	0.332	0.329	0.340	0.33
1.000	0.385	0.392*	0.309	0.339*
$\delta = 0.125$				
β	teste alternativo	teste usual	teste alternativo	teste usual
0.900	0.478	0.473	0.463	0.454
0.925	0.487	0.473	0.449	0.430
0.950	0.486	0.459	0.468	0.438
0.975	0.493	0.458	0.464	0.437
0.990	0.516	0.497	0.410	0.403
1.000	0.463	0.459	0.418	0.418
$\delta = 0.15$				
	teste alternativo	teste usual	teste alternativo	teste usual
0.900	0.615	0.612	0.573	0.566
0.925	0.613	0.607	0.595	0.583
0.950	0.647	0.633	0.609	0.586
0.975	0.619	0.592	0.607	0.582
0.990	0.626	0.593	0.600	0.576
1.000	0.617	0.600	0.597	0.570

Tabela 3.13: Valores do poder empírico, com a variância desconhecida (continuação)

	$\mathbf{X}_3$		$\mathbf{X}_4$	
$\delta = 0.1$				
β	teste alternativo	teste usual	teste alternativo	teste usual
0.900	0.554	0.550	0.204	0.186
0.925	0.551	0.545	0.169	0.147
0.950	0.555	0.539	0.177	0.166
0.975	0.603	0.567	0.164	0.162
0.990	0.576	0.547	0.167	0.179
1.000	0.563	0.546	0.153	0.174*
$\delta = 0.125$				
β	teste alternativo	teste usual	teste alternativo	teste usual
0.900	0.677	0.678	0.218	0.200
0.925	0.696	0.695	0.233	0.215
0.950	0.660	0.655	0.211	0.195
0.975	0.710	0.694	0.237	0.238
0.990	0.718	0.693	0.200	0.218*
1.000	0.708	0.693	0.185	0.211*
$\delta = 0.15$				
	teste alternativo	teste usual	teste alternativo	teste usual
0.900	0.839	0.839	0.296	0.278
0.925	0.824	0.824	0.302	0.279
0.950	0.830	0.828	0.317	0.290
0.975	0.856	0.848	0.279	0.266
0.990	0.867	0.852	0.285	0.287*
1.000	0.861	0.847	0.293	0.300*

Tabela 3.14: Fator de Bayes para a comparação de parâmetros de regressão, quando a variância é desconhecida

Probabilidades a priori		Fator de Bayes
π_1	π_2	
0.05	0.95	2.71
0.25	0.75	2.97
0.50	0.50	2.86
0.75	0.25	2.89
0.95	0.05	2.94

3.4.1 Enfoque bayesiano

Como σ^2 é desconhecida, consideramos para o parâmetro de precisão $\tau = \sigma^{-2}$ uma distribuição a priori gama de parâmetros $v_1 = 1$ e $v_2 = 0.0169$, portanto $E(\tau) = v_1/v_2 = 59.17$, esse valor é igual à estimativa de mínimos quadrados de τ . Não consideramos distribuições “pseudoprioris” para σ^2 que é um parâmetro comum nos dois modelos (ver Carlin e Chib (1995)). Para obter uma aproximação das “pseudo-priori” para os parâmetros de regressão, como foi feito no caso da variância conhecida, avaliamos cada modelo separadamente e usamos as distribuições a posteriori resultantes como distribuições “pseudo-prioris” para os parâmetros de cada modelo. Para avaliar a convergência utilizamos o método de Geweke (1992), descrito no Apêndice B, utilizando o CODA (Cowles, 1994).

A Tabela 3.14 apresenta o fator de Bayes para comparar os modelos 1 e 2 definidos em (3.20) e (3.21) para diferentes valores de probabilidades a priori de opção de cada modelo, π_j $j = 1, 2$, considerando 1000 iterações iniciais e atualizando o processo com 10000 iterações.

Em todos os casos o fator de Bayes favorece a hipótese alternativa. Os valores apresentados na Tabela 3.14 são maiores que os valores apresentados na Tabela 3.6, e portanto, o fato de termos variância desconhecida aumenta a evidência contra a hipótese nula. A Tabela 3.15 apresenta os valores empíricos da média e do desvio padrão a posteriori de cada parâmetro quando os modelos são igualmente prováveis a priori; estes valores são

Tabela 3.15: Valores empíricos a posteriori, sob o espaço paramétrico restrito, quando a variância é desconhecida

Parâmetro	Média	Desvio padrão
β	0.891	0.027
β_1	0.947	0.029
β_2	0.828	0.035
τ	6.236	2.120

Tabela 3.16: Fatores de Bayes correspondentes a distribuição a priori normal truncada, para o caso em que a variância é desconhecida, calculados com a metodologia dada por Carlin e Chib

Parâmetros		Precisão	Fator de Bayes
v_1	v_2	$(v_1/v_2^2)^{-1}$	B_{21}
1	0.0169	0.0003	2.86
10	0.1690	0.0030	2.97
100	1.6900	0.0286	2.97
1000	16.9000	0.2856	2.92

próximos dos valores empíricos para o caso em que a variância é conhecida, apresentados na Tabela 3.8.

Calculamos o fator de Bayes para diferentes pares de valores (v_1, v_2) , satisfazendo $E(\tau) = v_1/v_2 = 59.17$ (estimativa de mínimos quadrados de τ), com precisão $(Var(\tau))^{-1} = (v_1/v_2^2)^{-1}$, considerando iguais probabilidades a priori de opção para os modelos. Os resultados são apresentados na Tabela 3.16. Em todos os casos rejeitou-se a hipótese nula, sugerindo que o fator de Bayes parece não ser muito sensível à escolha destes hiperparâmetros.

Também calculamos o fator de Bayes considerando o espaço paramétrico irrestrito com probabilidades a priori $\pi_1 = \pi_2 = 0.5$, obtendo $B_{21} = 3.24$, que é maior que o correspondente fator de Bayes para o caso em que o espaço paramétrico é restrito ($B_{21} = 2.86$), ou seja que a restrição não aumenta a evidência contra a hipótese nula, contrariamente ao esperado.

A Tabela 3.17 apresenta as estimativas a posteriori para os parâmetros de cada modelo no caso em que o espaço paramétrico é irrestrito; elas são próximas aos estimativas

Tabela 3.17: Valores empíricos a posteriori, sob o espaço paramétrico irrestrito, quando a variância é desconhecida

Parâmetro	Média	Desvio Padrão
β	0.891	0.027
β_1	0.954	0.035
β_2	0.828	0.036
τ	6.219	2.107

a posteriori, também sob o espaço paramétrico irrestrito, no caso em que a variância é conhecida, dados na Tabela 3.9. Além disso, as estimativas a posteriori dos parâmetros, no caso do espaço paramétrico restrito são também próximas as estimativas a posteriori quando o espaço paramétrico é irrestrito. Concluimos analogamente, que quando a variância é conhecida, a restrição sobre o espaço paramétrico não modifica a estimação da média e desvio padrão a posteriori.

Para exemplificar novamente a versatilidade da técnica dada por Irony e Pereira (1995), reanalisamos os dados supondo que os parâmetros β_1 e β_2 , têm distribuições a priori beta e que $\tau = \sigma^{-2}$ com distribuição a priori gamma de parâmetros (ν_1, ν_2) . Os integrais (3.27) e (3.28) não se modificam.

Denotando $s(\beta_1, \beta_2) = (\sum_{i=1}^2 (s_{i1} - \beta_i s_{i2} + \beta_i^2 s_{i3}))$, com s_{i1} , s_{i2} e s_{i3} definidos em (3.23), (3.24) e (3.25) respectivamente, o numerador e denominador do cociente que define a função preditiva $f_1(d)$ (2.26) são dados pelos seguintes integrais

$$\begin{aligned}
& \int g(\beta, \beta) l(\beta, \beta) d\bar{\Omega}_1 \\
&= C \frac{\nu_2^{\nu_1}}{2\pi\Gamma(\nu_1)} \int_0^1 \beta^{A-1} (1-\beta)^{B-1} \left[\int_0^\infty \tau^{\frac{n_1+n_2}{2}+\nu_2-1} \exp\left(-\tau\left(\frac{1}{2}s(\beta, \beta) + \nu_2\right)\right) d\tau \right] d\beta \\
&= C \frac{\nu_2^{\nu_1}}{2\pi\Gamma(\nu_1)} \int_0^1 \beta^{A-1} (1-\beta)^{B-1} \left[\int_0^\infty \frac{z^{\frac{n_1+n_2}{2}+\nu_2-1}}{\left(\frac{1}{2}s(\beta, \beta) + \nu_2\right)^{\frac{n_1+n_2}{2}+\nu_2}} \exp(-z) dz \right] d\beta \\
&= \Gamma\left(\frac{n_1+n_2}{2} + \nu_2\right) C \frac{\nu_2^{\nu_1}}{2\pi\Gamma(\nu_1)} \int_0^1 \frac{\beta^{A-1} (1-\beta)^{B-1}}{\left(\frac{1}{2}s(\beta, \beta) + \nu_2\right)^{\frac{n_1+n_2}{2}+\nu_2}} d\beta,
\end{aligned}$$

$$\begin{aligned}
& \int g(\beta_1, \beta_2) l(\beta_1, \beta_2) d\Omega_2 \\
&= \Gamma\left(\frac{n_1 + n_2}{2} + \nu_2\right) C \frac{\nu_2^{\nu_1}}{2\pi\Gamma(\nu_1)} \\
& \int_0^1 \beta_1^{a_1-1} (1 - \beta_1)^{b_1-1} \int_0^{\beta_1} \beta_2^{a_2-1} (1 - \beta_2)^{b_2-1} \frac{d\beta_2}{\left(\frac{1}{2}s(\beta_1, \beta_2) + \nu_2\right)^{\frac{n_1+n_2}{2} + \nu_2}} d\beta_1.
\end{aligned}$$

Ao considerar distribuições a priori uniformes (0,1) isto é $a_1 = b_1 = 1$, em e $a_2 = b_2 = 1$, em (3.26) e τ com distribuição gama com parâmetros $\nu_1 = 1$ e $\nu_2 = 0.0169$ obtivemos $B_{21} = 0$.

Suponhamos agora que β_1 e β_2 têm distribuições normais truncadas em $[0,1]$ com parâmetros (ξ_1, γ_1) e (ξ_2, γ_2) , respectivamente, ou seja, para as quais as funções de densidades são dadas por (3.31).

Calculamos os integrais que definem as funções preditivas $f_i(d)$ (2.26) para obter o fator de Bayes. Os denominadores dos cocientes que definem $f_1(d)$ e $f_2(d)$, definidos pelos integrais (3.34) e (3.37), respectivamente, não se modificam. Resta assim calcular os numeradores dos cocientes que definem $f_1(d)$ e $f_2(d)$. O numerador de $f_1(d)$ é

$$\begin{aligned}
& \int g(\beta, \beta) l(\beta, \beta) d\bar{\Omega}_1 \\
&= C_1 \frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \frac{\nu_2^{\nu_1}}{2\pi\Gamma(\nu_1)} \\
& \int_0^1 \exp\left(-\left(\frac{\gamma_1}{2}(\beta - \xi_1)^2 + \frac{\gamma_2}{2}(\beta - \xi_2)^2\right)\right) \\
& \left[\int_0^\infty \tau^{\frac{n_1+n_2}{2} + \nu_2 - 1} \exp\left(-\tau\left(\frac{1}{2}s(\beta, \beta) + \nu_2\right)\right) d\tau \right] d\beta \\
&= \Gamma\left(\frac{n_1 + n_2}{2} + \nu_2\right) \frac{\nu_2^{\nu_1}}{2\pi\Gamma(\nu_1)} C_1 \frac{\sqrt{\gamma_1 \gamma_2}}{2\pi} \\
& \int_0^1 \frac{\exp\left(-\left(\frac{\gamma_1}{2}(\beta - \xi_1)^2 + \frac{\gamma_2}{2}(\beta - \xi_2)^2\right)\right)}{\left(\frac{1}{2}s(\beta, \beta) + \nu_2\right)^{\frac{n_1+n_2}{2} + \nu_2}} d\beta.
\end{aligned} \tag{3.47}$$

Tabela 3.18: Fatores de Bayes correspondentes à distribuição a priori normal truncada, para o caso em que a variância é desconhecida, calculados com a metodologia dada por Irony e Pereira

v_1	v_2	$(v_1/v_2^2)^{-1}$	B_{21}
1	0.0169	0.0003	2.26
10	0.1690	0.0030	2.20
100	1.6900	0.0286	1.87
1000	16.9000	0.2856	1.47

O denominador do cociente que define $f_2(d)$ é dado por

$$\begin{aligned}
 & \int g(\beta_1, \beta_2) l(\beta_1, \beta_2) d\Omega_2 \\
 = & \Gamma\left(\frac{n_1 + n_2}{2} + \nu_2\right) \frac{\nu_2^{\nu_1}}{2\pi\Gamma(\nu_1)} C_1 \frac{\sqrt{\gamma_1\gamma_2}}{2\pi} \\
 & \int_0^1 \exp\left(\frac{-\gamma_1}{2}(\beta_1 - \xi_1)^2\right) \left(\int_0^{\beta_1} \frac{\exp\left(\frac{-\gamma_2}{2}(\beta_2 - \xi_2)^2\right)}{\left(\frac{1}{2}s(\beta_1, \beta_2) + \nu_2\right)^{\frac{n_1+n_2}{2} + \nu_2}} d\beta_2 \right) d\beta_1.
 \end{aligned} \tag{3.48}$$

Para analisar a sensibilidade do fator de Bayes com respeito à escolha dos hiperparâmetros da distribuição a priori de τ , obtivemos os fatores de Bayes para diferentes pares de valores (v_1, v_2) , tais que $E(\tau) = v_1/v_2 = 59.17$ (estimativa de mínimos quadrados de τ) com precisão $(Var(\tau))^{-1} = (v_1/v_2^2)^{-1}$, considerando (ξ_1, γ_1) e (ξ_2, γ_2) iguais a suas estimativas de mínimos quadrados, ou seja, $\xi_1 = 0.114$, $\gamma_1 = 337.933$, $\xi_2 = 0.116$ e $\gamma_2 = 349.484$. Os resultados são apresentados na Tabela 3.18.

Em todos os casos o fator de Bayes evidencia a rejeição da hipótese nula, e é sensível à escolha dos hiperparâmetros da distribuição a priori de τ ; o fator de Bayes decresce quando a precisão aumenta, ao contrário ao observado em outras situações. Esta sensibilidade não é observada ao calcular o fator de Bayes com a metodologia proposta por Carlin e Chib (ver Tabela 3.16).

3.5 Discussão

Todas as conclusões apresentadas na Seção 2.5, continuam sendo válidas para o caso de comparação dos parâmetros de regressão de um modelo linear. Entre elas podemos destacar a sensibilidade do fator de Bayes à escolha das distribuições a priori dos parâmetros, e à técnica empregada para calculá-lo. Em alguns casos chegamos a conclusões que contradizem a intuição, como por exemplo que a restrição sob o espaço paramétrico não aumenta a evidência contra a hipótese nula. Ao aplicar a técnica frequentista, i.e., o teste alternativo, não chegamos a essa conclusão pois o valor crítico do teste alternativo é sempre menor que o valor crítico do teste usual.

O teste alternativo não requer de distribuições a priori, mas de um estimador do parâmetro sob a hipótese nula o que leva a uma perda do poder do teste. Como tópico de futuras pesquisas propomos o estudo do desempenho de outros estimadores de β em relação ao poder do teste alternativo.

Tabela 3.19: quantis amostrais para $\sigma = 0.13$

β	quantil(%)	X_1			X_2		
		teste baseado em			teste baseado em		
		s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
0.80	10.00	-2.682	-2.682	-2.867	-2.499	-2.499	-2.672
	20.00	-2.265	-2.265	-2.421	-2.275	-2.275	-2.432
	25.00	-2.172	-2.172	-2.322	-2.231	-2.231	-2.385
	50.00	-0.013	-0.013	-0.013	-0.004	-0.004	-0.004
	75.00	0.691	0.691	0.739	0.650	0.650	0.695
	90.00	1.333	1.333	1.425	1.312	1.312	1.403
	95.00	1.747	1.747	1.867	1.778	1.778	1.900
	97.50	2.213	2.213	2.366	2.169	2.169	2.319
	99.00	2.723	2.723	2.911	2.561	2.561	2.738
	99.50	3.002	3.002	3.210	3.084	3.084	3.297
	99.75	3.432	3.432	3.669	3.757	3.757	4.016
	99.90	3.821	3.821	4.084	3.874	3.874	4.141
	99.95	4.145	4.145	4.432	4.096	4.096	4.379
0.90	10.00	-2.479	-2.544	-2.650	-2.586	-2.586	-2.764
	20.00	-2.193	-2.267	-2.344	-2.165	-2.170	-2.314
	25.00	-2.060	-2.144	-2.202	-2.043	-2.045	-2.184
	50.00	-0.021	-0.021	-0.023	-0.029	-0.029	-0.031
	75.00	0.657	0.658	0.702	0.678	0.682	0.725
	90.00	1.290	1.311	1.379	1.173	1.231	1.254
	95.00	1.664	1.715	1.779	1.457	1.529	1.558
	97.50	2.062	2.134	2.205	1.671	1.735	1.787
	99.00	2.460	2.518	2.630	2.046	2.135	2.188
	99.50	2.732	2.791	2.920	2.253	2.418	2.409
	99.75	3.102	3.113	3.316	2.730	3.028	2.918
	99.90	3.381	3.558	3.615	2.845	3.275	3.042
	99.95	3.602	4.124	3.851	3.275	3.403	3.501
0.95	10.00	-1.981	-2.031	-2.118	-2.299	-2.500	-2.458
	20.00	-1.700	-1.861	-1.818	-1.927	-2.033	-2.060
	25.00	-1.579	-1.778	-1.688	-1.779	-1.927	-1.902
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.543	0.606	0.581	0.542	0.583	0.579
	90.00	1.032	1.126	1.104	0.950	1.054	1.015
	95.00	1.333	1.531	1.426	1.114	1.316	1.191
	97.50	1.643	1.765	1.756	1.311	1.447	1.402
	99.00	1.917	2.176	2.049	1.595	1.731	1.706
	99.50	2.176	2.407	2.327	1.734	1.945	1.853
	99.75	2.274	2.917	2.431	1.926	2.039	2.059
	99.90	2.344	3.051	2.505	1.934	2.237	2.068
	99.95	2.362	4.457	2.525	1.970	3.301	2.106

Tabela 3.20: quantis amostrais para $\sigma = 0.13$ (continuação)

β	quantil(%)	X_1			X_2		
		teste baseado em			teste baseado em		
		s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
0.99	10.00	-1.653	-1.855	-1.767	-1.628	-2.268	-1.740
	20.00	-1.389	-1.670	-1.485	-1.446	-1.792	-1.546
	25.00	-1.277	-1.577	-1.365	-1.404	-1.628	-1.501
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.277	0.352	0.296	0.208	0.242	0.222
	90.00	0.698	0.879	0.746	0.584	0.698	0.624
	95.00	1.002	1.254	1.071	0.760	0.934	0.813
	97.50	1.326	1.563	1.418	0.891	1.147	0.953
	99.00	1.636	1.966	1.748	1.098	1.376	1.173
	99.50	2.026	2.226	2.166	1.400	1.550	1.496
	99.75	2.226	2.349	2.380	1.584	1.720	1.693
	99.90	2.239	2.491	2.393	1.704	2.070	1.822
	99.95	2.288	2.612	2.446	1.720	2.352	1.838
1.00	10.00	-1.424	-1.840	-1.522	-1.678	-1.987	-1.794
	20.00	-1.166	-1.528	-1.247	-1.389	-1.612	-1.485
	25.00	-1.087	-1.406	-1.162	-1.298	-1.498	-1.387
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.160	0.202	0.171	0.143	0.173	0.153
	90.00	0.564	0.690	0.603	0.471	0.621	0.504
	95.00	0.915	1.116	0.978	0.678	0.849	0.725
	97.50	1.097	1.380	1.173	0.885	1.013	0.946
	99.00	1.435	1.683	1.534	1.073	1.269	1.147
	99.50	1.620	1.893	1.732	1.143	1.414	1.222
	99.75	1.773	2.178	1.896	1.465	1.631	1.566
	99.90	1.776	2.503	1.899	1.515	1.736	1.620
	99.95	1.924	2.689	2.057	1.800	1.800	1.924

Tabela 3.21: quantis amostrais para $\sigma = 0.13$ (continuação)

β	quantil(%)	X_3			X_4		
		teste baseado em			teste baseado em		
		s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
0.80	10.00	-2.427	-2.427	-2.595	-2.632	-2.648	-2.814
	20.00	-1.976	-1.976	-2.112	-2.177	-2.177	-2.327
	25.00	-1.908	-1.908	-2.039	-2.099	-2.127	-2.244
	50.00	0.049	0.049	0.052	0.064	0.064	0.069
	75.00	0.739	0.739	0.790	0.699	0.699	0.748
	90.00	1.379	1.379	1.474	1.372	1.374	1.466
	95.00	1.765	1.765	1.887	1.783	1.783	1.907
	97.50	2.108	2.108	2.254	2.014	2.014	2.153
	99.00	2.799	2.799	2.992	2.507	2.507	2.680
	99.50	3.354	3.354	3.585	2.889	2.889	3.089
	99.75	3.729	3.729	3.986	3.231	3.512	3.454
	99.90	4.561	4.561	4.876	3.512	3.821	3.755
	99.95	5.023	5.023	5.369	3.821	3.831	4.085
0.90	10.00	-2.483	-2.500	-2.654	-2.300	-2.300	-2.458
	20.00	-2.254	-2.312	-2.410	-1.858	-1.892	-1.987
	25.00	-2.183	-2.183	-2.334	-1.714	-1.844	-1.832
	50.00	0.021	0.021	0.023	0.005	0.005	0.005
	75.00	0.695	0.695	0.743	0.595	0.616	0.636
	90.00	1.357	1.357	1.451	1.065	1.109	1.139
	95.00	1.921	1.921	2.054	1.352	1.418	1.445
	97.50	2.312	2.312	2.472	1.696	1.730	1.813
	99.00	2.604	2.653	2.783	2.246	2.310	2.401
	99.50	2.827	2.841	3.022	2.423	2.648	2.591
	99.75	3.352	3.352	3.583	2.921	2.921	3.122
	99.90	3.548	3.548	3.793	3.738	3.738	3.996
	99.95	3.930	3.930	4.202	4.367	4.438	4.668
0.95	10.00	-2.042	-2.305	-2.183	-1.684	-1.947	-1.800
	20.00	-1.813	-1.904	-1.938	-1.466	-1.722	-1.567
	25.00	-1.704	-1.837	-1.822	-1.382	-1.636	-1.477
	50.00	-0.018	-0.018	-0.020	0.000	0.000	0.000
	75.00	0.627	0.657	0.670	0.485	0.530	0.519
	90.00	1.098	1.220	1.174	0.984	1.060	1.052
	95.00	1.481	1.625	1.583	1.282	1.443	1.371
	97.50	1.811	2.000	1.936	1.643	1.864	1.756
	99.00	2.228	2.258	2.382	2.066	2.189	2.209
	99.50	2.474	2.571	2.645	2.295	2.430	2.453
	99.75	2.685	2.685	2.870	2.447	2.863	2.616
	99.90	2.834	3.437	3.030	2.548	2.867	2.724
	99.95	3.791	3.791	4.053	3.382	3.382	3.616

Tabela 3.22: quantis amostrais para $\sigma = 0.13$ (continuação)

β	quantil(%)	X_3			X_4		
		teste baseado em			teste baseado em		
		s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
0.99	10.00	-1.560	-1.780	-1.668	-1.426	-1.620	-1.525
	20.00	-1.263	-1.576	-1.350	-1.175	-1.457	-1.256
	25.00	-1.199	-1.478	-1.281	-1.126	-1.395	-1.204
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.277	0.381	0.296	0.297	0.342	0.317
	90.00	0.723	0.969	0.773	0.715	0.832	0.764
	95.00	0.987	1.248	1.055	0.995	1.145	1.064
	97.50	1.173	1.551	1.254	1.303	1.515	1.393
	99.00	1.441	1.732	1.541	1.773	2.002	1.895
	99.50	1.666	1.854	1.781	1.893	2.206	2.024
	99.75	1.782	2.005	1.906	2.141	2.437	2.289
	99.90	1.848	2.080	1.976	2.317	2.447	2.477
	99.95	1.854	2.089	1.982	2.437	3.225	2.605
1	10.00	-1.467	-1.711	-1.569	-1.416	-1.685	-1.514
	20.00	-1.146	-1.525	-1.225	-1.165	-1.452	-1.245
	25.00	-1.036	-1.460	-1.108	-1.034	-1.360	-1.105
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.206	0.296	0.221	0.235	0.272	0.251
	90.00	0.597	0.753	0.639	0.660	0.788	0.706
	95.00	0.801	1.096	0.856	0.930	1.110	0.994
	97.50	0.993	1.443	1.062	1.159	1.424	1.239
	99.00	1.194	1.798	1.277	1.558	1.780	1.665
	99.50	1.684	2.073	1.800	1.652	2.054	1.766
	99.75	1.901	2.191	2.032	1.998	2.226	2.136
	99.90	2.191	2.377	2.342	2.226	2.434	2.380
	99.95	2.404	2.404	2.570	2.434	2.499	2.602

Tabela 3.23: quantis amostrais para $\sigma = 0.5$

β	quantil(%)	X_1			X_2		
		teste baseado em			teste baseado em		
		s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
0.80	10.00	-2.345	-2.397	-2.507	-2.244	-2.305	-2.399
	20.00	-2.053	-2.165	-2.195	-1.962	-1.998	-2.098
	25.00	-1.961	-2.062	-2.096	-1.890	-1.890	-2.020
	50.00	0.012	0.012	0.013	0.000	0.000	0.000
	75.00	0.638	0.665	0.682	0.556	0.582	0.594
	90.00	1.253	1.297	1.339	1.107	1.163	1.183
	95.00	1.658	1.728	1.772	1.400	1.417	1.496
	97.50	1.989	2.062	2.126	1.626	1.819	1.738
	99.00	2.360	2.369	2.523	1.876	2.144	2.005
	99.50	2.473	2.473	2.644	2.060	2.307	2.202
	99.75	2.703	2.988	2.890	2.303	2.345	2.462
	99.90	2.988	3.313	3.194	2.314	2.466	2.473
	99.95	3.333	3.333	3.563	2.437	2.473	2.605
0.90	10.00	-1.795	-1.996	-1.919	-2.123	-2.350	-2.270
	20.00	-1.538	-1.626	-1.644	-1.776	-1.891	-1.899
	25.00	-1.450	-1.540	-1.550	-1.715	-1.764	-1.834
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.423	0.461	0.452	0.343	0.359	0.367
	90.00	0.952	1.057	1.018	0.755	0.819	0.807
	95.00	1.281	1.370	1.369	1.054	1.137	1.126
	97.50	1.507	1.614	1.611	1.302	1.436	1.391
	99.00	1.836	1.848	1.962	1.579	1.811	1.689
	99.50	2.117	2.445	2.263	1.788	2.066	1.911
	99.75	2.468	2.966	2.638	1.895	2.210	2.025
	99.90	2.595	3.025	2.775	1.906	2.226	2.038
	99.95	3.270	3.812	3.496	2.318	3.163	2.478
0.95	10.00	-1.780	-1.948	-1.903	-1.891	-2.134	-2.022
	20.00	-1.567	-1.767	-1.675	-1.639	-1.865	-1.752
	25.00	-1.511	-1.596	-1.615	-1.586	-1.777	-1.695
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.374	0.410	0.400	0.290	0.307	0.310
	90.00	0.862	0.927	0.921	0.690	0.720	0.737
	95.00	1.086	1.233	1.161	0.929	1.023	0.993
	97.50	1.411	1.488	1.509	1.074	1.233	1.148
	99.00	1.683	1.774	1.799	1.380	1.544	1.475
	99.50	1.938	2.083	2.072	1.554	1.804	1.662
	99.75	2.141	2.231	2.289	1.840	1.982	1.967
	99.90	2.502	2.657	2.675	2.012	2.683	2.151
	99.95	3.623	3.623	3.874	2.287	3.357	2.445

Tabela 3.24: quantis amostrais para $\sigma = 0.5$ (continuação)

β	quantil(%)	$\mathbf{X}_1$			$\mathbf{X}_2$		
		teste baseado em			teste baseado em		
		s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
0.99	10.00	-1.515	-1.817	-1.619	-1.664	-1.915	-1.779
	20.00	-1.395	-1.576	-1.491	-1.400	-1.633	-1.497
	25.00	-1.317	-1.459	-1.408	-1.365	-1.559	-1.459
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.271	0.299	0.290	0.203	0.224	0.217
	90.00	0.681	0.740	0.728	0.602	0.652	0.644
	95.00	0.886	1.001	0.947	0.792	0.852	0.847
	97.50	1.142	1.271	1.221	0.968	1.034	1.035
	99.00	1.554	1.780	1.661	1.164	1.311	1.244
	99.50	1.914	2.001	2.046	1.283	1.475	1.372
	99.75	2.001	2.237	2.140	1.347	1.502	1.441
	99.90	2.081	2.287	2.225	1.375	1.507	1.470
	99.95	2.501	2.501	2.673	2.044	2.453	2.185
1.00	10.00	-1.453	-1.596	-1.554	-1.789	-1.867	-1.912
	20.00	-1.275	-1.439	-1.363	-1.468	-1.626	-1.570
	25.00	-1.213	-1.330	-1.297	-1.385	-1.598	-1.481
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.231	0.248	0.247	0.212	0.238	0.226
	90.00	0.666	0.713	0.712	0.575	0.616	0.615
	95.00	0.861	0.988	0.920	0.797	0.897	0.852
	97.50	1.218	1.346	1.302	1.006	1.163	1.075
	99.00	1.544	1.744	1.651	1.331	1.596	1.423
	99.50	1.842	2.136	1.969	1.623	2.006	1.735
	99.75	2.251	2.728	2.407	1.684	2.338	1.800
	99.90	3.043	3.122	3.253	1.981	2.361	2.118
	99.95	3.186	3.186	3.406	2.210	2.592	2.362

Tabela 3.25: quantis amostrais para $\sigma = 0.5$ (continuação)

