

UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG
INSTITUTO DE CIÊNCIAS EXATAS - ICE_x
DEPARTAMENTO DE ESTATÍSTICA

Tese de Doutorado:

**Inferência e interpretação de parâmetros em classes mais
flexíveis de modelos logísticos**

Cristiano de Carvalho Santos
Orientadora: Rosângela Helena Loschi

CRISTIANO DE CARVALHO SANTOS

Inferência e interpretação de parâmetros em classes mais flexíveis de modelos logísticos

Tese apresentada ao Programa de Pós-Graduação em Estatística do Departamento de Estatística da Universidade Federal de Minas Gerais como parte dos requisitos para a obtenção do grau de Doutor em Estatística.

Orientadora: Prof^a. Dr^a. Rosangela H. Loschi.

Data da defesa: 24/04/2015

Universidade Federal de Minas Gerais
Belo Horizonte
Maio de 2015

Dedico este trabalho
aos meus pais, Lucia e Geraldo.

Agradecimentos

Agradeço a Deus pelas oportunidades e conquistas que tem me proporcionado.

Aos meus pais, Lucia e Geraldo, pelo apoio e incentivo em minhas decisões e por sempre acreditarem no meu sucesso. Por terem me ensinado a buscar meus objetivos com dedicação, honestidade e humildade.

À minha família e amigos pelo apoio, incentivo e momentos de descontração.

Aos amigos da UFMG, principalmente colegas de doutorado, que estiveram presentes e me ajudaram a superar os momentos complicados nas disciplinas e no desenvolvimento da tese. Neste caso, cabe um agradecimento especial aos amigos Fábio Rocha, Fábio Demarqui, Isabel, José Luiz e Rivert pelas discussões científicas e pelo conhecimento compartilhado.

À minha orientadora, Professora Rosangela Helena Loschi pela confiança, paciência, dedicação, comprometimento e, principalmente, por não me deixar desanimar nos momentos mais difíceis, períodos em que a tese ficou estagnada devido a inúmeros problemas na implementação computacional dos métodos propostos.

Aos membros da banca examinadora, Prof. Flávio Bambirra Gonçalves (UFMG), Prof. Vinícius Diniz Mayrink (UFMG), Prof. Hélio dos Santos Migon (UFRJ) e Prof. Víctor Hugo Lachos Dávila (UNICAMP) pelas contribuições, correções e sugestões na tese.

Aos professores do departamento de Estatística pelos conhecimentos compartilhados durante mais de 10 anos de convívio.

Aos funcionários da UFMG, em especial àqueles da secretária de Pós-Graduação em Estatística, pelo bom trabalho executado facilitando a minha caminhada na pós-graduação.

À CAPES e à FAPEMIG pelo apoio financeiro com, respectivamente, a bolsa de doutorado e diversos auxílios financeiros para a participação em eventos.

Enfim, agradeço a todos que contribuíram para esta conquista.

Muito Obrigado!
Cristiano de Carvalho Santos

Resumo

Nesta tese apresentamos extensões do modelo logístico com efeitos mistos ao considerarmos distribuições Normais Assimétricas e Elípticas para modelarmos o comportamento dos efeitos aleatórios. Em ambos os casos, o modelo logístico com efeitos aleatórios normalmente distribuídos é um caso particular. Consideramos o modelo logístico com efeitos aleatórios para representar o efeito de *clusters* previamente conhecidos. Focamos na interpretação dos efeitos fixos e na medição da heterogeneidade entre os *clusters* utilizando a razão de chances. Como mostrado por Larsen *et al.* (2000), no modelo logístico com efeitos mistos a razão de chances é aleatória. Estendemos os resultados obtidos por Larsen *et al.* (2000), sob normalidade para os efeitos aleatórios, para os modelos propostos neste trabalho. Implementamos, através da abordagem bayesiana, o modelo logístico misto considerando as distribuições Normal Assimétrica Univariada e Multivariada para os efeitos aleatórios. Em um estudo com dados simulados, avaliamos o efeito da má-especificação da distribuição dos efeitos aleatórios na estimação dos parâmetros e da razão de chances, concluindo que o parâmetro de assimetria é mal estimado e o maior impacto da má-especificação ocorre nas estimativas dos efeitos aleatórios e, conseqüentemente, da razão de chances. Em um estudo Monte Carlo, investigamos o efeito de fixarmos o parâmetro de assimetria da distribuição Normal Assimétrica dos efeitos aleatórios e constatamos que a melhor estratégia é estimar tal parâmetro. Também aplicamos os modelos propostos a dois conjuntos de dados reais referentes a atividade teratogênica em ovos de galinha e a contaminação de suínos por *Ascaris Suum*. Construímos o modelo logístico misto em que os efeitos aleatórios têm distribuição na família Normal Independente. Neste caso, implementamos três tipos de algoritmos EM Monte Carlo para obtermos as estimativas de máxima verossimilhança. Discutimos aspectos referentes a geração de valores das variáveis latentes, número de amostras geradas e convergência nos algoritmos propostos. Realizamos um estudo com dados simulados para avaliar a implementação do algoritmo EM Monte Carlo e um estudo Monte Carlo para avaliarmos a qualidade da inferência dos modelos propostos, no qual concluímos que o principal ganho em especificar corretamente a distribuição dos efeitos aleatórios está na estimação da heterogeneidade entre os *clusters*. Reanalizamos os dados reais sobre a contaminação de suínos por *Ascaris Suum*.

Palavras-chave: Algoritmo EM Monte Carlo, *cluster*, distribuições Elípticas, distribuições Normais Independentes, distribuição *Skew-Normal*, métodos Monte Carlo via Cadeias de Markov, regressão Logística com efeitos mistos, razão de chances aleatória.

Abstract

In this dissertation we present some extensions of the logistic model with mixed effects by considering Skew-Normal and Elliptic Normal distributions to model the random effects behavior. In both cases, the normal mixed logistic model arises as a particular model. We consider the logistic model with random effects to represent the effect of previously known clusters. We mainly focus on the interpretation of fixed effects and the measure of heterogeneity among the clusters. For that proposal we consider the odds ratio, which is a random quantity, under the mixed logistic model. We extend the results obtained by Larsen *et al.* (2000) under normality for models proposed in this work. We implement, through of the Bayesian approach, the mixed logistic model considering both the univariate and multivariate Skew-Normal distributions for the random effects. Considering simulated datasets, we evaluated the effect of misspecifying the distribution of the random effects in the estimation of the parameters and the odds ratio, concluding that the skewness parameter is usually poorly estimated. Moreover, the biggest impact of such misspecification occurs in the estimation of the random effects and, consequently, of the odds ratio. A Monte Carlo study is performed to investigate the effect of fixing the skewness parameter in the Skew-Normal distribution of random effects. We concluded that the best strategy is to estimate this parameter. We also apply the proposed models in two case studies: teratogenic activity in chicken eggs and contamination of pigs by *Ascaris Suum*. We also introduced the mixed logistic model with random effects being distributed according to the Normal Independent family of distributions. We implemented three Monte Carlo EM algorithm in order to obtain the maximum likelihood estimates. We discuss aspects regarding the generation of the latent variable values, number of samples generated and convergence of the algorithms. We conducted a study considering simulated datasets to evaluate the efficiency of the Monte Carlo EM algorithms. A Monte Carlo simulation study is also performed to evaluate the quality of the estimates under the proposed models. We concluded that the main gain in correctly specify the distribution of the random effects is in the estimating the heterogeneity among the clusters. We reanalyzed the real data about the contamination of pigs by *Ascaris Suum*.

Keywords: Cluster, Elliptical distributions, Logistic regression with mixed effects, Markov chain Monte Carlo methods, Monte Carlo EM algorithm, Normal Independent distributions, random odds ratio, *Skew*-Normal distribution.

Sumário

	Página
Introdução	1
1 Modelo Logístico com Intercepto Aleatório	4
1.1 Interpretação dos parâmetros	5
2 Modelo Logístico Misto <i>Skew</i>-Normal	9
2.1 Introdução	9
2.2 Interpretação dos parâmetros	11
2.2.1 Resultados <i>a priori</i> para a razão de chances: Caso independente .	11
2.2.2 Resultados <i>a priori</i> para a razão de chances: Caso correlacionado	18
2.3 Inferência bayesiana no modelo logístico misto <i>Skew</i> -Normal	26
2.3.1 Razão de chances <i>a posteriori</i>	28
2.3.2 Métodos de seleção de modelos	29
2.4 Estudo de dados simulados: O efeito da má-especificação da distribuição dos efeitos aleatórios	30
2.5 Estudo Monte Carlo: O efeito da distribuição <i>a priori</i> de λ	38
2.5.1 Estimativas dos parâmetros	39
2.5.2 Estimativas dos efeitos aleatórios	45
2.5.3 Comparação de modelos	47
2.6 Estudo de caso 1: Atividade teratogênica	49
2.7 Estudo de caso 2: <i>Ascaris suum</i>	53
2.8 Conclusão	57
3 Modelo Logístico Misto Elíptico	59
3.1 Introdução	59
3.2 Distribuições Elípticas	60
3.3 Modelo proposto	66
3.4 Resultados para a razão de chances	67
3.4.1 Razão de chances no modelo logístico elíptico	70

3.5	Métodos de máxima verossimilhança no modelo logístico misto NI	74
3.5.1	Algoritmo EM Monte Carlo no modelo Logístico-NI	75
3.5.2	Matriz de informação de Fisher observada	86
3.6	Estudo em dados simulados: Implementação do algoritmo EMMC	88
3.6.1	Cenário 1: Comparação dos algoritmos para gerar de γ	89
3.6.2	Cenário 2: Modelo Logístico-T	95
3.6.2.1	Avaliando o efeito de n_{ij} em cenários com mesmo número de <i>clusters</i>	95
3.6.2.2	Avaliando o efeito de fixarmos ν	100
3.6.2.3	Avaliando o efeito da existência de β_0 no modelo	101
3.6.2.4	Avaliando os efeitos de k , n_i e do aumento de ν	103
3.6.3	Cenário 3: Modelo Logístico-SL	108
3.6.3.1	Avaliando o efeito do aumento de ν : Algoritmo EMMC-M	109
3.6.3.2	Avaliando o efeito do aumento de ν : Algoritmo EMMC-Q	113
3.6.4	Cenário 4: Modelo Logístico-NC	117
3.6.5	Discussão sobre a implementação do algoritmo EMMC	122
3.7	Estudo Monte Carlo: Avaliação da qualidade de estimação	123
3.7.1	Dados simulados do modelo Logístico-N	125
3.7.2	Dados simulados do modelo Logístico-T	130
3.7.2.1	Qualidade das estimativas dos parâmetros	131
3.7.2.2	A qualidade das estimativas da razão de chances mediana	143
3.7.2.3	Comparação de modelos	147
3.7.3	Dados simulados do modelo Logístico-SL	149
3.7.3.1	A qualidade das estimativas dos parâmetros	149
3.7.3.2	A qualidade das estimativas da razão de chances mediana	159
3.7.3.3	Comparação de modelos	162
3.7.4	Dados simulados do modelo Logístico-NC	163
3.7.4.1	A qualidade das estimativas dos parâmetros e da razão de chances mediana	163
3.7.4.2	Comparação de modelos	167
3.7.5	Discussão sobre o estudo Monte Carlo	168
3.8	Estudo de caso: <i>Ascaris suum</i>	169
3.9	Conclusões e trabalhos futuros	173
A Proposições Auxiliares		176
B Distribuições de Probabilidade		178

C	Distribuições condicionais completas no modelo logístico misto com efeitos aleatórios com distribuição <i>Skew</i> -Normal multivariada	183
D	Gráficos de viés e erro quadrático médio: Seção 2.5	185
E	Gráficos dos Estudos com dados simulados da Seção 3.6	189
F	Gráficos do Estudo Monte Carlo da Seção 3.7	204
G	Gráficos da análise de dados reais da Seção 3.8	212
	Referências Bibliográficas	218

Lista de Figuras

Figura	Página
1	Distribuições <i>a priori</i> de W_{12} e OR para o caso independente com $\xi = 0$, $\sigma^2 = 1$, $\kappa_{12} = 0$ e $\lambda = -8,5, -2, -1, -0,5$ e 0 17
2	Distribuições <i>a priori</i> de W_{12} e OR para o caso dependente, $\alpha_{i_2} = 0$ e -5 , $\sigma^2 = 1$, $\kappa_{12} = 0$ e $\alpha_{i_1} = 0, -5, -2, 2$ e 5 25
3	MOR_{12} versus α_{i_1} , para $\alpha_{i_2} = -5, \alpha_{i_2} = -2, \alpha_{i_2} = 0, \alpha_{i_2} = 2$ e $\alpha_{i_2} = 5$ 26
4	Histogramas e distribuição teórica, QQ -plot e Funções de distribuição acumulada teórica e empírica para os efeitos aleatórios gerados nos Dados 1, 2, 3 e 4. 31
5	Distribuições <i>a priori</i> para λ nos Modelos 2, 3 e 4. 32
6	Histogramas dos efeitos aleatórios gerados, f.d.p real dos efeitos aleatórios e f.d.p <i>plug-in</i> usando, <i>a posteriori</i> , média, mediana e moda, sob os Modelos 1 a 4, para Dados l , $l = 1, \dots, 4$ (linha l). 35
7	Distribuição <i>a priori</i> de λ , distribuições Normal e <i>Skew</i> -Normal. 39
8	Erro quadrático médio (EQM) com as médias <i>a posteriori</i> para os efeitos fixos e variância dos efeitos aleatórios, nos cenários com $k = 25$ e $k = 100$ e verdadeiros valores $\lambda = 0, \lambda = 5$ e $\lambda = 30$ 40
9	Vício com as médias <i>a posteriori</i> para os efeitos fixos e variância dos efeitos aleatórios, nos cenários com $k = 25$ e $k = 100$ e verdadeiros valores $\lambda = 0, \lambda = 5$ e $\lambda = 30$ 41
10	<i>Boxplots</i> das médias <i>a posteriori</i> para os cenários com $k = 25$ clusters, Cenário 1a, Cenário 2a e Cenário 3a. 43
11	<i>Boxplots</i> das médias <i>a posteriori</i> para os cenários com $k = 100$ clusters, Cenário 1b, Cenário 2b e Cenário 3b. 44
12	Porcentagem de vezes que cada modelo apresenta o menor e o segundo menor EQM_γ , cenários com $k = 25$ e $k = 100$ clusters, para $\lambda = 0, \lambda = 5$ e $\lambda = 30$ 46

13	<i>Boxplot</i> do EQM_γ para os cenários com $k = 25$ e $k = 100$ clusters, para $\lambda = 0$, $\lambda = 5$ e $\lambda = 30$	47
14	Distribuições <i>a priori</i> para σ^2 , $\sigma^2 \sim GI(2,001; 1)$ e $\sigma^2 \sim GI(0,01; 0,01)$	54
15	Densidade <i>a posteriori</i> de λ nos Modelos 3 e 4.	56
16	Curvas de nível da f.d.p das distribuições Normais bivariadas esférica e elíptica.	63
17	Densidade das distribuições <i>t</i> -Student e <i>Slash</i> . No gráfico à esquerda as distribuições Normal padrão, $T_1(0, 1, 1)$ e $T_1(0, 1, 4)$. No gráfico à direita as distribuições Normal padrão, $SL_1(0, 1, 1)$ e $SL_1(0, 1, 4)$	65
18	Densidade da distribuição Normal Contaminada. No gráfico à esquerda as distribuições Normal padrão, $NC_1(0; 1; 0,7; 0,3)$ e $NC_1(0; 1; 0,3; 0,3)$. No gráfico à direita as distribuições Normal padrão, $NC_1(0; 1; 0,5; 0,05)$ e $NC_1(0; 1; 0,5; 0,3)$	65
19	Gráfico da função $Q_L(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(m)})$ para $\bar{u}^{(m+1)} = 70$ visto por dois ângulos distintos.	84
20	Gráficos das estimativas dos parâmetros <i>versus</i> iterações, e média das estimativas nas últimas 100 iterações, obtidas com os algoritmos EMMC-M, EMMC-PA e EMMC-Q para Dados 1 no Cenário 1.	91
21	Gráficos das estimativas dos parâmetros <i>versus</i> iterações, e média das estimativas nas últimas 100 iterações, obtidas com os algoritmos EMMC-M, EMMC-PA e EMMC-Q para Dados 2 no Cenário 1.	92
22	Gráficos das estimativas dos parâmetros <i>versus</i> iterações, e média das estimativas nas últimas 100 iterações, obtidas com os algoritmos EMMC-M, EMMC-PA e EMMC-Q para Dados 5 no Cenário 1.	93
23	Gráficos das estimativas dos parâmetros <i>versus</i> iterações, e média das estimativas nas últimas 100 iterações, obtidas com os algoritmos EMMC-M, EMMC-PA e EMMC-Q para Dados 7 no Cenário 1.	94
24	Comparação entre as f.d.p das distribuições de probabilidade Normal padrão, $T(0; 0,5; 4)$ e $T(0; 0,75; 8)$ em (a) e um <i>zoom</i> da cauda a direita em (b).	95
25	Histograma e f.d.p teórica dos efeitos aleatórios gerados no Cenário 2 para Dados 1 a 12.	96
26	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-Q e média das estimativas nas últimas 100 iterações para Dados 1 a 4 no Cenário 2.	97

27	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-Q e média das estimativas nas últimas 100 iterações para Dados 5 a 8 no Cenário 2.	99
28	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-Q e média das estimativas nas últimas 100 iterações para Dados 9 a 12 no Cenário 2.	100
29	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-Q e média das estimativas nas últimas 100 iterações para Dados 1 a 4 no Cenário 2 ao considerarmos $\nu = 4$	101
30	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-Q e média das estimativas nas últimas 100 iterações para Dados 5 a 8 no Cenário 2 ao considerarmos $\beta_0 = 0$	102
31	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-Q e média das estimativas nas últimas 100 iterações para Dados 13 a 16 no Cenário 2.	104
32	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-Q e média das estimativas nas últimas 100 iterações para Dados 17 a 20 no Cenário 2.	105
33	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-Q e média das estimativas nas últimas 100 iterações para Dados 21 a 24 no Cenário 2.	106
34	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC e média das estimativas nas últimas 100 iterações para Dados 25 a 28 no Cenário 2.	107
35	Comparação entre as f.d.p das distribuições de probabilidade Normal padrão, $SL(0; 0,333; 1,5)$, $SL(0; 0,667; 3)$ e $SL(0; 0,875; 8)$ em (a) e uma visualização mais próxima da cauda a direita em (b).	108
36	Histograma e f.d.p teórica dos efeitos aleatórios gerados no Cenário 3 para Dados 1 a 12.	109
37	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-M e média das estimativas nas últimas 100 iterações para Dados 1 a 4 no Cenário 3.	110
38	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-M e média das estimativas nas últimas 100 iterações para Dados 5 a 8 no Cenário 3.	111
39	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-M e média das estimativas nas últimas 100 iterações para Dados 9 a 12 no Cenário 3.	112

40	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-Q com constante $c = 0,1, 0,3, 0,5$ e 1 e média das estimativas nas últimas 100 iterações com Dados 4 no Cenário 3.	114
41	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-Q com constante $c = 0,1, 0,3, 0,5$ e 1 e média das estimativas nas últimas 100 iterações com Dados 5 no Cenário 3.	115
42	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-Q com constante $c = 0,1, 0,3, 0,5$ e 1 e média das estimativas nas últimas 100 iterações com Dados 9 no Cenário 3.	116
43	Comparação entre as f.d.p das distribuições de probabilidade Normal padrão, $NC(0; 0,714; 0,10; 0,20)$, $NC(0; 0,555; 0,20; 0,20)$ e $NC(0; 0,385; 0,40; 0,20)$ em (a) e um <i>zoom</i> da cauda a direita em (b).	117
44	Histograma e f.d.p teórica dos efeitos aleatórios gerados no Cenário 4 para Dados 1 a 12.	118
45	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-M e média das estimativas nas últimas 100 iterações com Dados 1 a 4 no Cenário 4.	119
46	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-M e média das estimativas nas últimas 100 iterações com Dados 5 a 8 no Cenário 4.	120
47	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-M e média das estimativas nas últimas 100 iterações com Dados 9 a 12 no Cenário 4.	121
48	Histograma das estimativas de σ^2 nos Cenários A1 a A8.	128
49	<i>Boxplots</i> das estimativas fornecidas para os efeitos fixos e para a variância dos efeitos aleatórios, e valores reais, nos Cenários A1 a A8 e B1 a B8.	130
50	<i>Boxplots</i> das estimativas fornecidas para ν e valor real, sob os Modelos M2 e M4, nos Cenários C1 a C8.	133
51	Histogramas das estimativas fornecidas para ν , sob o Modelo M2, nos Cenários C1 a C8.	134
52	<i>Boxplots</i> das estimativas fornecidas para a variância dos efeitos aleatórios e valor real, sob todos os modelos, nos Cenários C1 a C8.	135
53	Histogramas das estimativas fornecidas para ν , sob o Modelo M2, nos Cenários D1 a D8.	138
54	<i>Boxplots</i> das estimativas fornecidas para a variância dos efeitos aleatórios e valor real, sob todos os modelos, nos Cenários D1 a D8.	139

55	Histogramas das estimativas fornecidas para ν , sob o Modelo M2, nos Cenários E1 a E8.	141
56	<i>Boxplots</i> das estimativas fornecidas para a variância dos efeitos aleatórios e valor real, sob todos os modelos, nos Cenários E1 a E8.	142
57	<i>Boxplots</i> das estimativas de $MOR_{ 12 }$ e valor real, sob todos os modelos, nos Cenários C1 a C8.	144
58	<i>Boxplots</i> das estimativas de $MOR_{ 12 }$ e valor real, sob todos os modelos, nos Cenários D1 a D8.	145
59	<i>Boxplots</i> das estimativas de $MOR_{ 12 }$ e valor real, sob todos os modelos, nos Cenários E1 a E8.	146
60	<i>Boxplots</i> das estimativas fornecidas para ν e valor real, sob os Modelos M2 e M4, nos Cenários F1 a F8.	151
61	Histogramas das estimativas fornecidas para ν , sob o Modelo M3, nos Cenários F1 a F8.	152
62	<i>Boxplots</i> das estimativas fornecidas para a variância dos efeitos aleatórios e valor real, sob todos os modelos, nos Cenários F1 a F8.	153
63	<i>Boxplots</i> das estimativas fornecidas para ν e valor real, sob os Modelos M2 e M4, nos Cenários G1 a G8.	156
64	Histogramas das estimativas fornecidas para ν , sob o Modelo M3, nos Cenários G1 a G8.	157
65	<i>Boxplots</i> das estimativas fornecidas para a variância dos efeitos aleatórios e valor real, sob todos os modelos, nos Cenários G1 a G8.	158
66	<i>Boxplots</i> das estimativas de $MOR_{ 12 }$ e valor real, sob todos os modelos, nos Cenários F1 a F8.	160
67	<i>Boxplots</i> das estimativas de $MOR_{ 12 }$ e valor real, sob todos os modelos, nos Cenários G1 a G8.	161
68	<i>Boxplots</i> das estimativas fornecidas para ν e valor real, sob os Modelos M2 e M3, nos Cenários H, I e J.	165
69	<i>Boxplots</i> das estimativas fornecidas para ν_1 e valor real, sob os Modelos M4, M5 e M6 nos Cenários H, I e J.	165
70	Histogramas das estimativas fornecidas para ν_1 , sob o Modelo M5, nos Cenários H, I e J.	166
71	<i>Boxplots</i> das estimativas da variância dos efeitos aleatórios e valor real, sob todos os modelos, nos Cenários H, I e J.	166

72	<i>Boxplots</i> das estimativas da $MOR_{ 12 }$ e valor real, sob todos os modelos, nos Cenários H, I e J.	167
73	Estimativas, sob todos os modelos ajustados, dos efeitos fixos β_0 e β_1 e da variância $V(\gamma_i)$ dos efeitos aleatórios, das medidas de razão de chances mediana MOR_{12} e $MOR_{ 12 }$ e do AIC, para dados de <i>Ascaris suum</i>	172
74	Densidade da distribuição <i>Skew-Normal</i> padrão com diferentes valores para o parâmetro λ . A esquerda a densidade de $SN(\lambda)$ com λ igual a -10, -2 e 0 e a direita com λ igual a 10, 2 e 0.	180
75	Densidade da distribuição <i>Skew-Normal</i> para diferentes valores de ξ e σ^2 . A esquerda a densidade de $SN(\xi, 1, 5)$ com ξ igual a -2, 0 e 2. A direita a densidade de $SN(0, \sigma^2, 5)$ com σ^2 igual a 1, 4 e 9.	180
76	Curvas de nível da f.d.p da distribuição <i>Skew-Normal</i> $SN_2(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$ com vetor de locação $\boldsymbol{\xi} = (0, 0)^t$, vetor de assimetria $\boldsymbol{\alpha} = (2, 2)^t$ e matriz de escala $\boldsymbol{\Sigma}$, em que $\boldsymbol{\Sigma} = \mathbf{I}_2$ ou $\boldsymbol{\Sigma}$ é uma matriz com diagonal principal com valor 1 e diagonal secundária com valor 0,3.	182
77	Erro quadrático médio (MSE) para as medianas <i>a posteriori</i> para os efeitos fixos e variância dos efeitos aleatórios, para os cenários com $k = 25$ e $k = 100$ e verdadeiros valores $\lambda = 0$, $\lambda = 5$ e $\lambda = 30$	185
78	Vício para as medianas <i>a posteriori</i> para os efeitos fixos e variância dos efeitos aleatórios, para os cenários com $k = 25$ e $k = 100$ e verdadeiros valores $\lambda = 0$, $\lambda = 5$ e $\lambda = 30$	186
79	Erro quadrático médio (MSE) para as modas <i>a posteriori</i> para os efeitos fixos e variância dos efeitos aleatórios, para os cenários com $k = 25$ e $k = 100$ e verdadeiros valores $\lambda = 0$, $\lambda = 5$ e $\lambda = 30$	187
80	Vício para as modas <i>a posteriori</i> para os efeitos fixos e variância dos efeitos aleatórios, para os cenários com $k = 25$ e $k = 100$ e verdadeiros valores $\lambda = 0$, $\lambda = 5$ e $\lambda = 30$	188
81	Histograma e f.d.p teórica dos efeitos aleatórios gerados no Cenário 1 para Dados 1 a 8.	189
82	Gráficos das estimativas dos parâmetros <i>versus</i> iterações, e média das estimativas nas últimas 100 iterações, obtidas com os algoritmos EMMC-M, EMMC-PA e EMMC-Q para Dados 3 e Cenário 1.	190
83	Gráficos das estimativas dos parâmetros <i>versus</i> iterações, e média das estimativas nas últimas 100 iterações, obtidas com os algoritmos EMMC-M, EMMC-PA e EMMC-Q para Dados 4 e Cenário 1.	191
84	Gráficos das estimativas dos parâmetros <i>versus</i> iterações, e média das estimativas nas últimas 100 iterações, obtidas com os algoritmos EMMC-M, EMMC-PA e EMMC-Q para Dados 6 e Cenário 1.	192

85	Gráficos das estimativas dos parâmetros <i>versus</i> iterações, e média das estimativas nas últimas 100 iterações, obtidas com os algoritmos EMMC-M, EMMC-PA e EMMC-Q para Dados 8 e Cenário 1.	193
86	Histograma e f.d.p teórica dos efeitos aleatórios gerados no Cenário 2 para Dados 13 a 28.	194
87	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-Q com constante $c = 0,1, 0,3, 0,5$ e 1 e média das estimativas nas últimas 100 iterações com Dados 1 no Cenário 3.	195
88	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMM-Q com constante $c = 0,1, 0,3, 0,5$ e 1 e média das estimativas nas últimas 100 iterações com Dados 2 no Cenário 3.	196
89	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-Q com constante $c = 0,1, 0,3, 0,5$ e 1 e média das estimativas nas últimas 100 iterações com Dados 3 no Cenário 3.	197
90	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-Q com constante $c = 0,1, 0,3, 0,5$ e 1 e média das estimativas nas últimas 100 iterações com Dados 6 no Cenário 3.	198
91	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-Q com constante $c = 0,1, 0,3, 0,5$ e 1 e média das estimativas nas últimas 100 iterações com Dados 7 no Cenário 3.	199
92	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-Q com constante $c = 0,1, 0,3, 0,5$ e 1 e média das estimativas nas últimas 100 iterações com Dados 8 no Cenário 3.	200
93	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-Q com constante $c = 0,1, 0,3, 0,5$ e 1 e média das estimativas nas últimas 100 iterações com Dados 10 no Cenário 3.	201
94	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-Q com constante $c = 0,1, 0,3, 0,5$ e 1 e média das estimativas nas últimas 100 iterações com Dados 11 no Cenário 3.	202
95	Gráficos das estimativas dos parâmetros <i>versus</i> iterações do algoritmo EMMC-Q com constante $c = 0,1, 0,3, 0,5$ e 1 e média das estimativas nas últimas 100 iterações com Dados 12 no Cenário 3.	203
96	<i>Boxplots</i> das estimativas de ν e valor real, sob os Modelos M2 e M4, nos Cenários D1 a D8.	204
97	<i>Boxplots</i> das estimativas de ν e valor real, sob os Modelos M2 e M4, nos Cenários E1 a E8.	205

98	Histogramas das estimativas fornecidas para ν , sob o Modelo M4, nos Cenários C1 a C8.	205
99	Histogramas das estimativas fornecidas para ν , sob o Modelo M4, nos Cenários D1 a D8.	206
100	Histogramas das estimativas fornecidas para ν , sob o Modelo M4, nos Cenários E1 a E8.	207
101	Histogramas das estimativas fornecidas para ν , sob o Modelo M2, nos Cenários F1 a F8.	208
102	Histogramas das estimativas fornecidas para ν , sob o Modelo M2, nos Cenários G1 a G8.	209
103	Histogramas das estimativas fornecidas para ν , sob o Modelo M2, nos Cenários H, I e J.	210
104	Histogramas das estimativas fornecidas para ν , sob o Modelo M3, nos Cenários H, I e J.	210
105	Histogramas das estimativas fornecidas para ν_1 , sob o Modelo M4, nos Cenários H, I e J.	211
106	Histogramas das estimativas fornecidas para ν_1 , sob o Modelo M6, nos Cenários H, I e J.	211
107	Gráficos das estimativas dos parâmetros sob o modelo Logístico-N <i>versus</i> iterações do algoritmo EMMC-Q e média das estimativas nas últimas 300 iterações com para dados de <i>Ascaris suum</i>	212
108	Gráficos das estimativas dos parâmetros sob o modelo Logístico-T <i>versus</i> iterações do algoritmo EMMC-Q e média das estimativas nas últimas 300 iterações com para dados de <i>Ascaris suum</i>	213
109	Gráficos das estimativas dos parâmetros sob o modelo Logístico-SL <i>versus</i> iterações do algoritmo EMMC-Q e média das estimativas nas últimas 300 iterações com para dados de <i>Ascaris suum</i>	213

Lista de Tabelas

Tabela	Página
2.1 Conjuntos de dados simulados	30
2.2 Comparação dos modelos	33
2.3 Estimativas <i>a posteriori</i> para β , $V(\gamma)$ e λ , todos os modelos e conjuntos de dados	34
2.4 <i>Clusters</i> selecionados para cálculo da <i>OR</i>	36
2.5 Estimativas <i>a posteriori</i> para a razão de chances	37
2.6 Covariáveis e número de observações em cada <i>cluster</i>	38
2.7 Porcentagem de vezes que cada modelo fornece o menor EQM_γ	45
2.8 Porcentagem de seleção de cada modelo	48
2.9 Dados Experimentais - Atividade teratogênica	49
2.10 Resumos <i>a posteriori</i> para os parâmetros no experimento sobre atividade teratogênica	51
2.11 Resumos <i>a posteriori</i> para algumas <i>OR</i> no experimento sobre atividade teratogênica	52
2.12 Dados dos Chiqueiros	53
2.13 Resumos <i>a posteriori</i> dos parâmetros no experimento dos chiqueiros	55
2.14 Casos abordados para Razão de Chances	56
2.15 Resumos <i>a posteriori</i> para algumas <i>OR</i> no experimento dos chiqueiros	57
3.1 Subclasses de distribuições esféricas <i>n</i> -variadas	62
3.2 Organização dos bancos de dados no Cenário 1	89
3.3 Média (desvio padrão) das estimativas fornecidas nas últimas 100 iterações dos algoritmos EMMC-M, EMMC-PA e EMMC-Q	90
3.4 Número L de réplicas Monte Carlo, ao longo das iterações, do algoritmo EMMC-M em cada modelo	124

3.5	Organização dos cenários gerados sob modelo Logístico-N	125
3.6	Viés, EQM e probabilidade de cobertura do IC (95%) para os efeitos fixos e variância dos efeitos aleatórios, nos Cenários A1 a A8, sob o modelo Logístico-N	127
3.7	Viés, EQM e probabilidade de cobertura do IC (95%) para os efeitos fixos e variância dos efeitos aleatórios, nos Cenários B1 a B8, sob o modelo Logístico-N	129
3.8	Organização dos cenários gerados sob modelo Logístico-T	131
3.9	Viés e EQM dos parâmetros de todos os modelos nos Cenários C1 a C8 .	132
3.10	Viés e EQM dos parâmetros de todos os modelos nos Cenários D1 a D8 .	137
3.11	Viés e EQM dos parâmetros de todos os modelos nos Cenários E1 a E8 .	140
3.12	Viés e EQM para $MOR_{ 12 }$ em todos os modelos nos cenários C1 a C8, D1 a D8 e E1 a E8	143
3.13	Proporção de escolha de cada modelo, usando o critério AIC, nos cenários gerados sob o modelo Logístico-T	148
3.14	Organização dos cenários gerados sob modelo Logístico-N	149
3.15	Viés e EQM dos parâmetros de todos os modelos nos Cenários F1 a F8 .	150
3.16	Viés e EQM dos parâmetros de todos os modelos nos Cenários G1 a G8 .	154
3.17	Viés e EQM para $MOR_{ 12 }$, em todos os modelos, nos Cenários F1 a F8 e G1 a G8	159
3.18	Proporção de escolha de cada modelo, usando o critério AIC, nos cenários gerados sob o modelo Logístico-SL	162
3.19	Viés e EQM dos parâmetros de todos os modelos nos Cenários H, I e J .	164
3.20	Proporção de escolha de cada modelo, usando o critério AIC, nos cenários gerados sob o modelo Logístico-NC	167
3.21	Estimativas dos parâmetros (erro padrão das estimativas), da variância dos efeitos aleatórios e da razão de chances sob todos os modelos	171

Introdução

O modelo de regressão logística é frequentemente utilizado para modelar dados que são binários ou binomiais. Nestes casos, em geral, o objetivo da modelagem estatística é explicar a probabilidade de ocorrência de um evento de interesse. A ocorrência de um evento de interesse é usualmente denominada por sucesso. O modelo de regressão logística é obtido ao descrevermos a relação entre a probabilidade de sucesso e as covariáveis observadas através da função de distribuição acumulada da distribuição logística. Esta relação pode ser modelada de outras formas como, por exemplo, através da função distribuição acumulada da distribuição Normal. Contudo, o modelo logístico tornou-se popular por apresentar uma forma simples de interpretar os efeitos fixos utilizando a razão de chances.