β	quantil(%)	X_3			X_4		
		teste baseado em			teste baseado em		
		s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
0.80	10.00	-2.022	-2.022	-2.162	-1.741	-1.916	-1.861
	20.00	-1.834	-1.881	-1.961	-1.602	-1.671	-1.713
	25.00	-1.761	-1.794	-1.883	-1.545	-1.604	-1.651
	50.00	-0.023	-0.023	-0.025	0.000	0.000	0.000
	75.00	0.504	0.544	0.539	0.541	0.554	0.579
	90.00	1.116	1.182	1.193	1.073	1.129	1.147
	95.00	1.495	1.556	1.598	1.396	1.435	1.492
	97.50	1.816	1.840	1.941	1.592	1.650	1.702
	99.00	2.150	2.379	2.299	1.880	2.075	2.010
	99.50	2.461	2.612	2.631	2.114	2.317	2.260
	99.75	2.736	2.736	2.925	2.487	2.653	2.659
	99.90	3.017	3.017	3.225	2.663	2.682	2.847
	99.95	3.237	3.237	3.460	3.060	3.612	3.271
0.90	10.00	-2.101	-2.308	-2.246	-1.653	-1.763	-1.767
	20.00	-1.732	-1.989	-1.852	-1.462	-1.540	-1.563
	25.00	-1.704	-1.819	-1.822	-1.332	-1.414	-1.424
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.483	0.510	0.517	0.397	0.403	0.424
	90.00	0.974	1.067	1.041	0.891	0.952	0.953
	95.00	1.321	1.421	1.412	1.223	1.252	1.307
	97.50	1.636	1.781	1.748	1.458	1.523	1.558
	99.00	1.990	2.117	2.128	1.783	1.966	1.906
	99.50	2.152	2.279	2.301	1.962	2.056	2.098
	99.75	2.279	2.565	2.437	2.314	2.412	2.474
	99.90	2.365	3.207	2.528	2.462	2.492	2.632
	99.95	2.501	3.357	2.674	2.722	2.722	2.910
0.95	10.00	-1.887	-1.925	-2.017	-1.780	-1.867	-1.903
	20.00	-1.472	-1.682	-1.573	-1.524	-1.577	-1.629
	25.00	-1.409	-1.628	-1.506	-1.375	-1.524	-1.470
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.391	0.429	0.418	0.329	0.345	0.352
	90.00	0.832	0.884	0.889	0.779	0.856	0.833
	95.00	1.141	1.304	1.220	1.018	1.107	1.088
	97.50	1.511	1.592	1.615	1.396	1.534	1.492
	99.00	1.928	2.016	2.061	1.686	1.798	1.802
	99.50	2.120	2.338	2.267	1.819	1.895	1.944
	99.75	2.555	2.670	2.732	2.072	2.218	2.215
	99.90	2.650	2.759	2.833	2.155	2.328	2.303
	99.95	3.187	3.187	3.407	2.708	2.727	2.895

Tabela 3.26: quantis amostrais para $\sigma = 0.5$ (continuação)

β	quantil(%)	X_3			X_4		
		teste baseado em			teste baseado em		
		s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
0.99	10.00	-1.454	-1.639	-1.555	-1.473	-1.586	-1.575
	20.00	-1.246	-1.369	-1.333	-1.195	-1.236	-1.278
	25.00	-1.198	-1.290	-1.281	-1.133	-1.186	-1.211
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.318	0.354	0.340	0.326	0.348	0.349
	90.00	0.778	0.878	0.832	0.718	0.802	0.768
	95.00	1.091	1.191	1.167	1.073	1.167	1.147
	97.50	1.358	1.523	1.452	1.454	1.512	1.555
	99.00	1.713	1.778	1.832	1.631	1.766	1.744
	99.50	1.821	2.225	1.946	1.763	1.877	1.885
	99.75	2.003	2.281	2.141	2.135	2.155	2.283
	99.90	2.131	2.655	2.278	2.171	2.315	2.321
	99.95	2.415	3.862	2.581	2.312	2.483	2.471
1.00	10.00	-1.484	-1.734	-1.587	-1.379	-1.480	-1.474
	20.00	-1.277	-1.456	-1.365	-1.156	-1.271	-1.236
	25.00	-1.225	-1.364	-1.309	-1.077	-1.190	-1.152
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.243	0.291	0.260	0.258	0.281	0.276
	90.00	0.653	0.731	0.698	0.747	0.772	0.799
	95.00	0.909	1.086	0.972	0.982	1.059	1.050
	97.50	1.177	1.327	1.259	1.212	1.302	1.296
	99.00	1.425	1.639	1.524	1.502	1.767	1.605
	99.50	1.644	1.742	1.757	1.754	1.945	1.875
	99.75	1.822	1.969	1.948	1.922	2.085	2.054
	99.90	1.826	1.970	1.952	2.013	2.430	2.152
	99.95	1.969	2.037	2.105	2.063	2.783	2.205

Tabela 3.27: quantis amostrais para $\sigma = 1.0$

β	quantil(%)	X_1			X_2		
		teste baseado em			teste baseado em		
		s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
0.80	10.00	-1.702	-1.876	-1.819	-1.819	-1.892	-1.945
	20.00	-1.502	-1.579	-1.605	-1.622	-1.706	-1.734
	25.00	-1.358	-1.474	-1.452	-1.506	-1.649	-1.610
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.479	0.507	0.512	0.389	0.407	0.416
	90.00	0.963	1.007	1.029	0.852	0.884	0.911
	95.00	1.284	1.340	1.373	1.103	1.148	1.179
	97.50	1.598	1.663	1.709	1.286	1.352	1.375
	99.00	2.032	2.089	2.173	1.501	1.552	1.604
	99.50	2.165	2.213	2.315	1.565	1.624	1.673
	99.75	2.244	2.407	2.399	1.731	2.007	1.851
	99.90	2.404	2.416	2.570	1.800	2.011	1.924
	99.95	2.696	2.710	2.882	2.157	2.157	2.306
0.90	10.00	-1.532	-1.638	-1.638	-1.741	-1.970	-1.861
	20.00	-1.401	-1.460	-1.498	-1.551	-1.728	-1.658
	25.00	-1.358	-1.407	-1.452	-1.489	-1.652	-1.592
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.340	0.357	0.363	0.272	0.275	0.291
	90.00	0.829	0.877	0.886	0.673	0.725	0.719
	95.00	1.144	1.211	1.223	0.917	0.968	0.980
	97.50	1.447	1.484	1.547	1.104	1.169	1.180
	99.00	1.621	1.729	1.733	1.358	1.390	1.452
	99.50	1.824	2.080	1.950	1.540	1.626	1.647
	99.75	2.183	2.422	2.333	1.627	1.758	1.739
	99.90	2.479	2.579	2.650	1.673	1.978	1.788
	99.95	2.553	2.584	2.729	2.163	2.163	2.313
0.95	10.00	-1.634	-1.674	-1.747	-1.574	-1.667	-1.682
	20.00	-1.331	-1.462	-1.423	-1.365	-1.517	-1.460
	25.00	-1.287	-1.379	-1.376	-1.327	-1.426	-1.418
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.361	0.390	0.386	0.279	0.291	0.299
	90.00	0.803	0.878	0.858	0.685	0.716	0.732
	95.00	1.124	1.242	1.202	0.920	0.959	0.983
	97.50	1.439	1.548	1.538	1.165	1.219	1.246
	99.00	1.687	1.823	1.803	1.390	1.485	1.486
	99.50	1.771	2.131	1.893	1.562	1.613	1.670
	99.75	2.131	2.264	2.278	1.645	1.709	1.759
	99.90	2.171	2.288	2.321	1.709	1.947	1.827
	99.95	2.202	2.565	2.354	2.063	2.063	2.205

Tabela 3.28: quantis amostrais para $\sigma = 1.0$ (continuação)

β	quantil(%)	X_1			X_2		
		teste baseado em			teste baseado em		
		s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
0.99	10.00	-1.500	-1.620	-1.603	-1.677	-1.859	-1.793
	20.00	-1.327	-1.398	-1.419	-1.484	-1.590	-1.586
	25.00	-1.250	-1.361	-1.337	-1.415	-1.536	-1.513
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.269	0.275	0.287	0.218	0.230	0.233
	90.00	0.757	0.798	0.809	0.596	0.629	0.637
	95.00	1.015	1.092	1.085	0.820	0.884	0.877
	97.50	1.306	1.344	1.396	0.990	1.063	1.059
	99.00	1.501	1.706	1.605	1.267	1.366	1.354
	99.50	1.683	1.939	1.799	1.673	1.673	1.788
	99.75	1.863	1.975	1.992	1.826	1.959	1.952
	99.90	1.865	2.074	1.993	2.090	2.352	2.235
	99.95	2.144	2.269	2.292	2.323	2.437	2.484
1.00	10.00	-1.560	-1.656	-1.668	-1.673	-1.929	-1.789
	20.00	-1.197	-1.353	-1.280	-1.496	-1.659	-1.599
	25.00	-1.170	-1.271	-1.251	-1.434	-1.592	-1.533
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.216	0.231	0.231	0.193	0.206	0.207
	90.00	0.675	0.719	0.722	0.556	0.593	0.594
	95.00	0.991	1.098	1.060	0.755	0.796	0.807
	97.50	1.330	1.466	1.422	0.923	1.046	0.987
	99.00	1.588	1.773	1.698	1.316	1.436	1.407
	99.50	1.729	1.992	1.849	1.523	1.590	1.628
	99.75	1.842	2.025	1.970	1.754	1.898	1.876
	99.90	2.015	2.041	2.154	2.175	2.176	2.326
	99.95	2.025	2.074	2.164	2.436	2.436	2.604

Tabela 3.29: quantis amostrais para $\sigma = 1.0$ (continuação)

β	quantil(%)	X_3			X_4		
		teste baseado em			teste baseado em		
		s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
0.80	10.00	-1.864	-1.990	-1.992	-1.163	-1.297	-1.244
	20.00	-1.719	-1.750	-1.838	-1.040	-1.125	-1.112
	25.00	-1.681	-1.697	-1.797	-1.029	-1.095	-1.100
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.498	0.521	0.532	0.372	0.392	0.398
	90.00	1.014	1.043	1.084	0.798	0.832	0.853
	95.00	1.382	1.429	1.477	0.917	0.981	0.980
	97.50	1.581	1.714	1.690	1.072	1.161	1.146
	99.00	2.193	2.485	2.344	1.255	1.344	1.341
	99.50	2.485	2.685	2.657	1.340	1.472	1.432
	99.75	2.695	2.979	2.881	1.479	1.582	1.581
	99.90	2.963	3.183	3.168	1.517	1.651	1.622
	99.95	3.831	4.528	4.096	1.549	1.673	1.656
0.90	10.00	-2.008	-2.024	-2.147	-1.185	-1.263	-1.267
	20.00	-1.542	-1.631	-1.648	-1.031	-1.126	-1.102
	25.00	-1.493	-1.516	-1.596	-0.977	-1.056	-1.044
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.457	0.487	0.489	0.369	0.383	0.394
	90.00	0.870	0.935	0.930	0.757	0.799	0.809
	95.00	1.121	1.234	1.198	0.900	0.981	0.962
	97.50	1.375	1.517	1.470	1.021	1.123	1.092
	99.00	1.753	2.006	1.874	1.116	1.266	1.193
	99.50	2.001	2.133	2.140	1.201	1.361	1.283
	99.75	2.133	2.354	2.280	1.256	1.500	1.342
	99.90	2.310	2.398	2.469	1.279	1.645	1.368
	99.95	2.708	2.708	2.895	1.569	1.667	1.677
0.95	10.00	-1.642	-1.769	-1.755	-1.178	-1.296	-1.259
	20.00	-1.492	-1.575	-1.596	-1.049	-1.179	-1.122
	25.00	-1.446	-1.500	-1.546	-1.024	-1.092	-1.095
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.338	0.361	0.362	0.319	0.338	0.341
	90.00	0.753	0.814	0.805	0.713	0.752	0.762
	95.00	1.028	1.097	1.099	0.860	0.926	0.919
	97.50	1.268	1.369	1.355	0.999	1.068	1.068
	99.00	1.634	1.781	1.746	1.161	1.301	1.241
	99.50	1.996	2.315	2.134	1.261	1.349	1.348
	99.75	2.551	2.551	2.727	1.335	1.400	1.427
	99.90	2.862	3.332	3.060	1.339	1.425	1.431
	99.95	3.103	3.482	3.318	1.355	1.498	1.449

Tabela 3.30: quantis amostrais para $\sigma = 1.0$ (continuação)

β	quantil(%)	X_3			X_4		
		teste baseado em			teste baseado em		
		s_{pr}	s_p	$\hat{\sigma}_R$	s_{pr}	s_p	$\hat{\sigma}_R$
0.99	10.00	-1.432	-1.599	-1.531	-1.108	-1.262	-1.185
	20.00	-1.212	-1.287	-1.296	-1.036	-1.106	-1.108
	25.00	-1.193	-1.250	-1.275	-1.011	-1.075	-1.080
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.288	0.318	0.308	0.270	0.283	0.289
	90.00	0.765	0.796	0.818	0.699	0.736	0.748
	95.00	1.028	1.176	1.099	0.895	0.957	0.956
	97.50	1.278	1.438	1.366	1.006	1.093	1.075
	99.00	1.634	1.814	1.746	1.181	1.284	1.263
	99.50	1.797	2.064	1.921	1.352	1.537	1.445
	99.75	2.065	2.248	2.208	1.378	1.624	1.473
	99.90	2.100	2.356	2.244	1.511	1.628	1.615
	99.95	3.012	3.510	3.220	1.519	1.734	1.623
1.00	10.00	-1.576	-1.813	-1.685	-1.046	-1.154	-1.118
	20.00	-1.337	-1.518	-1.430	-0.997	-1.075	-1.066
	25.00	-1.231	-1.304	-1.316	-0.972	-1.044	-1.039
	50.00	0.000	0.000	0.000	0.000	0.000	0.000
	75.00	0.310	0.334	0.331	0.298	0.312	0.318
	90.00	0.770	0.847	0.824	0.721	0.758	0.771
	95.00	1.039	1.132	1.111	0.894	0.940	0.956
	97.50	1.239	1.393	1.325	0.991	1.059	1.059
	99.00	1.560	1.674	1.667	1.161	1.303	1.241
	99.50	1.699	1.982	1.817	1.227	1.346	1.312
	99.75	1.982	2.472	2.119	1.331	1.447	1.423
	99.90	2.392	2.573	2.557	1.459	1.465	1.560
	99.95	2.497	3.093	2.670	1.738	1.746	1.858

Capítulo 4

Testes de hipóteses em condições não regulares

Neste capítulo estudamos testes de hipóteses sobre os elementos da matriz de covariância num modelo de efeitos aleatórios utilizando a estatística da razão de verossimilhanças apresentada por Vu e Zhou (1997). Esses autores estendem os resultados obtidos por Self e Liang (1987) sobre o teste da razão de verossimilhanças em condições não regulares para o caso em que as variáveis aleatórias não são identicamente distribuídas.

Na realidade, Vu e Zhou (1997) generalizam os resultados para estatísticas definidas como a diferença entre os valores de uma função de estimação avaliada nos máximos locais de dois subconjuntos diferentes do espaço paramétrico, assumindo que o verdadeiro valor do parâmetro está possivelmente na fronteira de cada um deles. Sob certas condições de regularidade, obtêm resultados sobre a existência e consistência de estimadores locais do verdadeiro valor do parâmetro. Esses resultados são descritos na Seção 4.1. Na Seção 4.2 analisamos as condições que o modelo de efeitos aleatórios deve satisfazer para que os resultados de Vu e Zhou (1997) possam ser aplicados com a finalidade de se testarem hipóteses como as mencionadas na Seção 1.2.

4.1 Generalização do Teste da Razão de Verossimilhanças sob Condições Não Regulares

4.1.1 Resultados Fundamentais

Consideremos uma amostra de n observações de variáveis aleatórias $Y_1, \dots, Y_n$ e suponhamos que a distribuição de Y_i pertence a uma família $F_i(y, \theta)$ em que $\theta \in \mathbb{R}^p$ é o parâmetro de interesse.

Além disso, seja $F_i(y, \theta_0)$ a distribuição de Y_i , em que $\theta_0 = (\theta_{10}, \dots, \theta_{p0})^t$ é chamado de *verdadeiro valor do parâmetro*. Consideramos uma função de estimação da forma

$$L_n(\theta) = \sum_{i=1}^n g(Y_i, \theta)$$

em que g é uma função de $\mathbb{R}^{p+1} \rightarrow \mathbb{R}$, por exemplo, a função de verossimilhança. Assumimos que $\theta \in \Theta \subseteq \mathbb{R}^p$, em que Θ é um cone da forma

$$\Theta = \left\{ \theta : \theta = \theta_0 + \tilde{\theta}_1 \mathbf{u}_1 + \dots + \tilde{\theta}_p \mathbf{u}_p, \tilde{\theta}_j \in I_j, j = 1, \dots, p \right\},$$

e $\mathbf{u}_j$ são vetores unitários, linearmente independentes, $j = 1, \dots, p$, e os I_j são intervalos fechados, abertos ou semiabertos que contêm o zero.

Seja

$$\mathbf{T} = [\mathbf{u}_1, \dots, \mathbf{u}_p],$$

a matriz $(p \times p)$ de direções que definem Θ .

Vamos definir as derivadas em Θ . Para cada $j = 1, \dots, p$, a derivada direcional de $L_n(\theta)$ na direção de $\mathbf{u}_j$ é denotada por $D_{\mathbf{u}_j} L_n(\theta)$ e é definida como

$$D_{\mathbf{u}_j} L_n(\theta) = \lim_{t \rightarrow 0} \frac{L_n(\theta + t\mathbf{u}_j) - L_n(\theta)}{t}, \quad (4.1)$$

se $\tilde{\theta}_j$ é ponto interior de I_j , ou seja, é a derivada direcional usual de $L_n(\theta)$ na direção de

$\mathbf{u}_j$ e como

$$D_{\mathbf{u}_j} L_n(\boldsymbol{\theta}) = \lim_{h \rightarrow 0, h + \tilde{\theta}_j \in I_j} D_{\mathbf{u}_j} L_n(\boldsymbol{\theta} + h \mathbf{u}_j), \quad (4.2)$$

se $\tilde{\theta}_j$ está na fronteira de I_j .

Analogamente podemos definir as derivadas segundas.

- Se $\tilde{\theta}_j$ é ponto interior de I_j , denotamos por $D_{\mathbf{u}_j} D_{\mathbf{u}_l} L_n(\boldsymbol{\theta})$ a derivada direcional usual de $D_{\mathbf{u}_l} L_n(\boldsymbol{\theta})$ na direção de $\mathbf{u}_j$.
- Se $\tilde{\theta}_j$ está na fronteira de I_j

$$D_{\mathbf{u}_j} D_{\mathbf{u}_l} L_n(\boldsymbol{\theta}) = \lim_{h \rightarrow 0, h + \tilde{\theta}_j \in I_j} D_{\mathbf{u}_j} D_{\mathbf{u}_l} L_n(\boldsymbol{\theta} + h \mathbf{u}_j). \quad (4.3)$$

Agora definimos a derivada de $L_n(\boldsymbol{\theta})$ com respeito a $\boldsymbol{\theta}$ pelo vetor

$$\mathbf{S}_n(\boldsymbol{\theta}) = \mathbf{T}^t [D_{\mathbf{u}_1} L_n(\boldsymbol{\theta}), \dots, D_{\mathbf{u}_p} L_n(\boldsymbol{\theta})]^t. \quad (4.4)$$

A matriz

$$\mathbf{F}_n(\boldsymbol{\theta}) = -\mathbf{T}^t [D_{\mathbf{u}_j} D_{\mathbf{u}_l} L_n(\boldsymbol{\theta})] \mathbf{T}, \quad (4.5)$$

é análoga à matriz hessiana no caso regular.

Definimos também

$$\mathbf{D}_n = E [\mathbf{S}_n(\boldsymbol{\theta}_0) \mathbf{S}_n^t(\boldsymbol{\theta}_0)] \quad (4.6)$$

e

$$\mathbf{G}_n = E(\mathbf{F}_n(\boldsymbol{\theta}_0)) \quad (4.7)$$

que é análoga à matriz de informação no caso regular.

Operações básicas como expansões de Taylor podem ser facilmente obtidas utilizando estas definições.

Consideremos as seguintes duas suposições sobre o espaço paramétrico e sobre a função $L_n(\boldsymbol{\theta})$.

A1) Para um entorno $\mathcal{N}$ de θ_0 , a função $L_n(\theta)$ é contínua em $\Theta \cap \mathcal{N}$. As derivadas direcionais de primeira e segunda ordem, $D_{u_j} L_n(\theta)$ e $D_{u_j} D_{u_l} L_n(\theta)$, $j, l = 1, \dots, p$ existem, são finitas e contínuas em $\Theta \cap \mathcal{N}$.

A2) Dado o conjunto Ω e o espaço paramétrico Θ , existe um cone fechado C_Ω com vértice em θ_0 tal que

$$C_\Omega \subseteq \Theta \text{ e } C_\Omega \cap \mathcal{N} = \Omega \cap \mathcal{N}$$

em que $\mathcal{N}$ é um entorno fechado de θ_0 .

Para obter resultados sobre a existência, a consistência e distribuição assintótica de um estimador de máxima verossimilhança, necessitamos as seguintes suposições sobre o comportamento assintótico das matrizes das derivadas primeiras e segundas e suas esperanças tomadas com respeito à verdadeira distribuição.¹

Primeiramente, para qualquer $A > 0$ fixo, definamos os seguintes subconjuntos de $\mathbb{R}^p$

$$N_n(A) = \{\theta \in \Theta: (\theta - \theta_0)^t \mathbf{G}_n(\theta - \theta_0) \leq A^2\}, \quad (4.8)$$

$$M_n(A) = \{\theta \in \Theta: (\theta - \theta_0)^t \mathbf{G}_n(\theta - \theta_0) = A^2\}. \quad (4.9)$$

Então consideremos as suposições

B1) $E(\mathbf{S}_n(\theta_0)) = 0$ e as matrizes $\mathbf{D}_n$ e $\mathbf{G}_n$ são finitas.

B2) $\lambda_{\min}(\mathbf{G}_n) \rightarrow \infty$. (Quando (B2) é verdadeira, $\mathbf{G}_n$ é positiva definida para n suficientemente grande).

$$\text{B3) } \sup_{\theta \in N_n(A)} \left\| \mathbf{G}_n^{-1/2} \mathbf{F}_n(\theta) \mathbf{G}_n^{-t/2} - \mathbf{I}_p \right\|_1 \xrightarrow{P} 0$$

¹Para qualquer matriz simétrica definida positiva $\mathbf{A}$, seja $\mathbf{A}^{1/2}(\mathbf{A}^{t/2})$ a “raiz quadrada” esquerda (direita), ou seja tais que $\mathbf{A}^{1/2} \mathbf{A}^{t/2} = \mathbf{A}$ e onde $\mathbf{A}^{t/2} = (\mathbf{A}^{1/2})^t$.

Denotamos por $|\mathbf{y}|$ o módulo de um vetor $\mathbf{y} \in \mathbb{R}^p$, $\|\mathbf{A}\|_1$ a soma dos valores absolutos dos elementos de uma matriz $\mathbf{A}$, e $\lambda_{\min}(\mathbf{A})$ e $\lambda_{\max}(\mathbf{A})$, os autovalores mínimo e máximo da matriz simétrica $\mathbf{A}$.

B4) Para alguma matriz definida positiva $\mathbf{V}$, $\left\| \mathbf{G}_n^{-1/2} \mathbf{D}_n \mathbf{G}_n^{-1/2} - \mathbf{V} \right\|_1 \xrightarrow{P} 0$. (Quando (B2) e (B4) são verdadeiras, $\mathbf{D}_n$ é positiva definida para n suficientemente grande).

B5) $\mathbf{D}_n^{-1/2} \mathbf{S}_n(\boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, \mathbf{I}_p)$.

Vu e Zhou (1997) mostram (Teorema 3.1) que para verificar B3), basta verificar que

$$\text{tr} \frac{1}{n} (\mathbf{F}_n(\boldsymbol{\theta}_0^*) - \mathbf{G}_n) = o_p(1), \quad (4.10)$$

$$\sup_{\boldsymbol{\theta}^* \in N_n(A)} \text{tr} \frac{1}{n} (\mathbf{F}_n(\boldsymbol{\theta}^*) - \mathbf{F}_n(\boldsymbol{\theta}_0^*)) = o_p(1). \quad (4.11)$$

Dizemos que uma seqüência de eventos $\{A_n\}$ acontece com probabilidade aproximada 1 (CPA1) se $P(A_n) \rightarrow 1$ quando $n \rightarrow \infty$. Um estimador $\hat{\boldsymbol{\theta}}_n$ de $\boldsymbol{\theta}_0$ é chamado estimador máximo (EM) em Ω se $L_n(\hat{\boldsymbol{\theta}}_n)$ é o valor máximo de $L_n(\boldsymbol{\theta})$ sobre a intersecção de Ω com um entorno aberto (possivelmente dependendo de n) de $\boldsymbol{\theta}_0$. Tal estimador será dito localmente único CPA1 se o evento *existe um único máximo nesta intersecção* acontece CPA1. Especificamente pode-se mostrar que o evento “existe um único máximo de $L_n(\boldsymbol{\theta})$ em $[N_n(A) \cap \Omega] \setminus M_n(A)$ ” acontece CPA1 para A suficientemente grande.

Para o primeiro teorema (que enunciamos adiante) sobre a existência e consistência de EM, podemos substituir (A2), (B3) e (B4) pelas seguintes condições mais fracas:

A2') Dados o conjunto Ω e o espaço paramétrico Θ então Ω contém $\boldsymbol{\theta}_0$ e a intersecção entre Ω e $\mathcal{N}$ (um entorno fechado $\boldsymbol{\theta}_0$) é um conjunto fechado de $\mathbb{R}^p$.

B3') Existe uma constante $c > 0$ tal que, para cada $A > 0$,

$$P \left(\inf_{\boldsymbol{\theta} \in N_n(A)} \lambda_{\min} (\mathbf{G}_n^{-1/2} \mathbf{F}_n(\boldsymbol{\theta}) \mathbf{G}_n^{-1/2}) \leq c \right) \rightarrow 0, \quad n \rightarrow \infty$$

B4') $\mathbf{G}_n^{-1/2} \mathbf{S}_n(\boldsymbol{\theta}_0)$, é justo (tight), isto é

$$\lim_{A \rightarrow \infty} \limsup_{n \rightarrow \infty} P(|\mathbf{G}_n^{-1/2} \mathbf{S}_n(\boldsymbol{\theta}_0)| > A) = 0.$$

Notemos que B4') implica que $\mathbf{G}_n^{-1/2} \mathbf{S}_n(\boldsymbol{\theta}_0) = O_p(1)$.

Teorema 4.1.1. *Seja Ω subconjunto de Θ que satisfaz (A2'). Se as condições (A1), (B1), (B2) e (B3'), (B4') são verdadeiras, então um estimador EM, $\hat{\boldsymbol{\theta}}_n$, de $\boldsymbol{\theta}_0$ sobre Ω existe, é localmente único CPA1 e é consistente para $\boldsymbol{\theta}_0$.*

É possível que existam varios estimadores máximos. O teorema acima diz que entre esses estimadores existe um EM que é consistente e localmente único CPA1. Este estimador EM é de fato um maximizador global dentro de um entorno de $\boldsymbol{\theta}_0$.

Sejam Ω e τ dois subconjuntos fixos de Θ que satisfazem (A2) com seus correspondentes cones C_Ω e C_τ . Sejam $\mathbf{T}_n$ matrizes ortogonais arbitrárias não estocásticas e definamos

$$\tilde{C}_{\Omega_n} = \left\{ \tilde{\boldsymbol{\theta}} : \tilde{\boldsymbol{\theta}} = \mathbf{T}_n \mathbf{G}_n^{t/2} (\boldsymbol{\theta} - \boldsymbol{\theta}_0), \boldsymbol{\theta} \in C_\Omega \right\}. \quad (4.12)$$

Similarmente, podemos definir $\tilde{C}_{\tau_n}$. Note que as matrizes $\mathbf{T}_n$, na definição de $\tilde{C}_{\Omega_n}$, podem ser diferentes daquelas da definição de $\tilde{C}_{\tau_n}$. Para o próximo teorema (enunciado adiante) necessitamos uma suposição adicional sobre o comportamento de Ω e τ .

A3) Dados o conjunto Ω e o espaço paramétrico Θ , existe um cone fechado $\tilde{C}_\Omega$ com vértice em $\mathbf{0}$, não dependendo de n , tal que os conjuntos $\tilde{C}_{\Omega_n}$ coincidem com $\tilde{C}_\Omega$ assintoticamente, no sentido que quando $n \rightarrow \infty$,

$$\sup_{|\boldsymbol{\beta}|=1} \left| \inf_{\boldsymbol{\theta} \in \tilde{C}_{\Omega_n}} |\boldsymbol{\beta} - \boldsymbol{\theta}|^2 - \inf_{\boldsymbol{\theta} \in \tilde{C}_\Omega} |\boldsymbol{\beta} - \boldsymbol{\theta}|^2 \right| \rightarrow 0.$$

Sejam $\hat{\boldsymbol{\theta}}_n^1$ e $\hat{\boldsymbol{\theta}}_n^2$ os máximos locais de $L_n(\boldsymbol{\theta})$ sobre Ω e τ respectivamente obtidos como no Teorema 4.1.1. Definimos uma estatística do tipo razão de verossimilhanças como

$$d_n = 2 \left[L_n(\hat{\boldsymbol{\theta}}_n^2) - L_n(\hat{\boldsymbol{\theta}}_n^1) \right]. \quad (4.13)$$

Denotamos por $\mathbf{N} = (N_1, \dots, N_p)^t$ um vetor aleatório distribuição normal multivariada com media zero e matriz de covariância $\mathbf{V}$.

Teorema 4.1.2. *Suponha que (A1), (A2) e (A3) são verdadeiros para Ω e τ . Suponhamos também que são válidos (B1)-(B5). Então a distribuição assintótica de d_n existe e é a mesma que a distribuição de*

$$\inf_{\theta \in \tilde{C}_\Omega} |\mathbf{N} - \theta|^2 - \inf_{\theta \in \tilde{C}_\tau} |\mathbf{N} - \theta|^2. \quad (4.14)$$

4.1.2 Aplicações

Queremos obter a distribuição assintótica da estatística da razão de verossimilhanças para o teste da hipóteses sobre a necessidade de inclusão do(s) efeito(s) aleatório(s) no modelo (1.9), i.e., o teste das hipóteses (1.12) e das hipóteses (1.13) e (1.14). Stram e Lee (1994) analisam estas hipóteses aplicando a metodologia dada por Self e Liang (1987) mas não levam em conta o fato que esses resultados valem só para variáveis aleatórias independentes e identicamente distribuídas, que em geral não é o caso para o modelo de efeitos aleatórios definido no Exemplo 1.3 por (1.8) e (1.9). De (1.8) e (1.9) temos

$$E(\mathbf{y}_i) = \mathbf{X}_i \boldsymbol{\alpha} \quad (4.15)$$

$$Var(\mathbf{y}_i) = \boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^t + \sigma^2 \mathbf{I}_{n_i} \quad (4.16)$$

Se as respostas associadas aos diferentes indivíduos são independentes, com as mesmas matrizes de especificação $\mathbf{X}_i$ e $\mathbf{Z}_i$, então $\mathbf{y}_1, \dots, \mathbf{y}_n$ podem ser consideradas como n observações independentes do mesmo vetor aleatório e desta forma é possível aplicar o Teorema 3 de Self e Liang (1987).

Em nosso modelo, para $i = 1, \dots, n$, $\mathbf{X}_i$ tem posto completo, e existem constantes K , $\tilde{K}$ e Υ_1 tais que

- $n_i \leq K$,

-

$$\tilde{K} = \max_{j,l=1,\dots,k', i=1,\dots,n} \left\{ \text{tr}(\mathbf{X}_i^t \mathbf{z}_{ij} \mathbf{z}_{il}^t \mathbf{X}_i) \right\}, \quad (4.17)$$

em que $\mathbf{z}_{ij}$ é a coluna j -ésima de $\mathbf{Z}_i$.

•

$$\max_{j,l=1,\dots,k', i=1,\dots,n} \{\tau_{ijl}\} = \Upsilon_1, \quad (4.18)$$

com $\text{tr}(\mathbf{z}_{ij}\mathbf{z}_{il}^t) = \tau_{ijl}$, $j, l = 1, \dots, k'$;

O logaritmo da função de verossimilhança correspondente ao modelo (1.9) é

$$l(\mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\theta}) = C - \frac{1}{2} \sum_{i=1}^n \ln |\boldsymbol{\Sigma}_i(\boldsymbol{\theta})| - \frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})^t \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})$$

com $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ definida em (1.10) e C denotando uma constante que não depende nem de $\boldsymbol{\alpha}$ nem de $\boldsymbol{\theta}$. $\mathbf{D}$ pode ser escrita na forma linear $\mathbf{D} = \sum_{g=1}^{k-1} \theta_g \mathbf{A}_g$ onde as matrizes $\mathbf{A}_g$ ($q \times q$) são matrizes conhecidas, $k = p - p'$ e os θ_g ($g = 1, \dots, k - 1$) são os componentes de $\boldsymbol{\theta}$. Assim cada $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ pode também ser escrita na forma linear

$$\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \sum_{g=1}^k \theta_g \mathbf{G}_{ig}$$

com $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{k-1}, \theta_k = \sigma^2)$ e

$$\mathbf{G}_{ig} = \begin{cases} \mathbf{Z}_i \mathbf{A}_g \mathbf{Z}_i^t & \text{para } g = 1, \dots, k-1 \\ \mathbf{I}_{n_i} & \text{para } g = k. \end{cases} \quad (4.19)$$

Definindo $\boldsymbol{\theta}^* = (\boldsymbol{\alpha}, \boldsymbol{\theta})$, de (4.4) temos $\mathbf{S}_n(\boldsymbol{\theta}^*) = (\mathbf{s}_{n\boldsymbol{\alpha}}^t(\boldsymbol{\theta}^*), \mathbf{s}_{n\boldsymbol{\theta}}^t(\boldsymbol{\theta}^*))^t$ com $\mathbf{s}_{n\boldsymbol{\alpha}}(\boldsymbol{\theta}^*)$ e $\mathbf{s}_{n\boldsymbol{\theta}}(\boldsymbol{\theta}^*)$ denotando vetores com dimensões $p' \times 1$ e $k \times 1$, respectivamente e dados por

$$\begin{aligned} \mathbf{s}_{n\boldsymbol{\alpha}}(\boldsymbol{\theta}^*) &= \partial l / \partial \boldsymbol{\alpha} = \sum_{i=1}^n \mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}) \mathbf{e}_i \\ &= \left(\sum_{i=1}^n \sum_{j=1}^{n_i} (\mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0))_{1j} e_{ij}, \dots, \sum_{i=1}^n \sum_{j=1}^{n_i} (\mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0))_{n_{ij}} e_{ij} \right) \\ &= \sum_{i=1}^n \mathbf{s}_{n\boldsymbol{\alpha}i}(\boldsymbol{\theta}^*), \end{aligned} \quad (4.20)$$

em que $\mathbf{e}_i = (e_{i1}, \dots, e_{in_i})^t = (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}),$ e

$$\mathbf{s}_{n\alpha i}(\boldsymbol{\theta}^*) = \mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}) \mathbf{e}_i, \quad (4.21)$$

e

$$\begin{aligned} \mathbf{s}_{n\theta}(\boldsymbol{\theta}^*) &= [(\partial l / \partial \theta_g)] \\ &= \left[\left\langle -\frac{1}{2} \sum_{i=1}^n \text{tr} \left(\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}) \dot{\boldsymbol{\Sigma}}_{i(g)}(\boldsymbol{\theta}) \right) + \frac{1}{2} \sum_{i=1}^n \mathbf{e}_i^t \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}) \dot{\boldsymbol{\Sigma}}_{i(g)}(\boldsymbol{\theta}) \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}) \mathbf{e}_i \right\rangle_g \right] \\ &= \left[\left\langle \sum_{i=1}^n \mathbf{s}_{n\theta i}(\boldsymbol{\theta}^*) \right\rangle_g \right] \end{aligned} \quad (4.22)$$

com $[\langle \rangle_a]$ denotando um vetor cujo a -ésimo elemento é o termo entre $\langle \rangle$,

$$\mathbf{s}_{n\theta i}(\boldsymbol{\theta}^*) = -\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}) \dot{\boldsymbol{\Sigma}}_{i(g)}(\boldsymbol{\theta}) \right) + \mathbf{e}_i^t \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}) \dot{\boldsymbol{\Sigma}}_{i(g)}(\boldsymbol{\theta}) \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}) \mathbf{e}_i \quad (4.23)$$

e $\dot{\boldsymbol{\Sigma}}_{i(g)}(\boldsymbol{\theta}) = \partial \boldsymbol{\Sigma}_i(\boldsymbol{\theta}) / \partial \theta_g$.

Analogamente de (4.5), temos

$$\mathbf{F}_n(\boldsymbol{\theta}^*) = \begin{bmatrix} \mathbf{F}_{n\alpha\alpha}(\boldsymbol{\theta}^*) & \mathbf{F}_{n\alpha\theta}(\boldsymbol{\theta}^*) \\ \mathbf{F}_{n\alpha\theta}^t(\boldsymbol{\theta}^*) & \mathbf{F}_{n\theta\theta}(\boldsymbol{\theta}^*) \end{bmatrix},$$

com

$$\mathbf{F}_{n\alpha\alpha}(\boldsymbol{\theta}^*) = \sum_{i=1}^n \mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}) \mathbf{X}_i, \quad (4.24)$$

$$\mathbf{F}_{n\alpha\theta}(\boldsymbol{\theta}^*) = \left[\left\langle \sum_{i=1}^n \mathbf{x}_{ij}^t \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}) \mathbf{G}_{ig} \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}) \mathbf{e}_i \right\rangle_{jg} \right], \quad (4.25)$$

e

$$\begin{aligned} & \mathbf{F}_{n\theta\theta}(\boldsymbol{\theta}^*) \\ &= \left[\left\langle \frac{1}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{G}_{ig} \Sigma_i^{-1}(\boldsymbol{\theta}) [2\mathbf{e}_i \mathbf{e}_i^t - \Sigma_i(\boldsymbol{\theta})] \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{G}_{ih} \right\rangle_{gh} \right], \end{aligned} \quad (4.26)$$

em que $\mathbf{x}_{ij}$ é a j -ésima coluna de $\mathbf{X}_i$, as matrizes $\mathbf{G}_{ig}$ e $\mathbf{G}_{ih}$ são dadas por (4.19) e $[\langle \rangle]_{ab}$ denota uma matriz cujo elemento de ordem (a, b) é dado pela expressão entre $\langle \rangle$.

Similarmente, de (4.7) temos

$$\mathbf{G}_n = \begin{bmatrix} \mathbf{G}_{n\alpha\alpha} & \mathbf{G}_{n\alpha\theta} \\ \mathbf{G}_{n\alpha\theta}^t & \mathbf{G}_{n\theta\theta} \end{bmatrix}$$

com

$$\mathbf{G}_{n\alpha\alpha} = \sum_{i=1}^n \mathbf{X}_i^t \Sigma_i^{-1}(\boldsymbol{\theta}_0) \mathbf{X}_i, \quad (4.27)$$

$$\mathbf{G}_{n\alpha\theta} = \mathbf{0}, \quad (4.28)$$

e

$$\mathbf{G}_{n\theta\theta} = \left[\left\langle \frac{1}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1}(\boldsymbol{\theta}_0) \mathbf{G}_{ig} \Sigma_i^{-1}(\boldsymbol{\theta}_0) \mathbf{G}_{ih} \right\rangle_{gh} \right]. \quad (4.29)$$

Detalhes sobre os cálculos envolvidos para a obtenção de (4.20) a (4.29) podem ser encontradas por exemplo, em Andreoni (1989).

Verificaremos inicialmente que as condições B1) e B4) são satisfeitas.

Como $\mathbf{e}_i \sim \mathbf{N}(\mathbf{0}, \Sigma_i)$, de (4.20), temos que $E(s_{n\alpha}(\boldsymbol{\theta}_0^*)) = \mathbf{0}$. Também podemos concluir de (4.22)-(4.23) que

$$\begin{aligned} E\left((s_{n\theta}(\boldsymbol{\theta}_0^*))_g\right) &= \frac{-1}{2} \sum_{i=1}^n \text{tr} \left\{ \Sigma_i^{-1}(\boldsymbol{\theta}_0) \dot{\Sigma}_{i(g)}(\boldsymbol{\theta}_0) \right\} \\ &+ \frac{1}{2} \sum_{i=1}^n \text{tr} \left\{ \Sigma_i^{-1}(\boldsymbol{\theta}_0) \dot{\Sigma}_{i(g)}(\boldsymbol{\theta}_0) \Sigma_i^{-1}(\boldsymbol{\theta}_0) \Sigma_i(\boldsymbol{\theta}_0) \right\} = 0. \end{aligned}$$

Logo $E(S_n(\theta_0^*)) = 0$.

Dado que $E(S_n(\theta_0^*)) = 0$, de (4.6) temos

$$\begin{aligned} \mathbf{D}_n &= E(S_n(\theta_0^*) S_n(\theta_0^*)^t) = Cov(S_n(\theta_0^*)) \\ &= \begin{bmatrix} Cov(s_{n\alpha}(\theta_0^*)) & Cov(s_{n\alpha}(\theta_0^*), s_{n\theta}(\theta_0^*)) \\ (Cov(s_{n\alpha}(\theta_0^*), s_{n\theta}(\theta_0^*)))^t & Cov(s_{n\theta}(\theta_0^*)) \end{bmatrix}. \end{aligned}$$

Notemos primeiramente que

$$\begin{aligned} Cov(s_{n\alpha}(\theta_0^*)) &= \sum_{i=1}^n \mathbf{X}_i^t \Sigma_i^{-1}(\theta_0) \Sigma_i(\theta_0) (\Sigma_i^{-1}(\theta_0))^t \mathbf{X}_i \\ &= \sum_{i=1}^n \mathbf{X}_i^t \Sigma_i^{-1}(\theta_0) \mathbf{X}_i, \end{aligned} \quad (4.30)$$

que é igual a $\mathbf{G}_{n\alpha\alpha}$ definida em (4.27). Considerando que dadas duas matrizes $\mathbf{A}$ e $\mathbf{B}$ simétricas e $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{V})$ então $Cov(\mathbf{B}\mathbf{x}, \mathbf{x}^t \mathbf{A}\mathbf{x}) = 2\mathbf{B}\mathbf{V}\mathbf{A}\boldsymbol{\mu}$ e $Cov(\mathbf{x}^t \mathbf{B}\mathbf{x}, \mathbf{x}^t \mathbf{A}\mathbf{x}) = 2\text{tr}(\mathbf{B}\mathbf{V}\mathbf{A}\mathbf{V}) + 4\boldsymbol{\mu}^t \mathbf{B}\mathbf{V}\mathbf{A}\boldsymbol{\mu}$ (ver por exemplo, Searle, 1971) e fazendo $\mathbf{B}_{(l)} = \Sigma_i^{-1}(\theta_0) \dot{\Sigma}_{i(l)}(\theta_0) \Sigma_i^{-1}(\theta_0)$ temos

$$\begin{aligned} &Cov(s_{n\theta}(\theta_0^*)) \\ &= \left[\left\langle Cov((s_{n\theta}(\theta_0^*))_g, (s_{n\theta}(\theta_0^*))_h) \right\rangle_{gh} \right] \\ &= \left[\left\langle \sum_{i=1}^n Cov\left(\frac{1}{2}\mathbf{e}_i^t \mathbf{B}_{(g)} \mathbf{e}_i, \frac{1}{2}\mathbf{e}_i^t \mathbf{B}_{(h)} \mathbf{e}_i\right) \right\rangle_{gh} \right] \\ &= \left[\left\langle \sum_{i=1}^n \frac{1}{2} \text{tr} \left\{ \Sigma_i^{-1}(\theta_0) \dot{\Sigma}_{i(g)}(\theta_0) \Sigma_i^{-1}(\theta_0) \dot{\Sigma}_{i(h)}(\theta_0) \right\} \right\rangle_{gh} \right] \\ &= \left[\left\langle \sum_{i=1}^n \frac{1}{2} \text{tr} \left\{ \Sigma_i^{-1}(\theta_0) \mathbf{G}_{ig} \Sigma_i^{-1}(\theta_0) \mathbf{G}_{ih} \right\} \right\rangle_{gh} \right], \end{aligned} \quad (4.31)$$

que é igual a $\mathbf{G}_{n\theta\theta}$ definida em (4.29). Por outro lado,

Fazendo $\tilde{k}_{ig} = -\frac{1}{2} \sum_{i=1}^n \text{tr} \left(\Sigma_i^{-1}(\theta_0) \dot{\Sigma}_{i(g)}(\theta_0) \right) = -\frac{1}{2} \sum_{i=1}^n \text{tr} \left(\Sigma_i^{-1}(\theta_0) \mathbf{G}_{ig} \right)$ e $\mathbf{A}_{ik} = \left((\mathbf{X}_i^t \Sigma_i^{-1}(\theta_0))_{k1}, \dots, (\mathbf{X}_i^t \Sigma_i^{-1}(\theta_0))_{kn_i} \right)$ temos

$$\begin{aligned}
& \text{Cov}(\mathbf{s}_{n\alpha}(\theta_0^*), \mathbf{s}_{n\theta}(\theta_0^*)) \\
&= \left[\left\langle \text{Cov} \left((\mathbf{s}_{n\alpha}(\theta_0^*))_k, (\mathbf{s}_{n\theta}(\theta_0^*))_g \right) \right\rangle_{kg} \right] \\
&= \left[\left\langle \text{Cov} \left(\sum_{i=1}^n \sum_{j=1}^{n_i} (\mathbf{X}_i^t \Sigma_i^{-1}(\theta_0))_{kj} e_{ij}, \tilde{k}_{ig} + \frac{1}{2} \sum_{i=1}^n \mathbf{e}_i^t \Sigma_i^{-1}(\theta_0) \dot{\Sigma}_{i(g)}(\theta_0) \Sigma_i^{-1}(\theta_0) \mathbf{e}_i \right) \right\rangle_{kg} \right] \\
&= \left[\left\langle \sum_{i=1}^n \text{Cov} \left(\sum_{j=1}^{n_i} (\mathbf{X}_i^t \Sigma_i^{-1}(\theta_0))_{kj} e_{ij}, \frac{1}{2} \mathbf{e}_i^t \Sigma_i^{-1}(\theta_0) \dot{\Sigma}_{i(g)}(\theta_0) \Sigma_i^{-1}(\theta_0) \mathbf{e}_i \right) \right\rangle_{kg} \right] \\
&= \left[\left\langle \sum_{i=1}^n \text{Cov} \left(\mathbf{A}_{ik} \mathbf{e}_i, \frac{1}{2} \mathbf{e}_i^t \Sigma_i^{-1}(\theta_0) \dot{\Sigma}_{i(g)}(\theta_0) \Sigma_i^{-1}(\theta_0) \mathbf{e}_i \right) \right\rangle_{kg} \right] \\
&= \left[\left\langle \sum_{i=1}^n 2\mathbf{A}_{ik} \Sigma_i(\theta_0) \Sigma_i^{-1}(\theta_0) \dot{\Sigma}_{i(g)}(\theta_0) \Sigma_i^{-1}(\theta_0) \mathbf{0} \right\rangle_{kg} \right] \\
&= \mathbf{0}
\end{aligned} \tag{4.32}$$

que é igual a $\mathbf{G}_{n\alpha\theta}$ definida em (4.28). Então a matriz $\mathbf{D}_n$ definida em (4.6) é igual à matriz $\mathbf{G}_n$, e portanto a condição (B4) está satisfeita. De (4.27), (4.28) e (4.29) observamos que $\mathbf{G}_n$ é uma matriz finita e como $\mathbf{D}_n = \mathbf{G}_n$, a condição B1) também está satisfeita.

As demais condições B2), B3) e B5) são consideradas em cada caso particular mencionados no Capítulo 1.

Notemos inicialmente que para verificar a condição (B5) podemos utilizar o teorema de Liapounov (ver, por exemplo, Sen and Singer, 1993, p. 108) bastando para isso provar que

$$\lambda_{\min}(\mathbf{D}_n) \geq c_0 n \tag{4.33}$$

com c_0 constante positiva e que

$$E \left(|S_{ni}(\theta_0^*)|^\delta \right) \leq L \quad (4.34)$$

com, por exemplo, $\delta = 4$, $1 \leq i \leq n$, para alguma constante L , e

$$S_{ni}(\theta_0^*) = (s_{n\alpha i}^t(\theta_0^*), s_{n\theta i}^t(\theta_0^*))^t = (s_{n\alpha i}^t(\theta_0^*), (s_{n\theta i}(\theta_0^*))_1, (s_{n\theta i}(\theta_0^*))_2)^t$$

com $s_{n\alpha i}(\theta^*)$ e $s_{n\theta i}(\theta^*)$ definidas em (4.21) e (4.23), respectivamente.

Para provar B5) aplicaremos o teorema de Cramér-Wold (ver, por exemplo, Sen and Singer, 1993, p.106), ou seja, $\mathbf{D}_n^{-1/2} \mathbf{S}_n(\theta_0) \xrightarrow{D} \mathbf{Z}$ com $\mathbf{Z} \sim N(0, \mathbf{I}_p)$, se para qualquer vetor unitário ξ_n , $\xi_n^t \mathbf{D}_n^{-1/2} \mathbf{S}_{ni}(\theta_0^*) \xrightarrow{D} \xi_n^t \mathbf{Z}$, com $\xi_n^t \mathbf{Z} \sim N(0, 1)$.

Seja ξ_n qualquer vetor unitário em $\mathbb{R}^p$; para $1 \leq i \leq n$, definimos $Y_{in} = \xi_n^t \mathbf{D}_n^{-1/2} \mathbf{S}_{ni}(\theta_0^*)$ e $\sigma_{in}^2 = \text{Var}(Y_{in}) = \xi_n^t \mathbf{D}_n^{-1/2} \mathbf{D}_{ni} \mathbf{D}_n^{-t/2} \xi_n$, com $\mathbf{D}_{ni} = \text{Cov}(\mathbf{S}_{ni}(\theta_0^*))$. Então Y_{in} , $1 \leq i \leq n$ são mutuamente independentes para cada n , $E(Y_{in}) = 0$ e $\sigma_n^2 = \sum_{i=1}^n \sigma_{in}^2 = \xi_n^t \mathbf{D}_n^{-1/2} (\sum_{i=1}^n \mathbf{D}_{ni}) \mathbf{D}_n^{-t/2} \xi_n = \xi_n^t \mathbf{D}_n^{-1/2} \mathbf{D}_n \mathbf{D}_n^{-t/2} \xi_n = 1$. Suponhamos que (4.34) é verdadeiro. Então

$$\begin{aligned} \sum_{i=1}^n E \left(|Y_{in}|^\delta \right) &= \sum_{i=1}^n E \left\{ \left(\xi_n^t \mathbf{D}_n^{-1/2} \mathbf{S}_{ni}(\theta_0^*) \mathbf{S}_{ni}^t(\theta_0^*) \mathbf{D}_n^{-t/2} \xi_n \right)^{\delta/2} \right\} \\ &\leq \sum_{i=1}^n E \left\{ \lambda_{\max}^{\delta/2} (\mathbf{S}_{ni}(\theta_0^*) \mathbf{S}_{ni}^t(\theta_0^*)) \left(\xi_n^t \mathbf{D}_n^{-1} \xi_n \right)^{\delta/2} \right\} \\ &\leq \sum_{i=1}^n E \left\{ \text{tr} (\mathbf{S}_{ni}(\theta_0^*) \mathbf{S}_{ni}^t(\theta_0^*)) \lambda_{\max}^{\delta/2} (\mathbf{D}_n^{-1}) \right\} \\ &\leq \sum_{i=1}^n \frac{E \left\{ |\mathbf{S}_{ni}(\theta_0^*)|^\delta \right\}}{\lambda_{\min}^{\delta/2} (\mathbf{D}_n^{-1})} \leq \frac{Ln}{(c_0 n)^{\delta/2}} \rightarrow 0. \end{aligned}$$

Assim pelo teorema de Liapounov, $\xi_n^t \mathbf{D}_n^{-1/2} \mathbf{S}_n(\theta_0^*) = \sum_{i=1}^n Y_{in}$ converge a uma distribuição normal padrão; logo se as condições (4.33) e (4.34) forem válidas B5) será também.

Verificação das condições B2), B3) e B5) para a hipótese de inclusão de um efeito aleatório.

Consideremos o caso particular definido por (1.8) e (1.9) com $\mathbf{Z}_i = \mathbf{z}_i = (z_{i1}, \dots, z_{in_i})$, $b_i = b_i$, escalar, com distribuição $N(0, \sigma_A^2)$, $\mathbf{D} = \sigma_A^2$, e $\boldsymbol{\theta} = (\sigma_A^2, \sigma^2)$. O espaço paramétrico é $\Omega = \mathbb{R}^{p'} \times [0, +\infty) \times (0, +\infty)$. Portanto as matrizes $\mathbf{G}_{ig}$ definidas em (4.19) são dadas por

$$\mathbf{G}_{ig} = \begin{cases} \mathbf{z}_i \mathbf{z}_i^t & \text{para } g = 1 \\ \mathbf{I}_{n_i} & \text{para } g = 2. \end{cases}$$

Como $\Sigma_i(\boldsymbol{\theta}) = \sigma_A^2 \mathbf{z}_i \mathbf{z}_i^t + \sigma^2 \mathbf{I}_{n_i}$, temos $\Sigma_i^{-1}(\boldsymbol{\theta}) = \sigma^{-2} \left[\mathbf{I}_{n_i} - \frac{\sigma_A^2}{\sigma^2 + \tau_i \sigma_A^2} \mathbf{z}_i \mathbf{z}_i^t \right]$ pelo resultado clássico apresentado em Graybill (1983), por exemplo. Sob a hipótese nula, $\Sigma_i^{-1}(\boldsymbol{\theta}_0) = \sigma^{-2} \mathbf{I}_{n_i}$. Assim de (4.20) temos

$$\begin{aligned} s_{n\alpha}(\boldsymbol{\theta}_0^*) &= \sum_{i=1}^n \sigma^{-2} \mathbf{X}_i^{-1} \mathbf{e}_i \\ &= \sum_{i=1}^n s_{n\alpha i}(\boldsymbol{\theta}_0^*), \end{aligned}$$

com

$$s_{n\alpha i}(\boldsymbol{\theta}_0^*) = \sigma^{-2} \mathbf{X}_i^{-1} \mathbf{e}_i. \quad (4.35)$$

Definindo $\tau_i = \tau_{i11} = \mathbf{z}_i^t \mathbf{z}_i = \text{tr}(\mathbf{z}_i \mathbf{z}_i^t)$ e observando que $\dot{\Sigma}_{i(1)} = \mathbf{z}_i \mathbf{z}_i^t$ e $\dot{\Sigma}_{i(2)} = \mathbf{I}_{n_i}$, de (4.22), podemos obter $s_{n\theta}(\boldsymbol{\theta}_0^*) = ((s_{n\theta}(\boldsymbol{\theta}_0^*))_1, (s_{n\theta}(\boldsymbol{\theta}_0^*))_2)$, com

$$(s_{n\theta}(\boldsymbol{\theta}_0^*))_1 = \left[\sum_{i=1}^n \frac{1}{2} (-\sigma^{-2} \tau_i + \sigma^{-4} \mathbf{e}_i^t \mathbf{z}_i \mathbf{z}_i^t \mathbf{e}_i) \right],$$

$$\begin{aligned} (s_{n\theta}(\boldsymbol{\theta}_0^*))_2 &= \frac{1}{2} \left[-\sum_{i=1}^n \sigma^{-2} n_i + \sigma^{-4} \sum_{i=1}^n \mathbf{e}_i^t \mathbf{e}_i \right], \\ &= \sum_{i=1}^n \frac{1}{2} [-n_i \sigma^{-2} + \sigma^{-4} \mathbf{e}_i^t \mathbf{e}_i]. \end{aligned}$$

De (4.27) e (4.29) temos

$$\mathbf{G}_{n\alpha\alpha} = \sigma^{-2} \sum_{i=1}^n \mathbf{X}_i^t \mathbf{X}_i,$$

$$\mathbf{G}_{n\theta\theta} = \left[\left\langle \frac{1}{2} \sum_{i=1}^n \text{tr} \{ \sigma^{-2} \mathbf{G}_{ig} \mathbf{G}_{ih} \} \right\rangle_{gh} \right].$$

Calculemos cada componente de $\mathbf{G}_{n\theta\theta}$

$$\begin{aligned} (\mathbf{G}_{n\theta\theta})_{11} &= \frac{1}{2} \sum_{i=1}^n \text{tr} \{ \sigma^{-4} \mathbf{z}_i \mathbf{z}_i^t \mathbf{z}_i \mathbf{z}_i^t \} = \frac{1}{2} \sum_{i=1}^n \text{tr} \{ \tau_i \sigma^{-4} \mathbf{z}_i \mathbf{z}_i^t \} \\ &= \frac{1}{2} \sum_{i=1}^n \tau_i^2 \sigma^{-4}, \\ (\mathbf{G}_{n\theta\theta})_{12} &= \frac{1}{2} \sum_{i=1}^n \text{tr} \{ \sigma^{-4} \mathbf{z}_i \mathbf{z}_i^t \} = \frac{1}{2} \sum_{i=1}^n \tau_i \sigma^{-4}, \\ (\mathbf{G}_{n\theta\theta})_{22} &= \frac{1}{2} \sum_{i=1}^n \text{tr} \{ \sigma^{-4} \mathbf{I}_{n_i} \} = \frac{1}{2} \sum_{i=1}^n n_i \sigma^{-4}. \end{aligned}$$

Logo a matriz $\mathbf{G}_n$ definida em (4.7) pode ser escrita como $\mathbf{G}_n = \sum_{i=1}^n \mathbf{G}_{ni}$ com

$$\mathbf{G}_{ni} = \begin{bmatrix} \sigma^{-2} \mathbf{X}_i^t \mathbf{X}_i & 0 \\ 0 & \mathbf{G}_{n\theta\theta i} \end{bmatrix}, \quad (4.36)$$

em que $\mathbf{G}_{n\theta\theta i} = \frac{1}{2\sigma^4} \begin{bmatrix} \tau_i^2 & \tau_i \\ \tau_i & n_i \end{bmatrix}$. Como $\det(\mathbf{G}_{n\theta\theta i}) = \frac{1}{4} \tau_i^2 \sigma^{-16} (n_i - 1)$ então $\det(\mathbf{G}_{ni}) = \det(\sigma^{-2} \mathbf{X}_i^t \mathbf{X}_i) \frac{1}{4} \tau_i^2 \sigma^{-16} (n_i - 1)$. Além disso, como $\mathbf{X}_i$ tem posto completo, então $\mathbf{X}_i^t \mathbf{X}_i$ é positiva definida logo $\det(\sigma^{-2} \mathbf{X}_i^t \mathbf{X}_i) > 0$. Portanto

$$\det(\mathbf{G}_{ni}) > 0 \text{ se } n_i > 1$$

logo existe $\tilde{c}_0 > 0$, tal que

$$\lambda_{\min}(\mathbf{G}_{ni}) \geq \tilde{c}_0, \text{ para todo } 1 \leq i \leq n \text{ tal que } n_i > 1,$$

e

$$\det(\mathbf{G}_{ni}) = 0 \text{ para todo } i \in I = \{1 \leq i \leq n : n_i = 1\}.$$

Sabemos que dada uma matriz $\mathbf{A}$, ao menos semidefinida positiva $q \times q$,

$$\lambda_{\min}(\mathbf{A}) = \min_{\mathbf{y} \neq \mathbf{0}} \left\{ \frac{\mathbf{y}^t \mathbf{A} \mathbf{y}}{\mathbf{y}^t \mathbf{y}} \right\}.$$

(ver por exemplo, Giri, 1977). Seja $\mathbf{B}$ uma matriz ao menos semidefinida positiva $q \times q$; então notemos que

$$\begin{aligned} \lambda_{\min}(\mathbf{A} + \mathbf{B}) &= \min_{\mathbf{y} \neq \mathbf{0}} \left\{ \frac{\mathbf{y}^t (\mathbf{A} + \mathbf{B}) \mathbf{y}}{\mathbf{y}^t \mathbf{y}} \right\} \\ &= \min_{\mathbf{y} \neq \mathbf{0}} \left\{ \frac{\mathbf{y}^t \mathbf{A} \mathbf{y}}{\mathbf{y}^t \mathbf{y}} + \frac{\mathbf{y}^t \mathbf{B} \mathbf{y}}{\mathbf{y}^t \mathbf{y}} \right\} \\ &\geq \min_{\mathbf{y} \neq \mathbf{0}} \left\{ \frac{\mathbf{y}^t \mathbf{A} \mathbf{y}}{\mathbf{y}^t \mathbf{y}} \right\} + \min_{\mathbf{y} \neq \mathbf{0}} \left\{ \frac{\mathbf{y}^t \mathbf{B} \mathbf{y}}{\mathbf{y}^t \mathbf{y}} \right\} \\ &\geq \lambda_{\min}(\mathbf{A}) + \lambda_{\min}(\mathbf{B}). \end{aligned}$$

Assim

$$\lambda_{\min}(\mathbf{G}_n) \geq \sum_{i=1}^n \lambda_{\min}(\mathbf{G}_{ni}) = \sum_{i \in I} \lambda_{\min}(\mathbf{G}_{ni}) \geq (n - np) \tilde{c}_0, \quad (4.37)$$

em que p denota a proporção de indivíduos com apenas uma observação; logo a cardinalidade de I é $n - np$ e conseqüentemente, a condição B2) está satisfeita.

Resta ainda provar as condições B3) e B5).