Os modelos logísticos assumem independência entre as observações e, por isso, não são apropriados para a análise de dados binários que apresentam estrutura de correlação. Dados correlacionados ocorrem, por exemplo, em estudos onde os indivíduos estão separados em *clusters* e também em estudos longitudinais. Em estudos com *clusters*, os indivíduos são observados dentro de grandes unidades, por exemplo, escolas, famílias, hospitais ou o local de trabalho. Em geral, as observações dentro de um *cluster* tendem a ser correlacionadas devido a várias características em comum. Exemplificando, se escolas são consideradas como *clusters*, observações de indivíduos dentro de uma mesma escola podem estar correlacionadas devido a características da prática de ensino que são particulares daquela escola. Já em estudos longitudinais, são feitas observações repetidas para cada indivíduo e, conseqüentemente, as observações para um mesmo indivíduo são correlacionadas. Estes modelos são conhecidos como modelos multiníveis ou hierárquicos.

Os modelos com efeitos mistos, frequentemente denominados por modelos mistos, surgiram como uma estratégia útil para modelarmos dados correlacionados. Nestes modelos, efeitos aleatórios são incluídos para introduzir a estrutura de correlação entre as observações. Com isso, ao condicionarmos no conhecimento dos efeitos aleatórios, as observações amostrais podem ser consideradas independentes. A adição dos termos de efeitos aleatórios no modelo também pode ser utilizada como alternativa para levarmos em consideração a ocorrência de superdispersão nos dados. Em geral, a superdispersão é causada pelo efeito de variáveis que não foram medidas, mas são importantes para explicar parte da variabilidade presente na variável resposta. Souza e Migon (2010) utilizaram um termo de efeito aleatório no intercepto do modelo de regressão logística para acomodar e detectar *outliers*. Em Gibbons *et al.* (1994), o efeito aleatório é usado para representar as características não observáveis e/ou não medidas que levam um médico a ter maior risco de sofrer um processo judicial por erro médico.

Em geral, os artigos científicos sobre modelos de regressão logística com efeitos mistos (ou modelo logístico misto, por simplicidade) assumem normalidade para os efeitos aleatórios. Por exemplo, Paulino *et al.* (2005) supõem normalidade para os efeitos

aleatórios ao abordar a análise de dados binários com erro de classificação sob o enfoque bayesiano. Todavia, a distribuição Normal pode ser muito restritiva por não ser capaz de representar comportamentos assimétricos ou com caudas pesadas. Liu e Dey (2008) comparam o ajuste do modelo logístico misto à dados reais ao considerar métodos não paramétricos e distribuições assimétricas para os efeitos aleatórios.

A interpretação dos parâmetros de efeitos fixos no modelo misto logístico não é tão simples quanto no modelo logístico usual pois a razão de chances (OR) é função tanto dos efeitos fixos quanto dos efeitos aleatórios (Larsen *et al.* (2000)). Sendo assim, a razão de chances é um objeto aleatório em ambas abordagens, Clássica e Bayesiana. Além disso, se os efeitos aleatórios forem estimados, a interpretação usual dos efeitos fixos através da razão de chances fica restrita apenas a comparação de indivíduos que possuem aqueles efeitos aleatórios considerados, isto é, a interpretação dos efeitos fixos é alterada ao mudarmos os efeitos aleatórios presentes na comparação. Supondo normalidade para os efeitos aleatórios, a razão de chances mediana foi introduzida por Larsen *et al.* (2000) como uma medida eficiente para interpretar os efeitos fixos e para medir a heterogeneidade causada devido aos efeitos aleatórios. Esta medida, em algumas comparações, tem a simples e conveniente interpretação em termos da razão de chances usual.

Neste trabalho consideramos distribuições mais flexíveis para representar o comportamento dos efeitos aleatórios. As distribuições consideradas são mais flexíveis no sentido de permitirem assimetria ou terem caudas mais pesadas do que a distribuição Normal. Para as distribuições assumidas estendemos os resultados propostos por Larsen *et al.* (2000) para interpretar os efeitos fixos e medir a heterogeneidade causada pelos efeitos aleatórios. No Apêndice B apresentamos a função de densidade de probabilidade e propriedades de algumas famílias de distribuições de probabilidade que são utilizadas ao longo deste trabalho, entre as quais estão as famílias de distribuições Normal Assimétrica (denotada ao longo do texto por distribuição *Skew-Normal*) univariada (Azzalini, 1985) e multivariada (Azzalini e Dalla-Valle, 1996) e a distribuição *Skew-Normal Unificada* introduzida por Arellano-Valle e Azzalini (2006). Os métodos estatísticos descritos são implementados no *software* gratuito OxEdit (Doornik, 2007) e os resultados obtidos são analisados utilizando o *software* gratuito R (R Development Core Team, 2012).

Este trabalho está organizado da seguinte forma. No Capítulo 1 apresentamos o modelo logístico com intercepto aleatório e os resultados propostos por Larsen *et al.* (2000) para interpretação dos parâmetros no modelo logístico misto e para medir a heterogeneidade resultando dos efeitos aleatórios.

No Capítulo 2 consideramos o modelo logístico com intercepto aleatório ao assumir a distribuição *Skew-Normal* para os efeitos aleatórios. A distribuição *Skew-Normal* é mais flexível do que a distribuição Normal no sentido de permitir assimetria para os efeitos aleatórios e, além disso, possui a distribuição Normal como um caso particular. Consideramos que os efeitos aleatórios são independentes e também correlacionados ao utilizarmos, respectivamente, a distribuição *Skew-Normal* univariada e multivariada. Ao estendermos os resultados propostos por Larsen *et al.* (2000) nestes modelos, obtivemos, como sub-produto, um resultado relacionado a distribuição de combinações lineares de variáveis aleatórias com distribuição *Skew-Normal*. Este resultado, até onde sabemos, é inédito na Literatura estatística. Os modelos propostos são implementados sob o enfoque bayesiano e aplicados tanto a dados simulados quanto a dados reais. O estudo com dados simulados nos permitirá avaliar, entre outras coisas, o efeito da má-especificação

do modelo nas estimativas dos parâmetros e na razão de chances. A extensão dos resultados propostos por Larsen *et al.* (2000) já havia sido tratada em Santos (2011) ao assumir a distribuição *Skew-Normal* para os efeitos aleatórios. No entanto, os resultados apresentados aqui são mais gerais do que os resultados em Santos (2011).

No Capítulo 3 assumimos que os efeitos aleatórios possuem uma distribuição da classe das distribuições elípticas. A classe de distribuições elípticas inclui a distribuição Normal e algumas com caudas mais pesadas do que a distribuição Normal, por exemplo, as distribuições *t-Student*, *Slash* e *Normal contaminada*. Neste capítulo consideramos um modelo logístico misto de forma mais geral do que no Capítulo 2, por assumir que os efeitos aleatórios também podem estar associados a covariáveis e não apenas ao intercepto. Apresentamos uma breve descrição da classe de distribuições elípticas e resultados gerais sobre a distribuição da razão de chances e sua mediana sem especificar a distribuição dos efeitos aleatórios. Obtemos resultados sobre a razão de chances para os casos particulares em que o comportamento dos efeitos aleatórios é modelado por alguma distribuição pertencente a classe das elípticas. Discutimos a implementação deste modelo sob o paradigma clássico utilizando o algoritmo EM Monte Carlo e avaliando a convergência do algoritmo em um estudo com dados simulados. Realizamos um estudo de simulação Monte Carlo para avaliar aspectos relacionados a qualidade de estimação dos parâmetros de interesse e má-especificação da distribuição dos efeitos aleatórios. Aplicamos os modelos propostos a um conjunto de dados reais considerados em Larsen *et al.* (2000), comparando-os com a análise realizada no Capítulo 2.

Capítulo 1

Modelo Logístico com Intercepto Aleatório

Neste capítulo, definimos o modelo logístico com intercepto aleatório e revisamos os resultados encontrados por Larsen *et al.* (2000) sobre a interpretação dos parâmetros deste modelo sob normalidade para os efeitos aleatórios.

Suponha que a população seja dividida em k clusters e que uma amostra de tamanho n_i seja selecionada dentro do i -ésimo cluster. Seja y_{ij} uma variável resposta binária que assume o valor 1 se ocorre um sucesso para o indivíduo j do cluster i , $i = 1, \dots, k$ e $j = 1, \dots, n_i$, e assume 0 caso contrário. Seja $\mathbf{x}_{ij} = (1, x_{ij1}, x_{ij2}, \dots, x_{ijp})^t$ o vetor de covariáveis $(p+1) \times 1$ associado ao indivíduo j dentro do i -ésimo cluster. Denote por $\mathbf{X} = (\mathbf{x}_{11}, \dots, \mathbf{x}_{1n_1}, \dots, \mathbf{x}_{k1}, \dots, \mathbf{x}_{kn_k})^t$ a matriz $N \times (p+1)$ com a informação relacionada as covariáveis para todos os $N = \sum_{i=1}^k n_i$ indivíduos observados.

Defina $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_k)^t \in \mathbb{R}^k$ como o vetor de efeitos aleatórios, em que γ_i denota o efeito aleatório para o i -ésimo cluster. Seja $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)^t \in \mathbb{R}^{p+1}$ o vetor de efeitos fixos. Denote por $\eta_{ij} = \mathbf{x}_{ij}^t \boldsymbol{\beta} + \gamma_i$ o preditor linear. Consequentemente, no modelo logístico misto com intercepto aleatório, $\pi_{ij} = P(y_{ij} = 1 | \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{X}) = \exp\{\eta_{ij}\} [1 + \exp\{\eta_{ij}\}]^{-1}$.

Para uma amostra de N indivíduos, assuma que dados $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ e \mathbf{X} , a variável resposta y_{ij} tem distribuição Bernoulli com probabilidade de sucesso π_{ij} . Como consequência, a função de verossimilhança é

$$f(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{X}) = \prod_{i=1}^k \prod_{j=1}^{n_i} \left[\frac{\exp\{\eta_{ij}\}}{1 + \exp\{\eta_{ij}\}} \right]^{y_{ij}} \left[\frac{1}{1 + \exp\{\eta_{ij}\}} \right]^{1-y_{ij}}, \quad (1.1)$$

em que $\mathbf{y} = (y_{11}, \dots, y_{1n_1}, \dots, y_{k1}, \dots, y_{kn_k})^t$.

Do ponto de vista clássico, fazer inferência em modelos mistos não é trivial na maioria dos casos. Para fazermos inferência sobre os efeitos fixos $\boldsymbol{\beta}$ que, em geral, são os parâmetros de interesse, assumimos para os efeitos aleatórios $\boldsymbol{\gamma}$ uma distribuição $f(\boldsymbol{\gamma} | \boldsymbol{\theta})$, indexada pelo vetor de hiperparâmetros $\boldsymbol{\theta}$, e integramos a função de verossimilhança conjunta para os dados e efeitos aleatórios de forma a obtermos uma função de verossimilhança livre dos efeitos aleatórios. A função de verossimilhança marginal é dada por

$$f(\mathbf{y} | \boldsymbol{\beta}, \mathbf{X}) = \int f(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{X}) f(\boldsymbol{\gamma} | \boldsymbol{\theta}) d\boldsymbol{\gamma}.$$

Após o procedimento de integração, adotamos os métodos usuais de estimação, como por exemplo, o método de máxima verossimilhança. Neste modelo, a inferência torna-se factível apenas em casos mais simples, em que a integral nos efeitos aleatórios pode ser analiticamente obtida ou em casos onde o gasto computacional para aproximá-la não é muito alto. Esta dificuldade no tratamento de modelos mistos nos conduz ao uso de distribuições nem sempre razoáveis para os efeitos aleatórios. Em geral, assumimos que os efeitos aleatórios γ são normalmente distribuídos. Esta suposição foi considerada, por exemplo, em Larsen *et al.* (2000), Larsen e Merlo (2005), Gibbons *et al.* (1994), entre outros. Supondo normalidade, este problema de integração é resolvido através do uso de técnicas de integração numérica, como por exemplo, a quadratura de Gauss-Hermite (McCulloch e Searle, 2001) ou a aproximação de Laplace. O caso não Normal requer um pouco mais de trabalho. De acordo com Nelson *et al.* (2006) e Liu e Yu (2008), uma transformação na verossimilhança conjunta é necessária para utilizar-se as aproximações numéricas usuais. Entre as possíveis alternativas para evitar o passo de integração estão os algoritmos EM (proposto por Dempster *et al.*, 1977), EM estocástico (Celeux e Diebolt, 1985), EM Monte Carlo (Wei e Tanner, 1990), ou alguma de suas variações, que são procedimentos muito utilizados para obtermos os estimadores de máxima verossimilhança, como pode ser visto em Kuhn e Lavielle (2005), em Chen *et al.* (2002) ou em Levine e Casella (2001).

No enfoque Bayesiano, como são desconhecidos, os efeitos aleatórios recebem o *status* de parâmetros a serem estimados e, com isso, utilizamos a função de verossimilhança em (1.1). Da mesma forma que na abordagem clássica, assumimos uma distribuição $f(\gamma|\theta)$ para γ e, para completar a modelagem Bayesiana, faz-se necessário especificar as distribuições *a priori* para os efeitos fixos β e para o vetor de parâmetros θ que indexam a distribuição dos efeitos aleatórios. Considerando a função de verossimilhança em (1.1) e as distribuições *a priori* eliciadas para os parâmetros de interesse, temos que a distribuição *a posteriori* para (β, γ, θ) é dada por

$$f(\beta, \gamma, \theta | \mathbf{y}, \mathbf{X}, \theta) = \frac{f(\mathbf{y}|\beta, \gamma, \mathbf{X})f(\beta)f(\gamma|\theta)f(\theta)}{\int f(\mathbf{y}|\beta, \gamma, \mathbf{X})f(\beta)f(\gamma|\theta)f(\theta)d\beta d\theta d\gamma}. \quad (1.2)$$

Em geral, a distribuição *a posteriori* em (1.2) não possui forma fechada, assim, os métodos Monte Carlo via Cadeias de Markov (MCMC) podem facilitar o processo de inferência. Esta facilidade computacional viabiliza o uso de distribuições mais apropriadas para descrever o comportamento dos efeitos aleatórios, possibilitando a utilização de distribuições que levam em conta assimetria e/ou caudas pesadas, entre outras características. Por exemplo, no modelo logístico misto com intercepto aleatório, Souza e Migon (2010) utilizam as distribuições *t*-Student e mistura de distribuições Normais independentes, para este fim, e Liu e Dey (2008), além de métodos não paramétricos, consideram as distribuições *Skew-Normal* e *Skew-t*.

1.1 Interpretação dos parâmetros

No modelo regressão logística ordinário, os efeitos fixos têm uma interpretação em termos da razão de chances entre indivíduos. Esta interpretação facilita a comunicação com pesquisadores de outra áreas e, na abordagem bayesiana, pode ajudar na construção de

distribuições *a priori* mais apropriadas para os efeitos fixos, uma vez que, é mais comum que o pesquisador tenha informação *a priori* sobre a razão de chances. A interpretação dos efeitos fixos no modelo logístico de regressão com efeitos aleatórios (MLREA) foi discutida pela primeira vez em Larsen *et al.* (2000). Larsen *et al.* (2000) mostram que a razão de chances depende tanto dos efeitos fixos quanto dos efeitos aleatórios e, conseqüentemente, várias interpretações úteis resultam desta quantidade. Reescrevendo os resultados introduzidos em Larsen *et al.* (2000), considere j_1 e j_2 dois indivíduos em diferentes *clusters* i_1 e i_2 , respectivamente. A razão de chances entre estes indivíduos é

$$OR_{i_1j_1, i_2j_2} = \exp \{ (\mathbf{x}_{i_1j_1}^t - \mathbf{x}_{i_2j_2}^t) \boldsymbol{\beta} + \gamma_{i_1} - \gamma_{i_2} \}. \quad (1.3)$$

Se a comparação ocorre entre indivíduos no mesmo *cluster* i , ou seja, tal que $i_1 = i_2 = i$ e que tenham covariáveis diferentes, a razão de chances torna-se $OR_{ij_1, ij_2} = \exp \{ (\mathbf{x}_{ij_1}^t - \mathbf{x}_{ij_2}^t) \boldsymbol{\beta} \}$. Neste caso, a razão de chances depende somente dos efeitos fixos e é exatamente a mesma que temos para o modelo de regressão logística ordinário. Para quantificar o impacto dos efeitos aleatórios, a comparação é feita assumindo dois indivíduos, j_1 e j_2 , com o mesmo vetor de covariáveis e que estão em diferentes *clusters* i_1 e i_2 . Neste caso, a razão de chances depende somente dos efeitos aleatórios e é dada por $OR_{i_1j_1, i_2j_2} = \exp \{ \gamma_{i_1} - \gamma_{i_2} \}$.

Quando a razão de chances entre dois indivíduos é uma quantidade dependente dos efeitos aleatórios surgem algumas dificuldades adicionais na interpretação dos parâmetros quando utilizamos a abordagem clássica. Usualmente, cada efeito fixo é interpretado avaliando o impacto na razão de chances quando comparamos indivíduos com valores distintos em apenas uma covariável. No entanto, quando a razão de chances depende dos efeitos aleatórios, para obtermos uma estimativa para a razão de chances é necessário obtermos estimativas também para os efeitos aleatórios. Estas estimativas para $\boldsymbol{\gamma}$, em geral, não são fornecidas diretamente pelos métodos clássicos de inferência, pois estes métodos consideram a função de verossimilhança marginal e nesta abordagem $\boldsymbol{\gamma}$ é uma variável aleatória. Além disto, note que mesmo se dispusermos de estimativas para $\boldsymbol{\gamma}$ e, conseqüentemente, para OR, a interpretação usual para os efeitos fixos através da razão de chances poderá ser utilizada apenas se fixarmos os *clusters* aos quais pertencem os indivíduos envolvidos na comparação. Isto é, quando usamos a razão de chances para interpretar os efeitos fixos, queremos obter uma medida que retrate a alteração na chance relativa de experimentar o evento de interesse devido a alteração do valor de uma covariável. No entanto, devido aos efeitos aleatórios, no modelo logístico misto não existirá uma interpretação única para os efeitos fixos através da razão de chances. Por exemplo, suponha que estejamos interessados em interpretar o efeito da idade na chance relativa de um paciente ter óbito após sofrer um infarto e, para isto, calculamos a razão de chances para indivíduos com todas as covariáveis iguais exceto a idade. Note que teremos uma estimativa para a razão de chances se os indivíduos pertencem ao mesmo *cluster* e outra estimativa diferente se os indivíduos são de *clusters* diferentes, a qual ainda dependerá de quais *clusters* estão envolvidos na comparação.

Na abordagem clássica, a OR é uma quantidade aleatória somente quando a comparação depender dos efeitos aleatórios. Como é aleatória, Larsen *et al.* (2000) propõe interpretar a OR através da mediana de sua distribuição. A então chamada razão de chances mediana é aqui nomeada por MOR. De acordo com Larsen *et al.* (2000), a MOR fornece uma interpretação sempre válida para os efeitos fixos, ou seja, ela não

se modifica ao trocarmos os *clusters* que estão envolvidos na comparação e, também, quantifica apropriadamente a heterogeneidade existente entre os diferentes *clusters*.

Para o caso geral, quando comparamos indivíduos com diferentes covariáveis em diferentes *clusters* a razão de chances mediana é definida como sendo

$$MOR_{i_1j_1, i_2j_2} = \text{med} \left\{ \exp \left\{ (\mathbf{x}_{i_1j_1}^t - \mathbf{x}_{i_2j_2}^t) \boldsymbol{\beta} + (\gamma_{i_1} - \gamma_{i_2}) \right\} \mid \boldsymbol{\beta}, \mathbf{X} \right\}. \quad (1.4)$$

É importante observar que a distribuição de $\gamma_{i_1} - \gamma_{i_2}$ é simétrica em torno de zero sob a suposição de efeitos aleatórios independentes e identicamente distribuídos (i.i.d) (ver Proposição A.1 no Apêndice A para situações mais gerais onde isto também ocorre). Nestes casos, independente dos *clusters* que são comparados, a *MOR*, definida em (1.4), assume sempre valor igual a 1 quando os indivíduos tem os mesmos valores observados para as covariáveis. Para quantificarmos apropriadamente o efeitos dos *clusters* nestes casos, Larsen *et al.* (2000) considera a mediana de $\exp \{ |\gamma_{i_1} - \gamma_{i_2}| \}$, ou seja, a razão de chances mediana é dada por

$$MOR_{|i_1j_1, i_2j_2|} = \text{med} \left\{ \exp \{ |\gamma_{i_1} - \gamma_{i_2}| \} \mid \boldsymbol{\beta}, \mathbf{X} \right\},$$

em que $|A|$ denota o valor absoluto de A . Para simplificarmos a notação, vamos denotar $OR_{12} = OR_{i_1j_1, i_2j_2}$, $MOR_{12} = MOR_{i_1j_1, i_2j_2}$ e $MOR_{|12|} = MOR_{|i_1j_1, i_2j_2|}$. Definimos também $\kappa_{12} = (\mathbf{x}_{i_1j_1}^t - \mathbf{x}_{i_2j_2}^t) \boldsymbol{\beta}$ e $W_{12} = \gamma_{i_1} - \gamma_{i_2}$.

Larsen *et al.* (2000) mostra que se $\boldsymbol{\gamma} \sim N_k(\mathbf{0}, \boldsymbol{\Sigma})$, então $W_{12} \sim N(0, \sigma_{12}^2)$, em que $\sigma_{12}^2 = (z_{i_1} - z_{i_2})^t \boldsymbol{\Sigma} (z_{i_1} - z_{i_2})$ denota a variância de W_{12} e z_i é um vetor de dimensão k com valor 1 na i -ésima posição e 0 nas demais posições. Consequentemente, a razão de chances em (1.3) pode ser reescrita como $OR_{12} = \exp \{ \kappa_{12} + W_{12} \}$. A partir daí, Larsen *et al.* (2000) obtém os resultados apresentados no Teorema 1.1 a seguir.

TEOREMA 1.1. *Se $\boldsymbol{\gamma}$ é um vetor aleatório com distribuição Normal com vetor de médias $\mathbf{0}$ e matriz de covariância $\boldsymbol{\Sigma}$ então, dados \mathbf{X} , $\boldsymbol{\beta}$ e $\boldsymbol{\Sigma}$, segue que*

a) *A quantidade aleatória OR_{12} tem distribuição Log-Normal com parâmetros de locação κ_{12} e escala σ_{12}^2 , cuja função de densidade de probabilidade (f.d.p) é dada por*

$$f_{OR_{12}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{X}}(r) = \frac{1}{\sqrt{2\pi r} \sigma_{12}} \exp \left\{ -\frac{1}{2\sigma_{12}^2} (\ln r - \kappa_{12})^2 \right\}, r \in \mathbb{R}_+, \quad (1.5)$$

b) *Para a mediana da distribuição em (1.5) segue que*

- (i) $MOR_{|12|} = \text{med} \{ \exp \{ |W_{12}| \} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{X} \} = \exp \{ \text{med} \{ |W_{12}| \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{X} \} \}$
 $= \exp \{ \sigma_{12} \Phi^{-1}(0,75) \},$
- (ii) $\text{med} \{ \exp \{ -|W_{12}| \} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{X} \} = \text{med} \{ \exp \{ |W_{12}| \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{X} \} \}^{-1},$
- (iii) $MOR_{12} = \text{med} \{ \exp \{ \kappa_{12} + W_{12} \} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{X} \} = \exp \{ \kappa_{12} \}.$

Note do item (i) que os termos $\Phi^{-1}(0,75) \approx 0,674$ e $\sigma_{12} \geq 0$, logo temos que a $MOR_{|12|} \geq 1$. Note que quanto maior o valor de σ_{12}^2 maior será $MOR_{|12|}$, indicando que maior variabilidade entre os efeitos aleatórios resulta em um aumento do valor de $MOR_{|12|}$. A $MOR_{|12|}$ nos mostra o quanto a chance relativa do evento de interesse de um

dos indivíduos é maior do que a chance relativa do outro indivíduo. O resultado do item (ii) do Teorema 1.1 é útil quando houver interesse em trocarmos a ordem dos indivíduos sob comparação, isto é, se houver o interesse de obtermos a razão de chances mediana entre a unidade de baixo risco e a unidade de alto risco, comparando os indivíduos exatamente nesta ordem. O item (iii) fornece uma medida para interpretarmos os efeitos fixos, a qual é independente dos *clusters* envolvidos na comparação, pois depende apenas das covariáveis e dos efeitos fixos β .

Do ponto de vista clássico, para exemplificarmos numericamente a interpretação através da *MOR*, suponha que em determinado estudo temos a estimativa $\widehat{MOR} = \exp\{\widehat{\sigma}_{12}\Phi^{-1}(0,75)\} = 3,06$. Isto significa que para indivíduos com valores iguais das covariáveis e que pertencem a *clusters* distintos, em termos medianos, o indivíduo de maior risco tem chance de experimentar o evento de interesse igual 3,06 vezes a chance do indivíduo com menor risco. Isto indicaria uma heterogeneidade substancial entre os *clusters* em estudo, pois este valor, relativamente distante de 1, é consequência apenas da diferença de efeito entre os *clusters*. Note que a interpretação feita desta forma é mais simples e intuitiva do que se avaliássemos a variância da distribuição dos efeitos aleatórios concluindo por grande heterogeneidade entre os *clusters* se esta variância fosse grande. Por outro lado, suponha que, em um outro estudo, estamos comparando um homem e uma mulher que pertencem a diferentes *clusters* e tais que as demais covariáveis do estudo tenham valores idênticos para ambos. Se a estimativa para *MOR* entre este homem e esta mulher for igual a 2,44, isto significa dizer que, em termos medianos, o homem tem chance de experimentar o evento de interesse igual a 2,44 vezes a chance da mulher, ou seja, com probabilidade de 50% esta razão de chances é maior do que 2,44.

Essa interpretação de *OR* foi considerada do ponto de vista clássico e também fornece uma boa interpretação *a priori* condicional para *OR*, dados a matriz \mathbf{X} , os efeitos fixos β e os hiperparâmetros da distribuição *a priori* de γ . Do ponto de vista bayesiano, contudo, a *OR* é uma quantidade aleatória que depende de ambos, os efeitos fixos β e/ou os efeitos aleatórios γ . Neste caso, os resultados obtidos por Larsen *et al.* (2000) podem ser utilizados apenas para descrever o conhecimento prévio do pesquisador ao eliciar a distribuição dos efeitos aleatórios. No processo de inferência consideramos a distribuição *a posteriori* da razão de chances e, também neste caso, a interpretação a *OR* poderá ser utilizada apenas se fixarmos os *clusters* aos quais pertencem os indivíduos envolvidos na comparação.

Capítulo 2

Modelo Logístico Misto *Skew-Normal*

2.1 Introdução

Neste capítulo, vamos considerar o modelo logístico misto apresentado no Capítulo 1 e consideraremos diferentes distribuições *Skew-Normal* (SN) para modelar o comportamento do efeito aleatório. Como em Santos (2011), o foco principal deste trabalho está na interpretação dos parâmetros. O uso da distribuição SN para modelar o comportamento dos efeitos aleatórios foi, anteriormente, considerado em Liu e Dey (2008) e Santos (2011). Através da abordagem bayesiana, Liu e Dey (2008) assumem que os efeitos aleatórios são independentes e identicamente distribuídos (i.i.d) com distribuição SN e também assume uma distribuição *a priori* discreta para o parâmetro de assimetria. Santos (2011), por sua vez, considera um modelo mais flexível do que o assumido em Liu e Dey (2008) e foca na interpretação dos parâmetros considerando que os efeitos aleatórios são i.i.d com distribuição *Skew-Normal* definida em Azzalini (1986) e, no caso correlacionado, assumindo o comportamento dos efeitos aleatórios pode ser descrito pela distribuição SN multivariada, proposta em Azzalini e Dalla-Valle (1996), com matriz de escala diagonal.

Dados biológicos, ambientais e financeiros são frequentemente assimétricos e/ou com caudas pesadas. Por terem esse tipo de comportamento, dados deste tipo não são, em geral, bem ajustados pela distribuição Normal. Pensando nesses exemplos, o atrativo de usar a família de distribuições SN está na flexibilidade em ajustar densidades de diferentes formas, entre elas uma distribuição perfeitamente simétrica (a distribuição Normal quando o parâmetro de assimetria é zero) ou uma distribuição com forte grau de assimetria (como a distribuição *half-Normal*). Além disso, as famílias de distribuições SN preservam algumas das boas propriedades da família Normal.

Neste capítulo assumiremos distribuições *Skew-Normais* independentes e correlacionadas mais gerais e flexíveis do que aquelas assumidas em Santos (2011) e Liu e Dey (2008) para representarmos o comportamento dos efeitos aleatórios. A f.d.p destas distribuições e algumas propriedades são apresentadas no Apêndice B. Na Seção 2.2, os resultados obtidos por Larsen *et al.* (2000) e Santos (2011) para distribuição da *OR* e sobre *MOR* são estendidos ao considerarmos estas classes de distribuições SN para os efeitos aleatórios. Em Santos (2011), no caso *Skew-Normal* independente, os resultados sobre *OR* e *MOR*

foram obtidos assumindo que efeitos aleatórios são independentes e identicamente distribuídos. Tais resultados são estendidos ao considerarmos apenas independência para os efeitos aleatórios. Sob tal suposição obtemos, como subproduto, um resultado fornecendo a distribuição de combinações lineares dos efeitos aleatórios. Ao reconhecermos que a distribuição para tais combinações lineares pertence a família *Skew-Normal Unificada* (SUN) proposta por Arellano-Valle e Azzalini (2006), os resultados em Santos (2011) para o caso i.i.d também foram melhorados e seguem como corolários dos casos mais gerais obtidos neste trabalho. Mesmo no caso i.i.d, resultados novos são obtidos a respeito da *MOR* proposta para quantificar a heterogeneidade entre os *clusters* (itens (2)(i) e (2)(ii) do Corolário 2.2). Apresentamos resultados para combinações lineares dos efeitos aleatórios, *OR* e *MOR* ao assumirmos a distribuição *Skew-Normal* multivariada para os efeitos aleatórios. Estes resultados são mais gerais que os apresentados em Santos (2011), uma vez que são obtidos sem assumirmos restrições para os parâmetros da distribuição *Skew-Normal* consideradas para modelarmos o comportamento dos efeitos aleatórios. Em Santos (2011) os resultados são obtidos considerando que a matriz de escala é uma matriz diagonal.

Na Seção 2.3 são apresentados detalhes sobre o processo de inferência bayesiana nos dois modelos logísticos mistos *Skew-Normal* utilizados. No primeiro caso, em que assumimos efeitos aleatórios *Skew-Normais* i.i.d, o modelo é um pouco mais geral do que o apresentado em Santos (2011), visto que consideramos também uma distribuição *a priori* *Skew-Normal* para representarmos a incerteza sobre o parâmetro de assimetria. Essa modificação permite elicitar uma distribuição *a priori* que favoreça valores positivos (ou negativos) para λ , tornando o modelo um pouco mais flexível. No segundo caso, assumimos a distribuição SN multivariada para representar o comportamento dos efeitos aleatórios. São usadas estruturas mais simples para a matriz de escala e para o vetor de assimetria da distribuição dos efeitos aleatórios, tornando o modelo mais parcimonioso. Este modelo é exatamente o utilizado em Santos (2011). Ainda na mesma seção, são apresentados detalhes de como é feito o processo de inferência *a posteriori* para a razão de chances, como descrito em Santos (2011), e são descritos os critérios de seleção de modelos utilizados nas seções posteriores.

Na Seção 2.4 realizamos um estudo com dados simulados para avaliarmos o impacto da má-especificação da distribuição dos efeitos aleatórios. Para isso, são considerados bancos de dados simulados em que os efeitos aleatórios são i.i.d e gerados de distribuições Normal e *Skew-Normal*. Os bancos de dados diferem não somente neste aspecto, mas também por considerarmos diferentes valores para a variância da distribuição dos efeitos aleatórios. Para a análise, ajustamos modelos que consideram efeitos aleatórios i.i.d tanto normalmente distribuídos quanto distribuídos segundo uma distribuição *Skew-Normal*. Na Seção 2.5 apresentamos um estudo de simulação Monte Carlo com o objetivo de avaliarmos a influência da distribuição *a priori* do parâmetro de assimetria na inferência *a posteriori*. Para tal, consideramos cenários com diferentes números de *clusters* e efeitos aleatórios com diferentes níveis de assimetria. Para cada cenário ajustamos modelos assumindo distribuições *a priori* degeneradas, Normal e *Skew-Normal* para o parâmetro de assimetria.

Nas Seções 2.6 e 2.7 os modelos propostos são aplicados a dados reais. Na primeira, reanalisamos o conjunto de dados apresentado em Liu e Dey (2008) sobre um experimento realizado para estudar a interação entre atividade teratogênica de dois análogos de nia-

cina *6-aminonicotinamide* (6AN) e *3-acetylpyridine* (3AP). Liu e Dey (2008) assumiu o modelo logístico misto com diferentes especificações para os efeitos aleatórios. Eles concluíram que o modelo logístico misto *Skew-Normal* é melhor que aquele que assume uma distribuição Normal para os efeitos aleatórios ou mesmo quando os efeitos aleatórios são tratados não parametricamente. Contudo, Liu e Dey (2008) consideram uma distribuição *a priori* com massa pontual para o parâmetro de assimetria λ , em que λ assumir um valor alto positivo ou alto negativo. Ao fazerem isto, eles assumem que os efeitos aleatórios são, necessariamente, todos positivos ou todos negativos. Na Seção 2.7, o banco de dados usado em Larsen *et al.* (2000), que trata da contaminação de suínos por *Ascaris Suum*, é reanalisado considerando tanto a distribuição Normal quanto a distribuição *Skew-Normal* para representar o comportamento dos efeitos aleatórios. A diferença entre a análise que aqui realizamos para aquela apresentada em Larsen *et al.* (2000) reside no fato de que esta última utiliza a abordagem clássica para o processo de inferência dos parâmetros do modelo e assume apenas a distribuição Normal para os efeitos aleatórios.

Finalizando este capítulo, a Seção 2.8 discorre sobre as principais conclusões tiradas ao considerarmos os modelos mais flexíveis introduzidos aqui.

2.2 Interpretação dos parâmetros

Para simplificarmos o processo de inferência no modelo logístico misto, os efeitos aleatórios são, usualmente, assumidos como sendo independentes com distribuição Normal. Tal suposição, no entanto, é questionável em alguns contextos como, por exemplo, no estudo com dados biológicos discutido em Liu e Dey (2008). Nesta seção, assumiremos distribuições mais flexíveis para modelar o comportamento dos efeitos aleatórios. Consideramos que tais efeitos aleatórios têm distribuição *Skew-Normal* e também assumiremos que são independentes (Azzalini, 1985) ou correlacionados (multivariados) (Azzalini e Dalla-Valle, 1996). Ambas famílias SN consideradas aqui incluem a Normal como caso particular. No entanto, diferentemente do que é observado para a distribuição Normal, o caso independente não segue diretamente do caso multivariado ao assumirmos uma matriz de escala diagonal. A independência na família SN decorre do caso multivariado se a matriz de escala é diagonal e se existe somente um componente não nulo no vetor de parâmetros de assimetria.