Provaremos agora (4.10), para isso notemos inicialmente que como $\mathbf{u}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i})$,

$$\frac{\text{tr}(\mathbf{u}_i \mathbf{u}_i^t)}{\sigma^2} = \frac{\mathbf{u}_i^t \mathbf{u}_i}{\sigma^2} \sim \chi_{n_i}^2$$

logo

$$\frac{1}{n^2} \sum_{i=1}^n E \left(\frac{\mathbf{u}_i^t \mathbf{u}_i}{\sigma^2} - n_i \right)^2 = \frac{1}{n^2} \sum_{i=1}^n 2n_i \xrightarrow{P} 0$$

então pela lei fraca dos grandes números de Markov (Sen and Singer, 1993, Teorema

2.3.7. por exemplo) temos que

$$\frac{1}{n} \sum_{i=1}^n \frac{\mathbf{u}_i^t \mathbf{u}_i}{\sigma^2} - \frac{1}{n} \sum_{i=1}^n n_i = \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{u}_i^t \mathbf{u}_i}{\sigma^2} - n_i \xrightarrow{P} 0. \quad (4.38)$$

Também notemos que

$$\text{tr}(\mathbf{z}_i \mathbf{u}_i^t) = \mathbf{u}_i^t \mathbf{z}_i = \mathbf{z}_i^t \mathbf{u}_i \sim N(0, \sigma^2 \mathbf{z}_i^t \mathbf{z}_i)$$

e

$$\frac{\text{tr}(\mathbf{z}_i \mathbf{z}_i^t \mathbf{u}_i \mathbf{u}_i^t \mathbf{z}_i \mathbf{z}_i^t)}{\sigma^2} = \frac{\mathbf{u}_i^t \mathbf{A}_i \mathbf{u}_i}{\sigma^2}$$

em que $\mathbf{A}_i = \mathbf{z}_i \mathbf{z}_i^t \mathbf{z}_i \mathbf{z}_i^t$, então

$$E\left(\frac{\mathbf{u}_i^t \mathbf{A}_i \mathbf{u}_i}{\sigma^2}\right) = \text{tr}(\mathbf{A}_i) \text{ e } \text{Var}\left(\frac{\mathbf{u}_i^t \mathbf{A}_i \mathbf{u}_i}{\sigma^2}\right) = 2 \text{tr}[(\mathbf{A}_i)^2]$$

Como

$$\frac{1}{n^2} \sum_{i=1}^n \text{Var}\left(\frac{\mathbf{u}_i^t \mathbf{A}_i \mathbf{u}_i}{\sigma^2}\right) = \frac{1}{n^2} \sum_{i=1}^n 2 \text{tr}[(\mathbf{A}_i)^2] \rightarrow 0$$

então

$$\frac{1}{n} \sum_{i=1}^n \frac{\text{tr}(\mathbf{z}_i \mathbf{z}_i^t \mathbf{u}_i \mathbf{u}_i^t \mathbf{z}_i \mathbf{z}_i^t)}{\sigma^2} - \frac{1}{n} \sum_{i=1}^n \text{tr}(\mathbf{z}_i \mathbf{z}_i^t \mathbf{z}_i \mathbf{z}_i^t) \xrightarrow{P} 0. \quad (4.39)$$

De (4.24) e (4.27) temos

$$\mathbf{F}_{n\alpha\alpha}(\boldsymbol{\theta}_0^*) - \mathbf{G}_{n\alpha\alpha} = \mathbf{0},$$

e de (4.26) e (4.29) temos

$$\begin{aligned} & \mathbf{F}_{n\theta\theta}(\boldsymbol{\theta}_0^*) - \mathbf{G}_{n\theta\theta} \\ &= \left[\left\langle \sum_{i=1}^n \text{tr} \left\{ \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0) \mathbf{G}_{ig} [\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0) \mathbf{e}_i \mathbf{e}_i^t - \mathbf{I}_{n_i}] \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0) \mathbf{G}_{ih} \right\} \right\rangle_{gh} \right]. \end{aligned} \quad (4.40)$$

Examinemos os termos da diagonal principal de (4.40)

$$\begin{aligned}
& (\mathbf{F}_{n\theta\theta}(\theta_0^*) - \mathbf{G}_{n\theta\theta})_{11} \\
&= \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \mathbf{z}_i \mathbf{z}_i^t \left[\sigma^{-2} \mathbf{e}_i \mathbf{e}_i^t - \mathbf{I}_{n_i} \right] \mathbf{z}_i \mathbf{z}_i^t \right\} \\
&= \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \mathbf{z}_i \mathbf{z}_i^t \left[\sigma^{-2} \mathbf{u}_i \mathbf{u}_i^t - \mathbf{I}_{n_i} \right] \mathbf{z}_i \mathbf{z}_i^t \right\} \\
&= \sum_{i=1}^n \sigma^{-4} \frac{\text{tr}(\mathbf{z}_i \mathbf{z}_i^t \mathbf{u}_i \mathbf{u}_i^t \mathbf{z}_i \mathbf{z}_i^t)}{\sigma^2} - \sigma^{-4} \text{tr}(\mathbf{z}_i \mathbf{z}_i^t \mathbf{z}_i \mathbf{z}_i^t), \tag{4.41}
\end{aligned}$$

e

$$\begin{aligned}
& (\mathbf{F}_{n\theta\theta}(\theta_0^*) - \mathbf{G}_{n\theta\theta})_{22} \\
&= \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \left[\sigma^{-2} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})^t - \mathbf{I}_{n_i} \right] \right\} \\
&= \sigma^{-4} \sum_{i=1}^n \text{tr} \left[\sigma^{-2} \mathbf{u}_i \mathbf{u}_i^t - \mathbf{I}_{n_i} \right] \\
&= \sigma^{-4} \sum_{i=1}^n \frac{\mathbf{u}_i^t \mathbf{u}_i}{\sigma^2} - n_i \tag{4.42}
\end{aligned}$$

assim de (4.38), (4.39), (4.41) e (4.42), temos

$$\begin{aligned}
& \text{tr} \frac{1}{n} (\mathbf{F}_n(\theta_0^*) - \mathbf{G}_n) \\
&= \frac{1}{n} ((\mathbf{F}_{n\theta\theta}(\theta_0^*) - \mathbf{G}_{n\theta\theta})_{11} + (\mathbf{F}_{n\theta\theta}(\theta_0^*) - \mathbf{G}_{n\theta\theta})_{22}) \\
&= \sigma^{-4} \left(\frac{1}{n} \sum_{i=1}^n \frac{\text{tr}(\mathbf{z}_i \mathbf{z}_i^t \mathbf{u}_i \mathbf{u}_i^t \mathbf{z}_i \mathbf{z}_i^t)}{\sigma^2} \right) - \sigma^{-4} \frac{1}{n} \sum_{i=1}^n \text{tr}(\mathbf{z}_i \mathbf{z}_i^t \mathbf{z}_i \mathbf{z}_i^t) + \sigma^{-4} \left(\frac{1}{n} \sum_{i=1}^n \frac{\mathbf{u}_i^t \mathbf{u}_i}{\sigma^2} - n_i \right) \\
&= o_p(1),
\end{aligned}$$

logo vale (4.10). Para verificar agora (4.11), necessitamos calcular

$$(\mathbf{F}_{n\alpha\alpha}(\theta^*) - \mathbf{F}_{n\alpha\alpha}(\theta_0^*)) = \left(\sum_{i=1}^n \mathbf{X}_i^t (\boldsymbol{\Sigma}_i^{-1}(\theta) - \boldsymbol{\Sigma}_i^{-1}(\theta_0)) \mathbf{X}_i \right). \tag{4.43}$$

Observemos que

$$\begin{aligned}
\text{tr} (\mathbf{F}_{n\alpha\alpha} (\boldsymbol{\theta}^*) - \mathbf{F}_{n\alpha\alpha} (\boldsymbol{\theta}_0^*)) &= \text{tr} \left(\sum_{i=1}^n \mathbf{X}_i^t (\boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\theta}) - \sigma^{-2} \mathbf{I}_{n_i}) \mathbf{X}_i \right) \\
&= \text{tr} \left(\sum_{i=1}^n \frac{-\sigma_A^2}{\sigma^2 + \tau_i \sigma_A^2} \mathbf{X}_i^t \mathbf{z}_i \mathbf{z}_i^t \mathbf{X}_i \right) \\
&= \sum_{i=1}^n \frac{-\sigma_A^2}{\sigma^2 + \tau_i \sigma_A^2} \text{tr} (\mathbf{X}_i^t \mathbf{z}_i \mathbf{z}_i^t \mathbf{X}_i). \tag{4.44}
\end{aligned}$$

Da condição (4.17) temos que para todo par (σ^2, σ_A^2) o termo $\frac{-\sigma_A^2}{\sigma^2 + \tau_i \sigma_A^2} \text{tr} (\mathbf{X}_i^t \mathbf{z}_i \mathbf{z}_i^t \mathbf{X}_i)$ é limitado. Das definições (4.8) e (4.7), e de (4.36) temos que

$$\begin{aligned}
N_n (A) &= \{ \boldsymbol{\theta}^* \in \Omega : (\boldsymbol{\theta}^* - \boldsymbol{\theta}_0^*)^t \mathbf{G}_n (\boldsymbol{\theta}^* - \boldsymbol{\theta}_0^*) \leq A^2 \} \\
&= \left\{ \boldsymbol{\theta}^* \in \Omega : \frac{\sigma_A^4}{2\sigma^4} \sum_{i=1}^n \tau_i^2 \leq A^2 \right\}, \tag{4.45}
\end{aligned}$$

Então

$$\begin{aligned}
\sup_{\boldsymbol{\theta}^* \in N_n(A)} \text{tr} \frac{1}{n} (\mathbf{F}_{n\alpha\alpha} (\boldsymbol{\theta}^*) - \mathbf{F}_{n\alpha\alpha} (\boldsymbol{\theta}_0^*)) &= \frac{1}{n} \sum_{i=1}^n \inf_{\boldsymbol{\theta}^* \in N_n(A)} \left(\frac{-\sigma_A^2}{\sigma^2 + \tau_i \sigma_A^2} \right) \text{tr} (\mathbf{X}_i^t \mathbf{z}_i \mathbf{z}_i^t \mathbf{X}_i) \\
&= O(1). \tag{4.46}
\end{aligned}$$

Notemos primeiro que

$$\begin{aligned}
&\mathbf{F}_{n\theta\theta} (\boldsymbol{\theta}^*)_{gg} \\
&= \frac{1}{2} \sum_{i=1}^n \text{tr} \{ \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\theta}) \mathbf{G}_{ig} \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\theta}) [2\mathbf{e}_i \mathbf{e}_i^t - \boldsymbol{\Sigma}_i (\boldsymbol{\theta})] \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\theta}) \mathbf{G}_{ig} \} \\
&= \sum_{i=1}^n \text{tr} \{ \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\theta}) \mathbf{G}_{ig} \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\theta}) \mathbf{e}_i \mathbf{e}_i^t \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\theta}) \mathbf{G}_{ig} \} - \frac{1}{2} \sum_{i=1}^n \text{tr} \{ \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\theta}) \mathbf{G}_{ig} \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\theta}) \mathbf{G}_{ig} \} \\
&= \sum_{i=1}^n \text{tr} \{ \mathbf{e}_i^t \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\theta}) \mathbf{G}_{ig} \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\theta}) \mathbf{G}_{ig} \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\theta}) \mathbf{e}_i \} - \frac{1}{2} \sum_{i=1}^n \text{tr} \{ \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\theta}) \mathbf{G}_{ig} \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\theta}) \mathbf{G}_{ig} \}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n \mathbf{e}_i^t \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{L}_{igh}(\boldsymbol{\theta}) \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{e}_i - \sum_{i=1}^n \text{tr} \{ \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{G}_{ig} \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{G}_{ig} \} \\
&\quad + \frac{1}{2} \sum_{i=1}^n \text{tr} \{ \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{G}_{ig} \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{G}_{ig} \}
\end{aligned} \tag{4.47}$$

Como $\mathbf{e}_i \sim N(0, \Sigma_i(\boldsymbol{\theta}))$, então $\Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{e}_i \sim N(0, \Sigma_i^{-1}(\boldsymbol{\theta}))$, dada a matriz simétrica $\mathbf{L}_{ig}(\boldsymbol{\theta}) = \mathbf{G}_{ig} \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{G}_{ig}$, temos que

$$\begin{aligned}
E(\mathbf{e}_i^t \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{L}_{ig}(\boldsymbol{\theta}) \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{e}_i) &= \text{tr} \{ \mathbf{L}_{ig}(\boldsymbol{\theta}) \Sigma_i^{-1}(\boldsymbol{\theta}) \} \\
&= \text{tr} \{ \mathbf{G}_{ig} \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{G}_{ig} \Sigma_i^{-1}(\boldsymbol{\theta}) \} \\
&= \text{tr} \left(\{ \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{G}_{ig} \}^2 \right) \\
\text{Var}(\mathbf{e}_i^t \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{L}_{ig}(\boldsymbol{\theta}) \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{e}_i) &= 2 \text{tr} \left(\{ \mathbf{L}_{ig}(\boldsymbol{\theta}) \Sigma_i^{-1}(\boldsymbol{\theta}) \}^2 \right) \\
K_r(\mathbf{e}_i^t \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{L}_{ig}(\boldsymbol{\theta}) \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{e}_i) &= 2^{r-1} (r-1)! \text{tr} \left(\{ \mathbf{L}_{ig}(\boldsymbol{\theta}) \Sigma_i^{-1}(\boldsymbol{\theta}) \}^r \right) \\
&= 2^{r-1} (r-1)! \text{tr} \left(\{ \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{G}_{ig} \}^{2r} \right)
\end{aligned}$$

em que K_r denota o r -ésimo cumulante (ver por exemplo, Searle, 1971, Teorema 1).

Assim pela lei fraca dos grandes números de Markov se por exemplo

$$\frac{1}{n^4} \sum_{i=1}^n K_4(\mathbf{e}_i^t \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{L}_{ig}(\boldsymbol{\theta}) \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{e}_i) \rightarrow 0, \tag{4.48}$$

temos que

$$\frac{1}{n} \sum_{i=1}^n \mathbf{e}_i^t \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{L}_{igh}(\boldsymbol{\theta}) \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{e}_i - \frac{1}{n} \sum_{i=1}^n \text{tr} \{ \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{G}_{ig} \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{G}_{ig} \} \xrightarrow{P} 0.$$

Se alem disso se satisfaz que

$$\frac{1}{2n} \sum_{i=1}^n \text{tr} \{ \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{G}_{ig} \Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{G}_{ig} \} = O(1), \tag{4.49}$$

de (4.47) temos que

$$\frac{1}{n} \mathbf{F}_{n\theta\theta}(\boldsymbol{\theta}^*)_{gg} = o_p(1) \quad (4.50)$$

Verifiquemos que (4.50) vale para $\mathbf{F}_{n\theta\theta}(\boldsymbol{\theta}^*)_{11}$

$$\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}) \mathbf{G}_{i1} = \sigma^{-2} \mathbf{z}_i \mathbf{z}_i^t \left[1 - \frac{\sigma_A^2 \tau_i}{\sigma^2 + \tau_i \sigma_A^2} \right].$$

Assim por (4.18) são satisfeitas (4.48) e (4.49) e portanto é satisfeita (4.50) para $g = 1$.

Calculemos agora

$$\begin{aligned} (\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}) \mathbf{G}_{i2})^2 &= \left(\sigma^{-2} \left[\mathbf{I}_{n_i} - \frac{\sigma_A^2}{\sigma^2 + \tau_i \sigma_A^2} \mathbf{z}_i \mathbf{z}_i^t \right] \right)^2 \\ &= \sigma^{-4} [\mathbf{I}_{n_i} + \varpi_i \mathbf{z}_i \mathbf{z}_i^t], \end{aligned}$$

com $\varpi_i = \left(\left(\frac{\sigma_A^2}{\sigma^2 + \tau_i \sigma_A^2} \right)^2 \tau_i - \frac{2\sigma_A^2}{\sigma^2 + \tau_i \sigma_A^2} \right)$.

Assim são satisfeitas (4.48) e (4.49) e portanto é satisfeita (4.50) para $g = 2$, logo temos que para $g = 1, 2$, então

$$\frac{1}{n} \text{tr} \left\{ (\mathbf{F}_{n\theta\theta}(\boldsymbol{\theta}^*))_{gg} - (\mathbf{F}_{n\theta\theta}(\boldsymbol{\theta}_0^*))_{gg} \right\} = o_p(1), \quad (4.51)$$

para todo $\boldsymbol{\theta}^* \in N_n(A)$, logo de (4.46) e (4.51) temos que (4.11) vale, i.e., B3) é válida.

Para verificar que a condição (B5) devemos verificar que valem (4.33) e (4.34). Notemos que (4.37) implica (4.33) considerando $c_0 = (1 - p) \tilde{c}_0$. Provaremos (4.34) com $\delta = 4$, i.e., provaremos que

$$E(|\mathbf{S}_{ni}(\boldsymbol{\theta}_0^*)|^4) < \infty. \quad (4.52)$$

Como

$$\begin{aligned} E(|\mathbf{S}_{ni}(\boldsymbol{\theta}_0^*)|^4) &= E\left((|\mathbf{S}_{ni}(\boldsymbol{\theta}_0^*)|^2)^2\right) \\ &= \text{Var}(|\mathbf{S}_{ni}(\boldsymbol{\theta}_0^*)|^2) + (E(|\mathbf{S}_{ni}(\boldsymbol{\theta}_0^*)|^2))^2, \end{aligned} \quad (4.53)$$

para verificar (4.52), provaremos que

$$(E(|S_{ni}(\theta_0^*)|^2)) < \infty, \quad (4.54)$$

e que

$$Var(|S_{ni}(\theta_0^*)|^2) < \infty. \quad (4.55)$$

Como $s_{n\alpha i}(\theta_0^*) = \mathbf{X}_i^t \Sigma_i^{-1}(\theta_0) \mathbf{e}_i$, temos que

$$\begin{aligned} & |S_{ni}(\theta_0^*)|^2 \\ &= \left| \left((\mathbf{X}_i^t \Sigma_i^{-1}(\theta_0) \mathbf{e}_i)^t, (s_{n\theta i}(\theta_0^*))_1, (s_{n\theta i}(\theta_0^*))_2 \right)^t \right|^2 \\ &= \left| \left(\sum_{j=1}^{n_i} (\mathbf{X}_i^t \Sigma_i^{-1}(\theta_0))_{1j} e_{ij}, \dots, \sum_{j=1}^{n_i} (\mathbf{X}_i^t \Sigma_i^{-1}(\theta_0))_{n_{ij}} e_{ij}, (s_{n\theta i}(\theta_0^*))_1, (s_{n\theta i}(\theta_0^*))_2 \right)^t \right|^2 \\ &= \sum_{k=1}^{n_i} Z_{ki}^2 + ((s_{n\theta i}(\theta_0^*))_1)^2 + ((s_{n\theta i}(\theta_0^*))_2)^2 \end{aligned} \quad (4.56)$$

em que $Z_{ki} = \sum_{j=1}^{n_i} (\mathbf{X}_i^t \Sigma_i^{-1}(\theta_0))_{kj} e_{ij} = (\mathbf{X}_i^t \Sigma_i^{-1}(\theta_0))_k \mathbf{e}_i$ é um variável aleatória com distribuição normal.

$$E(|S_{ni}(\theta_0^*)|^2) = E(|s_{n\alpha i}(\theta_0^*)|^2) + E(((s_{n\theta i}(\theta_0^*))_1)^2) + E((s_{n\theta i}(\theta_0^*))_2)^2. \quad (4.57)$$

Provaremos que cada termo do segundo membro de (4.57) é finito. O primeiro termo do segundo membro de (4.57) é

$$\begin{aligned} E(|s_{n\alpha i}(\theta_0^*)|^2) &= E \left(\sum_{k=1}^{n_i} \left(\sum_{j=1}^{n_i} (\mathbf{X}_i^t \Sigma_i^{-1}(\theta_0))_{kj} e_{ij} \right)^2 \right) \\ &= \sum_{k=1}^{n_i} E \left(\left(\sum_{j=1}^{n_i} (\mathbf{X}_i^t \Sigma_i^{-1}(\theta_0))_{kj} e_{ij} \right)^2 \right) \\ &= \sum_{k=1}^{n_i} E(Z_{ki}^2) \end{aligned}$$

$$< \infty. \quad (4.58)$$

Para provar que o segundo e o terceiro termos do segundo membro de (4.57) são finitos, notemos que para $g = 1, 2$,

$$\left| (s_{n\theta i}(\theta^*))_g \right|^2 = \left(\tilde{k}_{ig} + \frac{1}{2} \mathbf{e}_i^t (\Sigma_i^{-1}(\theta) \mathbf{G}_{ig} \Sigma_i^{-1}(\theta_0)) \mathbf{e}_i \right)^2,$$

com a constante $\tilde{k}_{ig} = -\frac{1}{2} \text{tr} \left(\Sigma_i^{-1}(\theta_0) \dot{\Sigma}_{i(g)}(\theta_0) \right) = -\frac{1}{2} \text{tr} \left(\Sigma_i^{-1}(\theta_0) \mathbf{G}_{ig} \right)$.

Em particular

$$\begin{aligned} E \left\{ |(s_{n\theta i}(\theta^*))_1|^2 \right\} &= E \left\{ \left(\tilde{k}_{i1} + \frac{1}{2} \mathbf{e}_i^t \Sigma_i^{-1}(\theta) \mathbf{z}_i \mathbf{z}_i^t \Sigma_i^{-1}(\theta_0) \mathbf{e}_i \right)^2 \right\} \\ &= E \left\{ \left(\tilde{k}_{i1} + \frac{\tau_i}{2} v_1 \right)^2 \right\} \\ &= \text{Var} \left(\left(\tilde{k}_{i1} + \frac{\tau_i}{2} v_1 \right) \right) + \left(E \left(\tilde{k}_{i1} + \frac{\tau_i}{2} v_1 \right) \right)^2 \\ &< \infty, \end{aligned} \quad (4.59)$$

já que $v_1 = \mathbf{e}_i^t \Sigma_i^{-1}(\theta) ((\tau_i)^{-1} \mathbf{z}_i \mathbf{z}_i^t) \Sigma_i^{-1}(\theta_0) \mathbf{e}_i \sim \chi^2(r(\mathbf{z}_i \mathbf{z}_i^t), 0)$, pois $((\tau_i)^{-1} \mathbf{z}_i \mathbf{z}_i^t)$ é uma matriz idempotente e $\Sigma_i^{-1}(\theta_0) \mathbf{e}_i \sim N(0, \Sigma_i^{-1}(\theta_0))$.

Da mesma forma dado que $v_2 = \mathbf{e}_i^t \Sigma_i^{-1}(\theta) \mathbf{I}_{n_i} \Sigma_i^{-1}(\theta_0) \mathbf{e}_i \sim \chi^2(n_i, 0)$ temos

$$\begin{aligned} E \left(|(s_{n\theta i}(\theta^*))_2|^2 \right) &= E \left(\left(\tilde{k}_{i2} + \frac{1}{2} \mathbf{e}_i^t \Sigma_i^{-1}(\theta) \Sigma_i^{-1}(\theta_0) \mathbf{e}_i \right)^2 \right) \\ &= E \left(\left(\tilde{k}_{i2} + \frac{1}{2} v_2 \right)^2 \right) \\ &= \text{Var} \left(\left(\tilde{k}_{i2} + \frac{1}{2} v_2 \right) \right) + \left(E \left(\tilde{k}_{i2} + \frac{1}{2} v_2 \right) \right)^2 \\ &< \infty, \end{aligned} \quad (4.60)$$

De (4.58), (4.59) e (4.60) segue (4.54).

Agora provaremos (4.55).

$$\begin{aligned}
& Var(|S_{ni}(\theta_0^*)|^2) \\
&= Var\left\{\sum_{k=1}^{n_i} Z_{ki}^2 + ((S_{n\theta i}(\theta^*))_1)^2 + ((S_{n\theta i}(\theta^*))_2)^2\right\} \\
&= Var\left\{\sum_{k=1}^{n_i} Z_{ki}^2\right\} + Var\{((S_{n\theta i}(\theta^*))_1)^2\} + Var\{((S_{n\theta i}(\theta^*))_2)^2\} \\
&\quad + 2Cov\left\{\sum_{k=1}^{n_i} Z_{ki}^2, ((S_{n\theta i}(\theta^*))_1)^2\right\} \\
&\quad + 2Cov\left\{\sum_{k=1}^{n_i} Z_{ki}^2, ((S_{n\theta i}(\theta^*))_2)^2\right\} \\
&\quad + 2Cov\{((S_{n\theta i}(\theta^*))_1)^2, ((S_{n\theta i}(\theta^*))_2)^2\}. \tag{4.61}
\end{aligned}$$

Provaremos que cada um dos termos do segundo membro de (4.61) é finito. Para o primeiro desses termos, temos

$$Var\left(\sum_{k=1}^{n_i} Z_{ki}^2\right) = \sum_{k=1}^{n_i} Var(Z_{ki}^2) + \sum_{k < l} Cov(Z_{ki}^2, Z_{li}^2). \tag{4.62}$$

Como

$$Var(Z_{ki}^2) = E(Z_{ki}^4) - [E(Z_{ki}^2)]^2 < \infty \tag{4.63}$$

pois Z_{ki} é uma variável aleatória com distribuição normal padrão e tem todos os momentos finitos, e

$$Cov(Z_{ki}^2, Z_{li}^2) \leq (E(Z_{ki}^4) E(Z_{li}^4))^{1/2} < \infty$$

pela desigualdade de Cauchy-Schwarz, podemos concluir que (4.62) é finito. Agora observemos que,

$$\begin{aligned}
& E \left\{ ((s_{n\theta i}(\boldsymbol{\theta}^*))_1)^4 \right\} \\
= & E \left\{ \left(\tilde{k}_{i1} + \frac{\tau_i}{2} v_1 \right)^4 \right\} \\
= & E \left\{ \left(\tilde{k}_{i1} + \frac{\tau_i}{2} v_1 \right)^2 \left(\tilde{k}_{i1} + \frac{\tau_i}{2} v_1 \right)^2 \right\} \\
= & E \left\{ \left(\tilde{k}_{i1}^2 + \tilde{k}_{i1} \tau_i v_1 + v_1^2 \right)^2 \right\} \\
= & E \left(\tilde{k}_{i1}^4 + 2\tilde{k}_{i1}^3 \tau_i v_1 + 2\tilde{k}_{i1}^2 v_1^2 + \tilde{k}_{i1}^2 \tau_i^2 v_1^2 + \tilde{k}_{i1} \tau_i v_1^2 + \tilde{k}_{i1} \tau_i v_1^3 + v_1^4 \right) \\
< & \infty
\end{aligned} \tag{4.64}$$

pois $v_1 \sim \chi^2(r(\mathbf{z}_i \mathbf{z}_i^t), 0)$ e portanto tem momentos de ordem 1, 2, 3 e 4. Esse resultado junto com (4.59) implica

$$Var \left([(s_{n\theta i}(\boldsymbol{\theta}^*))_1]^2 \right) = E \left([(s_{n\theta i}(\boldsymbol{\theta}^*))_1]^4 \right) - \left\{ E \left([(s_{n\theta i}(\boldsymbol{\theta}^*))_1]^2 \right) \right\}^2 < \infty \tag{4.65}$$

o que mostra que o segundo termo do segundo membro de (4.61) é finito.

De forma análoga, por (4.60) e pelo fato de

$$E \left([(s_{n\theta i}(\boldsymbol{\theta}^*))_2]^4 \right) < \infty, \tag{4.66}$$

temos que o terceiro termo do segundo membro de (4.61) é

$$Var \left([(s_{n\theta i}(\boldsymbol{\theta}^*))_2]^2 \right) = E \left([(s_{n\theta i}(\boldsymbol{\theta}^*))_2]^4 \right) - \left[E \left([(s_{n\theta i}(\boldsymbol{\theta}^*))_2]^2 \right) \right]^2 < \infty. \tag{4.67}$$

O quarto termo do segundo membro de (4.61) é

$$\begin{aligned}
& Cov \left(\sum_{k=1}^{n_i} Z_{ki}^2, [(s_{n\theta i}(\boldsymbol{\theta}^*))_1]^2 \right) \\
= & E \left(\left[\sum_{k=1}^{n_i} Z_{ki}^2 \right] [(s_{n\theta i}(\boldsymbol{\theta}^*))_1]^2 \right) - E \left(\sum_{k=1}^{n_i} Z_{ki}^2 \right) E \left([(s_{n\theta i}(\boldsymbol{\theta}^*))_1]^2 \right)
\end{aligned} \tag{4.68}$$

Por (4.58) e (4.59) basta provar que o primeiro termo do segundo membro de (4.68) é finito, ou seja que

$$\left[E \left(\left[\sum_{k=1}^{n_i} Z_{ki}^2 \right] [(s_{n\theta i}(\theta^*))_1]^2 \right) \right]^2 \leq E \left(\left[\sum_{k=1}^{n_i} Z_{ki}^2 \right]^2 \right) E \left([(s_{n\theta i}(\theta^*))_1]^4 \right) < \infty. \quad (4.69)$$

De (4.63) segue que

$$E \left(\left[\sum_{k=1}^{n_i} Z_{ki}^2 \right]^2 \right) < \infty, \quad (4.70)$$

e esse fato conjuntamente com (4.65) fazem válida (4.69), o que implica que (4.68) é finito.

De maneira análoga prova-se que

$$Cov \left(\sum_{k=1}^{n_i} Z_{ki}^2, [(s_{n\theta i}(\theta^*))_2]^2 \right) < \infty. \quad (4.71)$$

Por (4.64) e (4.66) e pela desigualdade de Cauchy-Schwarz, temos

$$\left\{ E \left([(s_{n\theta i}(\theta^*))_1]^2 [(s_{n\theta i}(\theta^*))_2]^2 \right) \right\}^2 \leq E \left([(s_{n\theta i}(\theta^*))_1]^4 \right) E \left([(s_{n\theta i}(\theta^*))_2]^4 \right) < \infty.$$

Então utilizando (4.59), (4.60), prova-se que

$$\begin{aligned} & Cov \left([(s_{n\theta i}(\theta^*))_1]^2, [(s_{n\theta i}(\theta^*))_2]^2 \right) \\ &= E \left([(s_{n\theta i}(\theta^*))_1]^2 [(s_{n\theta i}(\theta^*))_2]^2 \right) - \left\{ E \left([(s_{n\theta i}(\theta^*))_1]^2 \right) \right\}^2 \left\{ E \left([(s_{n\theta i}(\theta^*))_2]^2 \right) \right\}^2 \\ &< \infty. \end{aligned} \quad (4.72)$$

A demonstração de (4.55) segue diretamente de (4.63), (4.65), (4.67), (4.68), (4.71) e (4.72). De (4.54) e (4.55) segue que (4.52) também é verdadeira. Logo existe uma constante L_i tal que

$$E \left(|S_{ni}(\theta_0^*)|^4 \right) < L_i.$$

Considerando $L = \max_{1 \leq i \leq n} \{L_i\}$, temos que (4.34) é válida.

Em síntese, as condições de regularidade requeridas neste caso estão satisfeitas. Considerando os cones $C_\Omega = \mathbb{R}^{p'} \times \{0\} \times \mathbb{R}$, $C_\tau = \mathbb{R}^{p'} \times [0, \infty) \times \mathbb{R}$, e as matrizes $\mathbf{T}_n = \mathbf{I}_3$, temos $\tilde{C}_\Omega = \tilde{C}_{\Omega_n} = \mathbf{T}_n \mathbf{G}_n^{t/2} (C_\Omega - \boldsymbol{\theta}_0) = \mathbb{R}^{p'} \times \{0\} \times \mathbb{R}$ e $\tilde{C}_\tau = \tilde{C}_{\tau_n} = \mathbf{T}_n \mathbf{G}_n^{t/2} (C_\tau - \boldsymbol{\theta}_0) = \mathbb{R}^{p'} \times [0, \infty) \times \mathbb{R}$. Então pelo Teorema 4.1.2, a distribuição de d_n existe e é a mesma que a distribuição de

$$\inf_{\boldsymbol{\theta} \in \tilde{C}_\Omega} |\mathbf{N} - \boldsymbol{\theta}|^2 - \inf_{\boldsymbol{\theta} \in \tilde{C}_\tau} |\mathbf{N} - \boldsymbol{\theta}|^2 = N_{p'+2}^2 - f(N_{p'+2}) = N_{p'+2}^2 I_{\{N_{p'+2} \geq 0\}}$$

em que $\mathbf{N} = (N_1, \dots, N_{p'}, N_{p'+1}, N_{p'+2})^t$ é um vetor aleatório com distribuição normal multivariada com media zero e matriz de covariância $\mathbf{I}_p$, e $f(N_{p'+2}) = \begin{cases} N_{p'+2}^2 & \text{se } N_{p'+2} \leq 0 \\ 0 & \text{se } N_{p'+2} \geq 0 \end{cases}$.

Logo a distribuição *assintótica* de d_n é uma mistura com pesos 0.5, de distribuições qui-quadrado com um grau de liberdade e com zero grau de liberdade (um ponto com a massa em zero).

Verificação das condições B2), B3) e B5) para a hipótese de inclusão de um segundo efeito aleatório.

Suponhamos agora que nosso interesse é testar a necessidade de inclusão de um segundo efeito aleatório no modelo (1.9), admitindo a existência de um primeiro efeito aleatório. A hipótese a testar será (1.13) contra a alternativa (1.14), ou equivalentemente (1.15) contra a alternativa (1.16).

Neste caso $\mathbf{b}_i$ tem distribuição $N(0, \mathbf{D})$, com

$$\mathbf{D} = \sigma_A^2 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \sigma_B^2 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \sigma_{AB} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

e $\sigma_A^2 > 0$, assim o espaço paramétrico considerado é $\Omega = \mathbb{R}^{p'} \times (0, +\infty) \times (0, +\infty) \times [0, +\infty) \times \mathbb{R}$, com $\boldsymbol{\theta} = (\sigma_A^2, \sigma_B^2, \sigma_{AB}, \sigma^2)$. Portanto as matrizes $\mathbf{G}_{ig}$ definidas em (4.19)

são dadas por

$$\begin{aligned}
\mathbf{G}_{i1} &= \mathbf{Z}_i \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \mathbf{Z}_i^t \\
&= \begin{bmatrix} \mathbf{z}_{i1} & \mathbf{z}_{i2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{z}_{i1}^t & \mathbf{z}_{i2}^t \end{bmatrix} \\
&= \mathbf{z}_{i1} \mathbf{z}_{i1}^t, \\
\mathbf{G}_{i2} &= \mathbf{z}_{i2} \mathbf{z}_{i2}^t, \\
\mathbf{G}_{i3} &= \begin{bmatrix} \mathbf{z}_{i1} & \mathbf{z}_{i2} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{z}_{i1}^t & \mathbf{z}_{i2}^t \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{z}_{i2} & \mathbf{z}_{i1} \end{bmatrix} \begin{bmatrix} \mathbf{z}_{i1}^t & \mathbf{z}_{i2}^t \end{bmatrix} \\
&= \mathbf{z}_{i1} \mathbf{z}_{i2}^t + \mathbf{z}_{i2} \mathbf{z}_{i1}^t.
\end{aligned}$$

Logo

$$\begin{aligned}
\Sigma_i(\boldsymbol{\theta}) &= \sigma^2 \mathbf{I}_{n_i} + \sigma_A^2 \mathbf{z}_{i1} \mathbf{z}_{i1}^t + \sigma_B^2 \mathbf{z}_{i2} \mathbf{z}_{i2}^t + \sigma_{AB} (\mathbf{z}_{i1} \mathbf{z}_{i2}^t + \mathbf{z}_{i2} \mathbf{z}_{i1}^t) \\
&= \sigma^2 \mathbf{I}_{n_i} + \sigma_A^2 \mathbf{z}_{i1} \mathbf{z}_{i1}^t + \sigma_B^2 \mathbf{z}_{i2} \mathbf{z}_{i2}^t \\
&\quad + \sigma_{AB} (\mathbf{z}_{i1} \mathbf{z}_{i1}^t \mathbf{z}_{i2} \mathbf{z}_{i2}^t + \mathbf{z}_{i2} \mathbf{z}_{i2}^t \mathbf{z}_{i1} \mathbf{z}_{i1}^t) \\
&= \sigma^2 \mathbf{I}_{n_i} + \sigma_A^2 \mathbf{G}_{i1} + \sigma_B^2 \mathbf{G}_{i2} + (\sigma_{AB}/k_i) (\mathbf{G}_{i1} \mathbf{G}_{i2} + \mathbf{G}_{i2} \mathbf{G}_{i1}).
\end{aligned}$$

Então é possível provar (ver Apêndice C) que a matriz inversa de $\Sigma_i(\boldsymbol{\theta})$ é

$$\Sigma_i^{-1}(\boldsymbol{\theta}) = \beta_0 \mathbf{I}_{n_i} + \beta_{1i} \mathbf{z}_{i1} \mathbf{z}_{i1}^t + \beta_{2i} \mathbf{z}_{i2} \mathbf{z}_{i2}^t + \beta_{3i} k_i (\mathbf{z}_{i1} \mathbf{z}_{i2}^t + \mathbf{z}_{i2} \mathbf{z}_{i1}^t) \quad (4.73)$$

com

$$k_i = \tau_{i12} = \mathbf{z}_{i1}^t \mathbf{z}_{i2} = \mathbf{z}_{i2}^t \mathbf{z}_{i1},$$

e

$$\beta_0 = \sigma^{-2}, \quad (4.74)$$

$$\beta_{1i}(\boldsymbol{\theta}) = -\frac{\sigma^{-2}}{m_{1i}} \left(\sigma_A^2 + \left(c_{1i}\sigma_A^2 - \frac{\sigma_{AB}m_{1i}}{k_i} \right) (d_{1i}m_{1i} - c_{1i}s_{1i})^{-1} s_{1i} \right), \quad (4.75)$$

$$\beta_{2i}(\boldsymbol{\theta}) = -\frac{\sigma^{-2}}{m_{2i}} \left(\sigma_B^2 + \left(c_{2i}\sigma_B^2 - \frac{\sigma_{AB}m_{2i}}{k_i} \right) (d_{2i}m_{2i} - c_{2i}s_{2i})^{-1} s_{2i} \right), \quad (4.76)$$

$$\begin{aligned} \beta_{3i}(\boldsymbol{\theta}) &= \sigma^{-2} \left(c_{1i}\sigma_A^2 - \frac{\sigma_{AB}m_{1i}}{k_i} \right) (d_{1i}m_{1i} - c_{1i}s_{1i})^{-1}, \\ &= \sigma^{-2} \left(c_{2i}\sigma_B^2 - \frac{\sigma_{AB}m_{2i}}{k_i} \right) (d_{2i}m_{2i} - c_{2i}s_{2i})^{-1}. \end{aligned} \quad (4.77)$$

com

$$\begin{aligned} k_{1i} &= \mathbf{z}_{i1}^t \mathbf{z}_{i1} = \text{tr} \{ \mathbf{z}_{i1} \mathbf{z}_{i1}^t \} \\ k_{2i} &= \mathbf{z}_{i2}^t \mathbf{z}_{i2} = \text{tr} \{ \mathbf{z}_{i2} \mathbf{z}_{i2}^t \}, \end{aligned}$$

e

$$\begin{aligned} m_{1i} &= \sigma^2 + \sigma_A^2 k_{1i} + \sigma_{AB} k_i, & m_{2i} &= \sigma^2 + \sigma_B^2 k_{2i} + \sigma_{AB} k_i, \\ s_{1i} &= \sigma_A^2 k^2 + \sigma_{AB} k_{2i} k_i, & s_{2i} &= \sigma_B^2 k^2 + \sigma_{AB} k_{1i} k_i, \\ c_{1i} &= \sigma_B^2 + (\sigma_{AB}/k_i) k_{1i}, & c_{2i} &= \sigma_A^2 + (\sigma_{AB}/k_i) k_{2i}, \\ d_{1i} &= \sigma^2 + \sigma_B^2 k_{2i} + \sigma_{AB} k_i, & d_{2i} &= \sigma^2 + \sigma_A^2 k_{1i} + \sigma_{AB} k_i. \end{aligned}$$

Sob a hipótese nula, $\sigma_B^2 = \sigma_{AB} = 0$, $\beta_0 = \sigma^{-2}$, $\beta_{1i}(\boldsymbol{\theta}) = -\frac{\sigma^{-2}}{\sigma^2 + \sigma_A^2 k_{1i}} \sigma_A^2$, $\beta_{2i}(\boldsymbol{\theta}) = \beta_{3i}(\boldsymbol{\theta}) = \beta_{4i}(\boldsymbol{\theta}) = 0$. Logo

$$\begin{aligned} \Sigma_i^{-1}(\boldsymbol{\theta}_0) &= \sigma^{-2} \mathbf{I}_{n_i} - \frac{\sigma^{-2}}{\sigma^2 + \sigma_A^2 k_{1i}} \sigma_A^2 \mathbf{z}_{i1} \mathbf{z}_{i1}^t \\ &= \sigma^{-2} \left(\mathbf{I}_{n_i} - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{z}_{i1} \mathbf{z}_{i1}^t \right). \end{aligned} \quad (4.78)$$

De (4.27) e (4.29) temos

$$\begin{aligned} \mathbf{G}_{n\alpha\alpha} &= \sum_{i=1}^n \mathbf{X}_i^t \Sigma_i^{-1}(\boldsymbol{\theta}_0) \mathbf{X}_i \\ &= \sum_{i=1}^n \sigma^{-2} \mathbf{X}_i^t \mathbf{X}_i - \frac{\sigma^{-2}}{\sigma^2 + \sigma_A^2 k_{1i}} \sigma_A^2 \mathbf{X}_i^t \mathbf{z}_{i1} \mathbf{z}_{i1}^t \mathbf{X}_i, \end{aligned} \quad (4.79)$$

e

$$\mathbf{G}_{n\theta\theta} = \left[\left\langle \frac{1}{2} \sum_{i=1}^n \text{tr} \Sigma_i^{-1}(\boldsymbol{\theta}_0) \mathbf{G}_{ig} \Sigma_i^{-1}(\boldsymbol{\theta}_0) \mathbf{G}_{ih} \right\rangle_{gh} \right], \quad (4.80)$$

A explicitação dos elementos de $\mathbf{G}_{n\theta\theta}$ está apresentada no Apêndice D.

Assim a matriz $\mathbf{G}_n$, definida em (4.7), pode ser escrita como $\mathbf{G}_n = \sum_{i=1}^n \mathbf{G}_{ni}$ com

$$\mathbf{G}_{ni} = \begin{bmatrix} \sigma^{-2} \mathbf{X}_i^t \mathbf{X}_i - \frac{\sigma^{-2}}{\sigma^2 + \sigma_A^2 k_{1i}} \sigma_A^2 \mathbf{X}_i^t \mathbf{z}_{i1} \mathbf{z}_{i1}^t \mathbf{X}_i & 0 \\ 0 & \mathbf{G}_{n\theta\theta i} \end{bmatrix}, \quad (4.81)$$

em que

$$\mathbf{G}_{n\theta\theta i} = \frac{\sigma^{-4}}{2} \times \begin{bmatrix} m_i^2 k_{1i} & m_i^2 k_i & 2m_i^2 k_{1i} k_i & m_i^2 k_{1i} \\ m_i^2 k_i^2 & (k_{2i} - j_i k_i^2)^2 & 2k_i (m_i k_{2i} - j_i k_i) & k_{2i} - j_i k_i^2 (2 - j_i k_{1i}) \\ 2m_i^2 k_{1i} k_i & 2k_i (m_i k_{2i} - j_i k_i) & 2m_i [k_{2i} k_{1i} + k_i^2 (1 - 2j_i k_{1i})] & 2m_i k_i \\ m_i^2 k_{1i} & k_{2i} - j_i k_i^2 (2 - j_i k_{1i}) & 2m_i k_i & n_i + j_i k_{1i} (m_i - 2) \end{bmatrix},$$

com $m_i = \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}}\right)$ e $j_i = \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}}$.

Por (4.31) $\mathbf{G}_{n\theta\theta i}$ é semidefinida positiva logo $\det(\mathbf{G}_{n\theta\theta i}) \geq 0$, porém como o posto da matriz $\mathbf{G}_{n\theta\theta i}$ é 4 (isto foi verificado com o pacote Mathematica 2.2.1) então $\det(\mathbf{G}_{n\theta\theta i}) \neq 0$, logo $\det(\mathbf{G}_{n\theta\theta i}) > 0$.

Como $\Sigma_i^{-1}(\boldsymbol{\theta}_0)$ é positiva definida existe uma matriz $\mathbf{B}$ $n_i \times n_i$ de posto n_i tal que $\Sigma_i^{-1}(\boldsymbol{\theta}_0) = \mathbf{B}^t \mathbf{B}$. Agora, $\text{posto}(\mathbf{X}_i^t) = \text{posto}(\mathbf{X}_i^t \mathbf{B}^t) = p'$ e logo $\mathbf{B} \mathbf{X}_i$ é uma matriz $n_i \times p'$ de posto p' tal que $\mathbf{X}_i^t \Sigma_i^{-1}(\boldsymbol{\theta}_0) \mathbf{X}_i = \mathbf{X}_i^t \mathbf{B}^t \mathbf{B} \mathbf{X}_i$. Portanto $\mathbf{X}_i^t \Sigma_i^{-1}(\boldsymbol{\theta}_0) \mathbf{X}_i$ é positiva definida

e temos

$$\det \left(\sigma^{-2} \mathbf{X}_i^t \mathbf{X}_i - \frac{\sigma^{-2}}{\sigma^2 + \sigma_A^2 k_{1i}} \sigma_A^2 \mathbf{X}_i^t \mathbf{z}_{i1} \mathbf{z}_{i1}^t \mathbf{X}_i \right) = \det \left(\mathbf{X}_i^t \Sigma_i^{-1}(\boldsymbol{\theta}_0) \mathbf{X}_i \right) > 0. \quad (4.82)$$

Logo $\det(\mathbf{G}_{ni}) = \det \left(\sigma^{-2} \mathbf{X}_i^t \mathbf{X}_i - \frac{\sigma^{-2}}{\sigma^2 + \sigma_A^2 k_{1i}} \sigma_A^2 \mathbf{X}_i^t \mathbf{z}_{i1} \mathbf{z}_{i1}^t \mathbf{X}_i \right) \det(\mathbf{G}_{n\boldsymbol{\theta}i}) > 0$ então existe constante $c_0 > 0$ tal que

$$\lambda_{\min}(\mathbf{G}_{ni}) \geq c_0, \quad 1 \leq i \leq n,$$

logo

$$\lambda_{\min}(\mathbf{G}_n) \geq \sum_{i=1}^n \lambda_{\min}(\mathbf{G}_{ni}) \geq c_0 n, \quad (4.83)$$

implicando B2).

Provaremos agora B3) e B5).

A condição (B3) é verificada se, por exemplo, (4.10) e (4.11) são válidas como já foi mencionado anteriormente.

Para provar (4.10) e (4.11), notemos primeiramente que

$$\Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{G}_{i1} \quad (4.84)$$

$$= \beta_0(\boldsymbol{\theta}) \mathbf{z}_{i1} \mathbf{z}_{i1}^t + \beta_{1i}(\boldsymbol{\theta}) k_{1i} \mathbf{z}_{i1} \mathbf{z}_{i1}^t + \beta_{2i}(\boldsymbol{\theta}) k_i \mathbf{z}_{i1} \mathbf{z}_{i2}^t + \beta_{3i}(\boldsymbol{\theta}) k_i (k_{1i} \mathbf{z}_{i1} \mathbf{z}_{i2}^t + k_i \mathbf{z}_{i1} \mathbf{z}_{i1}^t)$$

$$\Sigma_i^{-1}(\boldsymbol{\theta}) \mathbf{G}_{i2} \quad (4.85)$$

$$= \beta_0(\boldsymbol{\theta}) \mathbf{z}_{i2} \mathbf{z}_{i2}^t + \beta_{1i}(\boldsymbol{\theta}) k_i \mathbf{z}_{i2} \mathbf{z}_{i1}^t + \beta_{2i}(\boldsymbol{\theta}) k_{2i} \mathbf{z}_{i2} \mathbf{z}_{i2}^t + \beta_{3i}(\boldsymbol{\theta}) k_i (k_i \mathbf{z}_{i2} \mathbf{z}_{i2}^t + k_{2i} \mathbf{z}_{i2} \mathbf{z}_{i1}^t)$$

$$\Sigma_i^{-1}(\theta) \mathbf{G}_{i3} \quad (4.86)$$

$$\begin{aligned} &= \beta_0(\theta) \mathbf{z}_{i1} \mathbf{z}_{i2}^t + \beta_{1i}(\theta) k_i \mathbf{z}_{i1} \mathbf{z}_{i1}^t + \beta_{2i}(\theta) k_{2i} \mathbf{z}_{i1} \mathbf{z}_{i2}^t \\ &\quad + \beta_{3i}(\theta) k_i (k_i \mathbf{z}_{i1} \mathbf{z}_{i2}^t + k_{2i} \mathbf{z}_{i1} \mathbf{z}_{i1}^t) + \beta_0(\theta) \mathbf{z}_{i2} \mathbf{z}_{i1}^t + \beta_{1i}(\theta) k_{1i} \mathbf{z}_{i2} \mathbf{z}_{i1}^t \\ &\quad + \beta_{2i}(\theta) k_i \mathbf{z}_{i2} \mathbf{z}_{i2}^t + \beta_{3i}(\theta) k_i (k_{1i} \mathbf{z}_{i2} \mathbf{z}_{i2}^t + k_i \mathbf{z}_{i2} \mathbf{z}_{i1}^t) \\ &\Sigma_i^{-1}(\theta) \mathbf{G}_{i4} \end{aligned} \quad (4.87)$$

$$= \beta_0(\theta) \mathbf{I}_{n_i} + \beta_{1i}(\theta) \mathbf{z}_{i1} \mathbf{z}_{i1}^t + \beta_{2i}(\theta) \mathbf{z}_{i2} \mathbf{z}_{i2}^t + \beta_{3i}(\theta) k_i (\mathbf{z}_{i1} \mathbf{z}_{i2}^t + \mathbf{z}_{i2} \mathbf{z}_{i1}^t).$$

Como de (4.24) e (4.27) temos $\mathbf{F}_{n\alpha\alpha}(\theta_0^*) - \mathbf{G}_{n\alpha\alpha} = \mathbf{0}$. Utilizando agora (4.40), calcularemos os elementos da diagonal da matriz $\mathbf{F}_{n\theta\theta}(\theta_0^*) - \mathbf{G}_{n\theta\theta}$, ou seja

$$\begin{aligned} &(\mathbf{F}_{n\theta\theta}(\theta_0^*) - \mathbf{G}_{n\theta\theta})_{gg} \\ &= \frac{1}{2} \sum_{i=1}^n \text{tr} \{ \Sigma_i^{-1}(\theta_0) \mathbf{G}_{ig} \Sigma_i^{-1}(\theta_0) [2\mathbf{e}_i \mathbf{e}_i^t - \Sigma_i(\theta_0)] \Sigma_i^{-1}(\theta_0) \mathbf{G}_{ig} \} \\ &\quad - \frac{1}{2} \sum_{i=1}^n \text{tr} \{ \Sigma_i^{-1}(\theta_0) \mathbf{G}_{ig} \Sigma_i^{-1}(\theta_0) \mathbf{G}_{ig} \} \\ &= \sum_{i=1}^n \text{tr} \{ \Sigma_i^{-1}(\theta_0) \mathbf{G}_{ig} \Sigma_i^{-1}(\theta_0) \mathbf{e}_i \mathbf{e}_i^t \Sigma_i^{-1}(\theta_0) \mathbf{G}_{ig} \} - \sum_{i=1}^n \text{tr} \{ \Sigma_i^{-1}(\theta_0) \mathbf{G}_{ig} \Sigma_i^{-1}(\theta_0) \mathbf{G}_{ig} \}. \end{aligned}$$

Sob a hipótese nula, $\mathbf{e}_i \sim N(\mathbf{0}, \Sigma_i(\theta_0))$ e então

$$\begin{aligned} &\sum_{i=1}^n \text{tr} \{ \Sigma_i^{-1}(\theta_0) \mathbf{G}_{ig} \Sigma_i^{-1}(\theta_0) \mathbf{e}_i \mathbf{e}_i^t \Sigma_i^{-1}(\theta_0) \mathbf{G}_{ig} \} \\ &= \sum_{i=1}^n \text{tr} \{ \mathbf{M}_{ig} \mathbf{e}_i \mathbf{e}_i^t \mathbf{V}_{ig}^t \} \\ &= \sum_{i=1}^n \mathbf{e}_i^t \mathbf{V}_{ig}^t \mathbf{M}_{ig} \mathbf{e}_i \end{aligned}$$

com $\mathbf{M}_{ig} = \Sigma_i^{-1}(\theta_0) \mathbf{G}_{ig} \Sigma_i^{-1}(\theta_0)$ e $\mathbf{V}_{ig} = \mathbf{G}_{ig} \Sigma_i^{-1}(\theta_0) = \Sigma_i(\theta_0) \mathbf{M}_{ig}$. Como $\mathbf{V}_{ig}^t \mathbf{M}_{ig} =$

$\mathbf{M}_{ig} \boldsymbol{\Sigma}_i(\boldsymbol{\theta}_0) \mathbf{M}_{ig} = (\mathbf{V}_{ig}^t \mathbf{M}_{ig})^t$ é uma matriz simétrica temos que

$$\begin{aligned} E(\mathbf{e}_i^t \mathbf{V}_{ig}^t \mathbf{M}_{ig} \mathbf{e}_i) &= \text{tr} \{ \mathbf{V}_{ig}^t \mathbf{M}_{ig} \boldsymbol{\Sigma}_i(\boldsymbol{\theta}_0) \} = \text{tr} \{ \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0) \mathbf{G}_{ig} \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0) \mathbf{G}_{ig} \} \\ \text{Var}(\mathbf{e}_i^t \mathbf{V}_{ig}^t \mathbf{M}_{ig} \mathbf{e}_i) &= 2(\text{tr} \{ \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0) \mathbf{G}_{ig} \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0) \mathbf{G}_{ig} \}) \\ K_r(\mathbf{e}_i^t \mathbf{V}_{ig}^t \mathbf{M}_{ig} \mathbf{e}_i) &= 2^{r-1} (r-1)! (\text{tr} \{ \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0) \mathbf{G}_{ig} \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0) \mathbf{G}_{ig} \})^r, \end{aligned}$$

em que K_r denota o r -ésimo cumulante (ver, por exemplo, Searle, 1971, Teorema 1).

Sob a hipótese nula, $\sigma_B^2 = \sigma_{AB} = 0$, $\beta_0 = \sigma^{-2}$, $\beta_{1i} = -\frac{\sigma^{-2}}{\sigma^2 + \sigma_A^2 k_{1i}} \sigma_A^2$, $\beta_{2i} = \beta_{3i} = \beta_{4i} = 0$. Logo de (4.84) a (4.87), temos

$$\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0) \mathbf{G}_{i1} = \sigma^{-2} \mathbf{z}_{i1} \mathbf{z}_{i1}^t + \beta_{1i} k_{1i} \mathbf{z}_{i1} \mathbf{z}_{i1}^t \quad (4.88)$$

$$\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0) \mathbf{G}_{i2} = \sigma^{-2} \mathbf{z}_{i2} \mathbf{z}_{i2}^t + \beta_{1i} k_{1i} \mathbf{z}_{i2} \mathbf{z}_{i1}^t \quad (4.89)$$

$$\begin{aligned} \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0) \mathbf{G}_{i3} &= \sigma^{-2} \mathbf{z}_{i1} \mathbf{z}_{i2}^t + \beta_{1i} k_{1i} \mathbf{z}_{i1} \mathbf{z}_{i1}^t \\ &\quad + \sigma^{-2} \mathbf{z}_{i2} \mathbf{z}_{i1}^t + \beta_{1i} k_{1i} \mathbf{z}_{i2} \mathbf{z}_{i1}^t \end{aligned} \quad (4.90)$$

$$\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}) \mathbf{G}_{i4} = \sigma^{-2} \mathbf{I}_{n_i} + \beta_{1i} \mathbf{z}_{i1} \mathbf{z}_{i1}^t. \quad (4.91)$$

Então de (4.18) e (4.88) a (4.91), temos que $K_r(\mathbf{e}_i^t \mathbf{V}_{ig}^t \mathbf{M}_{ig} \mathbf{e}_i) < \infty$, e para $\delta = 3$

$$\frac{1}{n^4} \sum_{i=1}^n E \left(\left| (\mathbf{e}_i^t \mathbf{V}_{ig}^t \mathbf{M}_{ig} \mathbf{e}_i) - [E(\mathbf{e}_i^t \mathbf{V}_{ig}^t \mathbf{M}_{ig} \mathbf{e}_i)] \right|^4 \right) \xrightarrow{n \rightarrow \infty} 0.$$

Logo pela lei fraca dos grandes números de Markov

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n \mathbf{e}_i^t \mathbf{V}_{ig}^t \mathbf{M}_{ig} \mathbf{e}_i - \frac{1}{n} \sum_{i=1}^n E(\mathbf{e}_i^t \mathbf{V}_{ig}^t \mathbf{M}_{ig} \mathbf{e}_i) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{e}_i^t \mathbf{V}_{ig}^t \mathbf{M}_{ig} \mathbf{e}_i - \frac{1}{n} \sum_{i=1}^n \text{tr} \{ \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0) \mathbf{G}_{ig} \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0) \mathbf{G}_{ig} \} \\ &= \frac{1}{n} (\mathbf{F}_{n\theta\theta}(\boldsymbol{\theta}_0^*) - \mathbf{G}_{n\theta\theta})_{gg} \xrightarrow{p} 0, \end{aligned}$$

para todo $g = 1, \dots, 4$; portanto vale (4.10).

Agora para verificar (4.11), notemos que

$$\begin{aligned}\Sigma_i^{-1}(\boldsymbol{\theta}) - \Sigma_i^{-1}(\boldsymbol{\theta}_0) &= (\beta_{1i}(\boldsymbol{\theta}) - \beta_{1i}(\boldsymbol{\theta}_0)) \mathbf{z}_{i1} \mathbf{z}_{i1}^t \\ &\quad + (\beta_{2i}(\boldsymbol{\theta}) - \beta_{2i}(\boldsymbol{\theta}_0)) \mathbf{z}_{i2} \mathbf{z}_{i2}^t \\ &\quad + (\beta_{3i}(\boldsymbol{\theta}) - \beta_{3i}(\boldsymbol{\theta}_0)) k_i (\mathbf{z}_{i1} \mathbf{z}_{i2}^t + \mathbf{z}_{i2} \mathbf{z}_{i1}^t) .\end{aligned}$$

Logo

$$\begin{aligned}\mathbf{F}_{n\alpha\alpha}(\boldsymbol{\theta}^*) - \mathbf{F}_{n\alpha\alpha}(\boldsymbol{\theta}_0^*) &= \sum_{i=1}^n \text{tr}(\mathbf{X}_i^t (\Sigma_i^{-1}(\boldsymbol{\theta}) - \Sigma_i^{-1}(\boldsymbol{\theta}_0)) \mathbf{X}_i) \\ &= \sum_{i=1}^n (\beta_{1i}(\boldsymbol{\theta}) - \beta_{1i}(\boldsymbol{\theta}_0)) \text{tr}(\mathbf{X}_i^t \mathbf{z}_{i1} \mathbf{z}_{i1}^t \mathbf{X}_i) \\ &\quad + (\beta_{2i}(\boldsymbol{\theta}) - \beta_{2i}(\boldsymbol{\theta}_0)) \text{tr}(\mathbf{X}_i^t \mathbf{z}_{i2} \mathbf{z}_{i2}^t \mathbf{X}_i) \\ &\quad + (\beta_{3i}(\boldsymbol{\theta}) - \beta_{3i}(\boldsymbol{\theta}_0)) k_i \text{tr}(\mathbf{X}_i^t \mathbf{z}_{i1} \mathbf{z}_{i2}^t \mathbf{X}_i) \\ &\quad + (\beta_{3i}(\boldsymbol{\theta}) - \beta_{3i}(\boldsymbol{\theta}_0)) k_i \text{tr}(\mathbf{X}_i^t \mathbf{z}_{i2} \mathbf{z}_{i1}^t \mathbf{X}_i) .\end{aligned}\quad (4.92)$$

Das definições (4.74) a (4.77), é claro que para todo $\boldsymbol{\theta}$ e para todo $1 \leq i \leq n$ existem constantes $G_1, G_2, G_3, G'_1, G'_2, G'_3$, tais que

$$G'_j \leq (\beta_{ji}(\boldsymbol{\theta}) - \beta_{ji}(\boldsymbol{\theta}_0)) \leq G_j \quad (4.93)$$

para $j = 1, 2, 3$. Notemos que quando σ_A^2, σ_B^2 ou σ_{AB} tendem a infinito m_{li}, d_{li}, c_{li} e s_{li} , para $l = 1, 2$ tendem a infinito, porém $\beta_{ji}(\boldsymbol{\theta}), j = 1, 2, 3$, é limitado, igualmente $\beta_{ji}(\boldsymbol{\theta}_0)$, de igual maneira são limitados quando σ_A^2, σ_B^2 ou σ_{AB} tendem a zero. De (4.17) temos

$$\frac{1}{n} (\mathbf{F}_{n\alpha\alpha}(\boldsymbol{\theta}^*) - \mathbf{F}_{n\alpha\alpha}(\boldsymbol{\theta}_0^*)) = O(1) .$$

Seja $\mathbf{F}_{n\theta\theta}(\boldsymbol{\theta}^*)$ é definida em (4.26), a expressão (4.47) é válida também neste caso, logo por argumentos similares aos que foram utilizados ao provar que (4.10) é válida quando testamos a inclusão de um efeito aleatório, e considerando as definições (4.74) a (4.77),

(4.84) a (4.87), (4.88) a (4.91) e a condição (4.18) temos que

$$\frac{1}{n} (\mathbf{F}_{n\theta\theta}(\boldsymbol{\theta}^*))_{gg} = o_p(1)$$

e

$$\frac{1}{n} (\mathbf{F}_{n\theta\theta}(\boldsymbol{\theta}_0^*))_{gg} = o_p(1),$$

portanto vale (4.11).