Ao longo deste trabalho denotamos por $\phi_n(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ a função densidade de probabilidade (f.d.p) associada com a distribuição multivariada $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, e por $\Phi_n(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ a correspondente função de distribuição acumulada (f.d.a). Se $\boldsymbol{\mu} = \mathbf{0}$ (respectivamente, $\boldsymbol{\mu} = \mathbf{0}$ e $\boldsymbol{\Sigma} = \mathbf{I}_n$) estas funções serão denotadas por $\phi_n(\mathbf{y} \mid \boldsymbol{\Sigma})$ e $\Phi_n(\mathbf{y} \mid \boldsymbol{\Sigma})$ (respectivamente, $\phi_n(\mathbf{y})$ e $\Phi_n(\mathbf{y})$). Por simplicidade, $\phi(\mathbf{y})$ e $\Phi(\mathbf{y})$ serão usadas no caso univariado.

2.2.1 Resultados *a priori* para a razão de chances: Caso independente

Assuma que, dados $\sigma_i^2 \in \mathbb{R}_+$ e $\xi_i, \lambda_i \in \mathbb{R}$, $\gamma_i \stackrel{ind.}{\sim} SN(\xi_i, \sigma_i^2, \lambda_i)$ para $i = 1, \dots, k$, em que ξ_i, σ_i^2 e λ_i são, respectivamente, os parâmetros de localização, escala e assimetria da distribuição SN. Isto é, $f(\gamma_i \mid \xi_i, \sigma_i^2, \lambda_i) = \frac{2}{\sigma_i} \phi\left(\frac{\gamma_i - \xi_i}{\sigma_i}\right) \Phi\left(\lambda_i \frac{\gamma_i - \xi_i}{\sigma_i}\right)$, $\gamma_i \in \mathbb{R}$.

Note que assumimos os efeitos aleatórios independentes, mas não necessariamente identicamente distribuídos e, por isso, os resultados obtidos aqui são mais gerais do que os obtidos em Santos (2011), onde se considera que os efeitos aleatórios são i.i.d com distribuição *Skew-Normal*.

Como apontado em (1.3), a OR_{12} depende dos efeitos aleatórios através $W_{12} = \gamma_{i_1} - \gamma_{i_2}$. Por isso, para obtermos a distribuição de OR_{12} , necessitamos encontrar a distribuição da combinação linear de variáveis aleatórias independentes SN. Exibimos este resultado na Proposição 2.1 a seguir. Para apresentá-lo necessitamos conhecer a família *Skew-Normal unificada* (SUN) introduzida em Arellano-Valle e Azzalini (2006). Sua definição e algumas de suas características são indicadas no Apêndice B. Como antes, seja $OR_{12} = \exp\{\kappa_{12} + W_{12}\}$, em que $\kappa_{12} = (\mathbf{x}_{i_1 j_1}^t - \mathbf{x}_{i_2 j_2}^t)\boldsymbol{\beta}$. Também considere $\delta_i = \lambda_i[1 + \lambda_i^2]^{-1/2}$.

PROPOSIÇÃO 2.1. *Se $\gamma_i \stackrel{ind}{\sim} SN(\xi_i, \sigma_i^2, \lambda_i)$, $\forall i = 1, \dots, k$, e $\mathbf{c}^t = (c_1, \dots, c_k)$ é um vetor de constantes reais de ordem $1 \times k$, então $\mathbf{c}^t \boldsymbol{\gamma} \sim SUN_{1,k}(\mathbf{c}^t \boldsymbol{\xi}, \mathbf{0}, [\mathbf{c}^t \boldsymbol{\Lambda}^t \boldsymbol{\Lambda} \mathbf{c}]^{1/2}, \boldsymbol{\Omega}^*)$, em que $\boldsymbol{\xi} = (\xi_1, \dots, \xi_k)$, $\boldsymbol{\Lambda} = \text{diag}\{\sigma_1, \dots, \sigma_k\}$, $\boldsymbol{\Omega}^* = \begin{pmatrix} I_k & \boldsymbol{\Gamma}^t \\ \boldsymbol{\Gamma} & 1 \end{pmatrix}$, $\boldsymbol{\Gamma}^t = (\mathbf{c}^t \boldsymbol{\Lambda}^t \boldsymbol{\Lambda} \mathbf{c})^{-1/2} \boldsymbol{\Lambda} \boldsymbol{\Delta} \mathbf{c}$, $\boldsymbol{\Delta} = \text{diag}\{\delta_1, \dots, \delta_k\}$, $\delta_i = \lambda_i[1 + \lambda_i^2]^{-1/2}$ e sua f.d.p é*

$$f_{\mathbf{c}^t \boldsymbol{\gamma} | \boldsymbol{\xi}, \boldsymbol{\Lambda}, \boldsymbol{\Omega}^*}(x) = 2^k \phi(x | \mathbf{c}^t \boldsymbol{\xi}, \mathbf{c}^t \boldsymbol{\Lambda}^t \boldsymbol{\Lambda} \mathbf{c}) \Phi_k \left(\frac{\boldsymbol{\Lambda} \boldsymbol{\Delta} \mathbf{c}}{\mathbf{c}^t \boldsymbol{\Lambda}^t \boldsymbol{\Lambda} \mathbf{c}} (x - \mathbf{c}^t \boldsymbol{\xi}) \mid I_k - \frac{\boldsymbol{\Lambda} \boldsymbol{\Delta} \mathbf{c} \mathbf{c}^t \boldsymbol{\Delta}^t \boldsymbol{\Lambda}^t}{\mathbf{c}^t \boldsymbol{\Lambda}^t \boldsymbol{\Lambda} \mathbf{c}} \right). \quad (2.1)$$

DEMONSTRAÇÃO.

Usando resultados conhecidos do cálculo de probabilidades segue que, a função geradora de momentos (f.g.m) de $\mathbf{c}^t \boldsymbol{\gamma}$ é dada por

$$\begin{aligned} M_{\mathbf{c}^t \boldsymbol{\gamma}}(r) &= \prod_{i=1}^k M_{\gamma_i}(rc_i) \\ &= \prod_{i=1}^k 2 \exp \left\{ rc_i \xi_i + \frac{r^2 c_i^2 \sigma_i^2}{2} \right\} \Phi \left(\frac{\lambda_i}{(1 + \lambda_i^2)^{1/2}} \sigma_i c_i r \right) \\ &= 2^k \exp \left\{ \sum_{i=1}^k rc_i \xi_i + \frac{1}{2} \sum_{i=1}^k r^2 c_i^2 \sigma_i^2 \right\} \prod_{i=1}^k \Phi \left(\frac{\lambda_i}{(1 + \lambda_i^2)^{1/2}} \sigma_i c_i r \right) \\ &= \exp \left\{ r \mathbf{c}^t \boldsymbol{\xi} + \frac{r^2}{2} \mathbf{c}^t \boldsymbol{\Lambda}^t \boldsymbol{\Lambda} \mathbf{c} \right\} \Phi_k(r \boldsymbol{\Lambda} \boldsymbol{\Delta} \mathbf{c} \mid I_k) [\Phi_k(\mathbf{0} \mid I_k)]^{-1}, \end{aligned}$$

que de acordo com Arellano-Valle e Azzalini (2006) (ver Apêndice B) é a f.g.m. da distribuição SUN citada em (2.1). □

Segue diretamente da Proposição 2.1 que W_{12} tem distribuição *Skew-Normal Unificada*, o que é dado no seguinte corolário.

COROLÁRIO 2.1. Se $\gamma_{i_l} \stackrel{ind}{\sim} SN(\xi_{i_l}, \sigma_{i_l}^2, \lambda_{i_l})$, $l = 1, 2$, então, dados os hiperparâmetros $\boldsymbol{\theta}_{12} = (\xi_{i_1}, \sigma_{i_1}^2, \lambda_{i_1}, \xi_{i_2}, \sigma_{i_2}^2, \lambda_{i_2})$, segue que $W_{12} \sim SUN_{1,2}(\xi_{i_1} - \xi_{i_2}, \mathbf{0}, [\sigma_{i_1}^2 + \sigma_{i_2}^2]^{\frac{1}{2}}, \boldsymbol{\Omega}_{12}^*)$, cuja f.d.p no ponto $w \in \mathbb{R}$ é dada por

$$f_{W_{12}|\boldsymbol{\theta}_{12}}(w) = 4\phi(w \mid \xi_{i_1} - \xi_{i_2}, \sigma_{i_1}^2 + \sigma_{i_2}^2) \Phi_2 \left(\frac{w - \xi_{i_1} + \xi_{i_2}}{\sigma_{i_1}^2 + \sigma_{i_2}^2} \boldsymbol{\epsilon} \mid \mathbf{I}_2 - \frac{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^t}{\sigma_{i_1}^2 + \sigma_{i_2}^2} \right), \quad (2.2)$$

em que $\boldsymbol{\epsilon} = (\sigma_{i_1}\delta_{i_1}, -\sigma_{i_2}\delta_{i_2})^t$, $\boldsymbol{\Omega}_{12}^* = \begin{pmatrix} \mathbf{I}_2 & \boldsymbol{\Gamma}_{12}^t \\ \boldsymbol{\Gamma}_{12} & 1 \end{pmatrix}$ e $\boldsymbol{\Gamma}_{12} = (\sigma_{i_1}^2 + \sigma_{i_2}^2)^{-\frac{1}{2}} \boldsymbol{\epsilon}^t$.

DEMONSTRAÇÃO.

Por simplicidade e sem perda de generalidade, o resultado será provado considerando que $i_1 = 1$ e $i_2 = 2$, ou seja, vamos considerar $W_{12} = \gamma_1 - \gamma_2$. Note que W_{12} pode ser escrito como $\mathbf{c}^t \boldsymbol{\gamma}$, em que $\boldsymbol{\gamma}^t = (\gamma_1, \dots, \gamma_k)$ e $\mathbf{c}^t = (1, -1, 0, \dots, 0)$. Assim, da Proposição 2.1 temos que

$$W_{12} \sim SUN_{1,k}(\mathbf{c}^t \boldsymbol{\xi}, \mathbf{0}, [\mathbf{c}^t \boldsymbol{\Lambda}^t \boldsymbol{\Lambda} \mathbf{c}]^{1/2}, \boldsymbol{\Omega}^*),$$

em que

$$\begin{aligned} \mathbf{c}^t \boldsymbol{\xi} &= \mathbf{c}^t (\xi_1, \dots, \xi_k) = \xi_1 - \xi_2, \\ \mathbf{c}^t \boldsymbol{\Lambda}^t \boldsymbol{\Lambda} \mathbf{c} &= \mathbf{c}^t \text{diag}^2 \{\sigma_1, \dots, \sigma_k\} \mathbf{c} = \sigma_1^2 + \sigma_2^2, \\ \boldsymbol{\Gamma}^t &= (\mathbf{c}^t \text{diag}^2 \{\sigma_1, \dots, \sigma_k\} \mathbf{c})^{-1/2} \text{diag} \{\sigma_1, \dots, \sigma_k\} \text{diag} \{\delta_1, \dots, \delta_k\} \mathbf{c} \\ &= (\sigma_1^2 + \sigma_2^2)^{-\frac{1}{2}} (\sigma_1 \delta_1, -\sigma_2 \delta_2, 0, \dots, 0)^t. \end{aligned}$$

Neste caso, uma vez que \mathbf{I}_k é uma matriz diagonal de ordem $k \times k$ e $\boldsymbol{\Gamma}^t$ tem $k - 2$ entradas nulas, a função geradora de momentos de W_{12} , no ponto $r \in \mathbb{R}$, é dada por

$$\begin{aligned} M_{W_{12}}(r) &= \exp \left\{ (\xi_1 - \xi_2)r + \frac{1}{2} r^2 (\sigma_1^2 + \sigma_2^2) \right\} \frac{\Phi_k(\boldsymbol{\Gamma}^t (\sigma_1^2 + \sigma_2^2)^{\frac{1}{2}} r \mid \mathbf{I}_k)}{\Phi_k(0 \mid \mathbf{I}_k)} \\ &= \exp \left\{ (\xi_1 - \xi_2)r + \frac{1}{2} r^2 (\sigma_1^2 + \sigma_2^2) \right\} \frac{\Phi_2(\boldsymbol{\Gamma}_{12}^t (\sigma_1^2 + \sigma_2^2)^{\frac{1}{2}} r \mid \mathbf{I}_2)}{\Phi_2(0 \mid \mathbf{I}_2)}, \end{aligned}$$

com $\boldsymbol{\Gamma}_{12}^t = (\sigma_1^2 + \sigma_2^2)^{-\frac{1}{2}} \boldsymbol{\epsilon}$ e $\boldsymbol{\epsilon} = (\sigma_1 \delta_1, -\sigma_2 \delta_2)^t$.

Portanto, $W_{12} \sim SUN_{1,2}(\xi_1 - \xi_2, \mathbf{0}_2, (\sigma_1^2 + \sigma_2^2)^{\frac{1}{2}}, \boldsymbol{\Omega}_{12}^*)$, com $\boldsymbol{\Omega}_{12}^* = \begin{pmatrix} \mathbf{I}_2 & \boldsymbol{\Gamma}_{12}^t \\ \boldsymbol{\Gamma}_{12} & 1 \end{pmatrix}$, concluindo a demonstração. □

Note que ao assumirmos $\xi_{i_1} = \xi_{i_2} = 0$, $\sigma_{i_1}^2 = \sigma_{i_2}^2 = \sigma^2$ e $\lambda_{i_1} = \lambda_{i_2} = \lambda$ em (2.2) obtemos o resultado apresentado em Santos (2011).

A próxima proposição fornece a distribuição *a priori* para OR_{12} e a sua mediana, dados os hiperparâmetros $\boldsymbol{\theta}_{12} = (\xi_{i_1}, \sigma_{i_1}^2, \lambda_{i_1}, \xi_{i_2}, \sigma_{i_2}^2, \lambda_{i_2})^t$ e os efeitos fixos $\boldsymbol{\beta}$, quando os efeitos aleatórios são independentes (ind.) com distribuição SN.

PROPOSIÇÃO 2.2. Se $\gamma_{i_l} \stackrel{ind}{\sim} SN(\xi_{i_l}, \sigma_{i_l}^2, \lambda_{i_l})$, $l = 1, 2$, então, dados \mathbf{X} , os efeitos fixos β e os hiperparâmetros $\theta_{12} = (\xi_{i_1}, \sigma_{i_1}^2, \lambda_{i_1}, \xi_{i_2}, \sigma_{i_2}^2, \lambda_{i_2})$, segue que

(i) a distribuição de OR_{12} é a distribuição Log-SUN, proposta por de Queiroz et al. (2015), com f.d.p

$$f_{OR_{12}|\beta, \theta_{12}, \mathbf{X}}(r) = \frac{4}{r} \phi(\ln r | \kappa_{12} + \xi_{i_1} - \xi_{i_2}, \sigma_{i_1}^2 + \sigma_{i_2}^2) \quad (2.3)$$

$$\times \Phi_2 \left(\frac{\ln r}{\sigma_{i_1}^2 + \sigma_{i_2}^2} \boldsymbol{\epsilon} \middle| \frac{\kappa_{12} + \xi_{i_1} - \xi_{i_2}}{\sigma_{i_1}^2 + \sigma_{i_2}^2} \boldsymbol{\epsilon}, \mathbf{I}_2 - \frac{\boldsymbol{\epsilon} \boldsymbol{\epsilon}^t}{\sigma_{i_1}^2 + \sigma_{i_2}^2} \right), \text{ para } r \in \mathbb{R}_+;$$

(ii) a razão de chances mediana é dada por

$$MOR_{12} = \exp \{ \kappa_{12} \} \exp \left\{ \Phi_{SUN_{1,2}}^{-1} \left(0,5 | \xi_{i_1} - \xi_{i_2}, \mathbf{0}, [\sigma_{i_1}^2 + \sigma_{i_2}^2]^{1/2}, \boldsymbol{\Omega}_{12}^* \right) \right\},$$

em que $\Phi_{SUN_{1,2}}(\cdot)$ denota a f.d.a da distribuição SUN dada em (2.2).

DEMONSTRAÇÃO.

(i) Utilizando o resultado da proposição anterior temos que, uma vez que $OR_{12} = \exp \{ \kappa_{12} + W_{12} \}$, segue que $W_{12} = \ln(OR_{12}) - \kappa_{12}$. Usando resultados do cálculo de probabilidades segue que

$$f_{OR_{12}|\beta, \theta_{12}, \mathbf{X}}(r) = f_{W_{12}|\theta_{12}}(\ln r - \kappa_{12}) \left| \frac{dW_{12}}{dr} \right|$$

$$= \frac{4}{r} \phi(\ln r - \kappa_{12} | \xi_{i_1} - \xi_{i_2}, \sigma_{i_1}^2 + \sigma_{i_2}^2)$$

$$\times \Phi_2 \left(\frac{\ln r - \kappa_{12} - (\xi_{i_1} - \xi_{i_2})}{\sigma_{i_1}^2 + \sigma_{i_2}^2} \boldsymbol{\epsilon} \middle| \mathbf{I}_2 - \frac{\boldsymbol{\epsilon} \boldsymbol{\epsilon}^t}{\sigma_{i_1}^2 + \sigma_{i_2}^2} \right)$$

$$= \frac{4}{r} \phi(\ln r | \kappa_{12} + \xi_{i_1} - \xi_{i_2}, \sigma_{i_1}^2 + \sigma_{i_2}^2)$$

$$\times \Phi_2 \left(\frac{\ln r}{\sigma_{i_1}^2 + \sigma_{i_2}^2} \boldsymbol{\epsilon} \middle| \frac{\kappa_{12} + \xi_{i_1} - \xi_{i_2}}{\sigma_{i_1}^2 + \sigma_{i_2}^2} \boldsymbol{\epsilon}, \mathbf{I}_2 - \frac{\boldsymbol{\epsilon} \boldsymbol{\epsilon}^t}{\sigma_{i_1}^2 + \sigma_{i_2}^2} \right).$$

(ii) Seja md a mediana da distribuição de OR_{12} . Considerando o resultado em (i) e denotando $\xi^* = \kappa_{12} + (\xi_{i_1} - \xi_{i_2})$ e $\Sigma^* = \sigma_{i_1}^2 + \sigma_{i_2}^2$, segue da definição de mediana que

$$F_{OR_{12}|\beta, \theta_{12}, \mathbf{X}}(md) = \int_0^{md} f_{OR_{12}|\beta, \sigma^2, \lambda}(r) dr = 0,5$$

$$\Leftrightarrow \int_0^{md} \frac{4}{r} \phi(\ln r - \xi^*, \Sigma^*) \Phi_2 \left(\frac{\ln r - \xi^*}{\Sigma^*} \boldsymbol{\epsilon} \middle| \mathbf{I}_2 - \frac{\boldsymbol{\Omega}_{12}}{\Sigma^*} \right) dr = 0,5.$$

Aplicando a transformação de variável $z = \ln r - \xi^*$, temos que $dz = dt/t$ e que

$$\int_0^{\ln md - \xi^*} 4 \phi(z, \Sigma^*) \Phi_2 \left(\frac{z}{\Sigma^*} \boldsymbol{\epsilon} \middle| \mathbf{I}_2 - \frac{\boldsymbol{\Omega}_{12}}{\Sigma^*} \right) dz = 0,5$$

$$\Leftrightarrow \Phi_{SUN_{1,2}}(\ln md - \kappa_{12} | \xi_{i_1} - \xi_{i_2}, \mathbf{0}, [\sigma_{i_1}^2 + \sigma_{i_2}^2]^{1/2}, \boldsymbol{\Omega}_{12}^*) = 0,5,$$

de onde segue que

$$\begin{aligned} \ln md - \kappa_{12} &= \Phi_{SUN_{1,2}}^{-1} \left(0,5 | \xi_{i_1} - \xi_{i_2}, \mathbf{0}, [\sigma_{i_1}^2 + \sigma_{i_2}^2]^{1/2}, \mathbf{\Omega}_{12}^* \right) \\ \Leftrightarrow md &= \exp \left\{ \kappa_{12} + \Phi_{SUN_{1,2}}^{-1} \left(0,5 | \xi_{i_1} - \xi_{i_2}, \mathbf{0}, [\sigma_{i_1}^2 + \sigma_{i_2}^2]^{1/2}, \mathbf{\Omega}_{12}^* \right) \right\}. \end{aligned}$$

Então, $md = \exp \{ \kappa_{12} \} \exp \left\{ \Phi_{SUN_{1,2}}^{-1} \left(0,5 | \xi_{i_1} - \xi_{i_2}, \mathbf{0}, [\sigma_{i_1}^2 + \sigma_{i_2}^2]^{1/2}, \mathbf{\Omega}_{12}^* \right) \right\}$,
concluindo a demonstração. □

Para reduzirmos o número de parâmetros as serem estimados, uma prática comum em modelos mistos é assumir efeitos aleatórios i.i.d. No próximo corolário apresentamos alguns resultados a respeito da distribuição de OR_{12} e das medidas MOR_{12} e $MOR_{|12|}$, os quais foram discutidos em Larsen *et al.* (2000) sob suposição de normalidade para os efeitos aleatórios.

COROLÁRIO 2.2. *Se $\gamma_{i_l} \stackrel{iid}{\sim} SN(\xi, \sigma^2, \lambda)$, $l = 1, 2$, então, dados \mathbf{X} , os efeitos fixos β e os hiperparâmetros $\theta = (\xi, \sigma^2, \lambda)$, segue que*

(1) *a distribuição de OR_{12} é a distribuição Log-SUN com f.d.p*

$$\begin{aligned} f_{OR_{12} | \beta, \theta, \mathbf{X}}(r) &= \frac{4}{r} \phi(\ln r | \kappa_{12}, 2\sigma^2) \\ &\times \Phi_2 \left(\frac{\delta \ln r}{2\sigma} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \middle| \frac{\delta \kappa_{12}}{2\sigma} \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \mathbf{I}_2 - \frac{\delta^2}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \right), \quad (2.4) \end{aligned}$$

para $r \in \mathbb{R}^+$;

(2) *a razão de chances mediana MOR_{12} é dada por*

- (i) $MOR_{12} = \text{med} \{ \exp \{ \kappa_{12} + W_{12} \} | \beta, \mathbf{X}, \theta \} = \exp \{ \kappa_{12} \}$;
- (ii) $MOR_{|12|} = \text{med} \{ \exp \{ |W_{12}| \} | \beta, \mathbf{X}, \theta \} = \exp \{ \text{med} \{ |W_{12}| | \beta, \mathbf{X}, \theta \} \}$
 $= \exp \left\{ \Phi_{SUN_{1,2}}^{-1} \left(0,75 | 0, \mathbf{0}, (2\sigma^2)^{1/2}, \mathbf{\Omega}_{12}^* \right) \right\}$, em que $\mathbf{\Omega}_{12}^* = \begin{pmatrix} \mathbf{I}_2 & \frac{\delta(1,-1)^t}{\sqrt{2}} \\ \frac{\delta(1,-1)}{\sqrt{2}} & 1 \end{pmatrix}$;
- (iii) $\text{med} \{ \exp \{ -|W_{12}| \} | \beta, \mathbf{X}, \theta \} = \text{med} \{ \exp \{ |W_{12}| \} | \beta, \mathbf{X}, \theta \}^{-1}$.

DEMONSTRAÇÃO.

- (1) O resultado segue diretamente do item (i) da Proposição 2.2 ao assumirmos $\xi_{i_1} = \xi_{i_2} = \xi$, $\sigma_{i_1}^2 = \sigma_{i_2}^2 = \sigma^2$ e $\lambda_{i_1} = \lambda_{i_2} = \lambda$.
- (2) Os próximos resultados a respeito da MOR_{12} decorrem do fato que, ao assumirmos efeitos aleatórios i.i.d, a distribuição de W_{12} é simétrica em torno de zero (Ver Proposição A.1, Apêndice A).

(i) Utilizando o resultado em (ii) da Proposição 2.2, temos que

$$\begin{aligned} MOR_{12} &= \text{med}\{\exp\{\kappa_{12} + W_{12}\} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}\} \\ &= \exp\{\kappa_{12}\} \exp\{\Phi_{SUN_{1,2}}^{-1}(0,5 \mid 0, \mathbf{0}, (2\sigma^2)^{1/2}, \boldsymbol{\Omega}_{12}^*)\}, \end{aligned} \quad (2.5)$$

$$\text{em que } \boldsymbol{\Omega}_{12}^* = \begin{pmatrix} \mathbf{I}_2 & \frac{\delta(1,-1)^t}{\sqrt{2}} \\ \frac{\delta(1,-1)}{\sqrt{2}} & 1 \end{pmatrix}.$$

Como $W_{12} \sim SUN_{1,2}(0, \mathbf{0}, (2\sigma^2)^{1/2}, \boldsymbol{\Omega}_{12}^*)$ a qual é uma distribuição simétrica em torno de zero, segue que $\Phi_{SUN_{1,2}}^{-1}(0,5 \mid 0, \mathbf{0}, (2\sigma^2)^{1/2}, \boldsymbol{\Omega}_{12}^*) = 0$. Logo, temos a simplificação da expressão em (2.5) e segue que $MOR_{12} = \exp\{\kappa_{12}\}$.

(ii) Seja $a_1 = \text{med}\{\exp\{|W_{12}|\} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}\}$. Assim, como a distribuição de W_{12} é simétrica em torno de zero,

$$\begin{aligned} P(\exp\{|W_{12}|\} \leq a_1 \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}) &= 0,5; \\ P(-\ln a_1 \leq W_{12} \leq \ln a_1 \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}) &= 0,5; \\ 2P(W_{12} \leq \ln a_1 \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}) - 1 &= 0,5; \\ P(W_{12} \leq \ln a_1 \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}) &= 0,75; \\ \Phi_{SUN_{1,2}}(\ln a_1 \mid 0, \mathbf{0}, (2\sigma^2)^{1/2}, \boldsymbol{\Omega}_{12}^*) &= 0,75. \end{aligned}$$

Desta forma, $a_1 = \Phi_{SUN_{1,2}}^{-1}(0,75 \mid 0, \mathbf{0}, (2\sigma^2)^{1/2}, \boldsymbol{\Omega}_{12}^*)$. Usando a Proposição A.2 no Apêndice A temos também que

$$\text{med}\{|W_{12}| \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}\} = \exp\{\Phi_{SUN_{1,2}}^{-1}(0,75 \mid 0, \mathbf{0}, (2\sigma^2)^{1/2}, \boldsymbol{\Omega}_{12}^*)\};$$

(iii) Seja $\text{med}\{\exp\{-|W_{12}|\} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}\} = h$. Assim,

$$\begin{aligned} P(\exp\{-|W_{12}|\} \leq h \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}) &= 0,5; \\ P(|W_{12}| \geq -\ln h \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}) &= 0,5; \\ P(W_{12} \geq -\ln h \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}) + P(W_{12} \leq \ln h \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}) &= 0,5; \\ 2P(W_{12} \leq \ln h \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}) &= 0,5; \\ \Phi_{SUN_{1,2}}(\ln h \mid 0, \mathbf{0}, (2\sigma^2)^{1/2}, \boldsymbol{\Omega}_{12}^*) &= 0,25. \end{aligned}$$

Desta forma,

$$\begin{aligned} \ln h &= \Phi_{SUN_{1,2}}^{-1}(0,25 \mid 0, \mathbf{0}, (2\sigma^2)^{1/2}, \boldsymbol{\Omega}_{12}^*); \\ \ln h &= -\Phi_{SUN_{1,2}}^{-1}(0,75 \mid 0, \mathbf{0}, (2\sigma^2)^{1/2}, \boldsymbol{\Omega}_{12}^*); \\ h &= \left[\exp\{\Phi_{SUN_{1,2}}^{-1}(0,75 \mid 0, \mathbf{0}, (2\sigma^2)^{1/2}, \boldsymbol{\Omega}_{12}^*)\} \right]^{-1}; \\ h &= [\text{med}\{\exp\{|W_{12}|\} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}\}]^{-1}. \end{aligned}$$

Logo, $\text{med}\{\exp\{-|W_{12}|\} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}\} = [\text{med}\{\exp\{|W_{12}|\} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}\}]^{-1}$, concluindo a demonstração.

□

Parte do resultado apresentado no Corolário 2.2 foi obtido por Santos (2011), a saber, a parte referente a distribuição de OR_{12} e sua mediana MOR proposta para quantificar os efeitos fixos. No entanto, em Santos (2011) a distribuição de W_{12} não havia sido reconhecida como membro da família de distribuições SUN e não havíamos provado que, no caso i.i.d, a distribuição de W_{12} se tratava de uma distribuição simétrica em torno de zero. Estes novos avanços permitiram melhorarmos também os resultados já apresentados no caso i.i.d e acrescentar novos itens referentes a MOR proposta para avaliar os efeitos aleatórios.

A Figura 1 apresenta a f.d.p de W_{12} e OR no caso i.i.d para diferentes valores de λ e assumindo $\sigma^2 = 1$ e $\kappa_{12} = 0$. Valores de λ iguais em módulo fornecem a mesma densidade. Notamos também que ao eliciarmos valores altos para $|\lambda|$, a OR tende a ser maior do que a assumida no modelo de Larsen *et al.* (2000).

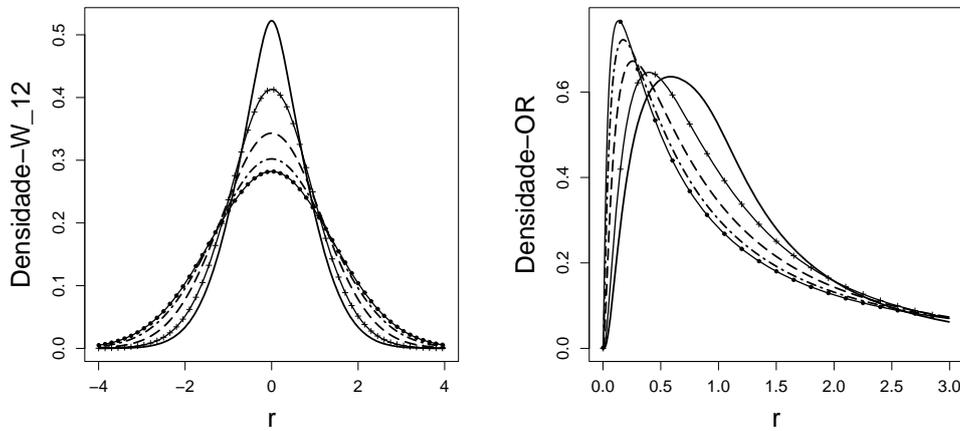


Figura 1: Distribuições *a priori* de W_{12} (esquerda) e OR (direita) para o caso independente com $\xi = 0$, $\sigma^2 = 1$, $\kappa_{12} = 0$ e $\lambda = -8,5$ (linha sólida), -2 (+), -1 (linha tracejada), $-0,5$ (linha tracejada com pontos), 0 (\bullet , modelo do Larsen *et al.* (2000)).

O primeiro gráfico da Figura 1 nos revela que quanto maior a assimetria assumida para a distribuição dos efeitos aleatórios mais leves serão as caudas da distribuição de W_{12} . Além disso percebemos que a f.d.p de W_{12} é simétrica em torno de zero. Em consequência disto, temos que a MOR_{12} é igual a 1 para qualquer valor de λ assumido, o que ilustra a necessidade de utilizarmos a $MOR_{|\lambda|}$ quando o interesse está em interpretar o efeito dos *clusters*. O segundo gráfico da Figura 1 mostra que se considerarmos a distribuição Normal ($\lambda = 0$) para os efeitos aleatórios esta revela que, mais provavelmente, ao compararmos dois indivíduos com as mesmas características e que estão em *clusters* diferentes, a razão de chances será menor que 1. No entanto, se considerarmos valores λ distantes de 0, a f.d.p da OR_{12} concentra maior massa de probabilidade em valores próximos de 1, isto é, retrata menor heterogeneidade entre os *clusters*. Para valores de λ menores do que $-8,5$ a f.d.p da variável aleatória OR_{12} é muito similar a obtida com $\lambda = -8,5$.

É importante notarmos (Corolário 2.2) que ao assumir efeitos aleatórios SN i.i.d a interpretação *a priori* dos efeitos fixos através da MOR não dependerá dos *clusters*

em que os indivíduos estão. Isto não é observado se os efeitos aleatórios não forem identicamente distribuídos. Um resultado similar foi obtido por Larsen *et al.* (2000) sob normalidade em um cenário mais geral, onde os efeitos aleatórios não são necessariamente identicamente distribuídos.

As distribuições de OR_{12} dadas em (2.3) e (2.4) pertencem a uma classe de distribuições Log-Skew-Normais diferentes da classe log-Skew-Normal introduzida por Marchenko e Genton (2010). Na classe de distribuições definida em Marchenko e Genton (2010), a função assimetrizadora é a f.d.a de uma distribuição Normal univariada enquanto que na família de distribuições definidas em (2.4) a função assimetrizadora é a f.d.a de uma distribuição Normal bivariada. A distribuição em (2.4) é obtida ao introduzirmos assimetria na distribuição em (1.5) e retornamos aos resultados de Larsen *et al.* (2000) se fizermos $\lambda = 0$ em (2.4).

2.2.2 Resultados *a priori* para a razão de chances: Caso correlacionado

Assuma que, dados ξ , Ω e α , $\gamma \sim SN_k(\xi, \Omega, \alpha)$ com densidade

$$f(\gamma \mid \xi, \Omega, \alpha) = 2\phi_k(\gamma \mid \xi, \Omega) \Phi(\alpha^t \omega^{-1}(\gamma - \xi)), \quad \gamma \in \mathbb{R}^k, \quad (2.6)$$

em que $\xi = (\xi_1, \dots, \xi_k)^t \in \mathbb{R}^k$ é o vetor de locação, $\alpha = (\alpha_1, \dots, \alpha_k)^t \in \mathbb{R}^k$ é o vetor de parâmetros de assimetria, Ω é uma $k \times k$ matriz de covariância positiva definida e $\omega = \text{diag}\{\sigma_1, \dots, \sigma_k\}$ é uma matriz diagonal formada pela raiz quadrada dos elementos na diagonal de Ω e é tal que $\Omega = \omega \bar{\Omega} \omega$ em que $\bar{\Omega}$ é uma matriz de correlação.

Para obtermos a distribuição da quantidade aleatória OR_{12} , inicialmente encontramos a distribuição de combinações lineares dos efeitos aleatórios. A distribuição de probabilidade de qualquer combinação linear dos efeitos aleatórios ao assumirmos a distribuição SN multivariada é apresentada na Proposição 2.3.