Para verificar que a condição (B5) é válida pelo teorema de Liapounov, por (4.83) basta provar (4.34) com, por exemplo, $\delta = 4$, $1 \leq i \leq n$, para alguma constante L , com o vetor

$$\begin{aligned} \mathbf{S}_{ni}(\boldsymbol{\theta}_0^*) &= (\mathbf{s}_{n\alpha i}(\boldsymbol{\theta}_0^*), \mathbf{s}_{n\theta i}(\boldsymbol{\theta}_0^*))^t \\ &= (\mathbf{s}_{n\alpha i}(\boldsymbol{\theta}_0^*), (\mathbf{s}_{n\theta i}(\boldsymbol{\theta}^*))_1, (\mathbf{s}_{n\theta i}(\boldsymbol{\theta}^*))_2, (\mathbf{s}_{n\theta i}(\boldsymbol{\theta}^*))_3, (\mathbf{s}_{n\theta i}(\boldsymbol{\theta}^*))_4)^t, \end{aligned}$$

e $\mathbf{s}_{n\alpha i}(\boldsymbol{\theta}^*)$ e $\mathbf{s}_{n\theta i}(\boldsymbol{\theta}^*)$ definidas em (4.21) e (4.23).

Com argumentos análogos aos considerados quando testamos a inclusão de um efeito aleatório, por (4.53) provaremos que

$$E(|\mathbf{S}_{ni}(\boldsymbol{\theta}_0^*)|^2) < \infty, \quad (4.94)$$

e

$$\text{Var}(|\mathbf{S}_{ni}(\boldsymbol{\theta}_0^*)|^2) < \infty. \quad (4.95)$$

Para provar (4.94) e (4.95), respectivamente, basta provar que para todo g valem

$$E\left(\left|(\mathbf{s}_{n\theta i}(\boldsymbol{\theta}^*))_g\right|^2\right) = E\left(\left[\tilde{k}_{ig} + \frac{1}{2}\mathbf{e}_i^t(\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta})\mathbf{G}_{ig}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0))\mathbf{e}_i\right]^2\right) < \infty. \quad (4.96)$$

e

$$E\left(\left|(\mathbf{s}_{n\theta i}(\boldsymbol{\theta}^*))_g\right|^4\right) = E\left(\left[\tilde{k}_g + \frac{1}{2}\mathbf{e}_i^t(\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta})\mathbf{G}_{ig}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0))\mathbf{e}_i\right]^4\right) < \infty, \quad (4.97)$$

com a constante $\tilde{k}_{ig} = -\frac{1}{2} \text{tr} \left(\Sigma_i^{-1}(\theta_0) \dot{\Sigma}_{i(g)}(\theta_0) \right) = -\frac{1}{2} \text{tr} \left(\Sigma_i^{-1}(\theta_0) \mathbf{G}_{ig} \right)$.

Para $g = 1, 2$ temos

$$\begin{aligned} E \left(\left| (s_{n\theta i}(\theta^*))_g \right|^2 \right) &= E \left(\left\{ \tilde{k}_{ig} + \frac{1}{2} \mathbf{e}_i^t \Sigma_i^{-1}(\theta) \mathbf{z}_{ig} \mathbf{z}_{ig}^t \Sigma_i^{-1}(\theta) \mathbf{e}_i \right\}^2 \right) \\ &= E \left(\left\{ \tilde{k}_{ig} + \frac{k_{gi}}{2} \mathbf{e}_i^t \Sigma_i^{-1}(\theta) [(k_{gi})^{-1} \mathbf{z}_{ig} \mathbf{z}_{ig}^t] \Sigma_i^{-1}(\theta) \mathbf{e}_i \right\}^2 \right). \end{aligned}$$

Como $\Sigma_i^{-1}(\theta_0) \mathbf{e}_i \sim N(0, \Sigma_i^{-1}(\theta_0) \Sigma_i(\theta) \Sigma_i^{-1}(\theta_0))$, e $(k_{gi})^{-1} \mathbf{z}_{ig} \mathbf{z}_{ig}^t$ é uma matriz idempotente, então $\mathbf{e}_i^t \Sigma_i^{-1}(\theta) [(k_{gi})^{-1} \mathbf{z}_{ig} \mathbf{z}_{ig}^t] \Sigma_i^{-1}(\theta_0) \mathbf{e}_i \sim \chi^2(r(\mathbf{z}_{ig} \mathbf{z}_{ig}^t), 0)$, assim (4.96) e (4.97) são válidas para $g = 1, 2$.

Para $g = 3$ temos

$$\begin{aligned} &E \left(\left| (s_{n\theta i}(\theta^*))_3 \right|^2 \right) \\ &= E \left(\left[\tilde{k}_{i3} + \frac{1}{2} \mathbf{e}_i^t \Sigma_i^{-1}(\theta) (\mathbf{z}_{i1} \mathbf{z}_{i2}^t + \mathbf{z}_{i2} \mathbf{z}_{i1}^t) \Sigma_i^{-1}(\theta_0) \mathbf{e}_i \right]^2 \right) \\ &= E \left(\left[\tilde{k}_{i3} + \frac{1}{2} \mathbf{e}_i^t \Sigma_i^{-1}(\theta) \mathbf{z}_{i1} \mathbf{z}_{i2}^t \Sigma_i^{-1}(\theta_0) \mathbf{e}_i + \frac{1}{2} \mathbf{e}_i^t \Sigma_i^{-1}(\theta) \mathbf{z}_{i2} \mathbf{z}_{i1}^t \Sigma_i^{-1}(\theta_0) \mathbf{e}_i \right]^2 \right) \\ &= E \left(\left[\tilde{k}_{i3} + \frac{k_i}{2} w_1 + \frac{k_i}{2} w_2 \right]^2 \right) \\ &= \text{Var} \left(\tilde{k}_3 + \frac{k_i}{2} \{w_1 + w_2\} \right) + \left[E \left(\tilde{k}_g + \frac{k_i}{2} \{w_1 + w_2\} \right) \right]^2, \end{aligned}$$

em que $w_1 = \mathbf{e}_i^t \Sigma_i^{-1}(\theta) [k_i^{-1} \mathbf{z}_{i1} \mathbf{z}_{i2}^t] \Sigma_i^{-1}(\theta_0) \mathbf{e}_i$ e $w_2 = \mathbf{e}_i^t \Sigma_i^{-1}(\theta) [k_i^{-1} \mathbf{z}_{i2} \mathbf{z}_{i1}^t] \Sigma_i^{-1}(\theta_0) \mathbf{e}_i$. Como $k_i^{-1} \mathbf{z}_{i1} \mathbf{z}_{i2}^t$ e $k_i^{-1} \mathbf{z}_{i2} \mathbf{z}_{i1}^t$ são matrizes idempotentes então w_1 e w_2 têm distribuição $\chi^2(r(\mathbf{z}_{i1} \mathbf{z}_{i2}^t), 0)$; logo têm momentos finitos e portanto vale (4.96). Pela desigualdade de Cauchy-Schwarz, é possível provar que

$$E(|w_1^d w_2^c|) \leq (E(w_1^{2d}) E(w_2^{2c}))^{1/2} < \infty,$$

com $d, c = 1, \dots, 3$, logo

$$\begin{aligned}
& E(|(\mathbf{s}_{n\theta i}(\boldsymbol{\theta}^*))_3|^4) \\
&= E\left(\left[\tilde{k}_{i3} + \frac{k_i}{2}\{w_1 + w_2\}\right]^4\right) \\
&= E\left(\left[\tilde{k}_{i3}^2 + 2\tilde{k}_{i3}\frac{k_i}{2}\{w_1 + w_2\} + \left(\frac{k_i}{2}\{w_1 + w_2\}\right)^2\right]^2\right) \\
&= E\left(\left[\tilde{k}_{i3}^2 + k_i\tilde{k}_{i3}\{w_1 + w_2\} + \frac{k_i^2}{4}w_1^2 + \frac{k_i^2}{4}w_2^2 + \frac{1}{2}k_i^2w_1w_2\right]^2\right) \\
&< \infty,
\end{aligned}$$

ou seja, que (4.97) vale para $g = 3$.

A prova de que (4.96) e (4.97) são válidas para $g = 4$ é direta, considerando que

$$E(|(\mathbf{s}_{n\theta i}(\boldsymbol{\theta}^*))_4|^2) = E\left(\left(\tilde{k}_{i4} + \frac{1}{2}v_4\right)^2\right),$$

com $v_4 = \mathbf{e}_i^t \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}) \mathbf{I}_{n_i} \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0) \mathbf{e}_i \sim \chi^2(n_i, 0)$.

Logo provamos que (4.96) e (4.97) são válidas para $g = 1, \dots, 4$, ou seja, provamos (4.94) e (4.95) e portanto (4.34) com $\delta = 4$, logo a condição B5) vale.

Como todas as condições de regularidade requeridas por Vu e Zhou (1997) estão satisfeitas sob as restrições (4.17) para as matrizes $\mathbf{X}_i$ e $\mathbf{z}_i$ podemos aplicar o Teorema 4.1.2.

Considerando os cones $C_\Omega = \mathbb{R}^{p'} \times (0, \infty) \times \{0\} \times \{0\} \times \mathbb{R}$,

$C_\tau = \mathbb{R}^{p'} \times (0, \infty) \times [0, \infty) \times \mathbb{R} \times \mathbb{R}$, e as matrizes $\mathbf{T}_n = \mathbf{I}_3$ temos $\tilde{C}_\Omega = \tilde{C}_{\Omega_n} = \mathbf{T}_n \mathbf{G}_n^{t/2}(C_\Omega - \boldsymbol{\theta}_0) = \mathbb{R}^{p'} \times (0, \infty) \times \{0\} \times \{0\} \times \mathbb{R}$ e $\tilde{C}_\tau = \tilde{C}_{\tau_n} = \mathbf{T}_n \mathbf{G}_n^{t/2}(C_\tau - \boldsymbol{\theta}_0) = \mathbb{R}^{p'} \times (0, \infty) \times [0, \infty) \times \mathbb{R} \times \mathbb{R}$. Pelo Teorema 4.1.2, a distribuição de d_n existe e é a

mesma que a distribuição de

$$\inf_{\boldsymbol{\theta} \in \tilde{C}_\Omega} |\mathbf{N} - \boldsymbol{\theta}|^2 - \inf_{\boldsymbol{\theta} \in \tilde{C}_\tau} |\mathbf{N} - \boldsymbol{\theta}|^2 = N_{p'+2}^2 - f(N_{p'+2}) + N_{p'+3}^2 = N_{p'+2}^2 I_{\{N_{p'+2} \geq 0\}} + N_{p'+3}^2,$$

em que $\mathbf{N} = (N_1, \dots, N_{p'}, N_{p'+1}, N_{p'+2}, N_{p'+3})^t$ é um vetor aleatório com distribuição normal multivariada com media zero e matriz de covariância $\mathbf{I}_p$, e $f(N_{p'+2}) = \begin{cases} N_{p'+2}^2 & \text{se } N_{p'+2} \leq 0 \\ 0 & \text{se } N_{p'+2} \geq 0 \end{cases}$

Logo a distribuição *assintótica* de d_n é uma mistura com pesos 0.5, de distribuições qui-quadrado com um grau de liberdade e com dois graus de liberdade.

4.2 Discussão

Neste trabalho consideramos a estatística da razão de verossimilhanças apresentada por Vu e Zhou (1997), para testar hipóteses sobre a inclusão de efeitos aleatórios em modelos lineares. Os resultados são válidos nos casos em que as variáveis aleatórias envolvidas não são identicamente distribuídas. Para resolver o mesmo problema Stram e Lee (1994) utilizam a teoria sobre a distribuição assintótica do teste da razão de verossimilhanças em condições não regulares desenvolvida por Self e Liang (1987), mas essa teoria é válida apenas para variáveis aleatórias identicamente distribuídas, o que não ocorre em geral, nos problemas abordados aqui.

A estatística da razão de verossimilhanças apresentada por Vu e Zhou (1997) tem condições de regularidade que são mais trabalhosas de serem verificadas quando o número de parâmetros aumenta. Além disso, a determinação dos cones que aproximam os espaços das hipóteses nula e alternativa são difíceis de serem especificados nestes casos. Nos casos mais gerais tratados por Stram e Lee (1994), esses autores também consideram cones para aproximar porém uma cuidadosa revisão dos resultados é necessária para saber se esses cones realmente aproximam os espaços e se as condições requeridas são satisfeitas. Como tópico de futuras pesquisas propomos o estudo de testes de hipóteses para componentes de variância em modelos de efeitos aleatórios longitudinais que considerem um maior número de parâmetros. Além disso pretendemos estudar outras estatísticas que precisem de condições de regularidade mais simples de serem verificadas.

Apêndice A

Definição de estimadores para uma média restrita

Seja $W_1, \dots, W_n$ uma amostra com distribuição normal com média ξ e variância σ^2 , existem varios estimadores propostos na literatura para a média quando está pertence a um intervalo aberto. Gupta e Roghatgi (1980) apresentam os seguintes estimadores

$$\hat{\xi} = f(\overline{W}), \quad (\text{A.1})$$

$$\hat{\xi}_1 = |\overline{W}|, \quad (\text{A.2})$$

$$\hat{\xi}_2 = \frac{1}{2} \left[\sum_{i=1}^n |W_i| \right], \quad (\text{A.3})$$

$$\hat{\xi}_3 = \frac{1}{2} \left[\sum_{i=1}^n f(W_i) \right], \quad (\text{A.4})$$

onde $f(x) = \max(0, x)$, juntamente com as expressões de seus respectivos vícios definidos por

$$vicio(\tilde{\xi}) = E(\tilde{\xi}) - \xi,$$

ou sejam

$$vicio(\widehat{\xi}) = \frac{\sigma}{\sqrt{n}}\varphi\left(\frac{\xi\sqrt{n}}{\sigma}\right) - \xi\Phi\left(-\frac{\xi\sqrt{n}}{\sigma}\right), \quad (A.5)$$

$$vicio(\widehat{\xi}_1) = \frac{2\sigma}{\sqrt{n}}\varphi\left(\frac{\xi\sqrt{n}}{\sigma}\right) - 2\xi\Phi\left(-\frac{\xi\sqrt{n}}{\sigma}\right), \quad (A.6)$$

$$vicio(\widehat{\xi}_2) = 2\sigma\varphi\left(\frac{\xi}{\sigma}\right) - 2\xi\Phi\left(-\frac{\xi}{\sigma}\right), \quad (A.7)$$

e

$$vicio(\widehat{\xi}_3) = \sigma\varphi\left(\frac{\xi}{\sigma}\right) - \xi\Phi\left(-\frac{\xi}{\sigma}\right), \quad (A.8)$$

em que φ é a função de densidade da distribuição normal padrão e erros quadráticos médios definidos por

$$MSE(\widetilde{\xi}) = E(\widetilde{\xi} - \xi)^2,$$

ou sejam

$$MSE(\widehat{\xi}) = \frac{\sigma^2}{n} \left\{ 1 + \left(\frac{n\xi^2}{\sigma^2} - 1 \right) \Phi\left(-\frac{\xi\sqrt{n}}{\sigma}\right) - \frac{\xi\sqrt{n}}{\sigma} \varphi\left(\frac{\xi\sqrt{n}}{\sigma}\right) \right\}, \quad (A.9)$$

$$MSE(\widehat{\xi}_1) = \frac{\sigma^2}{n} \left\{ 1 + \frac{4\xi^2}{\sigma^2} \Phi\left(-\frac{\xi\sqrt{n}}{\sigma}\right) - \frac{4\xi\sqrt{n}}{\sigma} \varphi\left(\frac{\xi\sqrt{n}}{\sigma}\right) \right\}, \quad (A.10)$$

$$MSE(\widehat{\xi}_2) = \frac{\sigma^2}{n} \left\{ 1 - \frac{4\xi}{\sigma} \left[\varphi\left(\frac{\xi}{\sigma}\right) - \frac{\xi}{\sigma} \Phi\left(-\frac{\xi}{\sigma}\right) \right] + 4(n-1) \left[\varphi\left(\frac{\xi}{\sigma}\right) - \frac{\xi}{\sigma} \Phi\left(-\frac{\xi}{\sigma}\right) \right]^2 \right\}, \quad (A.11)$$

$$MSE(\widehat{\xi}_3) = \frac{\sigma^2}{n} \left\{ \Phi\left(\frac{\xi}{\sigma}\right) - \frac{\xi}{\sigma} \left[\varphi\left(\frac{\xi}{\sigma}\right) - \frac{\xi}{\sigma} \Phi\left(-\frac{\xi}{\sigma}\right) \right] + (n-1) \left[\varphi\left(\frac{\xi}{\sigma}\right) - \frac{\xi}{\sigma} \Phi\left(-\frac{\xi}{\sigma}\right) \right]^2 \right\}. \quad (A.12)$$

O estimador $\widehat{\xi}$ (A.1) foi considerado por Heiny e Sidiqui (1970) que também provaram que não existe estimador não viciado de ξ .

Para tener uma ideia da magnitud os vícios para o estimador $\widehat{\xi}$, calculados por (A.5)

Tabela A.1: vícios para o estimador ξ

ξ	$\sigma = 0.5$			$\sigma = 1.0$		
	$n = 5$	$n = 10$	$n = 15$	$n = 5$	$n = 10$	$n = 15$
0.0	0.0892	0.0631	0.0515	0.1784	0.1262	0.1030
0.1	0.0480	0.0253	0.0162	0.1328	0.0824	0.0606
0.2	0.0227	0.0077	0.0034	0.0960	0.0506	0.0324
0.3	0.0093	0.0018	0.0004	0.0671	0.0290	0.0156
0.4	0.0033	0.0003	0.0000	0.0453	0.0155	0.0068
0.5	0.0009	0.0000	0.0000	0.0296	0.0077	0.0026
ξ	$\sigma = 1.5$			$\sigma = 2.0$		
	$n = 5$	$n = 10$	$n = 15$	$n = 5$	$n = 10$	$n = 15$
0.0	0.2676	0.1892	0.1545	0.3568	0.2523	0.2060
0.1	0.2206	0.1434	0.1096	0.3090	0.2055	0.1598
0.2	0.1794	0.1058	0.0747	0.2657	0.1648	0.1212
0.3	0.1439	0.0759	0.0487	0.2267	0.1301	0.0898
0.4	0.1138	0.0528	0.0303	0.1919	0.1012	0.0650
0.5	0.0887	0.0356	0.0180	0.1611	0.0773	0.0457

podemos observar a Tabela 1.

Notemos que como função de $\Delta = \xi/\sigma$,

$$vício(\xi) = B\left(\widehat{\xi}, \xi/\sigma\right) = \frac{\sigma}{\sqrt{n}} \left[\varphi(\Delta\sqrt{n}) - \Delta\sqrt{n}\Phi(-\Delta\sqrt{n}) \right]$$

é uma função positiva e monótona decrescente a zero para $\Delta \rightarrow \infty$.

Stroud (1974) propõe o estimador $\widehat{\xi}_3$ e sugere que se leve em consideração o número de observações negativas na amostra para escolher entre $\widehat{\xi}$ e $\widehat{\xi}_3$, mas não obteve nenhuma conclusão satisfatória.

Gupta e Rohatgi (1980) comparam os estimadores $\widehat{\xi}$ e $\widehat{\xi}_3$, $\widehat{\xi}_1$ e $\widehat{\xi}_2$, baseando-se no erro quadrático médio. Podemos observar das fórmulas para MSE que estas contêm um fator proporcional a σ^2 ; por este motivo, as comparações foram feitas em termos de MSE/σ^2 .

Gupta e Rohatgi (1980) sintetizam o desempenho destes estimadores através dos resultados expressos na Tabela A.2. As conclusões obtidas foram:

Tabela A.2: Desempenho dos estimadores da média restrita

Dominio	Estimador Preferido	Tamanho da Amostra
$\xi/\sigma = 0$	$\hat{\xi}$	3, 4,5, 10, 25
$0.25 \leq \xi/\sigma \leq 0.5$	$\hat{\xi}_1^a$	10, 25
$\xi/\sigma = 0.75$	$\hat{\xi}_3$	3, 4,5, 10, 25
$1.0 \leq \xi/\sigma \leq 2.0$	$\hat{\xi}_2$	3, 4,5, 10, 25

a Com poucas exceções

- i) O estimador $\hat{\xi}_2$ é melhor que seus concorrentes só para valores grandes de ξ/σ . Para valores menores de ξ/σ , a escolha deve ser feita entre os outros estimadores.
- ii) Para tamanhos de amostras pequenos e valores moderados em torno de 0.5 de ξ/σ , o estimador $\hat{\xi}_1$ é o preferido, enquanto que para valores de ξ/σ próximos de zero, $\hat{\xi}$ é melhor. Para valores próximos a 0.75, $\hat{\xi}_3$ é o escolhido.

Apêndice B

Elementos básicos de inferência bayesiana

B.1 Teorema de Bayes e Famílias Conjugadas

Consideremos $\mathbf{x} = (x_1, \dots, x_n)$ uma amostra aleatória de uma variável X com função densidade de probabilidade $f(\mathbf{x}|\boldsymbol{\theta})$, em que $\boldsymbol{\theta}$ é um vetor de parâmetros sobre o qual se deseja fazer inferência. O enfoque bayesiano deste problema envolve dois elementos principais: a distribuição das observações $f(\mathbf{x}|\boldsymbol{\theta})$ e a distribuição a priori do vetor $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$, $\pi(\boldsymbol{\theta})$. A função de verossimilhança é denotada por $l(\boldsymbol{\theta}) = f(\mathbf{x}|\boldsymbol{\theta})$.

As constantes que especificam a distribuição $\pi(\boldsymbol{\theta})$ são chamadas de hiperparâmetros.

A distribuição de probabilidades de $\boldsymbol{\theta}$ após a observação de valor do $\mathbf{x}$ é chamada de distribuição a posteriori e pode ser obtida através do Teorema de Bayes, ou seja

$$p(\boldsymbol{\theta}|\mathbf{x}) = \frac{f(\mathbf{x}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta})}{f(\mathbf{x})},$$

em que $f(\mathbf{x}) = \int f(\mathbf{x}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d(\boldsymbol{\theta})$ é uma constante. Logo podemos escrever o Teorema

de Bayes em forma reduzida como

$$p(\boldsymbol{\theta}) \propto l(\boldsymbol{\theta}) \pi(\boldsymbol{\theta}),$$

em que $p(\boldsymbol{\theta})$ denota $p(\boldsymbol{\theta}|\mathbf{x})$.

A densidade marginal a posteriori de θ_i é dada por

$$p(\theta_i|\mathbf{x}) = \int p(\theta_1, \dots, \theta_p|\mathbf{x}) d\boldsymbol{\theta}_{-i} \quad (\text{B.1})$$

em que $\boldsymbol{\theta}_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p)$ é o vetor $\boldsymbol{\theta}$ com sua i -ésima componente removida, $i = 1, \dots, p$.

Uma família de distribuições é conjugada se as distribuições a priori e a posteriori pertencem a uma mesma classe de distribuições. A utilização de análises com distribuições conjugadas simplifica os cálculos, permitindo que a atualização de $\boldsymbol{\theta}$ seja feita rotineiramente sem que a introdução de novas observações em seqüência cause maiores transtornos. No entanto, nem sempre uma distribuição a priori conjugada representa o problema adequadamente.

O modelo linear usual

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n) \quad (\text{B.2})$$

é completado sob o ponto de vista Bayesiano com distribuições de probabilidades a priori para os parâmetros $(\boldsymbol{\beta}, \sigma^2)$, por exemplo,

$$\boldsymbol{\beta} \sim N(\mathbf{b}_0, \mathbf{B}_0) \text{ e } \sigma^2 \sim GI\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right) \text{ independentes}$$

em que $\mathbf{b}_0, \mathbf{B}_0, n_0/2$ e $n_0 S_0/2$ são os hiperparâmetros correspondentes, e GI denota a função de distribuição gama inversa.

Reparametrizando o problema ou seja, utilizando o parâmetro $\tau = \sigma^{-2}$ ao invés de σ^2 temos que

$$\tau \sim G\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right),$$

em que G denota a distribuição gama, a densidade conjunta de β e τ é

$$\begin{aligned}\pi(\beta, \tau) &= \pi(\beta) \pi(\tau) \\ &\propto |\mathbf{B}_0|^{-1} \exp \left\{ -\frac{\mathbf{B}_0}{2} (\beta - \mathbf{b}_0)^t \mathbf{B}_0^{-1} (\beta - \mathbf{b}_0) \right\} \\ &\quad \tau^{(n_0/2)-1} \exp \left\{ -\frac{\tau}{2} n_0 S_0 \right\}.\end{aligned}\quad (\text{B.3})$$

A verossimilhança de (B.2) é dada por

$$\begin{aligned}l(\beta, \tau) &= f(\mathbf{y}|\beta, \tau) \\ &\propto \tau^{n/2} \exp \left\{ -\frac{\tau}{2} (\mathbf{y} - \mathbf{X}\beta)^t (\mathbf{y} - \mathbf{X}\beta) \right\}.\end{aligned}\quad (\text{B.4})$$

Por (B.3) e (B.4) a função de distribuição a posteriori é dada por

$$\begin{aligned}p(\beta, \tau|\mathbf{y}) &\propto l(\beta, \tau) \pi(\beta, \tau) \\ &\propto \tau^{(n+n_0)/2-1} |\mathbf{B}_0|^{-1} \exp \left\{ -\frac{1}{2} (\beta - \mathbf{b}_0)^t \mathbf{B}_0^{-1} (\beta - \mathbf{b}_0) \right\} \\ &\quad \exp \left\{ -\frac{\tau}{2} (\beta - \hat{\beta})^t \mathbf{X}^t \mathbf{X} (\beta - \hat{\beta}) + n_0 S_0 + S_e \right\},\end{aligned}\quad (\text{B.5})$$

em que $\hat{\beta} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y}$, e $S_e = (\mathbf{y} - \mathbf{X}\hat{\beta})^t (\mathbf{y} - \mathbf{X}\hat{\beta})$. Não existe conjugação porque (B.5) não permite reconhecer alguma forma tratável analiticamente.

No caso de um vetor paramétrico $\theta = (\theta_1, \dots, \theta_p)$, diz-se que uma componente θ_i , ($1 \leq i \leq p$) exibe conjugação condicional se a distribuição a priori condicional completa $p(\theta_i|\theta_{-i})$ e a distribuição a posteriori condicional completa $p(\theta_i|\theta_{-i}, \mathbf{y})$ pertencem à mesma família de distribuições.

As distribuições condicionais completas a posteriori, no modelo de regressão não são difíceis de serem obtidas apesar da inexistência de conjugação. Note que de (B.5) segue que

$$p(\beta|\tau, \mathbf{y}) \propto \exp \left(-\frac{1}{2} (\beta - \mathbf{b}_\tau)^t \mathbf{B}_\tau^{-1} (\beta - \mathbf{b}_\tau) \right),$$

com $\mathbf{b}_\tau = \mathbf{B}_\tau (\mathbf{B}_0^{-1} \mathbf{b}_0 + \tau \mathbf{X}^t \mathbf{y})$ e $\mathbf{B}_\tau^{-1} = \mathbf{B}_0^{-1} + \tau \mathbf{X}^t \mathbf{X}$. Logo a distribuição a posteriori $\beta|\tau \sim N(\mathbf{b}_\tau, \mathbf{B}_\tau)$, tem a mesma forma da distribuição a priori $\beta|\tau$, que pela independência entre β e τ é a distribuição a priori de β , ou seja, $N(b_0, B_0)$.

Analogamente, a distribuição condicional a posteriori de $\tau|\beta$ é dada por

$$p(\tau|\beta, \mathbf{y}) = \tau^{(n+n_0)/2-1} \exp \left\{ -\frac{\tau}{2} (\beta - \hat{\beta})^t \mathbf{X}^t \mathbf{X} (\beta - \hat{\beta}) + n_0 S_0 + S_e \right\}$$

e portanto, $\tau|\beta \sim G(n_1/2, S_\beta)$ com $n_1 = n + n_0$ e $S_\beta = (\beta - \hat{\beta})^t \mathbf{X}^t \mathbf{X} (\beta - \hat{\beta})$. Logo a distribuição a posteriori tem a mesma forma da distribuição a priori. Assim, ambas as componentes, β e τ , exibem, neste caso, conjugação condicional.

B.2 Cadeias de Markov e amostrador de Gibbs

Sejam a sequência de variáveis aleatórias $\{X_t\}$, $t \in T$, onde T é o espaço de índices ou parâmetros e a coleção $\{X(t, w) : t \in T, w \in \Omega\}$ tais que

$$X(t, \cdot) : \Omega \rightarrow S \subset \mathbb{R} \text{ é uma variável aleatória}$$

e

$$X(\cdot, w) : T \rightarrow S \subset \mathbb{R} \text{ é uma função de } t,$$

com S denotando o espaço de estados. Dizemos então que $\{X_t\}$, $t \in T$, é um processo estocástico em S .

Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidades com uma família distinta não decrescente $(\mathcal{F}_n)$ de σ -álgebras, $\mathcal{F}_n = \sigma(X_s, s \leq n)$ e $(S, \mathcal{S}, P_X)$ o espaço de probabilidade induzido pela variável aleatória X .

Um processo $\{X_n\}$, $n \in \mathbb{N}$, é chamado processo de Markov se

$$P(X_n \in B | \mathcal{F}_m) = P(X_n \in B | X_m) \quad (\text{B.6})$$

para todo $n > m \geq 0$ e todo $B \in \mathcal{S}$.

A propriedade de Markov (B.6) pode também ser estabelecida equivalentemente como

$$P(X_{n+1} = y | X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x) = P(X_{n+1} = y | X_n = x)$$

para todo $x_0, \dots, x_{n-1}, x, y \in S$. Essencialmente, processos de Markov são tais que, dado o presente, o resto do passado é irrelevante para se prever a sua posição num instante futuro.

Definimos uma cadeia de Markov como um processo de Markov onde $S = \mathbb{R}$ e $T = \mathbb{N}$.

Uma função $P : S \times S \rightarrow \mathbb{R}$ é dita função de probabilidade de transição se:

1. para cada $x \in S$, $P(x, \cdot)$ é uma medida de probabilidade em S ,
2. para cada $A \in \mathcal{S}$, a função $x \rightarrow P(x, A)$ é mensurável.

Para uma cadeia de Markov $\{X_n\}$, $n \in \mathbb{N}$, as probabilidades condicionais $P(X_{n+1} = y | X_n = x)$ são denominadas probabilidades de transição da cadeia e quando independem de n dizemos que são estacionárias; nesse caso, a cadeia é chamada homogênea.