PROPOSIÇÃO 2.3. *Considere a combinação linear $c^t \gamma$ em que $c = (c_1, \dots, c_k)^t$ é um vetor de constantes de ordem $k \times 1$. Se $\gamma \sim SN_k(\xi, \Omega, \alpha)$, então $c^t \gamma \sim SN(\xi^t c, c^t \Omega c, \lambda)$, em que $\lambda = [\alpha^t \bar{\Omega} \omega c] [(1 + \alpha^t \bar{\Omega} \alpha)(c^t \Omega c) - (\alpha^t \bar{\Omega} \omega c)^2]^{-\frac{1}{2}}$.*

DEMONSTRAÇÃO.

A função geradora de momentos do vetor aleatório γ aplicada no ponto $r = (r_1, \dots, r_k)^t$, obtida por Azzalini (2005), é dada por

$$M_\gamma(r) = 2 \exp \left\{ \xi^t r + \frac{1}{2} r^t \Omega r \right\} \Phi \left(\frac{\alpha^t \bar{\Omega}}{(1 + \alpha^t \bar{\Omega} \alpha)^{\frac{1}{2}}} \omega r \right),$$

em que $\omega = \text{diag}\{\sigma_1, \dots, \sigma_k\}$ e $\bar{\Omega} = \omega^{-1} \Omega \omega^{-1}$.

Além disso, a função geradora de momentos de $c^t \gamma$ pode ser escrita como

$$\begin{aligned} M_{c^t \gamma}(r) &= M_\gamma(r c) \\ &= 2 \exp \left\{ r \xi^t c + \frac{1}{2} r^2 c^t \Omega c \right\} \Phi \left(\frac{\alpha^t \bar{\Omega}}{(1 + \alpha^t \bar{\Omega} \alpha)^{\frac{1}{2}}} \omega c r \frac{(c^t \Omega c)^{\frac{1}{2}}}{(c^t \Omega c)^{\frac{1}{2}}} \right). \end{aligned}$$

Fazendo $\sigma^2 = \mathbf{c}^t \boldsymbol{\Omega} \mathbf{c}$, podemos reescrever como

$$M_{\mathbf{c}^t \boldsymbol{\gamma}}(r) = 2 \exp \left\{ r \boldsymbol{\xi}^t \mathbf{c} + \frac{1}{2} r^2 \sigma^2 \right\} \Phi \left(r \sigma \frac{\boldsymbol{\alpha}^t \bar{\boldsymbol{\Omega}} \boldsymbol{\omega} \mathbf{c}}{((1 + \boldsymbol{\alpha}^t \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha})(\mathbf{c}^t \boldsymbol{\Omega} \mathbf{c}))^{\frac{1}{2}}} \right).$$

A função acima é igual a função geradora de momentos da distribuição $SN(\xi, \sigma, \lambda)$, em que $\xi = \boldsymbol{\xi}^t \mathbf{c}$, $\sigma^2 = \mathbf{c}^t \boldsymbol{\Omega} \mathbf{c}$ e

$$\begin{aligned} \frac{\lambda}{\sqrt{1 + \lambda^2}} &= \frac{\boldsymbol{\alpha}^t \bar{\boldsymbol{\Omega}} \boldsymbol{\omega} \mathbf{c}}{((1 + \boldsymbol{\alpha}^t \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha})(\mathbf{c}^t \boldsymbol{\Omega} \mathbf{c}))^{\frac{1}{2}}} \\ \Rightarrow \lambda &= \frac{(\boldsymbol{\alpha}^t \bar{\boldsymbol{\Omega}} \boldsymbol{\omega} \mathbf{c})}{((1 + \boldsymbol{\alpha}^t \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha})(\mathbf{c}^t \boldsymbol{\Omega} \mathbf{c}) - (\boldsymbol{\alpha}^t \bar{\boldsymbol{\Omega}} \boldsymbol{\omega} \mathbf{c})^2)^{\frac{1}{2}}} \end{aligned} \quad (2.7)$$

Logo, $\mathbf{c}^t \boldsymbol{\gamma} \sim SN(\boldsymbol{\xi}^t \mathbf{c}, \mathbf{c}^t \boldsymbol{\Omega} \mathbf{c}, \lambda)$, concluindo a demonstração. \square

Como na seção anterior, é nosso interesse encontrarmos a distribuição de $W_{12} = \gamma_{i_1} - \gamma_{i_2}$. Apresentamos este resultado no Corolário 2.3.

COROLÁRIO 2.3. *Se $\boldsymbol{\gamma} \sim SN_k(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ então $W_{12} \sim SN(\xi_{i_1} - \xi_{i_2}, \sigma^{*2}, \lambda_{12}^*)$, em que*

$$\lambda_{12}^* = \left[\sum_{l=1}^k \frac{\alpha_l [\sigma_{i_1 l} - \sigma_{i_2 l}]}{\sigma_l} \right] \left[\left(1 + \sum_{l,m=1}^k \frac{\alpha_m \alpha_l \sigma_{lm}}{\sigma_l \sigma_m} \right) \sigma^{*2} - \left(\sum_{l=1}^k \frac{\alpha_l [\sigma_{i_1 l} - \sigma_{i_2 l}]}{\sigma_l} \right)^2 \right]^{-1/2} e$$

$$\sigma^{*2} = \sigma_{i_1}^2 + \sigma_{i_2}^2 - 2\sigma_{i_1 i_2}.$$

DEMONSTRAÇÃO.

Considerando o vetor \mathbf{c} de forma que assuma valor 1 na posição i_1 , valor -1 na posição i_2 e zero nas demais posições. Utilizando os resultados obtidos na Proposição 2.3 temos que

$$\mathbf{c}^t \boldsymbol{\gamma} = W_{12} \sim SN(\boldsymbol{\xi}^t \mathbf{c}, \mathbf{c}^t \boldsymbol{\Omega} \mathbf{c}, \lambda),$$

e resolvendo as contas matriciais, obtemos que

$$\begin{aligned} \mathbf{c}^t \boldsymbol{\Omega} \mathbf{c} &= \sigma_{i_1}^2 + \sigma_{i_2}^2 - \sigma_{i_2 i_1}; \\ \boldsymbol{\xi}^t \mathbf{c} &= \xi_{i_1} - \xi_{i_2}; \\ \boldsymbol{\alpha}^t \bar{\boldsymbol{\Omega}} \boldsymbol{\omega} \mathbf{c} &= \sum_{l=1}^k \alpha_l \left[\frac{\sigma_{i_1 l}}{\sigma_{i_1} \sigma_l} \sigma_{i_1} - \frac{\sigma_{i_2 l}}{\sigma_{i_2} \sigma_l} \sigma_{i_2} \right] = \sum_{l=1}^k \frac{\alpha_l}{\sigma_l} [\sigma_{i_1 l} - \sigma_{i_2 l}]; \\ \boldsymbol{\alpha}^t \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha} &= \sum_{l,m=1}^k \frac{\alpha_m \alpha_l \sigma_{lm}}{\sigma_l \sigma_m}. \end{aligned}$$

Substituindo os resultados na expressão de λ em (2.7), segue que $W_{12} \sim SN(\xi_{i_1} - \xi_{i_2}, \sigma^{*2}, \lambda_{12}^*)$, em que

$$\begin{aligned} \lambda_{12}^* &= \left[\sum_{l=1}^k \frac{\alpha_l [\sigma_{i_1 l} - \sigma_{i_2 l}]}{\sigma_l} \right] \left[\left(1 + \sum_{l,m=1}^k \frac{\alpha_m \alpha_l \sigma_{lm}}{\sigma_l \sigma_m} \right) \sigma^{*2} - \left(\sum_{l=1}^k \frac{\alpha_l [\sigma_{i_1 l} - \sigma_{i_2 l}]}{\sigma_l} \right)^2 \right]^{-1/2} e \\ \sigma^{*2} &= \sigma_{i_1}^2 + \sigma_{i_2}^2 - 2\sigma_{i_1 i_2}, \text{ o que conclui a demonstração.} \end{aligned}$$

\square

Estruturas mais simples para a matriz $\mathbf{\Omega}$ podem ser assumidas e assim modelos mais parcimoniosos são construídos. Também, dependendo da estrutura assumida para $\mathbf{\Omega}$, preserva-se a interpretação da razão de chances obtida sob normalidade. Esta invariância ocorre quando no Corolário 2.3 a distribuição de W_{12} é Normal, ou seja, para $\lambda_{12}^* = 0$.

Como discutido em Santos (2011), considere por exemplo que $\mathbf{\Omega} = \text{diag}(\sigma_1^2, \dots, \sigma_k^2)$ e $\sigma_i^2 \in \mathbb{R}_+$ em (2.6). Para algum vetor $\mathbf{c} \in \mathbb{R}^k$ que é ortogonal ao vetor $\mathbf{\Omega}\boldsymbol{\omega}^{-1}\boldsymbol{\alpha}$, digo, tal que $\sum_{i=1}^k \alpha_i \sigma_i c_i = 0$, temos que

$$\mathbf{c}^t \boldsymbol{\gamma} \sim N \left(\sum_{i=1}^k \xi_i c_i, \sum_{i=1}^k c_i^2 \sigma_i^2 \right), \quad (2.8)$$

que é equivalente ao resultado obtido sob a suposição de efeitos aleatórios independentes e normalmente distribuídos.

Uma escolha naturalmente feita é considerarmos $\sigma_i^2 = \sigma^2$ para todo i . Se $\sum_{i=1}^k \alpha_i c_i \neq 0$ e $\mathbf{\Omega} = \sigma^2 \mathbf{I}_k$, a distribuição para $\mathbf{c}^t \boldsymbol{\gamma}$ é

$$\mathbf{c}^t \boldsymbol{\gamma} \sim SN \left(\sum_{i=1}^k \xi_i c_i, \sigma^2 \sum_{i=1}^k c_i^2, \lambda^* \right),$$

em que $\lambda^* = [\sum_{i=1}^k \alpha_i c_i] [(1 + \sum_{i=1}^k \alpha_i^2) (\sum_{i=1}^k c_i^2) - (\sum_{i=1}^k \alpha_i c_i)^2]^{-1/2}$. Sob esta particular estrutura para $\mathbf{\Omega}$, o Corolário 2.4 apresenta dois resultados interessantes relacionados a distribuição de W_{12} , que foram obtidos em Santos (2011).

COROLÁRIO 2.4.

(i) Se $\boldsymbol{\gamma} \sim SN_k(\boldsymbol{\xi}, \sigma^2 \mathbf{I}_k, \boldsymbol{\alpha})$ então

$$W_{12} \sim SN \left(\xi_{i_1} - \xi_{i_2}, 2\sigma^2, [\alpha_{i_1} - \alpha_{i_2}] \left[2 \left(1 + \sum_{l=1}^k \alpha_l^2 \right) - (\alpha_{i_1} - \alpha_{i_2})^2 \right]^{-1/2} \right);$$

(ii) Se $\boldsymbol{\gamma} \sim SN_k(\boldsymbol{\xi}, \sigma^2 \mathbf{I}_k, \alpha \mathbf{1}_k)$ então $W_{12} \sim N(\xi_{i_1} - \xi_{i_2}, 2\sigma^2)$.

DEMONSTRAÇÃO.

(i) A demonstração segue diretamente da Proposição 2.3 ao considerarmos $\mathbf{\Omega} = \sigma^2 \mathbf{I}_k$ e o vetor \mathbf{c} assumindo os valores 1 na posição i_1 , -1 na posição i_2 e zero nas demais posições. Neste caso, basta notarmos que $\bar{\mathbf{\Omega}} = \mathbf{I}_k$. Resolvendo as operações matriciais temos que

$$\begin{aligned} \mathbf{c}^t \mathbf{\Omega} \mathbf{c} &= 2\sigma^2; \\ \boldsymbol{\xi}^t \mathbf{c} &= \xi_{i_1} - \xi_{i_2}; \\ \boldsymbol{\alpha}^t \bar{\mathbf{\Omega}} \boldsymbol{\omega} \mathbf{c} &= \boldsymbol{\alpha}^t \mathbf{I}_k \sigma \mathbf{I}_k \mathbf{c} = \sigma(\alpha_{i_1} - \alpha_{i_2}); \\ \boldsymbol{\alpha}^t \bar{\mathbf{\Omega}} \boldsymbol{\alpha} &= \boldsymbol{\alpha}^t \mathbf{I}_k \boldsymbol{\alpha} = \sum_{l=1}^k \alpha_l^2. \end{aligned}$$

Então $W_{12} \sim SN(\xi_{i_1} - \xi_{i_2}, 2\sigma^2, \lambda_{12}^*)$, em que

$$\lambda_{12}^* = \frac{\sigma(\alpha_{i_1} - \alpha_{i_2})}{\left[(1 + \sum_{l=1}^k \alpha_l^2) (2\sigma^2) - (\sigma(\alpha_{i_1} - \alpha_{i_2}))^2 \right]^{1/2}} = \frac{\alpha_{i_1} - \alpha_{i_2}}{\left[2(1 + \sum_{l=1}^k \alpha_l^2) - (\alpha_{i_1} - \alpha_{i_2})^2 \right]^{1/2}}.$$

- (ii) A demonstração segue diretamente de (i). Note que se $\alpha_{i_1} = \alpha_{i_2} = \alpha$ então $\lambda_{12}^* = 0$. Logo, temos que $W_{12} \sim N(\xi_{i_1} - \xi_{i_2}, 2\sigma^2)$, concluindo a demonstração. □

Então os resultados estabelecidos em Larsen *et al.* (2000) para OR_{12} e MOR_{12} também seguem se $\gamma \sim SN_k(\boldsymbol{\xi}, \sigma^2 \mathbf{I}_k, \alpha \mathbf{1}_k)$ e sob aquelas condições que levam ao resultado em (2.8). Em todos estes casos, a OR_{12} tem a mesma distribuição Log-Normal obtida em (1.5). Na Proposição 2.4, fornecemos a distribuição da OR_{12} e sua mediana para o caso geral.

PROPOSIÇÃO 2.4. *Se os efeitos aleatórios são tais que $\gamma \sim SN_k(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ com f.d.p dada em (2.6) então, dados $\boldsymbol{\beta}, \mathbf{X}$ e $\boldsymbol{\theta} = (\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$, segue que:*

- (i) a variável aleatória $OR_{12} = \exp\{\kappa_{12} + W_{12}\}$ tem f.d.p

$$f_{OR_{12}|\boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}}(r) = \frac{2}{r} \phi(\ln r \mid \kappa_{12} + \xi_{i_1} - \xi_{i_2}, \sigma^{*2}) \times \Phi\left(\frac{\lambda_{12}^* [\ln r - (\kappa_{12} + \xi_{i_1} - \xi_{i_2})]}{\sigma^*}\right), \quad (2.9)$$

$r \in \mathbb{R}_+$, em que $\sigma^{*2} = \sigma_{i_1}^2 + \sigma_{i_2}^2 - 2\sigma_{i_1 i_2}$ e λ_{12}^* é definido como no Corolário 2.3, e

- (ii) a mediana da distribuição em (2.9) é

$$MOR_{12} = \exp\{\kappa_{12}\} \exp\{\Phi_{SN}^{-1}(0,5 \mid \lambda_{12}^*) \sigma^* + \xi_{i_1} - \xi_{i_2}\},$$

em que $\Phi_{SN}(\cdot \mid \lambda_{12}^*)$ denota a f.d.a da distribuição SN padrão com parâmetro de assimetria λ_{12}^* dado no Corolário 2.3.

- (iii) se $\xi_{i_1} - \xi_{i_2} = 0$ segue que

- (1) $MOR_{|12|} = \text{med}\{\exp\{|W_{12}|\} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}\} = \exp\{\text{med}\{|W_{12}| \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}\}\} = \exp\{\sigma^* \Phi^{-1}(0,75)\};$
- (2) $\text{med}\{\exp\{-|W_{12}|\} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}\} = [\text{med}\{\exp\{|W_{12}|\} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}\}]^{-1}.$

DEMONSTRAÇÃO.

- (i) Utilizando a Proposição 2.3 temos que $W_{12} \sim SN(\xi^*, \sigma^{*2}, \lambda_{12}^*)$, em que $\xi^* = \xi_{i_1} - \xi_{i_2}$, e σ^{*2} e λ_{12}^* são como no Corolário 2.3. Defina a variável aleatória $T_{12} = \kappa_{12} + W_{12}$. Assim, segue que $T_{12} \sim SN(\xi^* + \kappa_{12}, \sigma^{*2}, \lambda_{12}^*)$. Consequentemente, a distribuição de $OR_{12} = \exp\{T_{12}\}$ é dada por

$$\begin{aligned} f_{OR_{12}|\boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}}(r) &= f_{T_{12}|\boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}}(\ln r) \left| \frac{dT}{dOR} \right| \\ &= \frac{2}{r\sigma^*} \phi(\ln r \mid \xi^* + \kappa_{12}, \sigma^{*2}) \Phi\left(\lambda_{12}^* \frac{\ln r - (\xi^* + \kappa_{12})}{\sigma^*}\right); \quad r > 0. \end{aligned}$$

(ii) Seja $a = \text{med}\{\text{OR}_{12}\}$. Logo, temos que

$$\begin{aligned} F_{\text{OR}_{12}|\beta, \mathbf{X}, \boldsymbol{\theta}}(a) &= \int_0^a f_{\text{OR}_{12}}(r) dr = 0,5; \\ &= \int_0^a \frac{2}{r\sigma^*} \phi\left(\frac{\ln r - (\xi^* + \kappa_{12})}{\sigma^*}\right) \Phi\left(\lambda_{12}^* \frac{\ln r - (\xi^* + \kappa_{12})}{\sigma^*}\right) dr = 0,5. \end{aligned}$$

Usando a transformação $z = \frac{\ln r - (\xi^* + \kappa_{12})}{\sigma^*}$, segue que

$$\begin{aligned} F_{\text{OR}_{12}|\beta, \mathbf{X}, \boldsymbol{\theta}}(a) &= \int_{-\infty}^{\frac{\ln a - (\xi^* + \kappa_{12})}{\sigma^*}} 2\phi(z)\Phi(\lambda_{12}^* z) dz = 0,5; \\ &= \Phi_{SN}\left(\frac{\ln a - (\xi^* + \kappa_{12})}{\sigma^*} \mid \lambda_{12}^*\right) = 0,5, \end{aligned}$$

em que $\Phi_{SN}(\cdot \mid \lambda_{12}^*)$ denota a f.d.a da distribuição SN padrão com parâmetro de assimetria λ_{12}^* dado no Corolário 2.3. Disto, segue que

$$\begin{aligned} \frac{\ln a - (\xi^* + \kappa_{12})}{\sigma^*} &= \Phi_{SN}^{-1}(0,5 \mid \lambda_{12}^*); \\ a &= \exp\{\Phi_{SN}^{-1}(0,5 \mid \lambda_{12}^*)\sigma^* + \xi^* + \kappa_{12}\}. \end{aligned}$$

Portanto, temos que $\text{med}\{\text{OR}_{12}\} = \exp\{\kappa_{12}\} \exp\{\Phi_{SN}^{-1}(0,5 \mid \lambda_{12}^*)\sigma^* + \xi^*\}$.

(iii) Considerando $\xi_{i_1} - \xi_{i_2} = 0$, segue que $W_{12} \sim SN(0, \sigma^*, \lambda_{12}^*)$, onde σ^{*2} e λ_{12}^* são definidos como em (i).

(1) Inicialmente vamos encontrar o valor de $\text{med}\{\exp\{|W_{12}|\} \mid \beta, \mathbf{X}, \boldsymbol{\theta}\}$. Para isso define $\text{OR}_{12}^* = \exp\{|W_{12}|\}$. Desta forma, temos que a f.d.a de OR_{12}^* é

$$\begin{aligned} F_{\text{OR}_{12}^*|\beta, \mathbf{X}, \boldsymbol{\theta}}(r) &= P(|W_{12}| \leq \ln r \mid \beta, \mathbf{X}, \boldsymbol{\theta}) \\ &= F_{W_{12}|\beta, \mathbf{X}, \boldsymbol{\theta}}(\ln r) - F_{W_{12}|\beta, \mathbf{X}, \boldsymbol{\theta}}(-\ln r). \end{aligned}$$

Como consequência temos que a f.d.p de OR_{12}^* é

$$\begin{aligned} f_{\text{OR}_{12}^*|\beta, \mathbf{X}, \boldsymbol{\theta}}(r) &= \frac{1}{r} [f_{W_{12}|\beta, \mathbf{X}, \boldsymbol{\theta}}(\ln r) + f_{W_{12}|\beta, \mathbf{X}, \boldsymbol{\theta}}(-\ln r)] \\ &= \frac{2}{r\sigma^*} \left[\phi\left(\frac{\ln r}{\sigma^*}\right) \Phi\left(\frac{\lambda_{12}^* \ln r}{\sigma^*}\right) + \phi\left(\frac{-\ln r}{\sigma^*}\right) \Phi\left(\frac{-\lambda_{12}^* \ln r}{\sigma^*}\right) \right] \\ &= \frac{2}{r\sigma^*} \phi\left(\frac{\ln r}{\sigma^*}\right) \\ &= \frac{2}{r\sigma^*} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \left(\frac{\ln r}{\sigma^*}\right)^2\right\}; \quad 1 \leq r < \infty. \end{aligned}$$

Sendo assim, se $a = \text{med}\{\text{OR}_{12}^* \mid \beta, \mathbf{X}, \boldsymbol{\theta}\}$, temos que

$$\int_1^a \frac{1}{r\sigma^*} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \left(\frac{\ln r}{\sigma^*}\right)^2\right\} dr = 0,25. \quad (2.10)$$

Usando a transformação $z = \ln r/\sigma^*$, segue de (2.10) que

$$\begin{aligned} \int_0^{\frac{\ln a}{\sigma^*}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\} dz &= 0,25; \\ \Phi\left(\frac{\ln a}{\sigma^*}\right) - \Phi(0) &= 0,25; \\ \Phi\left(\frac{\ln a}{\sigma^*}\right) &= 0,75, \end{aligned}$$

de onde decorre que $a = \text{med}\{\text{OR}_{12}^* \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}\} = \exp\{\sigma^*\Phi^{-1}(0,75)\}$.

A outra parte da igualdade em (1) segue ao usarmos a Proposição A.2. Desta forma,

$$\text{med}\{\exp\{|W_{12}|\} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}\} = \exp\{\text{med}\{|W_{12}| \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}\}\} = \exp\{\sigma^*\Phi^{-1}(0,75)\}.$$

(2) Seja $Y = \exp\{-|W_{12}|\}$. Assim, a f.d.a de Y é

$$\begin{aligned} F_{Y|\boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}}(y) &= P(\exp\{-|W_{12}|\} \leq y \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}) \\ &= 1 - F_{W_{12}|\boldsymbol{\theta}}(-\ln y) + F_{W_{12}|\boldsymbol{\theta}}(\ln y) \end{aligned}$$

Assim, temos que a f.d.p de Y é

$$\begin{aligned} f_{Y|\boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}}(y) &= \frac{1}{y} [f_{W_{12}|\boldsymbol{\theta}}(-\ln y) + f_{W_{12}|\boldsymbol{\theta}}(\ln y)] \\ &= \frac{2}{y\sigma^*} \left[\phi\left(\frac{-\ln y}{\sigma^*}\right) \Phi\left(\frac{-\lambda_{12}^* \ln y}{\sigma^*}\right) + \phi\left(\frac{\ln y}{\sigma^*}\right) \Phi\left(\frac{\lambda_{12}^* \ln y}{\sigma^*}\right) \right] \\ &= \frac{2}{y\sigma^*} \phi\left(\frac{\ln y}{\sigma^*}\right), \quad 0 < y \leq 1. \end{aligned}$$

Desta forma, se $c = \text{med}\{\exp\{-|W_{12}|\} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}\}$, temos que

$$\int_0^c \frac{2}{\sigma^* y} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\ln y}{\sigma^*}\right)^2\right\} dy = 0,5; \quad (2.11)$$

Usando a transformação $t = \ln y/\sigma^*$ em (2.11), segue que

$$\begin{aligned} \int_{-\infty}^{\frac{\ln c}{\sigma^*}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}t^2\right\} dt &= 0,25; \\ \Phi\left(\frac{\ln c}{\sigma^*}\right) &= 0,25. \end{aligned}$$

Consequentemente, temos que

$$\begin{aligned} c &= \exp\{\sigma^* \Phi^{-1}(0,25)\}; \\ &= \exp\{-\sigma^* \Phi^{-1}(0,75)\}; \\ &= [\text{med}\{\exp\{|W_{12}|\} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}\}]^{-1}, \end{aligned}$$

o que conclui a prova. □

O Corolário 2.5 que apresentamos a seguir mostra o resultado quando assumimos um caso mais parcimonioso, isto é, quando assumimos que o efeitos aleatórios são tais que sua distribuição tem matriz de escala $\sigma^2 \mathbf{I}_k$.

COROLÁRIO 2.5. *Se $\boldsymbol{\gamma} \sim SN_k(\boldsymbol{\xi}, \sigma^2 \mathbf{I}_k, \boldsymbol{\alpha})$, em que $\boldsymbol{\xi} = (\xi_1, \dots, \xi_k)^t$ e $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k)^t$ então, dados $\boldsymbol{\beta}$, σ^2 e $\boldsymbol{\alpha}$, segue que*

(i) a variável aleatória $OR_{12} = \exp\{\kappa_{12} + W_{12}\}$ tem f.d.p

$$\begin{aligned} f_{OR_{12}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma^2, \mathbf{X}}(r) &= \frac{\sqrt{2}}{\sigma r} \phi\left(\frac{1}{\sqrt{2}\sigma} [\ln r - (\kappa_{12} + \xi_{i_1} - \xi_{i_2})]\right) \\ &\quad \Phi\left(\frac{\lambda_{12}^*}{\sqrt{2}\sigma} [\ln r - (\kappa_{12} + \xi_{i_1} - \xi_{i_2})]\right), \quad r \in \mathbb{R}_+, \end{aligned} \quad (2.12)$$

(ii) a mediana da distribuição em (2.12) é

$$MOR_{12} = \exp\{\kappa_{12}\} \exp\{\Phi_{SN}^{-1}(0,5 \mid \lambda_{12}^*) \sqrt{2}\sigma + \xi_{i_1} - \xi_{i_2}\}, \quad (2.13)$$

em que $\lambda_{12}^* = [\alpha_{i_1} - \alpha_{i_2}][2(1 + \sum_{i=1}^k \alpha_i^2) - (\alpha_{i_1} - \alpha_{i_2})^2]^{-1/2}$.

DEMONSTRAÇÃO.

(i) Segue da Proposição 2.4 que a variável aleatória OR_{12} tem distribuição dada em (2.9). Ao assumirmos $\boldsymbol{\Omega} = \sigma^2 \mathbf{I}_k$ em (2.9) temos que

$$\begin{aligned} \sigma^{*2} &= 2\sigma^2 \\ \lambda_{12}^* &= [\alpha_{i_1} - \alpha_{i_2}][2(1 + \sum_{i=1}^k \alpha_i^2) - (\alpha_{i_1} - \alpha_{i_2})^2]^{-1/2}, \end{aligned} \quad (2.14)$$

o que fornece a distribuição em (2.12).

(ii) A demonstração deste resultado segue diretamente de (ii) na Proposição 2.4 ao considerar σ^{*2} e λ_{12}^* dados em (2.14). □

Os resultados obtidos em Santos (2011) são casos particulares do Corolário 2.5, onde foram somente mostrados resultados nos quais as suposições feitas sobre a distribuição dos efeitos aleatórios conduzem a uma distribuição Normal para W_{12} .

A Figura 2 mostra a f.d.p de W_{12} e OR_{12} quando comparamos os dois indivíduos com mesmos valores para as covariáveis ($\kappa_{12} = 0$) e quando $\xi_{i_1} = \xi_{i_2}$, $\sigma^2 = 1$ e $\alpha_{i_2} = 0$ e -5 . Diferentes valores de α_{i_1} são considerados.

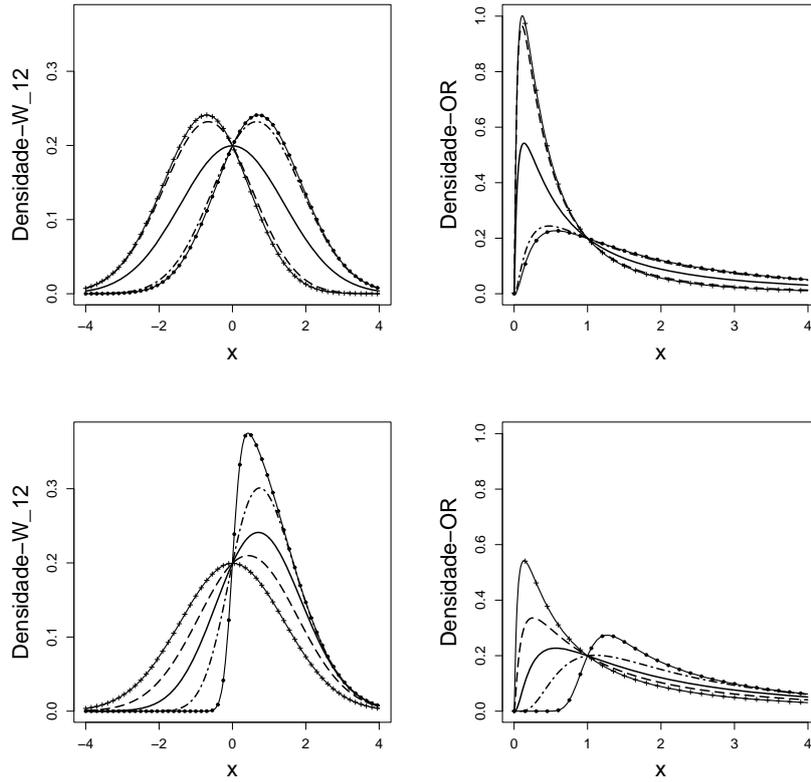


Figura 2: Distribuições *a priori* de W_{12} (esquerda) e OR (direita) para o caso dependente, $\alpha_{i_2} = 0$ (topo) e -5 (em baixo), $\sigma^2 = 1$, $\kappa_{12} = 0$ e $\alpha_{i_1} = 0$ (linha sólida), -5 (+), -2 (linha tracejada), 2 (linha tracejada com pontos), 5 (\bullet).

Se a opinião prévia de uma especialista revela que OR_{12} é, provavelmente, maior do que um, da Figura 2 percebemos que a distribuição Normal ($\alpha_{i_1} = \alpha_{i_2} = 0$) pode não ser uma escolha apropriada para descrever a incerteza sobre os efeitos aleatórios. Neste caso, a distribuição multivariada SN com $\alpha_{i_1} = 5$ e $\alpha_{i_2} = -5$ pode produzir uma melhor descrição da incerteza dele.

As distribuições em (2.9) e (2.12) pertencem à classe de distribuições Log-Skew-Elíptica introduzida por Marchenko e Genton (2010). Como para o caso independente, a MOR_{12} depende dos efeitos fixos através do termo $\exp\{\kappa_{12}\}$, termo este que é a razão de chances quando os efeitos aleatórios não são incluídos no modelo. Contudo, em ambos os casos, $\exp\{\kappa_{12}\}$ é multiplicado por um termo que depende da f.d.a. de alguma distribuição SN. Se, para o caso em (2.13), o parâmetro de assimetria do *cluster* i_2 é $\alpha_{i_2} = 0$ e para o *cluster* i_1 temos um forte grau de assimetria ($\alpha_{i_1} \rightarrow \infty$), então MOR_{12} é maior do que a observada para os efeitos aleatórios Normais. Por outro lado, se $\alpha_{i_1} = 0$ e $\alpha_{i_2} \rightarrow \infty$, então a MOR_{12} é menor que a observada para o caso Normal. A Figura 3 mostra o comportamento de MOR_{12} versus α_{i_1} para $\kappa_{12} = 0$ e $\sigma = 1$ assumindo diferentes valores para o parâmetro de assimetria no *cluster* i_2 . Vemos que ao alterar os valores de α_{i_1} e α_{i_2} diversos comportamentos para MOR_{12} são obtidos, sendo eles muito diferentes do que seria obtido sob normalidade em que teríamos $MOR_{12} = 1$ para o caso exemplificado. Por simplicidade, nas Figuras 2 e 3 assumimos $\alpha_{i_j} = 0$ para todo j exceto para $j = 1, 2$.

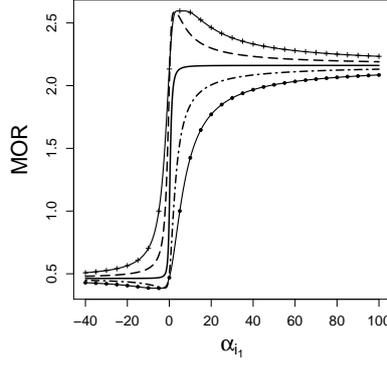


Figura 3: MOR_{12} versus α_{i_1} , para $\alpha_{i_2} = -5$ (+), $\alpha_{i_2} = -2$ (linha tracejada), $\alpha_{i_2} = 0$ (linha s3lida), $\alpha_{i_2} = 2$ (linha tracejada com pontos), $\alpha_{i_2} = 5$ (•).

2.3 Infer4ncia bayesiana no modelo log3stico misto *Skew-Normal*

Do ponto de vista bayesiano, a infer4ncia para os modelos mistos 3 mais simples dado que os efeitos aleat3rios $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_k)^t$, $\boldsymbol{\gamma} \in \mathbb{R}^k$, s3o consideradas como quantidades desconhecidas a serem estimadas. Assuma a verossimilhança em (1.1). Para completar a especificaça3o do modelo, devemos eliciar a distribuiça3o *a priori* para os par3metros $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$. Consideraremos dois modelos hier3rquicos diferentes. Primeiramente, assumimos que os efeitos aleat3rios s3o i.i.d com distribuiça3o *Skew-Normal* univariada, isto 3, $\gamma_i \stackrel{iid}{\sim} SN(-\delta\sigma\sqrt{2/\pi}, \sigma^2, \lambda)$, dados σ^2 e λ , em que $\delta = \lambda[1 + \lambda^2]^{-1/2}$. Tamb3m consideraremos efeitos aleat3rios correlacionados tal que, dados σ^2 e α , $\boldsymbol{\gamma} \sim SN_k(-\Delta\sigma\sqrt{2/\pi}\mathbf{1}_k, \sigma^2\mathbf{I}_k, \alpha\mathbf{1}_k)$ em que $\Delta = \alpha[1 + k\alpha^2]^{-1/2}$. Em ambos os modelos, centramos as distribuiça33es SN em zero para evitar problemas de falta de identificabilidade do modelo, isto 3, assumimos $E(\boldsymbol{\gamma}) = 0$. Para os efeitos fixos, assumimos a distribuiça33o *a priori* $\boldsymbol{\beta} \sim N_{p+1}(\mathbf{m}, b^2\mathbf{I}_{p+1})$, e para os hiperpar3metros consideramos $\sigma^2 \sim GI(a, d)$, $\lambda \sim SN(h, \tau^2, \theta)$ e $\alpha \sim N(h, \tau^2)$, em que $\mathbf{m} \in \mathbb{R}^{p+1}$, $h, \theta \in \mathbb{R}$, e τ^2 , b^2 , a e d s3o n3meros reais positivos. Consequentemente, segue que $E(\sigma^2) = d(a - 1)^{-1}$ e $V(\sigma^2) = d[(a - 1)^2(a - 2)]^{-1}$. A estrat3gia de considerar uma distribuiça33o *a priori Skew-Normal* para λ j3 foi considerada, por exemplo, em Arellano-Valle *et al.* (2009) e Arellano-Valle *et al.* (2013).

Seja $\mathbf{D} = (\mathbf{y}, \mathbf{X})$. Como consequ4ncia das suposiça33es anteriores, para o caso i.i.d a distribuiça33o *a posteriori* conjunta para $\boldsymbol{\vartheta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \lambda)$ 3

$$\begin{aligned}
 f(\boldsymbol{\vartheta}|\mathbf{D}) &\propto \left[\prod_{i=1}^k \prod_{j=1}^{n_i} (\pi_{ij})^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}} \right] \left(\frac{1}{\sigma^2} \right)^{(d+2)/2} \left(\frac{1}{\tau^2} \right)^{1/2} \exp \left\{ -\frac{a}{2\sigma^2} \right\} \\
 &\times \exp \left\{ -\frac{(\boldsymbol{\beta} - \mathbf{m})^t(\boldsymbol{\beta} - \mathbf{m})}{2b^2} - \frac{(\boldsymbol{\gamma} + \delta\sigma\sqrt{2/\pi}\mathbf{1}_k)^t(\boldsymbol{\gamma} + \delta\sigma\sqrt{2/\pi}\mathbf{1}_k)}{2\sigma^2} \right\} \\
 &\times \exp \left\{ -\frac{(\lambda - h)^2}{2\tau^2} \right\} \Phi_k \left(\frac{\lambda(\boldsymbol{\gamma} + \mathbf{1}_k\delta\sigma\sqrt{2/\pi})}{\sigma} \right) \Phi \left(\frac{\theta(\lambda - h)}{\tau} \right). \tag{2.15}
 \end{aligned}$$

A distribuição *a posteriori* em (2.15) não possui forma fechada conhecida, assim devemos considerar um esquema MCMC para gerarmos amostras desta distribuição. Para obtermos as distribuições condicionais completas *a posteriori* (d.c.c) dos efeitos aleatórios, consideramos a representação estocástica de Henze (Henze, 1986) para a distribuição SN univariada. Este resultado estabelece que, se $T_i \sim SN(\lambda)$ então $T_i \stackrel{d}{=} \delta|U_i| + (1 - \delta^2)^{1/2}V_i$, em que U_i e V_i são variáveis aleatórias i.i.d com distribuição Normal padrão. Por simplicidade, seja $\psi_i = |U_i|$. Visto que $\gamma_i \sim SN(-\delta\sigma\sqrt{2/\pi}, \sigma^2, \lambda)$, dados σ^2 e λ , segue que $\gamma_i \stackrel{d}{=} \delta\sigma\left(\psi_i - \sqrt{2/\pi}\right) + \sigma(1 - \delta^2)^{1/2}V_i$. Similarmente, se $\lambda \sim SN(h, \tau^2, \theta)$ obtemos que, dados h , τ^2 e θ , $\lambda \stackrel{d}{=} h + \frac{\theta\tau}{\sqrt{1+\theta^2}}M_3 + \frac{\tau}{\sqrt{1+\theta^2}}M_2$, em que M_1 e M_2 são i.i.d tal que $M_1 \stackrel{d}{=} M_2 \sim N(0, 1)$ e $M_3 = |M_1|$.

Para simplificar a notação, definimos $\vartheta^* = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \lambda, \boldsymbol{\psi}, M_3)$ e $\vartheta_{-\boldsymbol{\gamma}}^*$ representa o vetor ϑ^* sem o componente $\boldsymbol{\gamma}$. Assim, as d.c.c *a posteriori* são

$$\begin{aligned}
f(\boldsymbol{\beta}|\mathbf{D}, \vartheta_{-\boldsymbol{\beta}}^*) &\propto \left[\prod_{i=1}^k \prod_{j=1}^{n_i} \left[\frac{\exp\{\eta_{ij}\}}{1 + \exp\{\eta_{ij}\}} \right]^{y_{ij}} \left[\frac{1}{1 + \exp\{\eta_{ij}\}} \right]^{1-y_{ij}} \right] \\
&\times \exp \left\{ -\frac{(\boldsymbol{\beta} - \mathbf{m})^t(\boldsymbol{\beta} - \mathbf{m})}{2b^2} \right\}, \\
f(\boldsymbol{\gamma}|\mathbf{D}, \vartheta_{-\boldsymbol{\gamma}}^*) &\propto \left[\prod_{i=1}^k \prod_{j=1}^{n_i} \left[\frac{\exp\{\eta_{ij}\}}{1 + \exp\{\eta_{ij}\}} \right]^{y_{ij}} \left[\frac{1}{1 + \exp\{\eta_{ij}\}} \right]^{1-y_{ij}} \right] \\
&\times \exp \left\{ -\frac{\sum_{i=1}^n \left(\gamma_i - \sigma\delta(\psi_i - \sqrt{2/\pi}) \right)^2}{2\sigma^2(1 - \delta^2)} \right\}, \\
f(\boldsymbol{\psi}|\mathbf{D}, \vartheta_{-\boldsymbol{\psi}}^*) &\propto \prod_{i=1}^k \exp \left\{ -\frac{\left(\psi_i - (\gamma_i + \sigma\delta\sqrt{2/\pi})\frac{\delta}{\sigma} \right)^2}{2(1 - \delta^2)} \right\} \mathbf{1}_{(0, \infty)}(\psi_i), \\
f(\sigma^2|\mathbf{D}, \vartheta_{-\sigma^2}^*) &\propto \left(\frac{1}{\sigma^2} \right)^{\frac{k}{2}+a+1} \exp \left\{ -\frac{\sum_{i=1}^n \left(\gamma_i - \sigma\delta(\psi_i - \sqrt{2/\pi}) \right)^2}{2\sigma^2(1 - \delta^2)} - \frac{d}{\sigma^2} \right\}, \\
f(\lambda|\mathbf{D}, \vartheta_{-\lambda}^*) &\propto (1 + \lambda^2)^{\frac{k}{2}} \exp \left\{ -\frac{(1 + \lambda^2)}{2\sigma^2} \sum_{i=1}^n \left(\gamma_i - \frac{\lambda\sigma(\psi_i - \sqrt{2/\pi})}{\sqrt{1 + \lambda^2}} \right)^2 \right\}, \\
&\times \exp \left\{ -\frac{(1 + \theta^2) \left(\lambda - \left(\frac{\theta\tau M_3}{\sqrt{1+\theta^2}} + h \right) \right)^2}{2\tau^2} \right\}, \\
f(M_3|\mathbf{D}, \vartheta_{-M_3}^*) &\propto \exp \left\{ -\frac{(1 + \theta^2)}{2} \left(M_3 - \frac{\theta}{\sqrt{1 + \theta^2}} \frac{(\lambda - h)}{\tau} \right)^2 \right\} \mathbf{1}_{[0, \infty)}\{M_3\},
\end{aligned}$$

em que $\mathbf{1}_A\{x\}$ é a função indicadora assumindo 1 se $x \in A$. Note que a distribuição condicional completa *a posteriori* para cada componente ψ_i de $\boldsymbol{\psi}$ é uma distribuição Normal truncada abaixo em zero, com locação $(\gamma_i + \sigma\delta\sqrt{2/\pi})\delta\sigma^{-1}$ e escala $1 - \delta^2$.

Para o caso dependente, também temos que a distribuição *a posteriori* não tem forma fechada conhecida. As condicionais completas, neste caso, foram obtidas em Santos (2011) e podem ser encontradas no Apêndice C.

Tanto no caso i.i.d como no caso correlacionado, as d.c.c para $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ são log-côncavas e o algoritmo da amostragem adaptativa (ARS) pode ser usado para gerar das duas distribuições. O método da Rejeição e o algoritmo da amostragem adaptativa com passo de Metropolis (ARMS) também são usados para gerar das demais distribuições *a posteriori*. Para detalhes destes algoritmos veja Gilks *et al.* (1995) e Gilks e Wild (1992) por exemplo.

Observação: Se $\boldsymbol{\gamma} \sim SN_k(\mathbf{0}, \sigma^2\mathbf{I}_k, \alpha\mathbf{1}_k)$, então qualquer sub-vetor de $\boldsymbol{\gamma}$ de ordem $r \times 1$ tem a seguinte distribuição r -variada $SN_r(\mathbf{0}, \sigma^2\mathbf{I}_r, \alpha^*\mathbf{1}_r)$, em que $\alpha^* = \alpha\sigma^{-1}[1 + \alpha^2(k-r)]^{-1/2}$. Consequentemente, a distribuição marginal de $\gamma_i, \forall i = 1, \dots, k$, tende a uma distribuição Normal se existir um número grande de *clusters*. Podemos notar que, o grau de assimetria na distribuição marginal pertence ao intervalo $(-\sigma^{-1}[k-r]^{-1/2}; \sigma^{-1}[k-r]^{-1/2})$. Para quaisquer dois componentes γ_i e γ_j de $\boldsymbol{\gamma}$, segue também que $Corr(\gamma_i, \gamma_j) = -2(\alpha^*)^2[\pi + 2(\alpha^*)^2(\pi - 2)]^{-1}$, em que $\alpha^* = \alpha\sigma^{-1}[1 + \alpha^2(k-r)]^{-1/2}$. Veja detalhes em Azzalini e Dalla-Valle (1996).

2.3.1 Razão de chances *a posteriori*

A *OR* é uma quantidade aleatória que depende dos parâmetros $\boldsymbol{\beta}$ e dos efeitos aleatórios $\boldsymbol{\gamma}$. Devido a complexidade da distribuição *a posteriori* no MLREA, é inviável obtermos analiticamente a distribuição *a posteriori* de *OR*. Por outro lado, abordagens computacionais tais como métodos MCMC fornecem boas aproximações para as distribuições *a posteriori*. Assim, como foi descrito em Santos (2011), podemos obter amostras aproximada da distribuição *a posteriori* de OR_{12} como se segue:

Passo 1: Gerar uma amostra $(\boldsymbol{\beta}^l, \boldsymbol{\gamma}^l)$ da distribuição *a posteriori*;

Passo 2: Calcular $OR_{12}^l = \exp\{(\mathbf{x}_{i_1j_1} - \mathbf{x}_{i_2j_2})^t\boldsymbol{\beta}^l + \gamma_{i_1}^l - \gamma_{i_2}^l\}$.

Os passos 1 e 2 devem ser repetidos até obtermos a convergência para todos os parâmetros e amostras *a posteriori* de tamanhos razoáveis. A expressão no Passo 2 deve ser mudada convenientemente de acordo com o interesse específico do pesquisador. Uma vantagem desta abordagem é que resumos *a posteriori* de *OR*, como mediana, moda e média, assim como intervalos de credibilidade HPD, são facilmente obtidos (ver Martins-Filho *et al.* (2010)). A razão de chances modal *a posteriori* é um estimador atraente para *OR*, uma vez que está na região de mais alta densidade *a posteriori*. Informalmente falando, podemos concluir, quando comparamos dois indivíduos de *clusters* aleatoriamente selecionados, que a razão de chances pode ser, mais provavelmente, a razão de chances modal. A mediana e os intervalos de credibilidade HPD também tem interpretações naturais em termos de probabilidade e são similares as propostas por Larsen *et al.* (2000) para *MOR* e *IOR* (*interval odds ratio*), respectivamente. Contudo, usando a distribuição *a posteriori* de *OR* podemos concluir sobre a significância de valores específicos para *OR*

usando, por exemplo, os intervalos HPD. Neste caso, o principal interesse está em testarmos a hipótese nula $H_0 : OR_{12} = 1$, que pode ser aceita se 1 pertencer ao intervalo HPD. Realizar este tipo de teste traz um ganho na análise de dados visto que podemos levar em consideração apenas comparações que se mostrem significativas.

2.3.2 Métodos de seleção de modelos

Para comparação de modelos consideramos a ordenada preditiva condicional (CPO), duas abordagens para o DIC (*Deviance Information Criterion*) e o fator de Bayes (BF).

A CPO é uma ferramenta usual para seleção de modelo que nos permite avaliar a qualidade do ajuste, bem como para realizar comparações entre modelos. Para a observação i , é definida como sendo a distribuição preditiva de y_i dado as outras observações da amostra, isto é, $CPO_i = f(y_i | \mathbf{y}_{(-i)})$. Uma aproximação para a CPO_i é dada por

$$\widehat{CPO}_i = \left[\frac{1}{L} \sum_{l=1}^L \frac{1}{f(y_i | \theta^{(l)})} \right]^{-1}, \quad (2.16)$$

em que $\theta^{(l)}$, $l = 1, \dots, L$, é a amostra da distribuição *a posteriori* de θ . Apesar de fornecer boas aproximações este critério, proposto por Newton e Raftery (1994), pode apresentar instabilidades numéricas. Para detalhes sobre outras aproximações ver por exemplo, Dey *et al.* (1997); Gelfand (1996); Gelfand e Dey (1994).

A informação que o CPO_i traz sobre o ajuste do modelo pode ser resumida usando a estatística $B = n^{-1} \sum_{i=1}^n \log(CPO_i)$, em que n é o tamanho amostral. Valores altos para B indicam um bom ajuste. Para uma discussão mais detalhada sobre CPO indicamos a leitura de Dey *et al.* (1997) e suas referências. Além disso, por simplicidade, a estatística B será chamada de CPO no restante do texto.

O Critério DIC usualmente falha na seleção de modelos que incluem variáveis latentes. Algumas abordagens para calcular o DIC nestes casos podem ser encontrados em Celeux *et al.* (2006). Consideramos duas abordagens para calcular o DIC que são descritas abaixo.

Seja $D = -2 \log(f(\mathbf{y} | \boldsymbol{\theta}))$ a *deviance* e denote por \bar{D} a média *a posteriori* de D . Assuma que $\hat{D} = -2 \log(f(\mathbf{y} | \hat{\boldsymbol{\theta}}))$, em que $\hat{\boldsymbol{\theta}}$ é alguma estimativa *a posteriori* de $\boldsymbol{\theta}$. Seja $pD = \bar{D} - \hat{D}$. Disto o DIC é dado por

$$DIC = \bar{D} + pD = \hat{D} + 2pD = 2\bar{D} - \hat{D}. \quad (2.17)$$

Um valor pequeno para o DIC indica o melhor ajuste. Consideramos duas estimativas *a posteriori* de $\boldsymbol{\theta}$ e denotamos por DIC1 e DIC2 o DIC calculado assumindo que $\hat{\boldsymbol{\theta}}$ é, respectivamente, a média *a posteriori* da distribuição conjunta e as modas *a posteriori* das distribuições marginais.

O Fator de Bayes (BF) mede a evidência em favor do modelo M_1 se comparado a outro modelo M_2 . Um alto valor do fator de Bayes traz evidência em favor de M_1 . O BF é a razão entre as distribuições preditivas *a priori* obtidas para os modelos sob comparação, isto é, $BF = f_1(\mathbf{y})/f_2(\mathbf{y})$. As distribuições preditivas podem ser estimadas por

$$\widehat{f}_i(\mathbf{y}) = \left[\frac{1}{L} \sum_{l=1}^L \frac{1}{f(\mathbf{y} | \theta^{(l)})} \right]^{-1}, \quad (2.18)$$

onde $\theta^{(l)}$ é como definido anteriormente. A aproximação em (2.18), proposta por Newton e Raftery (1994), pode produzir estimativas instáveis dado que $f(\mathbf{y}|\theta^{(l)})$ pode assumir valores muito pequenos. Uma abordagem alternativa para obter a distribuição preditiva *a priori* pode ser encontrada em Gelfand e Dey (1994).

Em nosso caso, $f(y|\boldsymbol{\theta})$ é a função de verossimilhança dada em (1.1) e $\boldsymbol{\theta}$ é o vetor em que os componentes são os efeitos fixos e aleatórios. Por simplicidade, nas Seções 2.4, 2.6 e 2.7 utilizamos apenas a primeira abordagem (DIC1) e então denotamos apenas por DIC.

2.4 Estudo de dados simulados: O efeito da má-especificação da distribuição dos efeitos aleatórios

Nesta seção, o principal objetivo é avaliarmos o impacto da má-especificação das distribuições dos efeitos aleatórios na razão de chances *a posteriori*. Para isso, foram gerados quatro conjuntos de dados partindo do modelo de regressão logística misto. Consideramos 25 *clusters*, duas covariáveis e sua interação. Os conjuntos de dados (Dados 1, Dados 2, Dados 3 e Dados 4) são apresentados a seguir.

Tabela 2.1: Conjuntos de dados simulados

	X_1	X_2	n_i	Número de sucessos			
				Dados 1	Dados 2	Dados 3	Dados 4
1	-0,429	-1,519	110	6	0	15	43
2	-0,429	0,261	108	30	35	24	90
3	-1,194	0,261	107	8	89	12	5
4	-0,429	0,854	107	14	0	76	56
5	-1,194	0,854	103	17	32	37	68
6	0,031	-1,519	98	41	3	18	24
7	0,031	0,261	102	87	5	64	64
8	0,031	0,854	106	37	82	41	20
9	1,868	-1,519	203	195	201	194	148
10	1,868	0,854	175	137	151	152	160
11	-1,194	0,854	180	87	157	36	115
12	0,337	-1,519	156	122	43	67	131
13	0,337	0,261	152	27	131	132	79
14	-1,194	0,261	148	21	134	21	27
15	0,337	0,854	151	90	53	143	42
16	-1,194	0,854	148	63	88	66	54
17	1,868	0,261	110	92	100	94	91
18	0,031	-1,519	108	27	47	53	6
19	0,337	0,854	107	68	79	43	36
20	-0,429	0,261	103	68	23	60	6
21	-1,194	-1,519	98	3	1	0	1
22	-0,429	0,854	102	90	42	70	102
23	0,031	0,261	106	81	104	96	42
24	0,337	-1,519	203	142	69	51	21
25	1,868	0,854	175	144	83	156	171

A diferença entre os conjuntos de dados apresentados na Tabela 2.1 reside na distribuição usada para gerarmos os efeitos aleatórios. Para Dados 1 e Dados 2, os efeitos

aleatórios foram gerados de distribuições Normais i.i.d, centradas em zero e com variâncias 1 e 4, respectivamente. Para Dados 3 e Dados 4, assumimos distribuições *Skew-Normal* i.i.d que colocam maior massa em valores positivos, isto é, em que os efeitos aleatórios são positivos com probabilidade muito alta. Assumimos $\gamma_i \stackrel{iid}{\sim} SN(-1,32; 2,75; 35)$ e $\gamma_i \stackrel{iid}{\sim} SN(-2,64; 10,99; 35)$, respectivamente, para Dados 3 e 4. Visando uma comparação, ambas distribuições *Skew-Normais* têm média zero, enquanto a que variância é igual a 1 para Dados 3 e igual a 4 para Dados 4.

Na Figura 4 são apresentados alguns gráficos visando avaliar o comportamento dos efeitos aleatórios gerados em cada um dos bancos de dados.

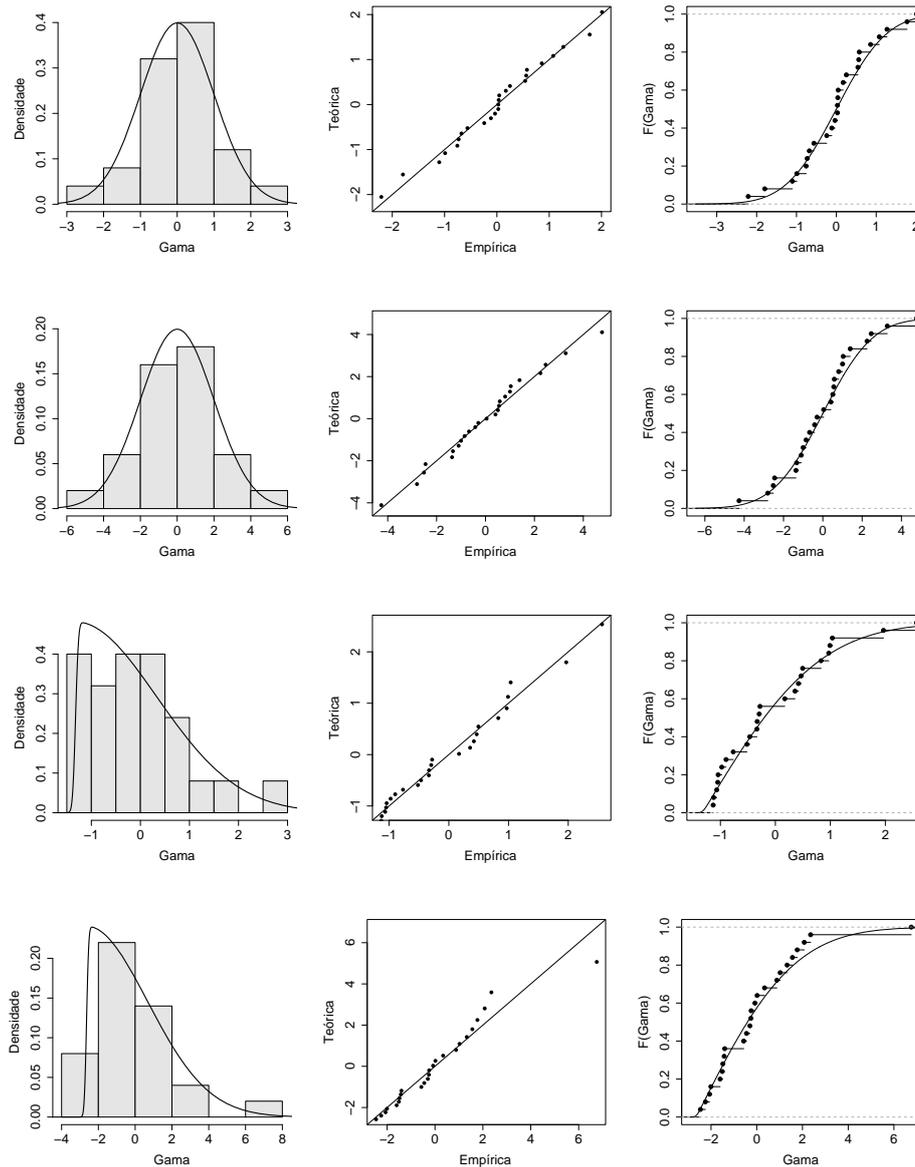


Figura 4: Histogramas e distribuição teórica (1° coluna), *QQ-plot* (2° coluna) e Funções de distribuição acumulada teórica e empírica (3° coluna) para os efeitos aleatórios gerados nos Dados 1 (1° linha), 2 (2° linha), 3 (3° linha) e 4 (4° linha).

Ao observarmos a Figura 4 notamos que, em cada um dos bancos de dados, os efeitos aleatórios gerados seguem corretamente o formato das distribuições teóricas propostas, mostrando que os efeitos aleatórios de fato tem as características de simetria (Dados 1 e 2) e assimetria (Dados 3 e 4) que desejávamos considerar neste estudo.

Em todos os bancos de dados gerados, ajustamos quatro modelos que diferem entre si por especificarmos diferentes distribuições *a priori* para os efeitos aleatórios. Por simplicidade, assumimos somente distribuições i.i.d para os efeitos aleatórios. No Modelo 1 consideramos que $\gamma_i \stackrel{iid}{\sim} N(0, \sigma^2)$. Nos Modelos 2, 3 e 4 temos que $\gamma_i \stackrel{iid}{\sim} SN(-\delta\sigma\sqrt{2/\pi}, \sigma^2, \lambda)$. Contudo, assumimos diferentes especificações *a priori* para o parâmetro de assimetria: $\lambda \sim N(0, 1000)$, no Modelo 2, $\lambda \sim SN(0; 2578, 21; 5)$, no Modelo 3, e no Modelo 4 consideramos $\lambda \sim SN(0; 2751, 46; 100)$. Em todos os casos a variância *a priori* para a distribuição de λ é 1000.

As distribuições *a priori* para λ nos Modelos 3 e 4 colocam maior parte de sua massa em valores positivos. A moda e a média *a priori* para λ no Modelo 3 (Modelo 4) são 18,81 (1,96) e 39,73 (41,85), respectivamente. A Figura 5 apresenta as densidades da distribuições *a priori* que assumimos para λ nos Modelos 2, 3 e 4.

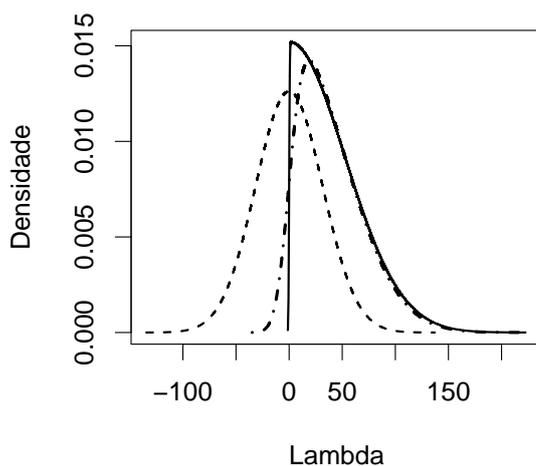


Figura 5: Distribuições *a priori* para λ nos Modelos 2 (linha tracejada), 3 (linha tracejada com pontos) e 4 (linha sólida).

Em nossa análise, também assumimos distribuições *a priori* pouco informativas para os efeitos fixos ($\beta_i \sim N(0, 10)$, $i = 0, \dots, 3$) e para σ^2 ($\sigma^2 \sim GI(2, 001; 1)$). A distribuição *a priori* para σ^2 foi escolhida de forma a garantir a existência da média e da variância *a priori*, as quais são, respectivamente, $E(\sigma^2) = 1$ e $V(\sigma^2) = 1000$.

Para o MCMC, geramos uma cadeia de tamanho 30.000 (50.000) para o caso Normal (*Skew-Normal*), descartamos as primeiras 20.000 amostras como o período de *burn-in* e usamos um *lag* de 10 (30) passos para diminuirmos a autocorrelação, obtendo uma amostra com tamanho final de 1.000. O algoritmo foi implementado utilizando o *software OxEdit* (Doornik, 2007).

Embora o DIC (*deviance information criterion*) e o CPO (*conditional predictive ordinate*) exibidos na Tabela 2.2 mostrem que o Modelo 1 é preferível para Dados 1 e Dados 2, e os Modelos 2 e 3 são os melhores para Dados 4 e Dados 3, respectivamente, estas medidas não fornecem evidências substanciais a favor de qualquer destes modelos, uma vez que todos os modelos fornecem valores similares. Além disto, observamos da Tabela 2.3 que os modelos também não fornecem estimativas pontuais substancialmente diferentes para os efeitos fixos e para $V(\gamma_i)$.

Por outro lado, os intervalos HPD *a posteriori* mostram ser uma boa ferramenta auxiliar na seleção do modelo. Para todos os parâmetros, os intervalos HPD de 95% *a posteriori* nos mostram que a incerteza *a posteriori* sobre os parâmetros é menor do que tínhamos *a priori*. No entanto, se a distribuição dos efeitos aleatórios não é bem especificada, as magnitudes dos intervalos HPD *a posteriori* para os efeitos fixos e para σ^2 tendem a ser altas.

O parâmetro de assimetria é em geral pobremente estimado, principalmente se existe grande assimetria nos efeitos aleatórios. Os intervalos HPD com probabilidade *a posteriori* de 95% revelam que a incerteza *a posteriori* sobre λ é ainda muito alta. Notamos também que, na maioria dos casos, λ não é significativamente diferente de zero, com exceção ocorrendo para Dados 3, sob os Modelos 3 e 4, e para Dados 4 sob o Modelo 4.

A Tabela 2.2 mostra o erro quadrático médio (EQM) associado às estimativas dos efeitos aleatórios em todos os casos, ao utilizarmos média, mediana e moda *a posteriori* para estimar os efeitos aleatórios.

Tabela 2.2: Comparação dos modelos

Modelo	EQM para γ			CPO	DIC
	Média	Mediana	Moda		
Dados 1					
1	3,05	2,95	2,98	-1631,84	3263,18
2	2,78	2,93	3,73	-1632,14	3263,91
3	3,61	3,85	4,86	-1632,17	3263,89
4	3,83	3,83	4,09	-1631,82	3263,21
Dados 2					
1	12,69	12,30	12,456	-1402,09	2803,93
2	20,13	19,66	19,12	-1402,46	2804,41
3	14,22	14,67	15,72	-1402,57	2804,86
4	13,35	13,26	14,43	-1402,77	2804,98
Dados 3					
1	5,03	5,12	5,30	-1568,09	3136,27
2	3,53	3,41	3,37	-1567,46	3135,01
3	3,17	3,01	3,12	-1566,81	3133,63
4	3,35	3,33	3,56	-1567,81	3135,63
Dados 4					
1	16,30	16,76	18,79	-1527,95	3056,05
2	12,01	11,40	11,30	-1526,32	3052,88
3	9,92	9,77	9,38	-1526,37	3053,15
4	10,26	10,68	11,74	-1526,61	3053,60

Notamos da Tabela 2.2, que o EQM mostra-nos que a má-especificação da distribuição dos efeitos aleatórios resulta em estimativas *a posteriori* pobres para os efeitos aleatórios (veja também a Figura 6). Este fato fica evidente, principalmente, para Dados 3 e 4 em que as estimativas sob o Modelo 1 têm o maior EQM. Consequentemente, pelo fato da *OR* depender de γ , obtemos estimativas *a posteriori* pobres também para a razão de chances (ver Tabela 2.5).

Tabela 2.3: Estimativas *a posteriori* para β , $V(\gamma)$ e λ , todos os modelos e conjuntos de dados

	Real	Média	Mediana	Moda	HPD		Real	Média	Mediana	Moda	HPD
Dados 1						Dados 2					
Modelo 1						Modelo 1					
β_0	-0,10	-0,15	-0,15	-0,17	[-0,52;0,35]	-0,10	-0,12	-0,12	-0,06	[-0,87;0,68]	
β_1	1,16	1,32	1,32	1,36	[0,86;1,73]	1,16	1,17	1,17	1,10	[0,26;2,00]	
β_2	0,48	0,32	0,32	0,34	[-0,14;0,76]	0,48	0,96	0,96	0,94	[-0,05;2,09]	
β_3	-0,78	-0,76	-0,77	-0,80	[-1,26;-0,32]	-0,78	-1,29	-1,27	-1,25	[-2,27;-0,49]	
$V(\gamma)$	1	1,11	1,04	0,95	[0,54;1,76]	4	4,31	4,01	3,60	[1,79;7,38]	
Modelo 2						Modelo 2					
β_0	-0,10	-0,17	-0,18	-0,19	[-0,58;0,29]	-0,10	-0,24	-0,23	-0,22	[-1,11;0,58]	
β_1	1,16	1,32	1,33	1,34	[0,84;1,83]	1,16	1,02	1,00	0,90	[0,15;1,82]	
β_2	0,48	0,36	0,36	0,39	[-0,10;0,82]	0,48	1,06	1,04	0,99	[0,14;2,11]	
β_3	-0,78	-0,74	-0,74	-0,77	[-1,31;-0,28]	-0,78	-1,34	-1,34	-1,17	[-2,38;-0,40]	
λ	0	0,48	0,33	0,19	[-52,41;39,83]	0	-9,95	-2,26	-0,70	[-43,10;3,08]	
$V(\gamma)$	1	1,22	1,15	1,06	[0,55;2,07]	4	4,87	4,50	3,62	[2,35;8,79]	
Modelo 3						Modelo 3					
β_0	-0,10	-0,22	-0,23	-0,22	[-0,72;0,24]	-0,10	-0,26	-0,27	-0,21	[-1,07;0,46]	
β_1	1,16	1,30	1,30	1,25	[0,84;1,71]	1,16	1,11	1,10	1,04	[0,33;2,15]	
β_2	0,48	0,28	0,26	0,19	[-0,08;0,73]	0,48	1,00	0,99	0,94	[0,18;1,86]	
β_3	-0,78	-0,70	-0,69	-0,67	[-1,23;-0,21]	-0,78	-1,28	-1,27	-1,17	[-2,15;-0,32]	
λ	0	16,97	4,03	1,12	[-7,25;74,12]	0	-0,53	-0,54	-0,06	[-12,80;3,35]	
$V(\gamma)$	1	1,23	1,14	1,01	[0,50;2,07]	4	4,59	4,28	3,75	[1,93;8,05]	
Modelo 4						Modelo 4					
β_0	-0,10	-0,20	-0,20	-0,19	[-0,68;0,22]	-0,10	-0,28	-0,28	-0,43	[-1,11;0,48]	
β_1	1,16	1,32	1,33	1,33	[0,84;1,78]	1,16	1,16	1,14	1,17	[0,36;2,01]	
β_2	0,48	0,28	0,27	0,24	[-0,12;0,79]	0,48	0,95	0,94	0,86	[0,10;1,78]	
β_3	-0,78	-0,69	-0,68	-0,68	[-1,13;-0,18]	-0,78	-1,39	-1,35	-1,28	[-2,42;-0,55]	
λ	0	3,37	1,70	0,84	[-0,94;14,21]	0	0,74	0,57	0,45	[-0,76;2,63]	
$V(\gamma)$	1	1,17	1,09	0,98	[0,52;1,99]	4	4,37	4,08	3,62	[2,08;7,34]	
Dados 3						Dados 4					
Modelo 1						Modelo 1					
β_0	-0,10	-0,04	-0,04	-0,04	[-0,42;0,37]	-0,10	-0,22	-0,24	-0,26	[-0,88;0,42]	
β_1	1,16	1,50	1,50	1,53	[1,10;1,94]	1,16	1,03	1,04	1,00	[0,34;1,79]	
β_2	0,48	0,82	0,82	0,86	[0,40;1,29]	0,48	0,92	0,92	0,96	[0,25;1,68]	
β_3	-0,78	-0,72	-0,71	-0,69	[-1,20;-0,22]	-0,78	-0,35	-0,38	-0,53	[-1,29;0,59]	
$V(\gamma)$	1	0,91	0,85	0,79	[0,39;1,55]	4	2,81	2,61	2,24	[1,34;4,95]	
Modelo 2						Modelo 2					
β_0	-0,10	-0,06	-0,07	-0,13	[-0,39;0,39]	-0,10	-0,25	-0,26	-0,25	[-0,94;0,42]	
β_1	1,16	1,45	1,44	1,38	[1,10;1,83]	1,16	1,23	1,24	1,12	[0,64;1,95]	
β_2	0,48	0,68	0,66	0,64	[0,32;1,07]	0,48	0,84	0,83	0,76	[0,29;1,48]	
β_3	-0,78	-0,69	-0,69	-0,69	[-1,05;-0,27]	-0,78	-0,33	-0,37	-0,40	[-0,98;0,50]	
λ	35	22,45	19,9	8,59	[-2,03;64,28]	35	24,63	20,81	5,78	[-1,15;64,28]	
$V(\gamma)$	1	0,83	0,78	0,70	[0,41;1,38]	4	2,90	2,71	2,37	[1,30;4,96]	
Modelo 3						Modelo 3					
β_0	-0,10	-0,05	-0,06	-0,10	[-0,35;0,36]	-0,10	-0,26	-0,27	-0,37	[-0,74;0,34]	
β_1	1,16	1,40	1,40	1,39	[1,10;1,72]	1,16	1,18	1,20	1,34	[0,56;1,84]	
β_2	0,48	0,64	0,63	0,60	[0,36;0,98]	0,48	0,82	0,79	0,73	[0,29;1,46]	
β_3	-0,78	-0,65	-0,66	-0,70	[-1,03;-0,28]	-0,78	-0,36	-0,35	-0,34	[-0,93;0,32]	
λ	35	48,63	42,63	21,22	[1,48;113,22]	35	43,08	38,14	23,00	[-0,51;102,06]	
$V(\gamma)$	1	0,84	0,79	0,74	[0,40;1,41]	4	2,84	2,73	2,80	[1,28;4,71]	
Modelo 4						Modelo 4					
β_0	-0,10	-0,05	-0,07	-0,08	[-0,42;0,29]	-0,10	-0,20	-0,23	-0,23	[-0,82;0,57]	
β_1	1,16	1,42	1,42	1,42	[1,10;1,77]	1,16	1,27	1,27	1,25	[0,52;1,94]	
β_2	0,48	0,68	0,66	0,66	[0,34;1,02]	0,48	0,82	0,80	0,77	[0,22;1,43]	
β_3	-0,78	-0,69	-0,69	-0,70	[-1,07;-0,28]	-0,78	-0,35	-0,36	-0,32	[-0,98;0,32]	
λ	35	12,62	12,96	20,33	[1,64;21,43]	35	15,25	16,86	21,39	[2,80;21,82]	
$V(\gamma)$	1	0,83	0,78	0,66	[0,41;1,50]	4	3,02	2,78	2,49	[1,28;5,10]	

A Figura 6 mostra a f.d.p dos efeitos aleatórios estimada usando os valores estimados dos parâmetros através da média (coluna 1), mediana (coluna 2) e moda (coluna 3) *a posteriori* para todos os modelos e conjuntos de dados utilizados. Os gráficos na linha l correspondem as estimativas para Dados l , $l = 1, \dots, 4$.