A função $P(x, y) = P(X_1 = y | X_0 = x)$ é chamada função de transição da cadeia e satisfaz:

1. $P(x, y) \geq 0$, $\forall x, y \in S$;
2. $\sum_{y \in S} P(x, y) = 1$, $\forall x \in S$.

O amostrador de Gibbs (Geman e Geman (1984)) é basicamente um processo iterativo de amostragem de uma cadeia de Markov, onde as probabilidades de transição são dadas pelas densidades condicionais completas a posteriori

$$p_i(v_i | \mathbf{y}) = p(v_i | v_{-i}, \mathbf{y})$$

em que $\mathbf{v}_i$ pode ser um escalar, um vetor ou uma matriz. Para descrever o algoritmo, suponha que a distribuição de interesse é $p(\mathbf{v}|\mathbf{y})$ com $\mathbf{v} = (\mathbf{v}_1, \dots, \mathbf{v}_p)$, e que a geração direta de amostras de $p(\mathbf{v}|\mathbf{y})$ é extremadamente complicada e/ou custosa mas que as gerações de amostras das densidades condicionais completas a posteriori $p_i(\mathbf{v}_i|\mathbf{y})$ podem ser obtidas mais facilmente. O algoritmo de Gibbs fornece uma forma alternativa de geração baseada em sucessivas gerações das distribuições condicionais completas e é dado por:

i) inicialize o contador das iterações da cadeia ($j = 1$) e escolha valores iniciais

$$\mathbf{v}^{(0)} = (\mathbf{v}_1^{(0)}, \dots, \mathbf{v}_p^{(0)});$$

ii) obtenha um novo valor $\mathbf{v}^{(j)} = (\mathbf{v}_1^{(j)}, \dots, \mathbf{v}_p^{(j)})$ a partir de $\mathbf{v}^{(j-1)}$ através de sucessivas gerações de valores

$$\begin{aligned} \mathbf{v}_1^{(j)} &\sim p(\mathbf{v}_1|\mathbf{v}_2^{(j-1)}, \dots, \mathbf{v}_p^{(j-1)}, \mathbf{y}) \\ \mathbf{v}_2^{(j)} &\sim p(\mathbf{v}_2|\mathbf{v}_1^{(j)}, \mathbf{v}_3^{(j-1)}, \dots, \mathbf{v}_p^{(j-1)}, \mathbf{y}) \\ &\vdots \\ \mathbf{v}_p^{(j)} &\sim p(\mathbf{v}_p|\mathbf{v}_1^{(j)}, \dots, \mathbf{v}_{p-1}^{(j)}, \mathbf{y}); \end{aligned}$$

iii) mude o contador j para $j + 1$ e retorne a (ii) até a convergência;

iv) verifique se o critério de convergência está satisfeito; em caso afirmativo interrompa o processo e em caso negativo retorne para ii).

Após o equilíbrio, todas as gerações de uma mesma cadeia são gerações da distribuição de equilíbrio e sucessivos valores dessa cadeia formam uma amostra de $p(\mathbf{v}|\mathbf{y})$.

Geweke (1992) propôs um diagnóstico de convergência baseado em métodos usuais de séries temporais apropriado para ser utilizado quando a convergência da média de alguma função das variáveis amostradas é de interesse. Consideremos uma função real

$t(\mathbf{v})$ e sua trajetória $t^{(1)}, t^{(2)}, \dots$ construída a partir de $t^{(j)} = t(\mathbf{v}^{(j)})$, $j = 1, 2, \dots, N$. Essa trajetória define uma série temporal. Técnicas usuais de séries temporais fornecem uma série de estimadores para $E(t(\mathbf{v}))$ cuja variância assintótica é dada por $S_t(0)/N$ em que $S_t(w)$ é a densidade espectral da série t , que em geral é desconhecida e é estimada através de análise espectral.

Após a observação da série t por um número suficientemente longo de iterações N , se obtém a média t_a das n_a primeiras iterações e a média t_b das n_b últimas iterações. Se a cadeia é estacionária, as médias dos valores das primeiras e das últimas iterações da sequência devem ser similares. Pode-se mostrar que se n_a/N e n_b/N são fixos e $N \rightarrow \infty$ então

$$\frac{t_a - t_b}{\left(\hat{S}_t^a(0)/n_a + \hat{S}_t^b(0)/n_b \right)} \rightarrow N(0, 1) \quad (\text{B.7})$$

em que $\hat{S}_t^a(0)$ e $\hat{S}_t^b(0)$ são estimadores independentes das variâncias assintóticas de t_a e t_b respectivamente. Então, se a diferença padronizada das médias for grande, dizemos que não há indicação de convergência. O pacote CODA (Best, Cowles and Wines, 1995) utiliza os valores os sugeridos por Geweke, ou seja, $n_a/N = 0.1$ e $n_b/N = 0.5$ como padrão.

B.3 Testes de hipóteses: enfoque Bayesiano para a escolha de modelos.

Uma possível técnica para comparar modelos é baseada no chamado fator de Bayes. Sejam M , um parâmetro com valores inteiros que indexa os modelos, e $\pi_j = p(M = j)$, $j = 1, 2$, as probabilidades a priori associadas aos modelos tais que $\pi_1 + \pi_2 = 1$. A razão entre as densidades marginais observadas dos dois modelos,

$$B_{21} = \frac{p(\mathbf{y}|M=2)}{p(\mathbf{y}|M=1)} = \frac{p(M=2|\mathbf{y})\pi_1}{p(M=1|\mathbf{y})\pi_2}, \quad (\text{B.8})$$

Tabela B.1: Regra de decisão

B_{21}	Evidência contra H_0
1 a 3	vaga
3 a 20	positiva
20 a 150	forte
>150	muito forte

é conhecida como fator de Bayes e pode ser utilizada para decidir a favor ou contra cada modelo. Kass e Raftery (1995) sugerem a regra apresentada na Tabela (B.1) para decidir por um ou outro modelo

Com o objetivo de calcular o fator de Bayes, necessitamos obter as probabilidades a posteriori de cada modelo $p(M = j|\mathbf{y})$, $j = 1, 2$. Com este intuito, Carlin e Chib (1995) propõem uma técnica que utiliza métodos Monte Carlo para Cadeias de Markov (MCMC). M é especificada como uma variável aleatória componente de $\mathbf{v} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, M)$, com $\boldsymbol{\theta}_j$, $j = 1, 2$, denotando os parâmetros do modelo j . Como a geração direta de amostras de $p(\mathbf{v}|\mathbf{y})$ é complicada e as gerações de amostras das densidades condicionais completas a posteriori $p_i(\mathbf{v}_i|\mathbf{y})$ são mais fáceis de ser realizadas, a realização de uma amostragem de tipo Gibbs gera um estimador de $p(M = j|\mathbf{y})$ dado pela a frequência de $M = j$ na amostra. Para isso, é necessário fazer uma especificação completa de cada modelo e verificar se as condições de convergência requeridas pelo método MCMC são satisfeitas.

Assumimos que as probabilidades a priori $p(\boldsymbol{\theta}_j|M = j)$ são próprias e que há independência completa entre os $\boldsymbol{\theta}_j$ dado o indicador do modelo, M . Logo

$$p(\boldsymbol{\theta}|M = j) = \prod_{i=1}^2 p(\boldsymbol{\theta}_i|M = j). \quad (\text{B.9})$$

Completamos as especificações do modelo com a escolha das chamadas distribuições “pseudo-prioris ” do modelo $p(\boldsymbol{\theta}_j|M \neq j)$, que deverão ser próprias. Devido à independência condicional, elas não interferem na expressão das densidades marginais. A

distribuição conjunta de $\mathbf{y}$ e $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ quando $M = j$ é dada por

$$\begin{aligned} p(\mathbf{y}, \boldsymbol{\theta}, M = j) &= f(\mathbf{y}|\boldsymbol{\theta}, M = j) p(\boldsymbol{\theta}|M = j) \\ &= f(\mathbf{y}|\boldsymbol{\theta}_j, M = j) \left\{ \prod_{i=1}^2 p(\boldsymbol{\theta}_i|M = j) \right\} \pi_j, \end{aligned}$$

com $f(\mathbf{y}|\boldsymbol{\theta}_j, M = j)$ denotando a função de verossimilhança. Para implementar a amostragem de Gibbs, precisamos da distribuição condicional completa de cada $\boldsymbol{\theta}_j$ e M ; sua expressão é a seguinte

$$p(\boldsymbol{\theta}_j|\boldsymbol{\theta}_{i \neq j}, M, \mathbf{y}) \propto \begin{cases} f(\mathbf{y}|\boldsymbol{\theta}_j, M = j) p(\boldsymbol{\theta}_j|M = j), & M = j \\ p(\boldsymbol{\theta}_j|M \neq j) & M \neq j \end{cases}, \quad (\text{B.10})$$

e portanto será gerada do modelo usual se $M = j$, e das “pseudo-prioris” quando $M \neq j$.

A distribuição condicional completa para M é dada por

$$p(M = j|\boldsymbol{\theta}, \mathbf{y}) = \frac{f(\mathbf{y}|\boldsymbol{\theta}_j, M = j) \left\{ \prod_{i=1}^2 p(\boldsymbol{\theta}_i|M = j) \right\} \pi_j}{\sum_{k=1}^2 f(\mathbf{y}|\boldsymbol{\theta}_k, M = k) \left\{ \prod_{i=1}^2 p(\boldsymbol{\theta}_i|M = k) \right\} \pi_k}. \quad (\text{B.11})$$

A cadeia de Markov é aquela cujo núcleo de transição é formado pelas distribuições condicionais completas e sob condições usuais de regularidade, converge para a distribuição a posteriori $p(\boldsymbol{\theta}, M|\mathbf{y})$. Assim, amostras desta distribuição podem ser geradas através da amostragem das distribuições condicionais completas (B.10) e (B.11), utilizando o algoritmo de Gibbs. Em cada iteração, uma amostra de tamanho 1; $\boldsymbol{\theta}_1^{(l)}, \boldsymbol{\theta}_2^{(l)}, M^{(l)}$, $l = 1, \dots, N$ é obtida. A razão

$$\hat{p}(M = j|\mathbf{y}) = \frac{\text{número de } M^{(l)} = j}{\text{número total de } M^{(l)}} = \frac{\text{número de } M^{(l)} = j}{N}, \quad j = 1, 2,$$

fornece um estimador simples de $p(M = j|\mathbf{y})$ que pode ser utilizado no cálculo do fator de Bayes.

Para obter uma aproximação das “pseudo-prioris” Carlin e Chib (1995) recomendam

avaliar cada modelo separadamente, ou seja, considerando $\pi_j = 1$ e $\pi_{i \neq j} = 0$, e usar a distribuição a posteriori resultante como a distribuição “pseudo-priori” para o parâmetro θ_j no modelo $M = i$.

Esta metodologia pode ser estendida para um número maior de modelos, mas se este número é grande ela pode não ser a mais adequada porque requer a especificação das distribuições “pseudo-priori”, na sua implementação.

B.4 Significado da constante c definida na Subseção

2.3.2

Um teste de hipótese é uma função binária dos dados tal que se $\delta(d) = 1$, rejeitamos H_1 (hipótese nula), e se $\delta(d) = 0$, não rejeitamos H_1 .

As probabilidades dos erros de tipo I e II associadas ao teste δ são denotadas por $\alpha(\delta)$ (a probabilidade de rejeitar H_1 quando é verdadeira), e $\beta(\delta)$ (a probabilidade de aceitar H_1 quando é falsa).

Seja $l(M, \delta)$ é função de perda associada ao teste definida como $l(1, 0) = 0$, $l(1, 1) = a$, $l(2, 0) = b$, e $l(2, 1) = 0$.

Os valores de a e b determinam o grau de importância (custo) das probabilidades dos erros de tipo I e II.

A perda ou risco esperado associado ao teste é dada por

$$r(\delta) = E(l(M, \delta)) = \pi_1 E(l(M, \delta) | M = 1) + \pi_2 E(l(M, \delta) | M = 2).$$

Como

$$E(l(M, \delta) | M = 1) = aP(\delta = 1 | M = 1) = a\alpha(\delta),$$

e

$$E(l(M, \delta) | M = 2) = bP(\delta = 0 | M = 2) = b\beta(\delta),$$

temos que $r(\delta) = \pi_1 a \alpha(\delta) + \pi_2 b \beta(\delta)$. Um teste que minimiza o risco $r(\delta)$, é um teste de Bayes.

DeGroot (1986) prova que um teste δ definido sob d , tal que

$$\text{se } R_{12}(d) \leq \frac{b}{a} = c \text{ então } \delta(d) = 1 \text{ (rejeitamos } H_1),$$

e

$$\text{se } R_{12}(d) \geq \frac{b}{a} = c \text{ então } \delta(d) = 0 \text{ (não rejeitamos } H_1),$$

minimiza o risco $r(\delta)$.

A constante c é escolhida levando-se em consideração o grau de importância dos erros.

Se $\pi_1 = \pi_2 = 0.5$ e $\frac{b}{a} = c = 1$, ambos os erros têm o mesmo custo, enquanto que, se $\frac{b}{a} = c < 1$, o erro de tipo II tem um custo c vezes maior que o erro de tipo I. Os papéis se invertem quando $\frac{b}{a} = c > 1$.

Apêndice C

Cálculo da matriz inversa

Seja a matriz

$$\Sigma_i(\boldsymbol{\theta}) = \sigma^2 \mathbf{I} + \sigma_A^2 \mathbf{G}_{i1} + \sigma_B^2 \mathbf{G}_{i2} + (\sigma_{AB}/k_i) (\mathbf{G}_{i1} \mathbf{G}_{i2} + \mathbf{G}_{i2} \mathbf{G}_{i1}),$$

com $\mathbf{G}_{i1} = \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t$, $\mathbf{G}_{i2} = \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t$, $k_i = \mathbf{Z}_{i1}^t \mathbf{Z}_{i2} = \mathbf{Z}_{i2}^t \mathbf{Z}_{i1}$.

Denotando por $k_{1i} = \mathbf{Z}_{i1}^t \mathbf{Z}_{i1}$ e $k_{2i} = \mathbf{Z}_{i2}^t \mathbf{Z}_{i2}$ as matrizes $\mathbf{G}_{i1}$ e $\mathbf{G}_{i2}$ satisfazem as seguintes relações

$$\begin{aligned} \mathbf{G}_{i1} \mathbf{G}_{i1} &= \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \\ &= k_{1i} \mathbf{G}_{i1} \\ \mathbf{G}_{i2} \mathbf{G}_{i2} &= \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t \\ &= k_{2i} \mathbf{G}_{i2} \\ \mathbf{G}_{i1} \mathbf{G}_{i2} \mathbf{G}_{i1} &= \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \\ &= k_i^2 \mathbf{G}_{i1} \\ \mathbf{G}_{i2} \mathbf{G}_{i1} \mathbf{G}_{i2} &= \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t \\ &= k_i^2 \mathbf{G}_{i2} \\ \mathbf{G}_{i1} \mathbf{G}_{i2} \mathbf{G}_{i1} \mathbf{G}_{i2} &= k_i^2 \mathbf{G}_{i1} \mathbf{G}_{i2} \end{aligned}$$

$$\begin{aligned}
\mathbf{G}_{i2}\mathbf{G}_{i1}\mathbf{G}_{i2}\mathbf{G}_{i1} &= k_i^2\mathbf{G}_{i2}\mathbf{G}_{i1} \\
\mathbf{G}_{i1}\mathbf{G}_{i2}\mathbf{G}_{i2}\mathbf{G}_{i1} &= k_{2i}\mathbf{G}_{i1}\mathbf{G}_{i2}\mathbf{G}_{i1} \\
&= k_{2i}k_i^2\mathbf{G}_{i1} \\
\mathbf{G}_{i2}\mathbf{G}_{i1}\mathbf{G}_{i1}\mathbf{G}_{i2} &= k_{1i}\mathbf{G}_{i2}\mathbf{G}_{i1}\mathbf{G}_{i2} \\
&= k_{1i}k_i^2\mathbf{G}_{i2}.
\end{aligned}$$

Suponhamos que $\Sigma_i(\boldsymbol{\theta})$ tem como matriz inversa a

$$\Sigma_i^{-1}(\boldsymbol{\theta}) = \beta_0\mathbf{I} + \beta_1\mathbf{G}_{i1} + \beta_2\mathbf{G}_{i2} + \beta_3\mathbf{G}_{i1}\mathbf{G}_{i2} + \beta_4\mathbf{G}_{i2}\mathbf{G}_{i1},$$

com $\beta_j = \beta_{ji}$, para simplificar a notação, para obter esses coeficientes resolveremos o sistema $\Sigma_i^{-1}(\boldsymbol{\theta})\Sigma_i(\boldsymbol{\theta}) = \mathbf{I}$. Verificaremos que esse sistema tem as mesmas soluções que as equações determinadas por o sistema $\Sigma_i(\boldsymbol{\theta})\Sigma_i^{-1}(\boldsymbol{\theta}) = \mathbf{I}$, e assim podremos concluir que $\Sigma_i^{-1}(\boldsymbol{\theta})$ é a inversa de $\Sigma_i(\boldsymbol{\theta})$.

$$\begin{aligned}
&\Sigma_i^{-1}(\boldsymbol{\theta})\Sigma_i(\boldsymbol{\theta}) \\
&= \beta_0\sigma^2\mathbf{I} + \beta_0\sigma_A^2\mathbf{G}_{i1} + \beta_0\sigma_B^2\mathbf{G}_{i2} + \beta_0(\sigma_{AB}/k_i)\mathbf{G}_{i1}\mathbf{G}_{i2} + \beta_0(\sigma_{AB}/k_i)\mathbf{G}_{i2}\mathbf{G}_{i1} \\
&\quad + \sigma^2\beta_1\mathbf{G}_{i1} + \beta_1\sigma_A^2k_{1i}\mathbf{G}_{i1} + \beta_1\sigma_B^2\mathbf{G}_{i1}\mathbf{G}_{i2} + \beta_1(\sigma_{AB}/k_i)k_{1i}\mathbf{G}_{i1}\mathbf{G}_{i2} \\
&\quad + \beta_1(\sigma_{AB}/k_i)k_i^2\mathbf{G}_{i1} \\
&\quad + \beta_2\sigma^2\mathbf{G}_{i2} + \beta_2\sigma_A^2\mathbf{G}_{i2}\mathbf{G}_{i1} + \beta_2\sigma_B^2k_{2i}\mathbf{G}_{i2} + \beta_2(\sigma_{AB}/k_i)k_i^2\mathbf{G}_{i2} \\
&\quad + \beta_2(\sigma_{AB}/k_i)k_{2i}\mathbf{G}_{i2}\mathbf{G}_{i1} \\
&\quad + \sigma^2\beta_3\mathbf{G}_{i1}\mathbf{G}_{i2} + \sigma_A^2\beta_3k_i^2\mathbf{G}_{i1} + \sigma_B^2\beta_3k_{2i}\mathbf{G}_{i1}\mathbf{G}_{i2} + (\sigma_{AB}/k_i)\beta_3k_i^2\mathbf{G}_{i1}\mathbf{G}_{i2} \\
&\quad + (\sigma_{AB}/k_i)\beta_3k_{2i}k_i^2\mathbf{G}_{i1} \\
&\quad + \sigma^2\beta_4\mathbf{G}_{i2}\mathbf{G}_{i1} + \sigma_A^2\beta_4k_{1i}\mathbf{G}_{i2}\mathbf{G}_{i1} + \sigma_B^2\beta_4k_i^2\mathbf{G}_{i2} + (\sigma_{AB}/k_i)\beta_4k_{1i}k_i^2\mathbf{G}_{i2} \\
&\quad + (\sigma_{AB}/k_i)\beta_4k_i^2\mathbf{G}_{i2}\mathbf{G}_{i1} \\
&= \mathbf{I}
\end{aligned}$$

Assim temos o sistema de equações

$$\left\{ \begin{array}{l} \beta_0 \sigma^2 \mathbf{I} = \mathbf{I} \\ (\beta_0 \sigma_A^2 + \sigma^2 \beta_1 + \beta_1 \sigma_A^2 k_{1i} + \beta_1 (\sigma_{AB}/k_i) k_i^2 + \sigma_A^2 \beta_3 k_i^2 + (\sigma_{AB}/k_i) \beta_3 k_{2i} k_i^2) \mathbf{G}_{i1} = 0 \\ (\beta_0 \sigma_B^2 + \sigma^2 \beta_2 + \beta_2 \sigma_B^2 k_{2i} + \beta_2 (\sigma_{AB}/k_i) k_i^2 + \sigma_B^2 \beta_4 k_i^2 + (\sigma_{AB}/k_i) \beta_4 k_{1i} k_i^2) \mathbf{G}_{i2} = 0 \\ (\beta_0 (\sigma_{AB}/k_i) + \beta_1 \sigma_B^2 + \beta_1 (\sigma_{AB}/k_i) k_{1i} + \sigma^2 \beta_3 + \sigma_B^2 \beta_3 k_{2i} + (\sigma_{AB}/k_i) \beta_3 k_i^2) \mathbf{G}_{i1} \mathbf{G}_{i2} = 0 \\ (\beta_0 (\sigma_{AB}/k_i) + \beta_2 \sigma_A^2 + \beta_2 (\sigma_{AB}/k_i) k_{2i} + \sigma^2 \beta_4 + \sigma_A^2 \beta_4 k_{1i} + (\sigma_{AB}/k_i) \beta_4 k_i^2) \mathbf{G}_{i2} \mathbf{G}_{i1} = 0 \end{array} \right.$$

ou seja, temos $\beta_0 = \sigma^{-2}$ e os sistemas

$$\left\{ \begin{array}{l} \beta_0 \sigma_A^2 + \beta_1 (\sigma^2 + \sigma_A^2 k_{1i} + (\sigma_{AB}/k_i) k_i^2) + \beta_3 (\sigma_A^2 k_i^2 + (\sigma_{AB}/k_i) k_{2i} k_i^2) = 0 \\ \beta_0 (\sigma_{AB}/k_i) + \beta_1 (\sigma_B^2 + (\sigma_{AB}/k_i) k_{1i}) + \beta_3 (\sigma^2 + \sigma_B^2 k_{2i} + (\sigma_{AB}/k_i) k_i^2) = 0 \end{array} \right. \quad (\text{C.1})$$

$$\left\{ \begin{array}{l} \beta_0 \sigma_B^2 + \beta_2 (\sigma^2 + \sigma_B^2 k_{2i} + (\sigma_{AB}/k_i) k_i^2) + \beta_4 (\sigma_B^2 k_i^2 + (\sigma_{AB}/k_i) k_{1i} k_i^2) = 0 \\ \beta_0 (\sigma_{AB}/k_i) + \beta_2 (\sigma_A^2 + (\sigma_{AB}/k_i) k_{2i}) + \beta_4 (\sigma^2 + \sigma_A^2 k_{1i} + (\sigma_{AB}/k_i) k_i^2) = 0 \end{array} \right. \quad (\text{C.2})$$

denotando

$$\begin{aligned} m_1 &= m_{1i} = \sigma^2 + \sigma_A^2 k_{1i} + \sigma_{AB} k_i, & m_2 &= m_{2i} = \sigma^2 + \sigma_B^2 k_{2i} + \sigma_{AB} k_i, \\ s_1 &= s_{1i} = \sigma_A^2 k_i^2 + \sigma_{AB} k_{2i} k_i, & s_2 &= s_{2i} = \sigma_B^2 k_i^2 + \sigma_{AB} k_{1i} k_i, \\ c_1 &= c_{1i} = \sigma_B^2 + (\sigma_{AB}/k_i) k_{1i}, & c_2 &= c_{2i} = \sigma_A^2 + (\sigma_{AB}/k_i) k_{2i}, \\ d_1 &= d_{1i} = \sigma^2 + \sigma_B^2 k_{2i} + \sigma_{AB} k_i, & d_2 &= d_{2i} = \sigma^2 + \sigma_A^2 k_{1i} + \sigma_{AB} k_i, \end{aligned}$$

temos que (C.1) e (C.2) são dados por

$$\left\{ \begin{array}{l} \beta_0 \sigma_A^2 + \beta_1 m_1 + \beta_3 s_1 = 0 \\ \beta_0 (\sigma_{AB}/k_i) + \beta_1 c_1 + \beta_3 d_1 = 0 \end{array} \right. \quad (\text{C.3})$$

$$\left\{ \begin{array}{l} \beta_0 \sigma_B^2 + \beta_2 m_2 + \beta_4 s_2 = 0 \\ \beta_0 (\sigma_{AB}/k_i) + \beta_2 c_2 + \beta_4 d_2 = 0 \end{array} \right. \quad (\text{C.4})$$

Do sistema (C.3) temos

$$\begin{aligned}\beta_1 &= -\beta_0\sigma_A^2/m_1 - \beta_3s_1/m_1 \\ \beta_0(\sigma_{AB}/k_i) + \beta_1c_1 + \beta_3d_1 &= 0 \\ \beta_0(\sigma_{AB}/k_i) - \beta_0c_1\sigma_A^2/m_1 - \beta_3c_1s_1/m_1 + \beta_3d_1 &= 0,\end{aligned}$$

assim,

$$\begin{aligned}\beta_3 &= \beta_0(c_1\sigma_A^2/m_1 - (\sigma_{AB}/k_i))(d_1 - c_1s_1/m_1)^{-1} \\ &= \beta_0(c_1\sigma_A^2 - (\sigma_{AB}/k_i)m_1)(d_1m_1 - c_1s_1)^{-1} \\ \beta_1 &= -\frac{\beta_0}{m_1}(\sigma_A^2 + (c_1\sigma_A^2 - (\sigma_{AB}/k_i)m_1)(d_1m_1 - c_1s_1)^{-1}s_1).\end{aligned}$$

Com cálculos análogos obtemos

$$\begin{aligned}\beta_4 &= \beta_0(c_1\sigma_B^2 - (\sigma_{AB}/k_i)m_2)(d_2m_2 - c_2s_2)^{-1} \\ \beta_2 &= -\frac{\beta_0}{m_2}(\sigma_B^2 + (c_2\sigma_B^2 - (\sigma_{AB}/k_i)m_2)(d_2m_2 - c_2s_2)^{-1}s_2).\end{aligned}$$

Planteamos agora o sistema

$$\begin{aligned}\Sigma_i(\boldsymbol{\theta})\Sigma_i^{-1}(\boldsymbol{\theta}) &= \mathbf{I} \\ &= \sigma^2\beta_0\mathbf{I} + \sigma^2\beta_1\mathbf{G}_{i1} + \sigma^2\beta_2\mathbf{G}_{i2} + \sigma^2\beta_3\mathbf{G}_{i1}\mathbf{G}_{i2} + \sigma^2\beta_4\mathbf{G}_{i2}\mathbf{G}_{i1} \\ &\quad + \beta_0\sigma_A^2\mathbf{G}_{i1} + \beta_1\sigma_A^2k_{1i}\mathbf{G}_{i1} + \beta_2\sigma_A^2\mathbf{G}_{i1}\mathbf{G}_{i2} + \beta_3\sigma_A^2k_{1i}\mathbf{G}_{i1}\mathbf{G}_{i2} + \beta_4\sigma_A^2k_i^2\mathbf{G}_{i1} \\ &\quad + \beta_0\sigma_B^2\mathbf{G}_{i2} + \beta_1\sigma_B^2\mathbf{G}_{i2}\mathbf{G}_{i1} + \beta_2\sigma_B^2k_{2i}\mathbf{G}_{i2} + \beta_3\sigma_B^2k_i^2\mathbf{G}_{i2} + \beta_4k_{2i}\sigma_B^2\mathbf{G}_{i2}\mathbf{G}_{i1} \\ &\quad + \beta_0(\sigma_{AB}/k_i)\mathbf{G}_{i1}\mathbf{G}_{i2} + \beta_1(\sigma_{AB}/k_i)k_i^2\mathbf{G}_{i1} + \beta_2(\sigma_{AB}/k_i)k_{2i}\mathbf{G}_{i1}\mathbf{G}_{i2} \\ &\quad + \beta_3(\sigma_{AB}/k_i)k_i^2\mathbf{G}_{i1}\mathbf{G}_{i2} + \beta_4(\sigma_{AB}/k_i)k_{2i}k_i^2\mathbf{G}_{i1} \\ &\quad + \beta_0(\sigma_{AB}/k_i)\mathbf{G}_{i2}\mathbf{G}_{i1} + \beta_1(\sigma_{AB}/k_i)k_{1i}\mathbf{G}_{i2}\mathbf{G}_{i1} + \beta_2(\sigma_{AB}/k_i)k_i^2\mathbf{G}_{i2} \\ &\quad + \beta_3(\sigma_{AB}/k_i)k_{1i}k_i^2\mathbf{G}_{i2} + \beta_4k_i^2(\sigma_{AB}/k_i)\mathbf{G}_{i2}\mathbf{G}_{i1}\end{aligned}$$

assim $\beta_0 = \sigma^{-2}$ e

$$\begin{cases} (\beta_0 \sigma_A^2 + \sigma^2 \beta_1 + \beta_1 \sigma_A^2 k_{1i} + \beta_1 (\sigma_{AB}/k_i) k_i^2 + \beta_4 \sigma_A^2 k_i^2 + \beta_4 (\sigma_{AB}/k_i) k_{2i} k_i^2) \mathbf{G}_{i1} = 0 \\ (\beta_0 \sigma_B^2 + \sigma^2 \beta_2 + \beta_2 \sigma_B^2 k_{2i} + \beta_2 (\sigma_{AB}/k_i) k_i^2 + \beta_3 \sigma_B^2 k_i^2 + \beta_3 (\sigma_{AB}/k_i) k_{1i} k_i^2) \mathbf{G}_{i2} = 0 \\ (\beta_0 (\sigma_{AB}/k_i) + \beta_2 \sigma_A^2 + \beta_2 (\sigma_{AB}/k_i) k_{2i} + \beta_3 \sigma_A^2 k_{1i} + \sigma^2 \beta_3 + \beta_3 (\sigma_{AB}/k_i) k_i^2) \mathbf{G}_{i1} \mathbf{G}_{i2} = 0 \\ (\beta_0 (\sigma_{AB}/k_i) + \beta_1 \sigma_B^2 + \beta_1 (\sigma_{AB}/k_i) k_{1i} + \sigma^2 \beta_4 + \beta_4 k_{2i} \sigma_B^2 + \beta_4 k_i^2 (\sigma_{AB}/k_i)) \mathbf{G}_{i2} \mathbf{G}_{i1} = 0, \end{cases}$$

logo temos os sistemas

$$\begin{cases} \beta_0 \sigma_A^2 + \beta_1 (\sigma^2 + \sigma_A^2 k_{1i} + (\sigma_{AB}/k_i) k_i^2) + \beta_4 (\sigma_A^2 k_i^2 + (\sigma_{AB}/k_i) k_{2i} k_i^2) = 0 \\ \beta_0 (\sigma_{AB}/k_i) + \beta_1 (\sigma_B^2 + (\sigma_{AB}/k_i) k_{1i}) + \beta_4 (\sigma^2 + \sigma_B^2 k_{2i} + (\sigma_{AB}/k_i) k_i^2) = 0 \end{cases} \quad (\text{C.5})$$

$$\begin{cases} \beta_0 \sigma_B^2 + \beta_2 (\sigma^2 + \sigma_B^2 k_{2i} + (\sigma_{AB}/k_i) k_i^2) + \beta_3 (\sigma_B^2 k_i^2 + (\sigma_{AB}/k_i) k_{1i} k_i^2) = 0 \\ \beta_0 (\sigma_{AB}/k_i) + \beta_2 (\sigma_A^2 + (\sigma_{AB}/k_i) k_{2i}) + \beta_3 (\sigma^2 + \sigma_A^2 k_{1i} + (\sigma_{AB}/k_i) k_i^2) = 0 \end{cases} \quad (\text{C.6})$$