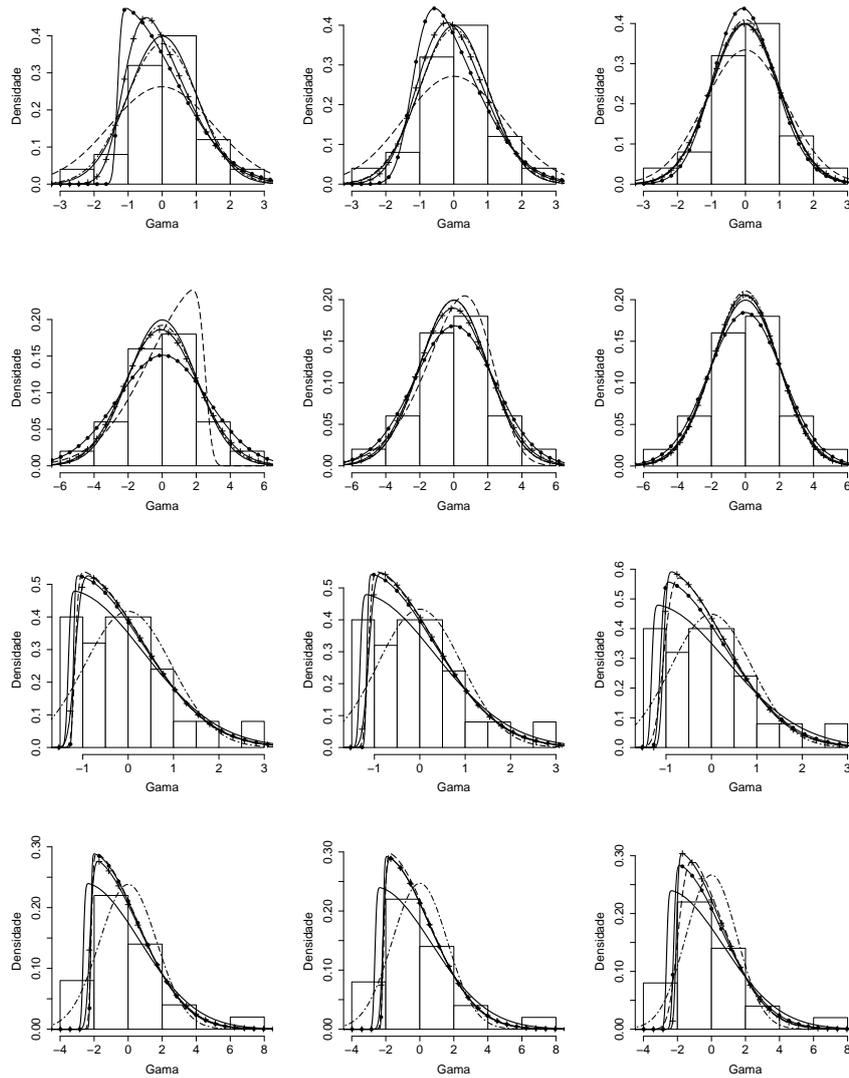


Figura 6: Histogramas dos efeitos aleatórios gerados, f.d.p real dos efeitos aleatórios (linha sólida) e f.d.p *plug-in* usando, *a posteriori*, média (1° coluna), mediana (2° coluna) e moda (3° coluna), sob Modelo 1 (linha tracejada com pontos), Modelo 2 (linha tracejada), Modelo 3 (●) e Modelo 4 (+), para Dados l , $l = 1, \dots, 4$ (linha l).

Através da Figura 6, podemos notar da coluna 3 que, usando a moda *a posteriori*, todos os modelos geram boas aproximações da densidade real para Dados 1 e Dados 2. Este resultado é esperado dado que a família de distribuições *Skew-Normal* inclui a distribuição Normal como um caso particular. Entretanto, isto não é necessariamente observado quando usamos como *plug-in* as estimativas obtidas a partir da média e da mediana *a posteriori*. Considerando a média *a posteriori*, por exemplo, os Modelos 3 e 4 para Dados 1 e o Modelo 2 para Dados 2 apontam grandes graus de assimetria para as distribuições dos efeitos aleatórios. Para Dados 3 e Dados 4, o Modelo 1 fornece estimativas ruins para a distribuição dos efeitos aleatórios, o que é esperado pois a distribuição

Normal não captura a assimetria existente. As estimativas *plug-in* obtidas sob os Modelos 2, 3 e 4 usando a média e a mediana, *a posteriori*, tendem a ser mais próximas da densidade real do que quando usamos a moda *a posteriori*.

A Tabela 2.5 apresenta alguns resumos *a posteriori* para razões de chances que comparam indivíduos em diferentes *clusters* com mesmas covariáveis (OR_1 , OR_2 e OR_3) e com diferentes covariáveis (OR_4 , OR_5 e OR_6). A descrição completa sobre as comparações realizadas está disponível na Tabela 2.4 a seguir.

Tabela 2.4: *Clusters* selecionados para cálculo da OR

OR	Dados 1		Dados 2		Descrição da comparação
	1° <i>Cluster</i>	2° <i>Cluster</i>	1° <i>Cluster</i>	2° <i>Cluster</i>	
1	22	4	2	20	diferentes <i>clusters</i> e mesmas covariáveis
2	16	5	11	5	diferentes <i>clusters</i> e mesmas covariáveis
3	6	18	19	15	diferentes <i>clusters</i> e mesmas covariáveis
4	9	17	4	1	diferentes <i>clusters</i> e covariáveis
5	12	21	23	2	diferentes <i>clusters</i> e covariáveis
6	10	24	8	21	diferentes <i>clusters</i> e covariáveis
OR	Dados 3		Dados 4		Descrição da comparação
	1° <i>Cluster</i>	2° <i>Cluster</i>	1° <i>Cluster</i>	2° <i>Cluster</i>	
1	18	6	5	11	diferentes <i>clusters</i> e mesmas covariáveis
2	12	24	2	20	diferentes <i>clusters</i> e mesmas covariáveis
3	25	10	22	4	diferentes <i>clusters</i> e mesmas covariáveis
4	23	2	17	9	diferentes <i>clusters</i> e covariáveis
5	4	1	20	3	diferentes <i>clusters</i> e covariáveis
6	10	24	14	24	diferentes <i>clusters</i> e covariáveis

Analisando a Tabela 2.5, notamos que as estimativas *a posteriori* para OR podem ser realmente diferentes de um modelo para o outro. A distribuição *a posteriori* para OR pode revelar forte assimetria, como é o caso da OR_6 em Dados 2, uma vez que resumos *a posteriori* desta distribuição são muito distantes um dos outros. Ao contrário do que foi sugerido em Larsen *et al.* (2000), a mediana *a posteriori* da distribuição de OR não necessariamente fornece a melhor estimativa *a posteriori* para OR . Notamos, por exemplo, que se existir grande variação entre os *clusters*, a média *a posteriori* tende a fornecer melhores estimativas para OR quando houver forte assimetria na distribuição dos efeitos aleatórios (Dados 4). Contudo, se existir simetria para a distribuição dos efeitos aleatórios (Dados 2), a média (mediana) *a posteriori* fornece melhores estimativas para indivíduos com mesmas (diferentes) covariáveis. Para conjuntos de dados onde há pequena variação entre os *clusters* (Dados 1 e 3), a mediana *a posteriori* não é o melhor estimador para OR . Nestes casos, as médias e modas *a posteriori* tendem a fornecer melhores estimativas. Se não existir assimetria entre os *clusters*, a moda *a posteriori* de OR tende a ser melhor se a comparação ocorre entre indivíduos com mesmas covariáveis. Nestes casos, se houver forte assimetria entre os *clusters* a média *a posteriori* revela-se melhor estimador de OR .

Tabela 2.5: Estimativas *a posteriori* para a razão de chances

	Real	Média	Mediana	Moda	HPD		Real	Média	Mediana	Moda	HPD
Dados 1											
Modelo 1											
OR_1	68,05	43,77	39,96	36,78	[15,06;79,95]		1,75	1,75	1,65	1,51	[0,91;3,00]
OR_2	2,91	3,69	3,53	3,35	[1,73;5,72]		14,41	15,94	15,20	13,10	[7,34;25,70]
OR_3	1,66	2,18	2,06	1,90	[1,13;3,60]		6,09	5,34	5,11	4,83	[2,91;8,46]
OR_4	5,69	5,97	5,22	4,26	[1,92;11,93]		1,60	9,61	2,51	0,66	[0;34,18]
OR_5	108,77	150,72	120,06	92,63	[34,77;338,70]		244,40	109,84	81,31	61,50	[19,73;282,15]
OR_6	1,58	1,63	1,57	1,49	[0,97;2,40]		538,73	1233,35	480,02	203,35	[54,56;4986,74]
Modelo 2											
OR_1	68,05	39,61	36,87	31,00	[15,67;69,72]		1,75	1,79	1,67	1,47	[0,78;2,88]
OR_2	2,91	3,79	3,58	3,34	[1,82;6,32]		14,41	16,16	15,36	13,39	[7,44;26,32]
OR_3	1,66	2,24	2,15	1,88	[1,00;3,53]		6,09	5,51	5,37	5,43	[2,92;8,41]
OR_4	5,69	5,70	5,08	4,43	[1,50;11,06]		1,60	136,09	2,04	0,35	[0,00;101,42]
OR_5	108,77	160,26	126,26	90,71	[28,89;386,59]		244,40	82,20	61,73	44,89	[14,12;200,95]
OR_6	1,58	1,60	1,55	1,45	[0,90;2,34]		538,73	1937,18	529,76	223,08	[27,57;5203,58]
Modelo 3											
OR_1	68,05	39,49	36,64	31,40	[14,96;69,86]		1,75	1,79	1,69	1,60	[0,87;3,04]
OR_2	2,91	3,74	3,55	3,37	[1,71;5,96]		14,41	15,82	15,01	13,48	[7,79;26,24]
OR_3	1,66	2,20	2,09	1,96	[1,08;3,68]		6,09	5,50	5,27	4,81	[2,69;8,45]
OR_4	5,69	5,80	5,17	4,40	[1,50;11,88]		1,60	21,43	2,49	0,56	[0,00;38,13]
OR_5	108,77	164,98	131,58	92,36	[21,66;395,93]		244,40	100,73	74,16	42,66	[15,25;266,41]
OR_6	1,58	1,61	1,55	1,45	[0,92;2,34]		538,73	1457,14	514,93	220,62	[40,66;4162,15]
Modelo 4											
OR_1	68,05	42,13	39,13	32,71	[17,05;75,78]		1,75	1,75	1,65	1,50	[0,79;2,88]
OR_2	2,91	3,73	3,57	3,39	[1,69;5,97]		14,41	15,89	15,09	13,24	[7,67;26,82]
OR_3	1,66	2,18	2,08	1,87	[1,07;3,65]		6,09	5,36	5,17	5,22	[2,96;8,50]
OR_4	5,69	5,92	5,14	4,27	[1,71;12,16]		1,60	24,91	3,03	0,79	[0,02;42,67]
OR_5	108,77	152,70	122,66	103,22	[24,92;358,69]		244,40	110,21	82,40	62,18	[15,17;265,10]
OR_6	1,58	1,61	1,56	1,40	[0,92;2,34]		538,73	1449,51	511,03	202,53	[39,81;4308,76]
Dados 3											
Modelo 1											
OR_1	4,03	4,29	4,05	3,68	[1,86;7,05]		1,98	1,13	1,09	1,00	[0,63;1,71]
OR_2	2,34	2,26	2,22	2,11	[1,28;3,24]		59,34	81,02	71,50	52,95	[24,99;166,57]
OR_3	1,85	1,31	1,23	1,09	[0,59;2,14]		1160,22	133,73	68,14	43,64	[9,58;439,28]
OR_4	53,48	28,57	26,29	21,42	[12,12;52,84]		2,10	1,88	1,77	1,72	[0,85;3,01]
OR_5	9,30	17,17	16,12	14,80	[7,18;30,10]		2,21	1,55	1,28	0,96	[0,16;3,51]
OR_6	15,77	21,12	20,28	19,75	[11,62;32,39]		1,82	1,96	1,85	1,66	[0,89;3,38]
Modelo 2											
OR_1	4,03	4,33	4,13	4,00	[1,98;7,24]		1,98	1,14	1,11	1,11	[0,64;1,78]
OR_2	2,34	2,20	2,16	2,00	[1,38;3,26]		59,34	64,75	58,72	52,84	[26,77;123,12]
OR_3	1,85	1,32	1,24	1,12	[0,60;2,22]		1160,22	331,00	115,55	55,39	[10,30;1194,05]
OR_4	53,48	30,93	28,10	25,10	[11,95;57,29]		2,10	1,87	1,79	1,70	[0,93;3,01]
OR_5	9,30	17,45	16,00	14,14	[6,32;29,29]		2,21	1,73	1,54	1,28	[0,52;3,37]
OR_6	15,77	19,63	18,76	17,20	[10,59;30,35]		1,82	1,97	1,91	1,89	[1,00;3,19]
Modelo 3											
OR_1	4,03	4,44	4,24	4,00	[2,21;7,19]		1,98	1,14	1,09	0,98	[0,61;1,72]
OR_2	2,34	2,17	2,13	2,11	[1,27;3,12]		59,34	60,24	54,93	45,56	[24,14;110,90]
OR_3	1,85	1,32	1,26	1,15	[0,65;2,18]		1160,22	1179,08	133,44	53,33	[13,74;3093,65]
OR_4	53,48	31,27	29,32	27,41	[13,73;57,79]		2,10	1,88	1,81	1,66	[0,86;2,98]
OR_5	9,30	17,48	16,49	14,18	[7,71;30,54]		2,21	1,79	1,65	1,50	[0,47;3,32]
OR_6	15,77	19,26	18,58	16,64	[10,31;29,25]		1,82	2,01	1,92	1,62	[0,96;3,30]
Modelo 4											
OR_1	4,03	4,31	4,14	3,58	[2,01;7,10]		1,98	1,11	1,09	1,08	[0,61;1,68]
OR_2	2,34	2,17	2,13	2,16	[1,34;3,20]		59,34	64,34	58,40	48,15	[22,01;117,00]
OR_3	1,85	1,32	1,24	1,07	[0,54;2,15]		1160,22	979,37	132,36	50,01	[13,28;3209,89]
OR_4	53,48	31,02	28,21	25,50	[12,98;59,37]		2,10	1,86	1,77	1,58	[0,89;3,03]
OR_5	9,30	17,36	16,23	13,36	[7,46;30,70]		2,21	1,74	1,55	1,32	[0,46;3,60]
OR_6	15,77	19,52	18,57	16,81	[10,66;30,74]		1,82	2,03	1,93	1,80	[0,96;3,46]

2.5 Estudo Monte Carlo: O efeito da distribuição *a priori* de λ

Na Seção 2.4 concluímos que usualmente o parâmetro de assimetria λ não é bem estimado no modelo de regressão logística com efeitos aleatórios *Skew-Normais*. Visto que a distribuição *a priori* dos efeitos aleatórios influencia a interpretação dos efeitos fixos, nesta seção, o nosso objetivo é considerarmos diferentes especificações *a priori* para o parâmetro de assimetria e avaliarmos se existe algum ganho em fixar o parâmetro de assimetria como é feito em Liu e Dey (2008).

Nesta seção, realizamos um estudo Monte Carlo em que consideramos 100 replicações do modelo logístico misto dado em (1.1), assumindo como efeitos fixos os valores $\beta = (\beta_0, \beta_1, \beta_2)^t = (-0,1; 1,2; 0,5)^t$. Consideramos 6 cenários que se diferenciam no número de *clusters* ($k = 25$ para os Cenários 1a, 2a e 3a e $k = 100$ para os Cenários 1b, 2b e 3b) e na distribuição usada para gerar os efeitos aleatórios. Em todos os cenários são gerados efeitos aleatórios γ_i i.i.d com distribuição *Skew-Normal* tal que $V(\gamma_i) = 2$. Para os Cenários 1a e 1b assumimos $\gamma_i \sim SN(0, 2, 0)$. Os efeitos aleatórios nos Cenários 2a e 2b são gerados de $SN(-1,777; 5,156; 5)$ e, nos Cenários 3a e 3b, assumimos que $\gamma_i \sim SN(-1,869; 5,493; 30)$. As covariáveis consideradas no estudo estão na Tabela 2.6. Para os conjuntos de dados com $k = 100$ *clusters* consideramos as mesmas covariáveis, mas o número de observações em cada *cluster* é assumido ser aproximadamente $n_i/4$, isto é, dividimos cada *cluster* na Tabela 2.6 em quatro *clusters*.

Tabela 2.6: Covariáveis e número de observações em cada *cluster*

<i>Cluster</i>	n_i	X_1	X_2	<i>Cluster</i>	n_i	X_1	X_2
1	110	-0,429	-1,519	14	148	-1,194	0,261
2	108	-0,429	0,261	15	151	0,337	0,854
3	107	-1,194	0,261	16	148	-1,194	0,854
4	107	-0,429	0,854	17	110	1,868	0,261
5	103	-1,194	0,854	18	108	0,031	-1,519
6	98	0,031	-1,519	19	107	0,337	0,854
7	102	0,031	0,261	20	103	-0,429	0,261
8	106	0,031	0,854	21	98	-1,194	-1,519
9	203	1,868	-1,519	22	102	-0,429	0,854
10	175	1,868	0,854	23	106	0,031	0,261
11	180	-1,194	0,854	24	203	0,337	-1,519
12	156	0,337	-1,519	25	175	1,868	0,854
13	152	0,337	0,261				

Em cada um destes cenários ajustamos sete diferentes modelos. Em todos eles assumimos distribuições *a priori* pouco informativas para os efeitos fixos ($\beta_i \sim N(0, 10)$, $i = 0, \dots, 2$) e para σ^2 por eliciarmos $\sigma^2 \sim IG(2,001; 1)$. Consequentemente, *a priori*, temos que $E(\sigma^2) = 1$ e $V(\sigma^2) = 1.000$. Assumimos também que $\gamma_i \stackrel{iid}{\sim} SN(-\delta\sigma\sqrt{2/\pi}, \sigma^2, \lambda)$. Os modelos diferem nas distribuições *a priori* para o parâmetro de assimetria. Os modelos M_1 a M_5 assumem distribuições *a priori* degeneradas para λ , isto é, o parâmetro de assimetria é fixado em $\lambda = -30, -5, 0, 5$ e 30 , respectivamente. Nos modelos M_6 e M_7 ,

assumimos distribuições *a priori* *Skew*-Normal não-degeneradas, com médias e variâncias iguais. Especificamente, consideramos $\lambda \sim N(0, 100)$ e $\lambda \sim SN(0; 257,821; 821,5)$ para os modelos M_6 e M_7 , respectivamente. As densidades destas distribuições *a priori* são apresentadas na Figura 7.

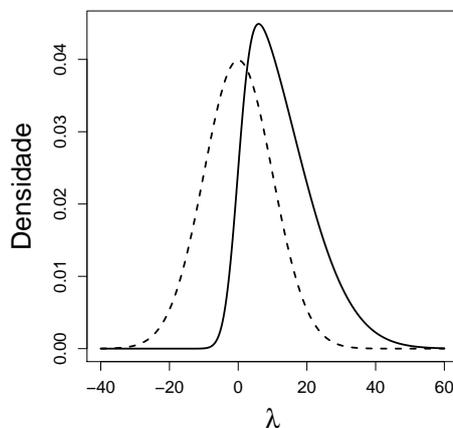


Figura 7: Distribuição *a priori* de λ , distribuições Normal (linha pontilhada) e *Skew*-Normal (linhas sólida).

Para o MCMC, geramos uma cadeia de tamanho 30.000 (45.000) para os modelos M_1 a M_5 (modelos M_6 e M_7), descartamos as primeiras 5.000 (20.000) amostras como o período de *burn-in* e usamos um *lag* de 50 passos para diminuir a autocorrelação, obtendo uma amostra da distribuição *a posteriori* de tamanho 500. O algoritmo foi implementado usando o *software* OxEdit.

2.5.1 Estimativas dos parâmetros

Nesta subseção analisamos o efeito da especificação *a priori* para λ na média *a posteriori* para os efeitos fixos, na variância de γ_i e no parâmetro de assimetria λ .

A Figura 8 mostra o erro quadrático médio (EQM) para as médias *a posteriori* sob todos os modelos e para todos os cenários. Como esperado, o EQM tende a ser menor quando o número de *clusters* é alto e tende a ser maior quando o valor para λ no modelo está longe do valor verdadeiro. O modelo M_3 é usualmente o melhor modelo se os efeitos aleatórios são gerados de uma distribuição Normal e também é um modelo razoavelmente bom nos outros dois casos. Os modelos M_4 a M_7 são modelos comparáveis se os efeitos aleatórios vêm de distribuições *Skew*-Normal com $\lambda = 5$ e 30. Nestes dois casos, os EQMs são muito próximos e tendem a ser menores para todos os parâmetros.

Em conclusão, se os efeitos aleatórios são normalmente distribuídos, usualmente obtemos estimativas ruins se especificamos incorretamente o valor *a priori* de λ . Contudo, se os efeitos aleatórios são gerados de distribuições na família *Skew*-Normal, o modelo que assume efeitos aleatórios normalmente distribuídos tem bom desempenho. Também concluímos que, embora o parâmetro de assimetria não seja muito bem estimado, assumir uma distribuição *a priori* não-degenerada para ele pode não ser uma boa estratégia.

Notamos que o EQM para os efeitos fixos e variância dos efeitos aleatórios são, em geral, menores sob os modelos M_6 e M_7 .

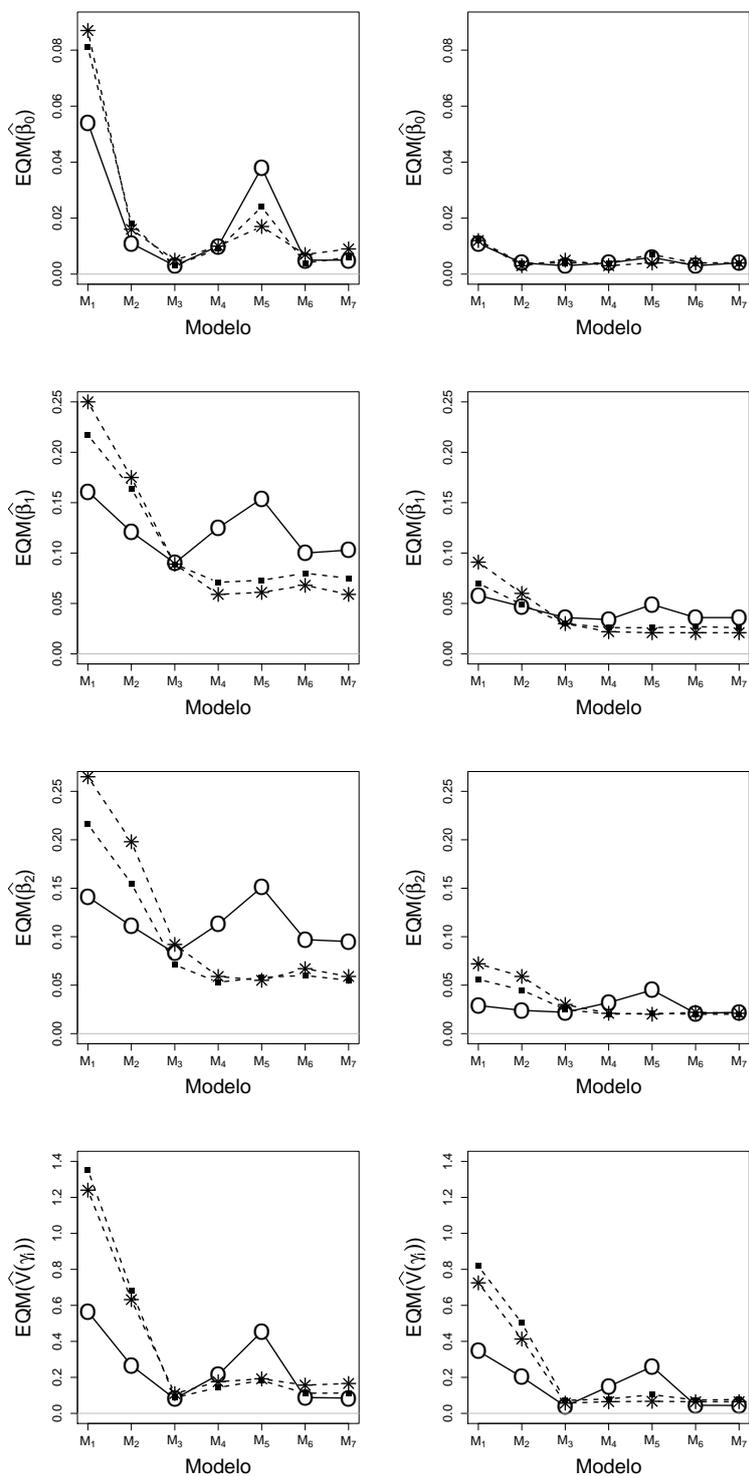


Figura 8: Erro quadrático médio (EQM) com as médias *a posteriori* para os efeitos fixos e variância dos efeitos aleatórios, nos cenários com $k = 25$ (esquerda) e $k = 100$ (direita) e verdadeiros valores $\lambda = 0$ (\circ), $\lambda = 5$ (\blacksquare) e $\lambda = 30$ ($*$).

Na Figura 9 apresentamos o vício obtido empiricamente para as estimativas dos parâmetros, sob todos os modelos e para todos os cenários, quando usamos as médias *a posteriori* como estimador.

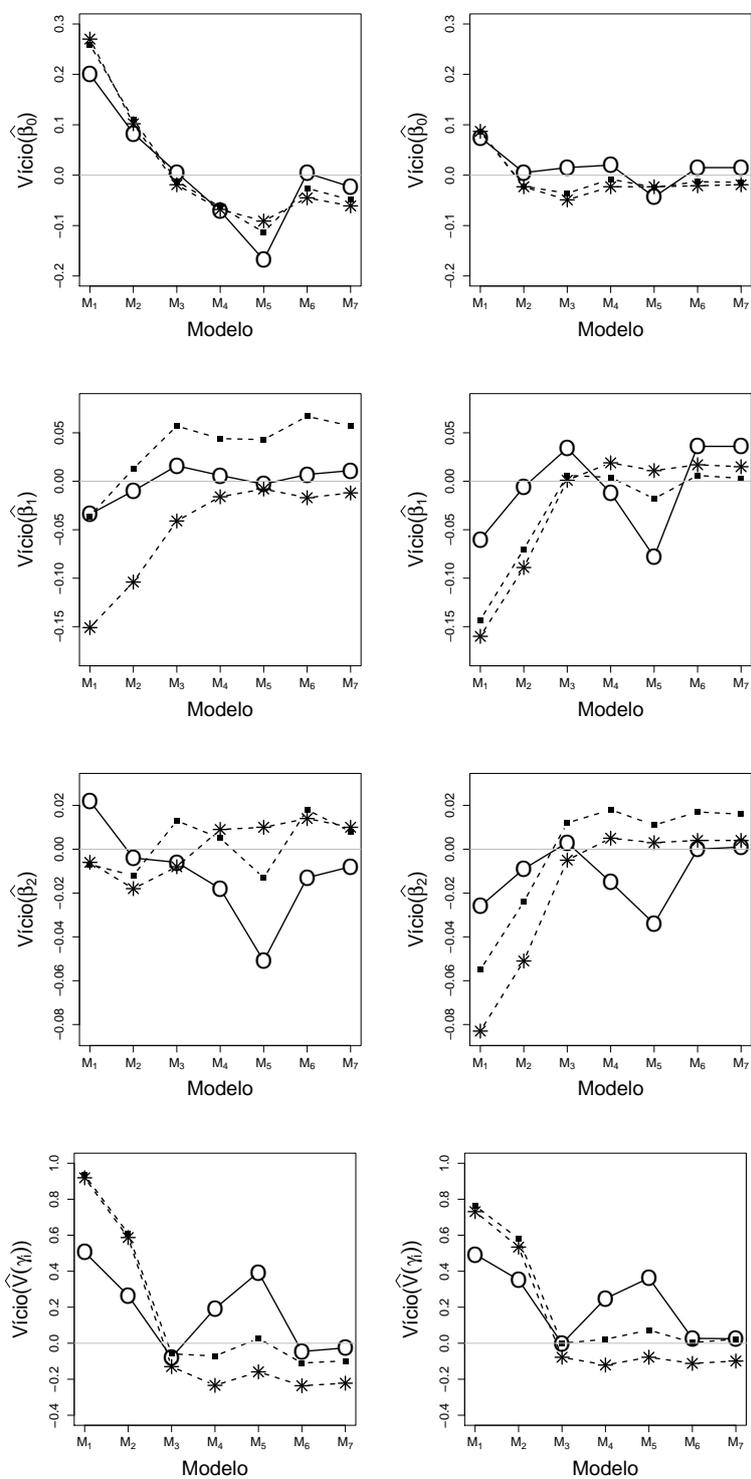


Figura 9: Vício com as médias *a posteriori* para os efeitos fixos e variância dos efeitos aleatórios, nos cenários com $k = 25$ (esquerda) e $k = 100$ (direita) e verdadeiros valores $\lambda = 0$ (\circ), $\lambda = 5$ (\blacksquare) e $\lambda = 30$ ($*$).

Da Figura 9 observamos que, em geral, os modelos em que se estima o parâmetro de assimetria λ , ou seja, os modelos M_6 e M_7 , estão entre os modelos com menor vício para todos os parâmetros. Para os cenários com assimetria positiva na distribuições dos efeitos aleatórios e $k = 100$ (Cenários 2b e 3b) notamos que, como seria esperado, os modelos M_1 e M_2 tem maior vício para quase todos os parâmetros. A exceção ocorre apenas para β_0 no modelo M_2 . Ainda para β_0 percebemos que para todos os modelos e valores de λ , o vício tende a diminuir quando o número de *clusters* aumenta de $k = 25$ para $k = 100$. Quando analisamos $V(\gamma_i)$, vemos que o modelo M_3 apresenta o menor vício em quase todos os cenários. Quando $\lambda = 0$, ou seja, nos cenários 1a e 1b, o vício de $V(\gamma_i)$ aumenta quando fixamos valores de λ mais distantes do valor real. Além disso, os modelos M_6 e M_7 tendem a fornecer bons resultados nestes cenários.

Conclusões similares podem ser obtidas tanto se consideramos EQM quanto o vício se a mediana ou a moda *a posteriori* são utilizadas para estimarem os efeitos fixos e a variância dos efeitos aleatórios (ver gráficos no Apêndice D).

As conclusões anteriores são corroboradas pelos gráficos nas Figuras 10 e 11 que apresentam o *Boxplot* das médias *a posteriori* de todos os parâmetros, sob todos os modelos e para todos os cenários. A Figura 10 mostra que as estimativas *a posteriori* para β_0 são verdadeiramente influenciadas pela especificação *a priori* para o parâmetro de assimetria. Frequentemente, nos três cenários, β_0 é melhor estimado sob o modelo M_3 , que assume $\lambda = 0$, e sob os modelos M_6 e M_7 , que assumem distribuições *a priori* mais flexíveis para λ . Se os efeitos aleatórios são normalmente distribuídos, os modelos M_1 , M_2 , M_4 e M_5 tendem a superestimar a variância dos efeitos aleatórios, o mesmo é observado para efeitos aleatórios *Skew-Normais* sob os modelos M_1 e M_2 . Também notamos que se a distribuição dos efeitos aleatórios possui forte assimetria (Cenário 3a), o parâmetro λ é subestimado nos modelos M_6 e M_7 e, sob o modelo M_7 , λ tende a ser superestimado nos Cenários 1a e 2a. Na Figura 11 observamos a diminuição na variabilidade das estimativas dos parâmetros quando o número de *clusters* aumenta de $k = 25$ para $k = 100$ e isto faz com que as estimativas obtidas sob os diferentes modelos se tornem bem próximas em vários casos. Para β_0 , ao contrário do que ocorreu quando $k = 25$, para os cenários com $k = 100$ temos que nem sempre o modelo M_3 é o melhor. Entre os modelos que consideram fixo o valor de λ , o modelo M_4 parece ser o melhor no Cenário 2b. Os modelos M_6 e M_7 estimam bem β_0 em todos os cenários. Ao avaliarmos as estimativas de β_1 e β_2 na Figura 11, vemos resultados muito similares aos observados quando $k = 25$. Em todos os cenários os modelos que fixam o valor de λ igual ao valor correto (M_3 quando λ real igual a 0, M_4 quando $\lambda = 5$ e M_5 quando $\lambda = 30$) apresentam melhores estimativas, mas estes modelos são seguidos de perto pelos modelos que fixam λ em valores próximos do valor real e pelos modelos em que distribuições *a priori* mais flexíveis para λ são assumidas (M_6 e M_7). Para as estimativas de $V(\gamma_i)$ e λ , observamos comportamento similar ao observado para os cenários com $k = 25$.

De forma geral, notamos melhores estimativas nos modelos em que o valor de λ é especificado corretamente e nos modelos em que λ é estimado. Visto que existe grande dificuldade em saber o valor correto de λ , estimar λ parece ser a melhor alternativa.

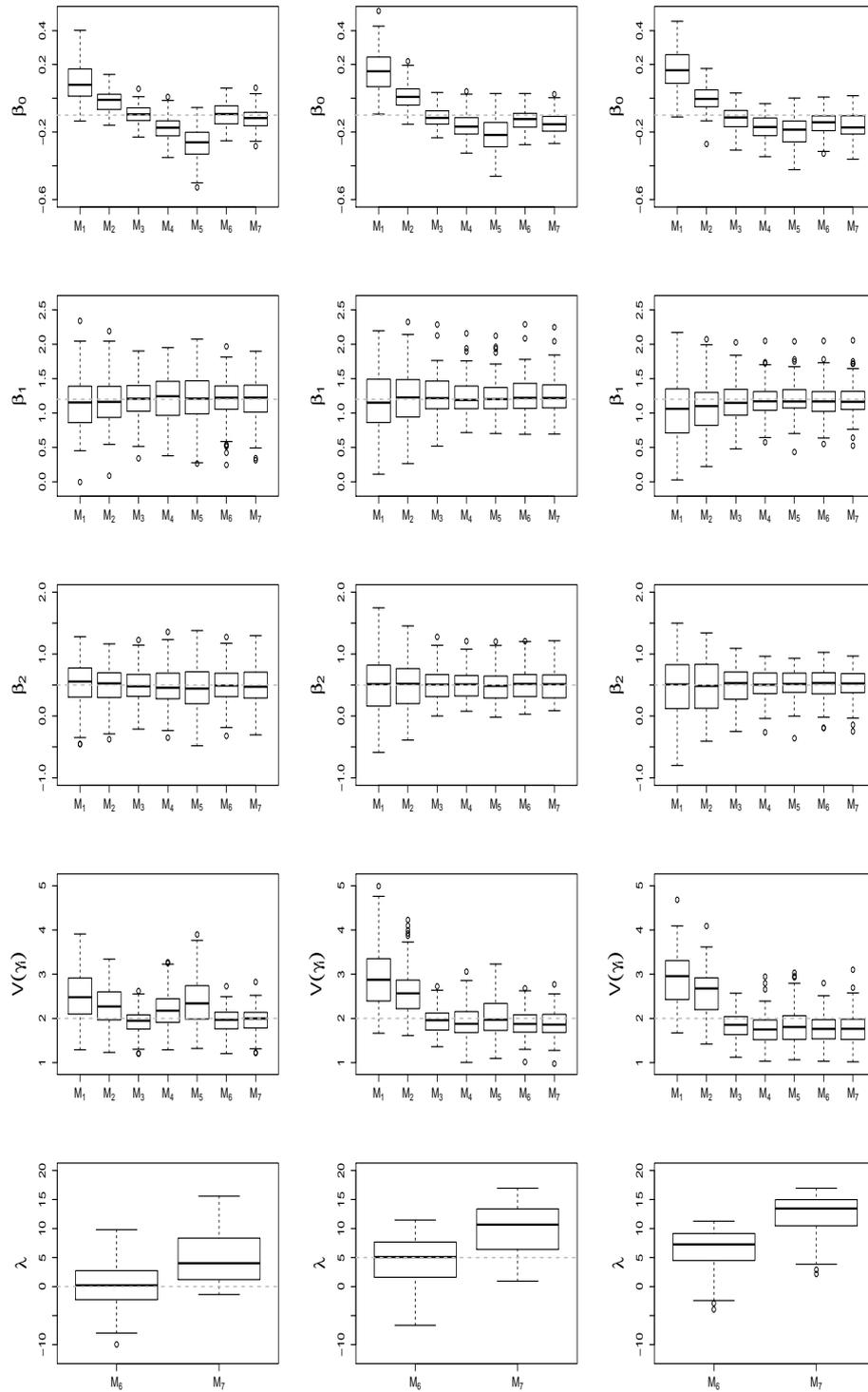


Figura 10: *Boxplots* das médias *a posteriori*, com valor real (linha pontilhada), para os cenários com $k = 25$ clusters, Cenário 1a ($\lambda = 0$, esquerda), Cenário 2a ($\lambda = 5$, centro) e Cenário 3a ($\lambda = 30$, direita).

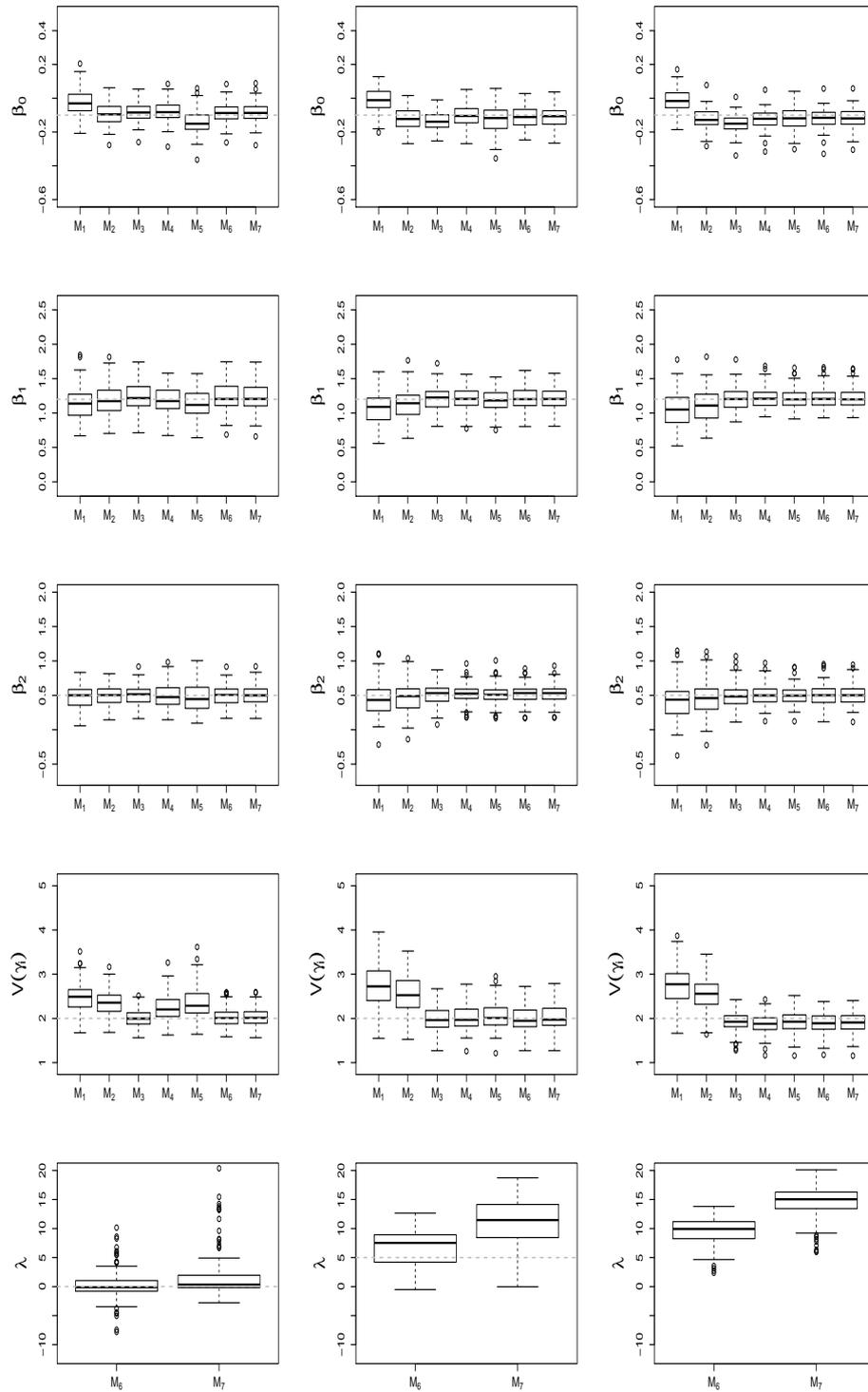


Figura 11: *Boxplots* das médias *a posteriori*, com valor real (linha pontilhada), para os cenários com $k = 100$ clusters, Cenário 1b ($\lambda = 0$, esquerda), Cenário 2b ($\lambda = 5$, centro) e Cenário 3b ($\lambda = 30$, direita).

2.5.2 Estimativas dos efeitos aleatórios

Como já foi mostrado anteriormente, os efeitos aleatórios tem importante papel na interpretação dos parâmetros. Agora buscamos avaliar o efeito da especificação *a priori* para λ nas estimativas dos efeitos aleatórios. Resumimos a informação sobre a qualidade destas estimativas através do erro quadrático médio $EQM_\gamma = (\gamma - \hat{\gamma})^t(\gamma - \hat{\gamma})/k$, em que $\hat{\gamma}$ é a média *a posteriori* do vetor γ .

A Tabela 2.7 mostra a porcentagem de vezes que cada modelo fornece o menor EQM_γ . Para todos os cenários, os efeitos aleatórios tendem a ser melhor estimados com o modelo correto mas, em geral, os modelos em que fixamos λ próximo de seu valor real também apresentam o menor EQM_γ em uma alta porcentagem de vezes. É importante notar que os modelos M_6 e M_7 , onde λ é estimado, frequentemente estimam os efeitos aleatórios bem, principalmente, se o número de *clusters* é alto.

Tabela 2.7: Porcentagem de vezes que cada modelo fornece o menor EQM_γ

Cenário	Modelo						
	M_1	M_2	M_3	M_4	M_5	M_6	M_7
1a	1	22	31	13	5	13	15
1b	0	7	37	12	3	25	16
2a	1	6	21	24	24	11	13
2b	0	0	15	33	21	9	22
3a	0	3	16	20	30	16	15
3b	0	1	1	17	39	19	23

A Figura 12 fornece a frequência em que cada modelo tem o menor EQM_γ , ou seja, é o melhor modelo. Dado o melhor modelo, ela também apresenta a frequência em que cada modelo tem o segundo melhor desempenho. Para o Cenário 1a, por exemplo, notamos que M_3 apresenta o menor EQM_γ em 31% das vezes e, nestas ocasiões, em aproximadamente metade das vezes (15% dos bancos de dados) o modelo M_6 foi o segundo melhor modelo. Para os Cenários 1a e 1b, notamos que das vezes em que o modelo M_3 é selecionado como o melhor modelo, o segundo melhor modelo é, em geral, um dos modelos que, *a priori*, não fixa λ . Além disso, quando M_6 e M_7 apresentam o menor EQM_γ , M_3 é geralmente o segundo melhor modelo. Conclusão similar pode ser obtida quando comparamos os modelos M_4 , M_6 e M_7 nos Cenários 2a e 2b, e os modelos M_5 , M_6 e M_7 nos Cenários 3a e 3b. Se, *a priori*, fixamos corretamente o valor de λ , então o EQM_γ para os efeitos aleatórios tende a ser o menor, principalmente, se o número de *clusters* é alto.

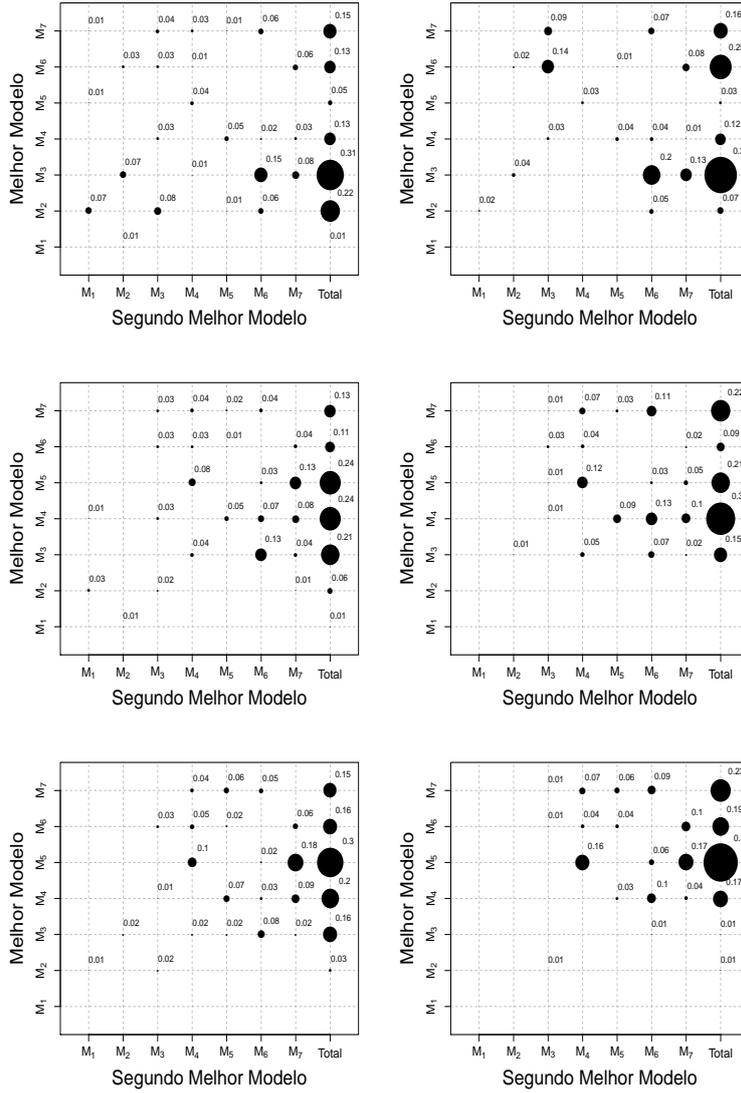


Figura 12: Porcentagem de vezes que cada modelo apresenta o menor e o segundo menor EQM_γ , cenários com $k = 25$ (esquerda) e $k = 100$ (direita) *clusters*, para $\lambda = 0$ (acima), $\lambda = 5$ (centro) e $\lambda = 30$ (abaixo).

A Figura 13 apresenta os *boxplots* para EQM_γ para todos os casos sob consideração. É notável que o comportamento do EQM_γ é similar se comparamos cenários com 25 e 100 *clusters*. Contudo, o EQM_γ tem menor variabilidade se o número de *clusters* é alto. Também notamos que o comportamento observado para o EQM_γ nos modelos M_6 e M_7 é similar ao que é observado no modelo em que λ é corretamente especificado.

Em resumo, se o objetivo é estimar os efeitos aleatórios, ou algumas quantidades que dependem deles, é melhor assumir distribuições *a priori* para λ se não podemos pré-especificar com precisão o seu valor.

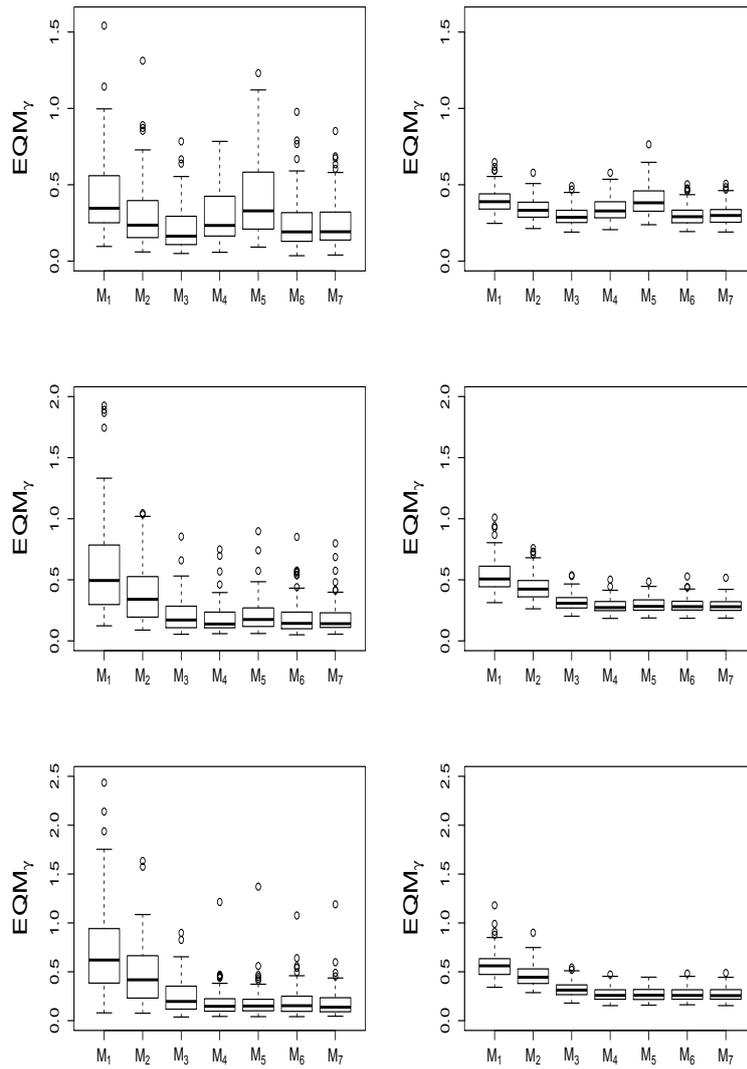


Figura 13: *Boxplot* do EQM_γ para os cenários com $k = 25$ (esquerda) e $k = 100$ (direita) *clusters*, para $\lambda = 0$ (acima), $\lambda = 5$ (centro) e $\lambda = 30$ (abaixo).

2.5.3 Comparação de modelos

A importância de se especificar corretamente o modelo está principalmente associada com qualidade das estimativas dos efeitos aleatórios e, conseqüentemente, com todas as quantidades que dependem deles, como ocorre com a razão de chances. Por esta razão, na Tabela 2.8, apresentamos a porcentagem de vezes que cada modelo foi selecionado como o melhor usando CPO, DIC1, DIC2 e o Fator de Bayes.

Tabela 2.8: Porcentagem de seleção de cada modelo

Cenário	Modelo						
	M_1	M_2	M_3	M_4	M_5	M_6	M_7
CPO							
1a	7	10	26	15	13	17	12
1b	2	7	30	6	3	30	22
2a	8	16	18	15	11	17	15
2b	0	1	34	20	10	15	20
3a	8	2	16	26	14	15	19
3b	0	0	15	24	14	27	20
DIC1							
1a	15	14	23	12	12	10	14
1b	5	10	38	8	8	17	14
2a	9	8	15	15	26	11	16
2b	1	2	10	36	31	13	7
3a	3	2	8	27	34	9	17
3b	0	0	4	13	47	15	21
DIC2							
1a	9	17	13	8	23	16	14
1b	0	8	27	3	1	32	29
2a	27	19	11	8	12	9	14
2b	0	2	27	31	3	24	13
3a	32	14	6	7	14	12	15
3b	0	1	13	29	9	25	23
Fator de Bayes							
1a	16	17	12	13	8	22	12
1b	1	10	37	5	4	21	22
2a	9	6	18	11	16	19	21
2b	0	4	28	26	4	17	21
3a	3	4	13	28	10	23	19
3b	0	0	22	21	18	22	17

Comparando somente os modelos M_1 a M_5 (aqueles que assumem valor fixo para λ) notamos que, para os cenários 1a e 1b, a CPO geralmente seleciona o modelo certo. Contudo, para os outros cenários, ele tende a selecionar os modelos que consideram menor assimetria que a verdadeira.

O DIC1 usualmente fornece bons resultados selecionando, mais frequentemente, o modelo correto em todos os cenários, exceto para o Cenário 2b. Nos cenários que consideramos efeitos aleatórios *Skew-Normais* esta medida também tende a selecionar os modelos que consideram assimetria. O DIC2 se mostrou ineficiente selecionando, em muitos casos, modelos em que o parâmetro de assimetria é muito maior do que o valor verdadeiro.

O fator de Bayes seleciona mais frequentemente o modelo correto somente no Cenário 1b. Nos cenários 2a, 2b, 3a e 3b ele tende a selecionar mais frequentemente o modelo com menor assimetria.

Em todos os casos, os modelos que assumem distribuições *a priori* não-degeneradas

para λ são selecionados uma alta porcentagem de vezes. Assim, se não pudermos definir, com precisão, o valor de λ no modelo, é uma boa estratégia estimá-lo. Note que as distribuições *a priori* assumidas aqui colocam pouca massa de probabilidade em valores próximos de $\lambda = 30$, o valor verdadeiro para o parâmetro de assimetria nos cenários 3a e 3b. Apesar disso, a porcentagem de vezes que os modelos M_6 e M_7 são selecionados como melhores nestes cenários é, relativamente, alta.

2.6 Estudo de caso 1: Atividade teratogênica

Nesta seção analisamos o conjunto de dados previamente considerado em Liu e Dey (2008). O objetivo do estudo é avaliarmos a interação entre atividade teratogênica de dois análogos de niacina *6-aminonicotinamide* (6AN) e *3 acetylpyridine* (3AP). Neste estudo, ovos foram incubados e, após 96 horas de incubação, ambas as substâncias foram dissolvidas juntas e injetadas em cada ovo. O experimento foi realizado em quatro meses de forma independente.

Consideramos um modelo Bernoulli em que a variável resposta assume o valor 1 se o ovo gera um frango anormal após o experimento. As covariáveis, padronizadas, são 6AN, 3AP e sua interação. Consideramos duas definições para os *clusters*, os meses (Dados 1) como em Liu e Dey (2008) e os *slots* (Dados 2). O número total de *clusters* nos Dados 1 e Dados 2 são 4 e 16, respectivamente.

Na Tabela 2.9 apresentamos os dados obtidos do experimento teratogênico. Os dados consistem de 2154 observações separadas em *slots*. As observações feitas em ovos alocados dentro de um mesmo *slot* possuem valores idênticos para as covariáveis. Isto ocorre pois os ovos alocados dentro de um mesmo *slot* receberam doses iguais, em microgramas (mcg), das substâncias 6AN e 3AP. Na Tabela 2.9, denotamos por n o número de observações e r o número observado de frangos anormais dentro de cada *slot*.

Tabela 2.9: Dados Experimentais - Atividade teratogênica

Datas	Slot	Dosagem de 6AN	Dosagem de 3AP	Sobreviventes (n)	Frangos anormais (r)	r/n
Nov. 1961	(1)	2,5	0	110	17	0,155
	(2)	2,5	375	108	28	0,259
	(3)	0	375	107	25	0,234
	(4)	2,5	500	107	43	0,402
	(5)	0	500	103	36	0,350
Dez. 1961	(1)	4	0	98	22	0,224
	(2)	4	375	102	39	0,382
	(3)	4	500	106	80	0,755
Jan. 1962	(1)	10	0	203	198	0,975
	(2)	10	500	175	153	0,875
	(3)	0	500	180	90	0,500
Abr. 1962	(1)	5	0	156	93	0,596
	(2)	5	375	152	41	0,270
	(3)	0	375	148	32	0,216
	(4)	5	500	151	106	0,702
	(5)	0	500	148	70	0,473

Ajustamos cinco modelos que diferem entre si em termos das distribuições assumidas para os efeitos aleatórios e também pelas distribuições *a priori* consideradas para os parâmetros de assimetria nos modelos SN. Estes modelos são descritos a seguir.

$$\text{Modelo 1: } \gamma_i \stackrel{iid}{\sim} N(0, \sigma^2);$$

Modelo 2: $\gamma_i \stackrel{iid}{\sim} SN(-\delta\sigma\sqrt{2/\pi}, \sigma^2, \lambda)$ e $\lambda \sim N(-8,5; 100)$;

Modelo 3: $\gamma_i \stackrel{iid}{\sim} SN(-\delta\sigma\sqrt{2/\pi}, \sigma^2, \lambda)$ e $\lambda \sim N(0, 100)$;

Modelo 4: $\gamma \sim SN_k(-\Delta\sigma\sqrt{2/\pi}\mathbf{1}_k, \sigma^2 I_k, \alpha\mathbf{1}_k)$ e $\alpha \sim N(-8,5; 100)$;

Modelo 5: $\gamma \sim SN_k(-\Delta\sigma\sqrt{2/\pi}\mathbf{1}_k, \sigma^2 I_k, \alpha\mathbf{1}_k)$ e $\alpha \sim N(0, 100)$.

Note que sob tais especificações para os Modelos 1, 3 e 5, como as distribuições para λ e α são centradas em zero as distribuições da *OR* dados κ_{12} e σ^2 , sob estes modelos teriam um comportamento similar *a priori*, colocando massa de probabilidade substancial em valores inferiores a 1,5 (ver Figura 1, caso $\lambda = 0$). Assim, tais especificações seriam razoáveis se a crença inicial sobre *OR* estivesse compatível com isto.

Para fins de comparação, consideramos a informação *a posteriori* fornecida em Liu e Dey (2008) para construir as distribuições *a priori* do parâmetro de assimetria nos Modelos 2 e 4. Nestes casos, assumimos *a priori* que λ e α são negativos com alta probabilidade e, conseqüentemente, temos uma forte evidência em favor de uma assimetria negativa para os efeitos aleatórios. Em todos os casos, contudo, consideramos distribuições *a priori* vagas para β_i e σ^2 , isto é, assumimos $\beta_i \sim N(0, 10)$, para todo $i = 0, \dots, 3$, e $\sigma^2 \sim GI(2,001; 1)$.

No método MCMC, geramos cadeias de tamanhos 30.000, 50.000 e 70.000, respectivamente, para os casos Normal, *Skew*-Normal independente e *Skew*-Normal correlacionada. Descartamos as primeiras 20.000 (casos Normal e *Skew*-Normal independente) e 40.000 (*Skew*-Normal correlacionada) amostras como período de *burn-in* e usamos *lag* de 10 (caso Normal) e 30 (casos *Skew*-Normal independente e correlacionado) passos para evitarmos amostras da distribuição *a posteriori* autocorrelacionadas. Com isto, obtivemos amostras da distribuição *a posteriori* de tamanho 1.000 em todos os modelos.

A Tabela 2.10 apresenta as estimativas pontuais *a posteriori* e os intervalos de credibilidade HPD com 95% de probabilidade para os efeitos fixos, variância dos efeitos aleatórios e o parâmetro de assimetria. Comparando os modelos, para ambos os conjuntos de dados, as estimativas pontuais *a posteriori* dos efeitos fixos sob os modelos SN são similares as obtidas com o modelo Normal, mas as estimativas para a variância da distribuição dos efeitos aleatórios $V(\gamma_i)$ diferem. As estimativas *a posteriori* indicam que os dois análogos de niacina, 6AN e 3AP, tem efeitos positivos na atividade teratogênica e sua interação tem efeito negativo. Além disso, o intercepto não é relevante para a explicação da atividade teratogênica. As variâncias da distribuição dos efeitos aleatórios $V(\gamma_i)$ são menores que 1, com alta probabilidade *a posteriori*, em todos os casos. Para facilitar a comparação, consideremos apenas os modelos em que a informação *a posteriori* fornecida por Liu e Dey (2008) não é utilizada. Neste casos, *a posteriori*, os intervalos de credibilidade HPD com 95% de probabilidade para os efeitos fixos e para $V(\gamma_i)$ sob os Modelos 3 e 5 são mais curtos para ambos os bancos de dados, fornecendo evidências contra a suposição de normalidade para os efeitos aleatórios. A distribuição adotada para descrever a incerteza *a priori* sobre o parâmetros de assimetria não exerce forte influência sobre a inferência *a posteriori* da variância dos efeitos aleatórios (por exemplo, a moda *a posteriori* de $V(\gamma_i)$ é 0,14 para o Modelo 2, e é 0,13 sob Modelo 3). Entretanto, as estimativas de $V(\gamma_i)$ tendem a aumentar com o número de *clusters*. Sob o Modelo 2, a moda *a posteriori* para $V(\gamma_i)$ nos Dados 1 e Dados 2 são 0,14 e 0,25, respectivamente.

Comportamentos similares são observados para os efeitos fixos e o parâmetro de assimetria. Uma grande redução na incerteza sobre os parâmetros de assimetria λ e α foram observados *a posteriori* em ambos os conjuntos de dados e todos os modelos.

Tabela 2.10: Resumos *a posteriori* para os parâmetros no experimento sobre atividade teratogênica

	Média	Mediana	Moda	HPD	Média	Mediana	Moda	HPD
Dados 1				Dados 2				
Modelo 1				Modelo 1				
DIC=2453,353; CPO= -1226,711				DIC=2361,254; CPO= -1180,818				
β_0	-0,22	-0,22	-0,29	[-0,78;0,37]	-0,30	-0,30	-0,32	[-0,64;0,06]
β_1	1,07	1,07	1,07	[0,91;1,23]	1,26	1,26	1,30	[0,86;1,67]
β_2	0,56	0,56	0,56	[0,42;0,70]	0,67	0,67	0,67	[0,30;1,05]
β_3	-0,71	-0,70	-0,69	[-0,93;-0,53]	-0,78	-0,78	-0,76	[-1,15;-0,38]
$V(\gamma_i)$	0,44	0,37	0,27	[0,10;0,97]	0,44	0,40	0,38	[0,17;0,79]
Modelo 2				Modelo 2				
DIC=2453,154; CPO= -1226,613				DIC=2360,162; CPO= -1180,198				
β_0	-0,28	-0,28	-0,29	[-0,83;0,16]	-0,29	-0,28	-0,24	[-0,60;-0,01]
β_1	1,08	1,07	1,07	[0,92;1,24]	1,32	1,32	1,34	[1,02;1,65]
β_2	0,57	0,57	0,55	[0,44;0,71]	0,68	0,68	0,69	[0,41;0,99]
β_3	-0,72	-0,72	-0,72	[-0,93;-0,56]	-0,77	-0,77	-0,84	[-1,10;-0,41]
λ	-7,88	-7,57	-6,61	[-27,89;11,58]	-11,83	-11,02	-11,84	[-27,55;3,19]
$V(\gamma_i)$	0,24	0,19	0,14	[0,04;0,54]	0,32	0,28	0,25	[0,12;0,60]
Modelo 3				Modelo 3				
DIC=2453,725; CPO= -1226,914				DIC=2360,621; CPO= -1180,419				
β_0	-0,22	-0,22	-0,29	[-0,68;0,18]	-0,29	-0,29	-0,27	[-0,59;0,06]
β_1	1,07	1,07	1,04	[0,90;1,23]	1,27	1,28	1,28	[0,82;1,63]
β_2	0,56	0,56	0,57	[0,42;0,69]	0,65	0,66	0,66	[0,26;0,96]
β_3	-0,72	-0,71	-0,71	[-0,93;-0,53]	-0,77	-0,76	-0,79	[-1,12;-0,41]
λ	1,14	2,03	5,68	[-19,11;20,06]	-6,26	-5,54	-3,78	[-22,07;9,34]
$V(\gamma_i)$	0,22	0,18	0,13	[0,05;0,50]	0,35	0,31	0,24	[0,11;0,70]
Modelo 4				Modelo 4				
DIC=2453,436; CPO= -1226,778				DIC=2361,036; CPO= -1180,699				
β_0	-0,26	-0,27	-0,30	[-0,58;0,18]	-0,30	-0,31	-0,34	[-0,59;-0,04]
β_1	1,07	1,07	1,09	[0,92;1,23]	1,26	1,26	1,24	[0,87;1,66]
β_2	0,57	0,56	0,55	[0,44;0,71]	0,64	0,65	0,68	[0,25;1,07]
β_3	-0,72	-0,71	-0,68	[-0,90;-0,51]	-0,76	-0,76	-0,74	[-1,14;-0,36]
α	-8,28	-8,61	-12,07	[-28,67;9,69]	-8,78	-8,75	-8,98	[-29,46;10,41]
$V(\gamma_i)$	0,38	0,31	0,24	[0,09;0,85]	0,43	0,39	0,33	[0,17;0,82]
Modelo 5				Modelo 5				
DIC=2453,095; CPO= -1226,604				DIC=2360,549; CPO= -1180,454				
β_0	-0,23	-0,24	-0,27	[-0,72;0,21]	-0,30	-0,30	-0,29	[-0,55;-0,06]
β_1	1,07	1,07	1,09	[0,92;1,24]	1,23	1,24	1,25	[0,84;1,59]
β_2	0,56	0,56	0,55	[0,42;0,69]	0,66	0,66	0,64	[0,19;1,04]
β_3	-0,72	-0,71	-0,73	[-0,89;-0,51]	-0,77	-0,76	-0,71	[-1,15;-0,35]
α	0,05	-0,07	2,14	[-17,9;20,76]	0,26	0,22	-0,85	[-18,43;20,28]
$V(\gamma_i)$	0,40	0,32	0,23	[0,08;0,91]	0,42	0,38	0,33	[0,16;0,76]

Considerando os intervalos HPD com 95% de probabilidade *a posteriori* e as estimativas pontuais de λ e α , temos forte evidência em favor de um grau de assimetria negativo para a distribuição dos efeitos aleatórios em Dados 1 (Modelos 2 e 4) e para Dados 2 (Modelos 2, 3 e 4). Isto vai de encontro com as conclusões apresentadas em Liu e Dey (2008). Para Dados 2 e todos os modelos notamos que a assimetria é forte, na análise *a posteriori*. Todavia, os intervalos HPD revelam que λ e α não são significativamente diferentes de zero, apontando para a distribuição Normal dos efeitos aleatórios. Apesar disto, levando em consideração o DIC e a CPO, os modelos *Skew-Normal* são selecionados como os melhores modelos (Modelo 2 para Dados 2 e Modelo 5 para Dados 1). A CPO e o DIC também indicam uma melhora no ajuste do modelo ao assumirmos os *slots* como *clusters*.

A Tabela 2.11 fornece os resumos das distribuições *a posteriori* da razão de chances para algumas comparações particulares, as quais estão descritas a seguir:

- OR_1 compara os ovos dos *slots* 4 e 5, Novembro de 1961;
- OR_2 compara os ovos dos *slots* 2 e 3, Abril de 1962;
- OR_3 compara os ovos do *slot* 3, Janeiro de 1962, e do *slot* 5, Novembro de 1961;
- OR_4 compara os ovos do *slot* 3, Abril de 1962, e do *slot* 5, Novembro de 1961;
- OR_5 compara os ovos do *slot* 1, Janeiro de 1962, e do *slot* 1, Dezembro de 1961;
- OR_6 compara os ovos do *slot* 2, Dezembro de 1961, e do *slot* 5, Abril de 1962.