Assim para que os sistemas (C.1), (C.2), (C.5) e (C.6) tenham as mesmas soluções devemos verificar que $\beta_3 = \beta_4$, ou seja

$$(c_1 \sigma_A^2 - (\sigma_{AB}/k_i) m_1) (d_2 m_2 - c_2 s_2) = (c_2 \sigma_B^2 - (\sigma_{AB}/k_i) m_2) (d_1 m_1 - c_1 s_1),$$

mas

$$\begin{aligned} c_1 \sigma_A^2 - (\sigma_{AB}/k_i) m_1 &= \sigma_A^2 \sigma_B^2 + \sigma_A^2 (\sigma_{AB}/k_i) k_{1i} \\ &\quad - (\sigma_{AB}/k_i) \sigma^2 - (\sigma_{AB}/k_i) \sigma_A^2 k_{1i} - (\sigma_{AB}/k_i) k_i^2 \\ &= \sigma_A^2 \sigma_B^2 - (\sigma_{AB}/k_i) \sigma^2 - (\sigma_{AB}/k_i) k_i^2 \\ &= c_2 \sigma_B^2 - (\sigma_{AB}/k_i) m_2. \end{aligned}$$

Por outro lado

$$\begin{aligned}s_1 &= k_i^2 c_2 \\s_2 &= k_i^2 c_1 \\c_2 s_2 &= k_i^2 c_1 c_2 = c_1 s_1,\end{aligned}$$

assim resta verificar $d_2 m_2 = d_1 m_1$ denotando por $\Delta = \sigma_A^2 k_{1i} - \sigma_B^2 k_{2i} = d_2 - m_2$ temos

$$\begin{aligned}d_1 &= d_2 - \Delta \\m_1 &= m_2 + \Delta \\d_1 m_1 &= (d_2 - \Delta)(m_2 + \Delta) \\&= d_2 m_2 - \Delta m_2 + \Delta d_2 - \Delta^2 \\&= d_2 m_2.\end{aligned}$$

Apêndice D

Cálculo dos elementos da matriz análoga à matriz de informação

Explicitaremos os elementos de $\mathbf{G}_{n\theta\theta}$

$$\begin{aligned} & (\mathbf{G}_{n\theta\theta})_{11} \\ &= \frac{1}{2} \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \left(\mathbf{I}_{n_i} - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \left(\mathbf{I}_{n_i} - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right\} \\ &= \frac{1}{2} \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \left(\mathbf{Z}_{i1} \mathbf{Z}_{i1}^t - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \left(\mathbf{Z}_{i1} \mathbf{Z}_{i1}^t - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \right\} \\ &= \frac{1}{2} \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right)^2 \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right\} \\ &= \frac{1}{2} \sum_{i=1}^n \sigma^{-4} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right)^2 k_{1i} \text{tr} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \\ &= \sum_{i=1}^n \frac{1}{2} \sigma^{-4} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right)^2 k_{1i}^2, \end{aligned} \tag{D.1}$$

$$\begin{aligned}
& (\mathbf{G}_{n\theta\theta})_{12} \\
&= \frac{1}{2} \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \left(\mathbf{Z}_{i2} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \right) \right\} \\
&= \frac{1}{2} \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \left(\mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \right) \right\} \\
&= \frac{1}{2} \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \left(k_i - \frac{\sigma_A^2 k_i k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \right\} \\
&= \frac{1}{2} \sum_{i=1}^n \sigma^{-4} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \left(k_i - \frac{\sigma_A^2 k_i k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \text{tr} \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \\
&= \sum_{i=1}^n \frac{1}{2} \sigma^{-4} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \left(k_i - \frac{\sigma_A^2 k_i k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_i \\
&= \sum_{i=1}^n \frac{1}{2} \sigma^{-4} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right)^2 k_i^2, \tag{D.2}
\end{aligned}$$

$$\begin{aligned}
& (\mathbf{G}_{n\theta\theta})_{13} \\
&= \frac{1}{2} \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \left(\mathbf{Z}_{i1} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \right) \right\} \\
&\quad + \text{tr} \left\{ \sigma^{-4} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \left(\mathbf{Z}_{i2} \mathbf{Z}_{i1}^t - \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \right\} \\
&= \frac{1}{2} \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \left(k_{1i} \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2 k_{1i}^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \right) \right\} \\
&\quad + \text{tr} \left\{ \sigma^{-4} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \left(k_i \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t - \frac{\sigma_A^2 k_i k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \right\} \\
&= \frac{\sigma^{-4}}{2} \sum_{i=1}^n \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \left\{ \left(k_{1i} - \frac{\sigma_A^2 k_{1i}^2}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_i + \left(k_i - \frac{\sigma_A^2 k_i k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_{1i} \right\} \\
&= \frac{\sigma^{-4}}{2} \sum_{i=1}^n \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \left\{ \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_{1i} k_i + \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_i k_{1i} \right\} \\
&= \sum_{i=1}^n \frac{\sigma^{-4}}{2} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right)^2 2k_{1i} k_i, \tag{D.3}
\end{aligned}$$

$$(\mathbf{G}_{n\theta\theta})_{21} = (\mathbf{G}_{n\theta\theta})_{12}$$

$$\begin{aligned}
& (\mathbf{G}_{n\theta\theta})_{22} \\
&= \frac{1}{2} \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \left(\mathbf{I}_{n_i} - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t \left(\mathbf{I}_{n_i} - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t \right\} \\
&= \frac{1}{2} \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \left(\mathbf{Z}_{i2} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t \right) \right. \\
&\quad \times \left. \left(\mathbf{Z}_{i2} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t \right) \right\} \\
&= \frac{1}{2} \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \left(\mathbf{Z}_{i2} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \left(\mathbf{Z}_{i2} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \right\} \\
&= \frac{\sigma^{-4}}{2} \sum_{i=1}^n \text{tr} \left\{ \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t \right. \\
&\quad \left. - \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t + \left(\frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \right)^2 \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right\} \\
&= \frac{\sigma^{-4}}{2} \sum_{i=1}^n k_{2i}^2 - \text{tr} \left\{ \frac{\sigma_A^2 k_i k_{2i}}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right\} \\
&\quad - \text{tr} \left\{ \frac{\sigma_A^2 k_i^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t \right\} + \text{tr} \left\{ \left(\frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \right)^2 k_i \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right\} \\
&= \sum_{i=1}^n \frac{\sigma^{-4}}{2} \left[k_{2i}^2 - 2 \frac{\sigma_A^2 k_i^2 k_{2i}}{\sigma^2 + \sigma_A^2 k_{1i}} + \left(\frac{\sigma_A^2 k_i^2}{\sigma^2 + \sigma_A^2 k_{1i}} \right)^2 \right] \\
&= \sum_{i=1}^n \frac{\sigma^{-4}}{2} \left[k_{2i} - \frac{\sigma_A^2 k_i^2}{\sigma^2 + \sigma_A^2 k_{1i}} \right]^2, \tag{D.4}
\end{aligned}$$

$$\begin{aligned}
& (\mathbf{G}_{n\theta\theta})_{23} \\
&= \frac{\sigma^{-4}}{2} \sum_{i=1}^n \text{tr} \left\{ \left(\mathbf{Z}_{i2} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t \right) \right. \\
&\quad \times \left(\mathbf{Z}_{i1} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \right) \Big\} \\
&\quad + \frac{\sigma^{-4}}{2} \sum_{i=1}^n \text{tr} \left\{ \left(\mathbf{Z}_{i2} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t \right) \right. \\
&\quad \times \left(\mathbf{Z}_{i2} \mathbf{Z}_{i1}^t - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t \right) \Big\} \\
&= \frac{\sigma^{-4}}{2} \sum_{i=1}^n \text{tr} \left\{ \left(\mathbf{Z}_{i2} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \right) \left(\mathbf{Z}_{i1} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \right) \right\} \\
&\quad + \frac{\sigma^{-4}}{2} \sum_{i=1}^n \text{tr} \left\{ \left(\mathbf{Z}_{i2} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \right) \left(\mathbf{Z}_{i2} \mathbf{Z}_{i1}^t - \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \right\} \\
&= \frac{\sigma^{-4}}{2} \sum_{i=1}^n \text{tr} \left\{ \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_i \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t \right\} \\
&\quad - \text{tr} \left\{ \left(\frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_i \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \right\} \\
&\quad + \frac{\sigma^{-4}}{2} \sum_{i=1}^n \text{tr} \left(\mathbf{Z}_{i2} \mathbf{Z}_{i2}^t \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t \right) - \text{tr} \left\{ \left(\frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t \right\} \\
&\quad - \frac{\sigma^{-4}}{2} \sum_{i=1}^n \text{tr} \left\{ \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right\} \\
&\quad + \frac{\sigma^{-4}}{2} \sum_{i=1}^n \text{tr} \left\{ \left(\frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \right)^2 \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right\} \\
&= \frac{\sigma^{-4}}{2} \sum_{i=1}^n \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_i k_{2i} - \left(\frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_i^2 \\
&\quad + \frac{\sigma^{-4}}{2} \sum_{i=1}^n k_{2i} k_i - \left(\frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_{1i} k_{2i} - \frac{\sigma_A^2 k_i^3}{\sigma^2 + \sigma_A^2 k_{1i}} + \left(\frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \right)^2 k_i k_{1i}
\end{aligned}$$

$$\begin{aligned}
& (\mathbf{G}_{n\theta\theta})_{23} \\
&= \frac{\sigma^{-4}}{2} \sum_{i=1}^n 2 \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_i k_{2i} - \left(\frac{\sigma_A^2 k_i^3}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \\
&\quad - \left(\frac{\sigma_A^2 k_i^3}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \left(1 + \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \\
&= \sum_{i=1}^n \frac{\sigma^{-4}}{2} \left\{ 2 \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_i k_{2i} - 2 \left(\frac{\sigma_A^2 k_i^3}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \right\}, \tag{D.5}
\end{aligned}$$

Calculamos primeiramente

$$\begin{aligned}
& \sigma^2 \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0) \mathbf{G}_{i3} \\
&= \left(\mathbf{I}_{n_i} - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) (\mathbf{Z}_{i1} \mathbf{Z}_{i2}^t + \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t) \\
&= \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t + \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t \\
&= \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t + \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \\
&= \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t + \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t - \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t,
\end{aligned}$$

assim

$$\begin{aligned}
& \sigma^4 \Sigma_i^{-1}(\theta_0) \mathbf{G}_{i3} \Sigma_i^{-1}(\theta_0) \mathbf{G}_{i3} \\
&= \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}}\right) \left(\left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}}\right) \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t + \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t \right) \\
& \quad \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}}\right) \left(-\frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \\
& \quad + \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}}\right) \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t + \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t - \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \\
& \quad - \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \left[\left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}}\right) \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t + \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t - \frac{\sigma_A^2 k_i k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right] \\
&= \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}}\right) \left(\left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}}\right) k_i \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t + k_{2i} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t - \frac{\sigma_A^2 k_i^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \\
& \quad + \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}}\right) k_{1i} \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t + k_i \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t - \frac{\sigma_A^2 k_i k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t \\
& \quad - \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \left[\left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}}\right) k_{1i} \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t + k_i \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t - \frac{\sigma_A^2 k_i k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right],
\end{aligned}$$

logo

$$\begin{aligned}
& \sigma^4 \Sigma_i^{-1}(\theta_0) \mathbf{G}_{i3} \Sigma_i^{-1}(\theta_0) \mathbf{G}_{i3} \\
&= \left\{ \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}}\right) \left(k_{2i} - \frac{\sigma_A^2 k_i^2}{\sigma^2 + \sigma_A^2 k_{1i}} \right) - \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \left(k_i - \frac{\sigma_A^2 k_i k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \right\} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \\
& \quad + \left\{ \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}}\right)^2 k_i - \frac{\sigma_A^2 k_i k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}}\right) \right\} \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \\
& \quad + \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}}\right) k_{1i} \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t \\
& \quad + \left(k_i - \frac{\sigma_A^2 k_i k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t \\
&= \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}}\right) \left(k_{2i} - 2 \frac{\sigma_A^2 k_i^2}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \\
& \quad \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}}\right) \left(1 - 2 \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}}\right) k_i \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \\
& \quad + \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}}\right) k_{1i} \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t + \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}}\right) k_i \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t,
\end{aligned}$$

$$\begin{aligned}
& (\mathbf{G}_{n\theta\theta})_{33} \\
&= \frac{\sigma^{-4}}{2} \sum_{i=1}^n \text{tr} \left(\Sigma_i^{-1} (\theta_0^*) \mathbf{G}_{i3} \Sigma_i^{-1} (\theta_0^*) \mathbf{G}_{i3} \right) \\
&= \frac{\sigma^{-4}}{2} \sum_{i=1}^n \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \left(k_{2i} - 2 \frac{\sigma_A^2 k_i^2}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_{1i} \\
&\quad + \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \left(1 - 2 \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_i^2 \\
&\quad + \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_{1i} k_{2i} + \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_i^2, \\
&= \frac{\sigma^{-4}}{2} \sum_{i=1}^n \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \left(k_{2i} k_{1i} - 2 \frac{\sigma_A^2 k_i^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} + k_i^2 \left(1 - 2 \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \right) \\
&\quad + \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) (k_{1i} k_{2i} + k_i^2) \\
&= \sum_{i=1}^n \frac{\sigma^{-4}}{2} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \left(2k_{2i} k_{1i} + 2k_i^2 \left(1 - 2 \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \right) \tag{D.6}
\end{aligned}$$

$$\begin{aligned}
& (\mathbf{G}_{n\theta\theta})_{41} \\
&= \frac{\sigma^{-4}}{2} \sum_{i=1}^n \text{tr} \left\{ \left(\mathbf{I}_{n_i} - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \left(\mathbf{Z}_{i1} \mathbf{Z}_{i1}^t - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \right\} \\
&= \frac{\sigma^{-4}}{2} \sum_{i=1}^n \text{tr} \left\{ \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \left(\mathbf{I}_{n_i} - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right\} \\
&= \frac{\sigma^{-4}}{2} \sum_{i=1}^n \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right)^2 \text{tr} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \\
&= \sum_{i=1}^n \frac{1}{2} \sigma^{-4} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right)^2 k_{1i}, \tag{D.7}
\end{aligned}$$

$$\begin{aligned}
& (\mathbf{G}_{n\theta\theta})_{42} \\
&= \frac{1}{2} \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \left(\mathbf{I}_{n_i} - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \mathbf{I}_{n_i} \left(\mathbf{I}_{n_i} - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t \right\} \\
&= \frac{1}{2} \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \left(\mathbf{I}_{n_i} - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \left(\mathbf{Z}_{i2} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \right) \right\} \\
&= \frac{1}{2} \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \left(\mathbf{Z}_{i2} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \right) \right\} \\
&\quad - \frac{\sigma^{-4}}{2} \sum_{i=1}^n \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \text{tr} \left(\mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i2} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \right) \\
&= \frac{\sigma^{-4}}{2} \sum_{i=1}^n k_{2i} - \frac{\sigma_A^2 k_i^2}{\sigma^2 + \sigma_A^2 k_{1i}} \\
&\quad - \frac{\sigma^{-4}}{2} \sum_{i=1}^n \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \left(k_i^2 - \frac{\sigma_A^2 k_i^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \\
&= \frac{\sigma^{-4}}{2} \sum_{i=1}^n k_{2i} - \frac{\sigma_A^2 k_i^2}{\sigma^2 + \sigma_A^2 k_{1i}} - \frac{\sigma_A^2 k_i^2}{\sigma^2 + \sigma_A^2 k_{1i}} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \\
&= \frac{\sigma^{-4}}{2} \sum_{i=1}^n k_{2i} - \frac{\sigma_A^2 k_i^2}{\sigma^2 + \sigma_A^2 k_{1i}} \left(2 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right), \tag{D.8}
\end{aligned}$$

$$\begin{aligned}
& (\mathbf{G}_{n\theta\theta})_{43} \\
&= \frac{1}{2} \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \left(\mathbf{I}_{n_i} - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right)^2 (\mathbf{Z}_{i1} \mathbf{Z}_{i2}^t + \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t) \right\} \\
&= \frac{1}{2} \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \left(\mathbf{I}_{n_i} - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \left(\mathbf{Z}_{i1} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \right) \right\} \\
&\quad + \frac{1}{2} \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \left(\mathbf{I}_{n_i} - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \left(\mathbf{Z}_{i2} \mathbf{Z}_{i1}^t - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t \right) \right\} \\
&= \frac{1}{2} \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \left(\mathbf{Z}_{i1} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \right) \right\} \\
&\quad - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \text{tr} \left\{ \sigma^{-4} \left(\mathbf{Z}_{i1} \mathbf{Z}_{i2}^t - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i1} \mathbf{Z}_{i2}^t \right) \right\} \\
&\quad + \sum_{i=1}^n \text{tr} \left\{ \sigma^{-4} \left(\mathbf{Z}_{i2} \mathbf{Z}_{i1}^t - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i2} \mathbf{Z}_{i1}^t \right) \right\} \\
&\quad + \frac{1}{2} \sum_{i=1}^n - \frac{\sigma_A^2 k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \text{tr} \left\{ \sigma^{-4} \left(\mathbf{Z}_{i1} \mathbf{Z}_{i1}^t - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \right\} \\
&= \frac{1}{2} \sum_{i=1}^n \sigma^{-4} \left(k_i - \frac{\sigma_A^2 k_i k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \left(k_i - \frac{\sigma_A^2 k_i k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \\
&\quad + \left(k_i - \frac{\sigma_A^2 k_i k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) - \frac{\sigma_A^2 k_i k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} + \left(\frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \right)^2 k_{1i}^2 \\
&= \frac{1}{2} \sum_{i=1}^n \sigma^{-4} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_i - \frac{\sigma_A^2 k_{1i} k_i}{\sigma^2 + \sigma_A^2 k_{1i}} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \\
&\quad + \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_i - \frac{\sigma_A^2 k_i k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} + \left(\frac{k_{1i} \sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \right)^2 \\
&= \frac{1}{2} \sum_{i=1}^n \sigma^{-4} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right)^2 k_i + \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_i \\
&\quad - k_i \left(\frac{k_{1i} \sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \left(1 - \frac{k_{1i} \sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \\
&= \frac{1}{2} \sum_{i=1}^n \sigma^{-4} \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right)^2 k_i + \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_i \left(1 - \left(\frac{k_{1i} \sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \right) \right) \\
&= \frac{1}{2} \sum_{i=1}^n \sigma^{-4} 2 \left(1 - \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \right)^2 k_i, \tag{D.9}
\end{aligned}$$

$$\begin{aligned}
& (\mathbf{G}_{n\theta\theta})_{44} \\
&= \frac{1}{2} \sum_{i=1}^n \text{tr } \sigma^{-4} \left(\mathbf{I}_{n_i} - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \mathbf{I}_{n_i} \left(\mathbf{I}_{n_i} - \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \mathbf{I}_{n_i} \\
&= \frac{1}{2} \sum_{i=1}^n \text{tr } \sigma^{-4} \left(\mathbf{I}_{n_i} - 2 \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t + \left(\frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \right)^2 k_{1i} \mathbf{Z}_{i1} \mathbf{Z}_{i1}^t \right) \\
&= \frac{1}{2} \sum_{i=1}^n \text{tr } \sigma^{-4} n_i - 2 \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} + \left(\frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \right)^2 k_{1i}^2 \\
&= \frac{1}{2} \sum_{i=1}^n \text{tr } \sigma^{-4} n_i + \frac{\sigma_A^2 k_{1i}}{\sigma^2 + \sigma_A^2 k_{1i}} \left(-2 + \left(\frac{\sigma_A^2}{\sigma^2 + \sigma_A^2 k_{1i}} \right) k_{1i} \right). \tag{D.10}
\end{aligned}$$

Bibliografia

- [1] Andreoni, S. (1989). *Modelos de Efeitos Aleatórios para Análise de Dados Longitudinais Não Balanceados em Relação ao Tempo*. Dissertação de Mestrado. IME. USP.
- [2] Aitkin, M. (1991). Posterior Bayes Factors. *Journal of the Royal Statistical Society B*, 53, 111-142.
- [3] Barlow, R. E., Bartholomew, D.J., Bremner, J. N. and Brunk, H. H. (1972). *Statistical Inference under Order Restrictions*. New York: John Wiley.
- [4] Bartholomew, D.J. (1959a). A Test of Homogeneity for Ordered Alternatives, I. *Biometrika*, 46, 36-48.
- [5] Bartholomew, D.J. (1959b). A Test of Homogeneity for Ordered Alternatives, II. *Biometrika*, 46, 328-335
- [6] Bartholomew, D.J. (1961). A Test of Homogeneity of Means Under Restricted Alternatives. *Journal of the Royal Statistical Society B*, 23, 239-281.
- [7] Bazaraa, M.S. and Shetty, C. M. (1979). *Nonlinear programming: theory and algorithms*. New York: John Wiley.
- [8] Berger, J. O. and Wolpert, R. L. (1984). *The Likelihood Principle*. IMS Monograph Vol. 6. Hayward, CA: Institute of Mathematical Statistics.

- [9] Berger, J. O. and Mortera, J. (1991). Interpreting the Stars in Precise Hypothesis Testing. *International Statistical Review*, 59, 337-353.
- [10] Berger, J. O. and Pericchi, L. (1996). The intrinsic Bayes factor for Model Selection and Prediction. *Journal of the American Statistical Association*, 91, 109-122.
- [11] Best, N. G., Cowles, M. K. and Wines, S. K. (1995). *CODA: Convergence Diagnostics and Output Analysis Software for Gibbs Sampler Output*: Versão 0.3, MRC Biostatistics Unit, Institute of Public Health Robinson Way, Cambridge, UK.
- [12] Carlin, B. P. and Chib, S. (1995). Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society B*, 57, 473-484.
- [13] Carlin, B. P. and Louis, T. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. New York: Chapman and Hall, .
- [14] Cowles, M. K. (1994). *Practical issues in Gibbs sampler implementation with application to bayesian hierarchical modeling of clinical trial data*. PhD thesis, Division of Biostatistics, University of Minnesota.
- [15] Cox, D. R. (1961). Tests of Separate Families of Hypotheses. *Proceedings of the Fourth Berkeley Symposium*, 1 , 105-123.
- [16] Cox, D. R. (1962). Further Results on Tests of Separate Families of Hypotheses. *Journal of the Royal Statistical Society B*, 24, 406-424.
- [17] Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman and Hall.
- [18] DeGroot, M. (1986). *Probability and Statistics*. 2. ed. . Menlo Park, Addison-Wesley.
- [19] De Santis, F. and Spezzaferri, F. (1997). Alternative Bayes factors for model selection. *The Canadian Journal of Statistics*, 4, 503-515.

- [20] DiCiccio, T., Kass, R. and Wasserman, L. (1997). Computing Bayes factors by combining simulation and asymptotic approximations. *Journal of the American Statistical Association*, 85, 398-409.
- [21] Edwards, W., Lindman, H. and Savadge, L. J. (1963). Bayesian Statistical Inference for Psychological Research. *Psychological Review*, 70, 193-242.
- [22] Chernoff, H. (1954). On the Distribution of the Likelihood Ratio. *The Annals of Mathematical Statistics*, 25, 573-578.
- [23] Judge, G. G. and Takayama, T. (1966). Inequality Restrictions in Regression Analysis. *Journal of the American Statistical Association*, 61, 166-181.
- [24] Gamerman, D. (1997). Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. London, Chapman & Hall.
- [25] Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association*, 70, 320-328.
- [26] Geisser, S. and Eddy, W. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 74, 153-160.
- [27] Gelfand, A. E., Dey, D. K. and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. In *Bayesian Statistics 4* (ed. J. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith) 147-168. Oxford University Press.
- [28] Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society B*, 56, 501-514.
- [29] Gelfand, A. E. and Ghosh, S. K. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika*, 85, 1-11.

- [30] Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments moments. In *Bayesian Statistics 4*, (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith). Oxford, Clarendon Press.
- [31] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741
- [32] Giri, N. C. (1977). *Multivariate Statistical Inference*. New York, Academic Press.
- [33] Gouriéroux, C., Holly, A. and Monfort, A. (1982). Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters. *Econometrica*, 50, 63-80.
- [34] Gouriéroux, G. and Monfort, A. (1995). *Statistics and Econometrics*, Vols 1 and 2. Cambridge: Cambridge University Press.
- [35] Graybill, F. A. (1976). *Theory and Application of the Linear Model*. North Scituate (Massachusetts): Duxbury Press.
- [36] Graybill, F. A. (1983). *Matrices with Applications in Statistics*. Belmont (California) Wadsworth.
- [37] Gupta, A.K. and Rohatgi, V. K. (1980). On the Estimation of Restricted Mean. *Journal of Statistical Planning and Inference*, 4, 369-379.
- [38] Heiny, R. L. and Siddiqui, M. M. (1970). Estimation of the Parameters of a Normal Distribution when the Mean is Restricted to an Interval. *Australian Journal of Statistics*, 12, 112-117.
- [39] Irony, T. Z. and Pereira, C. A. B. (1995). Bayesian Hypothesis Test: Using Surface Integrals to Distribute Prior Information Among the Hypotheses. *Resenhas IME-USP*, 27-46.

- [40] Iwata, S. (1994). Lower bounds on Bayes factor for a linear regression model. *Communications in Statistics- Theory and Methods*, 23, 6, 1825-1834.
- [41] Kass, R. and Raftery, A. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90, 777-795.
- [42] Kudô, A. (1963). A Multivariate Analogue of the One-Sided Test. *Biometrika*, 50, 403-418.
- [43] Liew, C.K. (1976). Inequality Constrained Least-Squares Estimation. *Journal of the American Statistical Association*, 71, 746-751.
- [44] Nüesch, P. E. (1966). On The Problem of Testing Location in Multivariate Populations for Restricted Alternatives. *The Annals of Mathematical Statistics*, 37, 113-119.
- [45] O'Hagan, A. (1995). Fractional Bayes factors for model comparison (with discussion). *Journal of the Royal Statistical Society B*, 57, 99-138.
- [46] Paula, G. A. (1997). *Estimação e Testes em Modelos de Regressão com Parâmetros Restritos*. São Paulo: Associação Brasileira de Estatística.
- [47] Pereira, C. A. B. (1985). *Teste de Hipóteses Definidas em Espaços de Diferentes Dimensões: Visão Bayesiana e Interpretação Clássica*. Tese de Livre Docência. IME-USP, 107p.
- [48] Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1994) *Numerical Recipes in C, The Art of Scientific Computing*. Second Edition. Cambridge: Cambridge University Press.
- [49] Raftery, A. E. and Akman (1986). Bayesian Analysis of a Poisson Process with a Change Point. *Biometrika*, 73, 85-89.
- [50] Raftery, A. E. (1993). Approximate Bayes Factor and Accounting for Model Uncertainty in Generalized Linear Models. Technical Report 255, University of Washington, Dept. of Statistics.

- [51] Raftery, A. E. (1995). Bayesian Model Selection in Social Research (with Discussion), *Sociological Methodology 1995*. edited by Peter V. Marsden. Crambridge, Mass: Blackwell Publishers, 111-195.
- [52] Rosenkranz, S. (1992). The Bayes Factor for Model Evaluation in Hierarchical Poisson Model for Area Counts. Ph.D. dissertation, University of Washington, Dept. of Biostatistics.
- [53] Searle, S. R. (1971). *Linear Models*. New York: John Wiley & Sons.
- [54] Self, S. G. and Liang, K. Y. (1987). Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Tests Under Nonstandard Conditions. *Journal of the American Statistical Association*, 82, 605-610.
- [55] Sen, P. K. and Singer, J. M. (1993). *Large Sample Methods in Statistics: An Introduction with Applications*. New York: Chapman and Hall.
- [56] Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review* 56, 49-62.
- [57] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461-464.
- [58] Silverman, B. W. (1986). *Density Esimation for Statistics and Data Analysis*. New York, Chapman and Hall.
- [59] Singer, J. M. and Andrade, D.F. (1997). Regression Models for Analysis of Pretest/Posttest Data. *Biometrics*, 53, 729-735.
- [60] Smith, A. F. M. (1991). Discussion of 'Posterior Bayes factors' by A. Aitken. *Journal of the Royal Statistical Society B*, 55, 3-24.
- [61] Stone, M. (1974). Cross-validatory choice and assesment of statistical predictions (with discussion). *Journal of the Royal Statistical Society B*, 36, 111-147.

- [62] Stone, M. (1997). An asymptotic equivalence of choice selection model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society B*, 39, 44-47.
- [63] Stram, D. O. and Lee, J. W. (1994). Variance Components Testing in Longitudinal Mixed Effects Model. *Biometrics*, 50, 1171-1177.
- [64] Stram, D. O. and Lee, J. W. (1995). Correction to Variance Components Testing in Longitudinal Mixed Effects Model. *Biometrics*, 51, 1196.
- [65] Tang, D. (1994) Uniformly More Powerful Tests in One-Sided Multivariate Problem. *Journal of the American Statistical Association*, 89, 1006-1011.
- [66] Thomas, A., Spielglharter, D. and Gilks, W. (1992). Bugs: a program to perform bayesian inference ousing Gibbs sampling. *Bayesian Statistics 4*, Claredon Press, Oxford, UK, 837-842.
- [67] Toutenburg, H. and Srivastava, V. K. (1996). Estimation of Regression Coefficients Subject to Interval Constraints. *Sankhyā A*, 58, 273-282
- [68] Vu, H. T. V. and Zhou, S. (1997). Generalization of Likelihood Ratio Tests under Nonstandard Conditions. *The Annals of Statistics*, 25, 897-916.
- [69] Wang, Y. and McDermott M. P. (1998) Conditional Likelihood Ratio Test for a Nonnegative Normal Mean Vector. *Journal of the American Statistical Association*, 93, 380-396.
- [70] Wolak, F. A. (1987) An Exact Test for Multiple Inequality and Equality Constraints in the Linear Regression Model. *Journal of the American Statistical Association*, 82, 782-793.
- [71] Wolak, F. A. (1989). Testing Inequality Constraints in Linear Econometric Models. *Journal of Econometrics*, 41, 205-235.