Tabela 2.11: Resumos *a posteriori* para algumas OR no experimento sobre atividade teratogênica

	Média	Mediana	Moda	HPD		Média	Mediana	Moda	HPD
Dados 1					Dados 2				
Modelo 1					Modelo 1				
OR_1	1,48	1,48	1,48	[1,34;1,63]	OR_1	1,32	1,27	1,15	[0,70;2,03]
OR_2	4,28	4,26	4,22	[3,51;5,34]	OR_2	1,55	1,49	1,42	[0,79;2,33]
OR_3	2,15	2,12	2,15	[1,46;2,87]	OR_3	1,85	1,78	1,53	[1,01;2,72]
OR_4	1,29	1,28	1,25	[0,95;1,57]	OR_4	0,97	0,93	0,86	[0,50;1,55]
OR_5	109,18	98,4	79,89	[38,59;206,49]	OR_5	170,55	143,00	103,94	[45,99;367,00]
OR_6	1,50	1,48	1,48	[1,09;1,97]	OR_6	0,77	0,75	0,74	[0,43;1,17]
Modelo 2					Modelo 2				
OR_1	1,48	1,48	1,48	[1,33;1,61]	OR_1	1,34	1,28	1,13	[0,69;2,07]
OR_2	4,32	4,30	4,30	[3,39;5,27]	OR_2	1,51	1,46	1,37	[0,78;2,22]
OR_3	2,04	2,01	1,97	[1,37;2,69]	OR_3	1,78	1,73	1,59	[1,04;2,62]
OR_4	1,28	1,28	1,25	[0,97;1,59]	OR_4	0,97	0,94	0,89	[0,51;1,53]
OR_5	109,31	99,08	79,19	[34,48;197,56]	OR_5	171,77	148,71	108,70	[36,37;354,32]
OR_6	1,49	1,47	1,41	[1,09;1,98]	OR_6	0,79	0,77	0,69	[0,44;1,17]
Modelo 3					Modelo 3				
OR_1	1,48	1,48	1,48	[1,33;1,63]	OR_1	1,34	1,27	1,17	[0,72;2,10]
OR_2	4,29	4,24	4,11	[3,38;5,26]	OR_2	1,52	1,49	1,54	[0,80;2,37]
OR_3	2,07	2,04	2,00	[1,39;2,72]	OR_3	1,80	1,73	1,60	[1,08;2,69]
OR_4	1,28	1,26	1,24	[1,01;1,66]	OR_4	0,96	0,94	0,84	[0,50;1,51]
OR_5	108,93	97,99	86,50	[36,48;198,44]	OR_5	168,80	146,97	122,37	[42,14;348,33]
OR_6	1,49	1,48	1,39	[1,05;1,96]	OR_6	0,77	0,75	0,75	[0,43;1,19]
Modelo 4					Modelo 4				
OR_1	1,48	1,48	1,48	[1,34;1,61]	OR_1	1,34	1,31	1,28	[0,73;2,14]
OR_2	4,27	4,26	4,32	[3,42;5,14]	OR_2	1,53	1,50	1,52	[0,80;2,35]
OR_3	2,14	2,10	1,96	[1,51;2,84]	OR_3	1,86	1,76	1,69	[1,12;2,94]
OR_4	1,29	1,28	1,26	[0,98;1,63]	OR_4	0,97	0,93	0,90	[0,45;1,54]
OR_5	110,91	98,41	85,80	[36,59;211,24]	OR_5	164,35	141,74	113,97	[44,90;344,45]
OR_6	1,49	1,48	1,45	[1,05;1,95]	OR_6	0,77	0,74	0,72	[0,41;1,16]
Modelo 5					Modelo 5				
OR_1	1,48	1,48	1,46	[1,34;1,62]	OR_1	1,31	1,26	1,08	[0,72;2,02]
OR_2	4,28	4,25	4,00	[3,33;5,17]	OR_2	1,57	1,51	1,31	[0,82;2,38]
OR_3	2,14	2,11	1,99	[1,41;2,82]	OR_3	1,85	1,80	1,78	[1,04;2,71]
OR_4	1,29	1,28	1,25	[0,95;1,58]	OR_4	0,96	0,93	0,84	[0,48;1,54]
OR_5	108,75	98,66	85,25	[42,09;202,69]	OR_5	168,96	142,91	108,15	[45,51;371,59]
OR_6	1,51	1,48	1,44	[1,13;1,96]	OR_6	0,76	0,73	0,70	[0,43;1,16]

Da Tabela 2.11 percebemos que a média, a mediana e a moda *a posteriori* para todas as razões de chance (exceto para OR_5), em cada um dos modelos e para os dois conjuntos de dados tendem a ser similares, indicando somente um grau suave de assimetria na sua distribuição *a posteriori*. Tal similaridade é mais forte em Dados 2. Para Dados 1, sob quase todos os modelos (exceto Modelo 3), podemos concluir que os ovos no *slot* 3, Abril de 1962, são comparáveis aos do *slot* 5, Novembro de 1961 (veja OR_4). Neste caso, o

valor 1 pertence ao intervalo HPD, indicando que a OR não é significativamente diferente de um, ou seja, os ovos no *slot* 3, incubados em Abril de 1962, e no *slot* 5, incubados em Novembro de 1961, têm a mesma chance de gerar frangos anormais. Conclusão similar pode ser obtida de OR_1 , OR_2 , OR_4 e OR_6 , para Dados 2, sob todos os modelos.

Considere Dados 1 e o melhor modelo (Modelo 5). Levando em conta a moda da distribuição *a posteriori* de OR_3 , por exemplo, concluímos que, para os ovos incubados em janeiro de 1962, que não receberam qualquer dose de 6AN e uma dose de 500 mcg de 3AP, a chance de gerarem aves anormais é provavelmente 1,99 vezes a chance para os ovos incubados em novembro de 1961. Se considerarmos a OR_3 mediana, a conclusão é que a chance é maior que 2,11 com probabilidade *a posteriori* de 50%. Para OR_5 , as estimativas *a posteriori* diferem substancialmente. Investigando os ovos que não receberam qualquer dosagem de 3AP, em média e mais provavelmente, as chances daqueles incubados em Janeiro de 1962, que receberam a dosagem de 10 mcg de 6AN, gerarem um frango anormal são, respectivamente, 108,8 e 85,3 a chance dos ovos que foram incubados em Dezembro de 1961 e que receberam a dosagem de 4 de 6AN. As outras comparações possíveis ocorrem de forma semelhante.

2.7 Estudo de caso 2: *Ascaris suum*

Nesta seção analisamos os dados apresentados em Larsen *et al.* (2000), oriundos de um estudo para investigar a ocorrência de *Ascaris suum* (lombriga) em suínos. Amostras de fezes foram retiradas de 1016 porcos de abate e leitoas na Dinamarca. Dois tipos diferentes de chiqueiros foram incluídos no estudo, denotados por chiqueiros convencional e SPF (*Specific Pathogen Free*). A Tabela 2.12 apresenta os dados de 72 chiqueiros convencionais e 36 chiqueiros do tipo SPF, sendo que o símbolo 1/15, por exemplo, é usado para indicar que determinado chiqueiro tinha 15 suínos dos quais apenas 1 foi detectado como contaminado por *Ascaris suum*.

Tabela 2.12: Dados dos Chiqueiros

SPF								
0/16	0/15	0/15	0/15	0/15	0/15	0/15	1/15	2/15
0/14	0/14	0/12	0/10	0/10	0/10	0/10	0/10	0/10
0/10	0/10	1/10	1/10	2/10	2/10	0/9	1/7	1/7
0/5	0/5	0/5	0/5	1/5	1/5	4/5	1/3	0/1
Convencional								
0/15	0/15	0/15	0/15	0/15	1/15	1/15	1/15	1/15
2/15	3/15	3/15	3/15	4/15	4/15	6/15	2/11	0/10
0/10	0/10	0/10	0/10	0/10	0/10	0/10	0/10	0/10
0/10	1/10	1/10	1/10	1/10	3/10	3/10	4/10	0/9
0/9	0/9	0/9	1/9	2/9	5/9	9/9	0/8	0/8
2/8	3/8	0/7	1/7	1/7	1/7	1/7	2/7	3/7
0/6	0/6	0/6	0/5	0/5	0/5	1/5	1/5	1/5
1/5	0/4	0/4	3/4	0/3	0/2	0/2	0/1	1/1

O principal objetivo deste estudo é investigar se suínos criados em chiqueiro SPF tem menor risco de serem infectados com *Ascaris Suum* do que suínos criados nos chiqueiros convencionais e, caso isso ocorra, quantificar a diferença.

Ao observarmos o banco de dados, notamos que é necessário considerar algum artifício no modelo para representar a influência de cada chiqueiro particular na probabilidade de

contaminação por *Ascaris suum*. Isto se faz necessário visto que chiqueiros de mesmo tipo, seja SPF ou convencional, apresentam grande variabilidade quanto a proporção de contaminação. Para tal, assim como em Larsen *et al.* (2000), aqui também consideramos cada chiqueiro como um *cluster*. Desta forma, assumimos um número k de *clusters* igual a 108. Quanto as covariáveis incluídas no modelo, utilizamos apenas o tipo de chiqueiro, sendo que tal covariável recebeu o valor 1 para chiqueiros SPF e 0 para chiqueiros convencionais.

Ajustamos 4 modelos, que diferem entre si tanto na distribuição assumida para os efeitos aleatórios quanto na distribuição *a priori* do parâmetro de escala σ^2 . Os modelos ajustados foram:

Modelo 1: $\gamma_i \stackrel{iid}{\sim} N(0, \sigma^2)$ e $\sigma^2 \sim GI(2,001; 1)$;

Modelo 2: $\gamma_i \stackrel{iid}{\sim} N(0, \sigma^2)$ e $\sigma^2 \sim GI(0,01; 0,01)$;

Modelo 3: $\gamma_i \stackrel{iid}{\sim} SN(-\delta\sigma\sqrt{2/\pi}, \sigma^2, \lambda)$ e $\sigma^2 \sim GI(2,001; 1)$;

Modelo 4: $\gamma_i \stackrel{iid}{\sim} SN(-\delta\sigma\sqrt{2/\pi}, \sigma^2, \lambda)$ e $\sigma^2 \sim GI(0,01; 0,01)$.

Nos Modelos 3 e 4 assumimos *a priori* que $\lambda \sim N(0, 100)$, uma distribuição pouco informativa para λ pois tem um valor relativamente alto para a variância. Como visto anteriormente, para σ^2 utilizamos duas distribuições *a priori* que diferem quanto aos valores escolhidos para os hiperparâmetros da distribuição Gama Inversa. Ambas são distribuições que trazem pouca informação sobre σ^2 . No primeiro caso, em que $\sigma^2 \sim GI(2,001; 1)$, os valores dos hiperparâmetros foram escolhidos de forma que a esperança e a variância da distribuição existissem e fossem aproximadamente iguais a 1 e 1.000, respectivamente. Na segunda opção, buscamos uma distribuição *a priori* mais vaga, com a qual o modelo pudesse fornecer estimativas mais próximas da abordagem clássica, permitindo fazer algum tipo de comparação com os resultados obtidos em Larsen *et al.* (2000). Nesta segunda escolha, a distribuição *a priori* Gama(0,01; 0,01) para σ^2 não possui esperança e variância. A Figura 14 ilustra a f.d.p das duas escolhas de distribuições *a priori* para σ^2 .

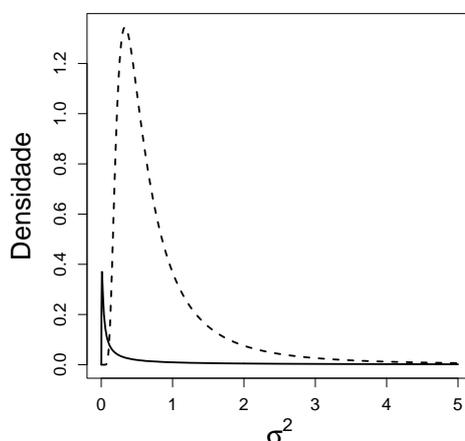


Figura 14: Distribuições *a priori* para σ^2 , $\sigma^2 \sim GI(2,001; 1)$ (linha tracejada) e $\sigma^2 \sim GI(0,01; 0,01)$ (linha sólida).

Em todos os modelos, para os efeitos fixos β_0 e β_1 assumimos distribuições *a priori* independentes e pouco informativas, a saber, distribuição Normal com média 0 e variância 1.000 para cada um dos parâmetros.

Para o MCMC, geramos uma cadeia de tamanho 35.000 para os Modelos 1 e 2, 250.000 para o Modelo 3 e 350.000 para o Modelo 4. Como período de *burn-in*, descartamos as primeiras 20.000 amostras nos Modelos 1 e 2, 50.000 no Modelo 3 e 150.000 no Modelo 4. Por fim, usamos um *lag* de 15 (200) nos Modelos 1 e 2 (3 e 4) obtendo uma amostra da distribuição *a posteriori* de tamanho 1.000.

Na Tabela 2.13, apresentamos as estimativas pontuais *a posteriori* e os intervalos de credibilidade HPD com 95% de probabilidade *a posteriori* dos efeitos fixos, da variância dos efeitos aleatórios e do parâmetro de assimetria. Ao compararmos o ajuste dos modelos, observamos um comportamento semelhante ao registrado na Seção 2.6. Notamos que as estimativas pontuais *a posteriori* obtidas com os modelos SN (Modelos 3 e 4) são relativamente próximas das obtidas com os modelos normais (Modelos 1 e 2). Contudo, as estimativas para a variância dos efeitos aleatórios $V(\gamma_i)$ tem maior discrepância ao mudarmos a distribuição dos efeitos aleatórios. Vemos também, como já era esperado, que as estimativas pontuais *a posteriori* e o comprimento do intervalo HPD para $V(\gamma_i)$ são maiores no Modelo 2 do que as obtidas no Modelo 1, e são maiores no Modelo 4 que no Modelo 3. Isto se justifica pelo fato dos Modelos 2 e 4 considerarem uma distribuição *a priori* para σ^2 menos informativa do que os Modelos 1 e 3. Ao compararmos os modelos, fixando a distribuição *a priori* para σ^2 , notamos que os modelos SN fornecem estimativas menores para $V(\gamma_i)$ do que os modelos Normais. As estimativas *a posteriori* para β indicam que chiqueiros do tipo SPF tem menor probabilidade de contaminação por *Ascaris Suum* do que os chiqueiros convencionais, uma vez que obtivemos estimativas negativas para β_1 . Além disto, em todos os modelos, os intervalos HPD para β_1 não contém o valor zero, apontando que o tipo de chiqueiro é uma covariável significativa para explicar a probabilidade de contaminação por *Ascaris Suum*. Conclusão similar é tirada sobre o intercepto em todos os casos.

Tabela 2.13: Resumos *a posteriori* dos parâmetros no experimento dos chiqueiros

	Média	Mediana	Moda	HPD	Média	Mediana	Moda	HPD
	Modelo 1				Modelo 2			
	DIC=565,851; CPO=-285,140				DIC=562,469; CPO=-284,436			
β_0	-2,50	-2,49	-2,52	[-3,05;-2,02]	-2,57	-2,56	-2,57	[-3,19;-2,07]
β_1	-1,08	-1,06	-1,02	[-2,08;-0,28]	-1,13	-1,10	-1,09	[-2,15;-0,21]
$V(\gamma)$	2,12	1,99	1,86	[0,87;3,50]	2,51	2,37	2,16	[1,11;4,31]
	Modelo 3				Modelo 4			
	DIC=563,397; CPO=-283,862				DIC=563,177; CPO=-284,429			
β_0	-2,42	-2,40	-2,38	[-2,86;-1,90]	-2,46	-2,44	-2,46	[-3,14;-1,95]
β_1	-1,08	-1,08	-1,07	[-1,87;-0,11]	-1,10	-1,08	-1,00	[-2,10;-0,25]
λ	5,53	3,86	0,66	[-2,39;18,31]	6,19	5,37	4,84	[-5,42;21,19]
$V(\gamma)$	1,67	1,53	1,44	[0,64;3,24]	2,09	1,86	1,54	[0,74;4,10]

Ao analisarmos as estimativas para o parâmetro de assimetria λ nos Modelos 3 e 4, observamos que estimativas pontuais *a posteriori* revelam assimetria positiva para a distribuição dos efeitos aleatórios, o que é uma indicação contra a suposição de normalidade para estes efeitos. Observando os intervalos HPD, notamos que houve grande redução na

incerteza sobre λ *a posteriori* e que, apesar dos intervalos HPD revelarem que λ não é significativamente diferente de zero, pois contêm o valor zero, as distribuições *a posteriori* revelam que existe grande evidência em favor de assimetria positiva para os efeitos aleatórios. Note da Figura 15 que as distribuições *a posteriori* de λ nos Modelos 3 e 4 colocam grande massa de probabilidade em valores positivos. Ao compararmos as curvas de densidades na Figura 15 e os intervalos HPD para λ percebemos que a distribuição *a priori* exerce influência sobre a inferência de λ . Utilizando DIC, o melhor modelo seria o Modelo 2, seguido pelos Modelos 4 e 3, nesta ordem. Quanto a CPO, a indicação seria favorável a assimetria para a distribuição dos efeitos aleatórios, sendo que o critério seleciona o Modelo 3 como o melhor e depois viriam os Modelos 4 e 2.

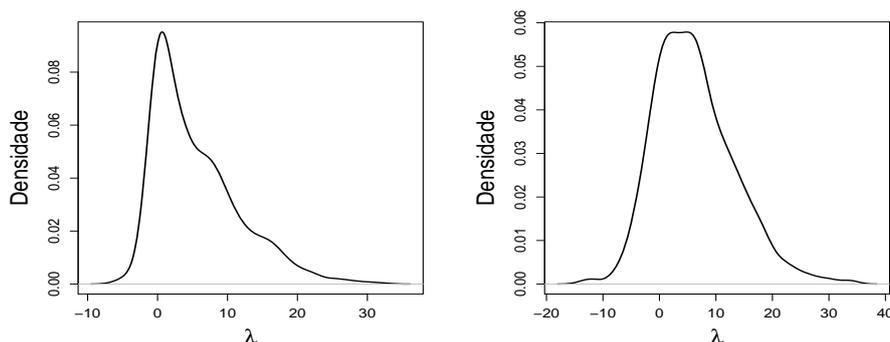


Figura 15: Densidade *a posteriori* de λ nos Modelos 3 (esquerda) e 4 (direita).

Como vimos, neste modelo, inúmeras comparações são possíveis. Entre estas várias possibilidades, optamos por algumas em particular que são descritas na Tabela 2.14, que também traz informações sobre o tipo de chiqueiro e proporção de contaminação do *cluster* ao qual cada suíno sob comparação pertence. Em alguns casos, como por exemplo OR_2 , OR_3 e OR_5 , os efeitos aleatórios parecem desempenhar papel importante, uma vez que essas comparações envolvem suínos em tipos de chiqueiros idênticos, mas que aparentam ter grandes diferenças quanto a chance de contaminação.

Tabela 2.14: Casos abordados para Razão de Chances

OR	1° Cluster	2° Cluster	Covariável para os suínos comparados
1	79 (9/9)	34 (4/5)	Convencional e SPF
2	79 (9/9)	76 (1/9)	Convencional e Convencional
3	79 (9/9)	78 (5/9)	Convencional e Convencional
4	37 (0/15)	1 (0/16)	Convencional e SPF
5	34 (4/5)	33 (1/5)	SPF e SPF
6	9 (2/15)	7 (0/15)	SPF e SPF

A Tabela 2.15 fornece os resumos das distribuições *a posteriori* da razão de chances para tais comparações. É importante estarmos atentos para o fato de que, em todos os casos, a comparação é feita com indivíduos de *clusters* diferentes, pois os indivíduos dentro do mesmo *cluster* sempre tem o mesmo valor para a covariável. Ao contrário do que foi visto na Seção 2.6, notamos que as estimativas obtidas via média, mediana e

moda *a posteriori* são muito diferentes para todas as razões de chances e sob todos os modelos. Isto indica um grau elevado de assimetria nas distribuições *a posteriori* das OR . Além disto, em todos os modelos e em quase todas as OR , exceção OR_2 , os intervalos de credibilidade HPD com 95% de probabilidade *a posteriori* contém o valor 1. Assim, usando o HPD, não podemos concluir que os indivíduos sob comparação tenham chances de contaminação por *Ascaris suum* significativamente diferentes. Notamos também que os intervalos HPD apresentam grande comprimento, revelando grande incerteza *a posteriori* quanto aos valores das razões de chance.

Tabela 2.15: Resumos *a posteriori* para algumas OR no experimento dos chiqueiros

	Média	Mediana	Moda	HPD	Média	Mediana	Moda	HPD
	Modelo 1				Modelo 2			
OR_1	9,87	5,08	1,83	[0,12;33,45]	9,43	5,01	1,93	[0,11;33,27]
OR_2	120,44	40,91	17,87	[1,38;384,17]	128,64	50,15	16,75	[2,08;422,15]
OR_3	10,37	5,57	1,85	[0,32;34,59]	11,07	5,76	2,92	[0,47;38,30]
OR_4	6,76	1,94	0,50	[0,01;25,12]	7,70	1,85	0,37	[0,01;29,27]
OR_5	37,83	10,63	2,59	[0,18;131,74]	39,12	11,86	3,77	[0,21;124,90]
OR_6	23,20	6,55	1,62	[0,02;96,46]	36,22	6,73	1,76	[0,08;113,59]
	Modelo 3				Modelo 4			
OR_1	11,88	4,87	1,73	[0,04;36,95]	11,85	4,84	2,11	[0,24;44,48]
OR_2	150,45	56,32	20,12	[2,76;580,26]	212,98	67,18	21,63	[1,33;685,79]
OR_3	14,42	6,08	2,34	[0,33;46,36]	17,09	6,92	2,54	[0,33;61,74]
OR_4	4,73	2,09	0,55	[0,00;15,53]	24,86	2,31	0,77	[0,00;20,72]
OR_5	34,79	11,83	5,53	[0,23;151,04]	45,80	14,64	3,85	[0,31;180,02]
OR_6	14,44	4,55	1,60	[0,10;39,57]	62,31	5,69	1,72	[0,11;55,14]

Vamos considerar apenas o Modelo 3 que foi o melhor modelo segundo a CPO. Levando em consideração a estimativa via moda *a posteriori* para OR_2 , por exemplo, concluímos que suínos do chiqueiro referente ao *cluster* 79 tem chance de contaminação por *Ascaris suum*, muito provavelmente, 20,12 maior que a chance de contaminação de um suíno no chiqueiro que forma o *cluster* 76. Se considerarmos a OR_2 mediana, a conclusão é que esta chance de contaminação no chiqueiro 79 é pelo menos 56,32 vezes a chance do chiqueiro 76, com probabilidade *a posteriori* de 50%. Consideramos agora a OR_4 , caso em que se compara dois chiqueiros de tipos diferentes e que não tiveram suíno contaminado. Para a comparação dada em OR_4 , segundo a moda da distribuição *a posteriori*, mais provavelmente a chance de contaminação de um suíno do *cluster* 37 é 0,55 vezes a chance de contaminação de um suíno do primeiro *cluster*. Se usamos a mediana *a posteriori*, por sua vez, concluímos que com probabilidade de 50%, esta chance é maior do que 2,09 vezes. As demais comparações são feitas da mesma forma.

2.8 Conclusão

Neste capítulo estendemos trabalhos anteriores que fornecem uma interpretação para efeitos fixos em termos da razão de chances no modelo de regressão logística com intercepto aleatório. Implementamos os modelos propostos sob o paradigma Bayesiano. Assumimos distribuições *Skew-Normal* para os efeitos aleatórios. Consideramos efeitos aleatórios tanto independentes quanto dependentes e, dados os efeitos fixos, obtivemos analiticamente as distribuições *a priori* para a razão de chances e suas medianas. Como um subproduto, também obtivemos resultados relacionados a combinações lineares de variáveis aleatórias *Skew-Normal* que, até onde sabemos, não haviam sido consideradas

na literatura. As distribuições *a posteriori* das razões de chance foram aproximadas usando métodos MCMC. Uma vantagem da interpretação da razão de chances *a posteriori* é que ela também fornece ferramentas para decidir sobre sua significância. Realizamos um estudo Monte Carlo para avaliar a influência da distribuição *a priori* do parâmetro de assimetria na inferência *a posteriori*. Analisamos conjuntos de dados simulados, um conjunto de dados reais apresentado em Liu e Dey (2008) de um experimento sobre atividade teratogênica e também um conjunto de dados reais apresentado em Larsen *et al.* (2000) sobre contaminação por *Ascaris suum* em suínos.

Em resumo, embora as estimativas pontuais para os efeitos fixos não sejam muito influenciadas pelas distribuições dos efeitos aleatórios, tais distribuições influenciam as estimativas para a razão de chances, bem como a magnitude dos intervalos HPD. Concluímos que a má especificação das distribuições de efeitos aleatórios pode levar a estimativas pobres para a razão de chances e a intervalos HPD com maior comprimento. Observamos também que, *a posteriori*, a razão de chances mediana não é necessariamente a melhor medida para estimar a razão de chances, como foi sugerido por Larsen *et al.* (2000). Em vários casos, a média e a moda *a posteriori* mostraram ser melhores estimadores. Além disso, a moda *a posteriori* para *OR* também tem uma atraente interpretação em termos de probabilidade. No estudo Monte Carlo concluímos que o uso de distribuições *a priori* degeneradas para o parâmetro de assimetria podem levar a estimativas pobres se tais distribuições são estabelecidas erroneamente. Se não temos informação precisa sobre o parâmetro de assimetria uma boa estratégia é estimá-lo. Para o experimento a respeito de atividade teratogênica consideramos modelos *Skew-Normal* mais flexíveis do que o considerado em Liu e Dey (2008). As estimativas *a posteriori* para o parâmetro de assimetria foram negativas na maioria dos modelos, levando as mesmas conclusões obtidas em Liu e Dey (2008). No estudo sobre contaminação por *Ascaris suum*, ajustamos modelos mais flexíveis que os adotados em Larsen *et al.* (2000), visto que em Larsen *et al.* (2000) considera-se apenas a distribuição Normal para representar comportamento dos efeitos aleatórios. Para os efeitos fixos, as conclusões foram similares as obtidas em Larsen *et al.* (2000), no entanto, os modelos *Skew-Normal* ajustados indicaram assimetria positiva na distribuição dos efeitos aleatórios.

Os resultados apresentados neste capítulo originaram o artigo intitulado “Parameters Interpretation in Skewed Logistic Regression with random intercept” (Santos *et al.*, 2013) e artigo “Aspects of Bayesian Inference in Skewed Mixed Logistic Regression Models” (Santos e Loschi, 2015) aceito para publicação como capítulo do livro “Current Trends in Bayesian Methodology with Applications”.

Capítulo 3

Modelo Logístico Misto Elíptico

3.1 Introdução

Neste capítulo consideraremos um modelo logístico misto mais geral do que o apresentado no Capítulo 1. A primeira generalização decorre do fato de que a partir de agora assumiremos que a variável resposta possui distribuição Binomial. Assim, o modelo Bernoulli discutido no Capítulo 1 pode ser visto como um caso particular deste que veremos aqui. Além disto, vamos considerar um modelo em que, além do intercepto aleatório, também considera efeitos aleatórios associados a algumas covariáveis. Isto é usualmente assumido na modelagem de dados longitudinais, nos quais observamos que os indivíduos, semelhantes em termos de covariáveis, possuem diferentes trajetórias para variável resposta ao longo do tempo (ver Laird e Ware, 1982, por exemplo). Desta forma, nos modelos longitudinais utilizamos efeitos aleatórios associados a covariável tempo para explicarmos a diferença de comportamento da variável resposta entre os indivíduos semelhantes. Esta estratégia é também utilizada fora do contexto longitudinal. Larsen *et al.* (2000), por exemplo, considera um banco de dados relacionado a um experimento sobre o acasalamento de salamandras (McCullagh e Nelder, 1989) onde o interesse está em modelarmos o resultado do acasalamento e os efeitos aleatórios são associados as espécies dos machos e das fêmeas em cada tentativa de acasalamento. Também em Larsen *et al.* (2000), outra aplicação apresenta um estudo sobre a prescrição de morfina em pacientes de um hospital da Dinamarca. Em tal análise, os sintomas apresentados por cada paciente são considerados tanto como efeitos fixos como também são associados a efeitos aleatórios.

Para modelarmos o comportamento dos efeitos aleatórios consideraremos distribuições na classe de distribuições elípticas. O principal atrativo desta classe é o fato de que inclui algumas distribuições com caudas mais pesadas do que a distribuição Normal, entre as quais estão as distribuições Normal Contaminada, *Slash* e *t*-Student. Distribuições com caudas mais pesadas são frequentemente utilizadas para modelar dados onde há a presença de valores atípicos como visto em Souza e Migon (2010). Sob o enfoque bayesiano, estes autores consideram o modelo logístico com intercepto aleatório visando acomodar e identificar valores atípicos na análise. Em Souza e Migon (2010), os efeitos aleatórios são considerados objetivando-se acomodar a variabilidade extra ocasionada pela presença de valores atípicos. Para os efeitos aleatórios, Souza e Migon (2010) consideraram como distribuições *a priori*, as distribuições Normal, Normal Contaminada e *t*-Student. Prates

et al. (2014) consideram o algoritmo EM para implementar um modelo misto para dados correlacionados binários em que os efeitos aleatórios possuem distribuição t -Student e a função de ligação é a f.d.a da distribuição t -Student.

Neste capítulo, consideraremos a classe de modelos logísticos mistos Binomial e assumiremos distribuições na classe elíptica para modelarmos o comportamento dos efeitos aleatórios. O foco principal está na interpretação dos parâmetros. Desta forma, estenderemos alguns resultados apresentados em Larsen *et al.* (2000) e Santos *et al.* (2013) relacionados à distribuição da OR. A inferência sobre o modelo é feita sob o paradigma clássico. Introduzimos um algoritmo tipo EM para obter os estimadores de Máxima Verossimilhança para o modelo, a saber, utilizamos o algoritmo EM Monte Carlo. Algumas alternativas de implementação do algoritmo proposto (McCulloch (1997); Quintana *et al.* (1999)), diferentes quanto ao passo de geração dos efeitos aleatórios, são comparadas. Faremos um estudo com dados simulados para avaliarmos a convergência das estimativas obtidas com o algoritmo e um estudo Monte Carlo para avaliarmos a qualidade das estimativas dos parâmetros e da razão de chances mediana. O modelos propostos são implementados no *software* OxEdit.

Este capítulo está organizado como segue. Na Seção 3.2 apresentamos uma revisão sobre a classe de distribuições elípticas e algumas de suas propriedades, as quais são de maior interesse para o trabalho. Na Seção 3.3 apresentamos o modelo logístico misto considerado ao longo do Capítulo 3. Na Seção 3.4 exibimos resultados para a razão de chances e para a razão de chances mediana sob suposições mais gerais para a distribuição dos efeitos aleatórios e, também, ao assumirmos distribuições elípticas para os efeitos aleatórios. Estes resultados são extensões dos resultados obtidos por Larsen *et al.* (2000) e englobam estes últimos como casos particulares. Na Seção 3.5 discutimos o processo de inferência no modelo logístico com intercepto aleatório, propondo uma implementação do algoritmo EM Monte Carlo, ao assumirmos uma distribuição na classe de distribuições Normal Independente para representar o comportamento dos efeitos aleatórios. Na Seção 3.6 apresentamos um estudo com dados simulados para avaliarmos o comportamento do algoritmo EM Monte Carlo proposto no que diz respeito a convergência das estimativas obtidas. Na Seção 3.7 avaliamos a qualidade das estimativas dos modelos propostos através de um estudo Monte Carlo. Na Seção 3.8, considerando os modelos propostos neste capítulo, reanalisamos o banco de dados usado em Larsen *et al.* (2000), que trata a respeito de contaminação de suínos por *Ascaris Suum* e que também foi analisado na Seção 2.7. Por fim, na Seção 3.9 apresentamos nossas conclusões finais sobre os resultados obtidos ao longo do capítulo.

3.2 Distribuições Elípticas

No decorrer deste capítulo vamos assumir distribuições elípticas para representar o comportamento dos efeitos aleatórios. Resultados importantes sobre as distribuições elípticas podem ser encontradas em Fang *et al.* (1990). Apresentamos uma breve introdução sobre tal família de distribuições e, para tal, iniciamos definindo a classe de distribuições esféricas.

Seja \mathbf{O}_n o grupo das matrizes $\mathbf{\Gamma}$ de ordem $n \times n$ que são ortogonais, isto é, são matrizes tais que $\mathbf{\Gamma}^t \mathbf{\Gamma} = \mathbf{I}_n$. O grupo \mathbf{O}_n é chamado grupo ortogonal e, aqui, é usado para definir

a classe de distribuições esféricas.

DEFINIÇÃO 3.1. *O vetor aleatório \mathbf{Z} de ordem $n \times 1$ tem distribuição esférica (n -variada) se para cada matriz $\mathbf{\Gamma}$ em \mathbf{O}_n , temos que*

$$\mathbf{Z} \stackrel{d}{=} \mathbf{\Gamma}\mathbf{Z}.$$

Como consequência da Definição 3.1, temos que um vetor \mathbf{Z} tem distribuição esférica se, e somente se, a função característica de \mathbf{X} é da forma

$$\varphi_{\mathbf{Z}}(\mathbf{r}) = \varphi(\mathbf{r}^t\mathbf{r}), \mathbf{r} \in \mathbb{R}^n,$$

para alguma função φ tal que $\varphi(u) \in \mathbb{R}$ para $u \geq 0$. Também segue da Definição 3.1 que, quando a f.d.p de \mathbf{Z} existir, esta é da forma

$$f_{\mathbf{Z}}(\mathbf{z}) = h(\mathbf{z}^t\mathbf{z}), \mathbf{z} \in \mathbb{R}^n, \quad (3.1)$$

para alguma função h tal que $h(u) \geq 0$ para $u \geq 0$. Kelker (1970) mostra que uma função não-negativa h pode ser usada para definir a densidade de alguma distribuição esférica n -dimensional se, e somente se, $\int_0^\infty r^{n-1}h(r^2)dr \leq \infty$. Ao longo deste trabalho somente distribuições esféricas que possuem f.d.p serão consideradas.

Na Definição 3.2 apresentamos a classe de distribuições elípticas a qual também pode ser construída a partir de uma transformação linear de uma variável aleatória esférica, como formalizaremos mais adiante. A classe de distribuições elípticas contém a classe de distribuições esféricas como caso particular.

DEFINIÇÃO 3.2. *O vetor aleatório \mathbf{X} , $n \times 1$, é dito ter distribuição elíptica com vetor de locação $\boldsymbol{\mu} \in \mathbb{R}^n$ e matriz de dispersão $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ positiva definida, denotada por $\mathbf{X} \sim El_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \varphi)$, se sua função característica é da forma*

$$\varphi_{\mathbf{X}}(\mathbf{r}) = \exp(i\mathbf{r}^t\boldsymbol{\mu})\varphi(\mathbf{r}^t\boldsymbol{\Sigma}\mathbf{r}), \mathbf{r} \in \mathbb{R}^n, \quad (3.2)$$

para alguma função φ tal que $\varphi(u) \in \mathbb{R}$ para $u \geq 0$.

De forma equivalente, podemos definir a distribuição de \mathbf{X} com função característica dada em (3.2) considerando a transformação linear

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + \mathbf{A}\mathbf{Z},$$

em que \mathbf{A} é uma matriz $n \times k$ tal que $rank(\mathbf{A}) = k$, $\mathbf{A}\mathbf{A}^t = \boldsymbol{\Sigma}$ e \mathbf{Z} é um vetor aleatório $k \times 1$ com distribuição esférica k -variada cuja função característica é $\varphi_{\mathbf{Z}}(\mathbf{s}^t\mathbf{s})$, $\mathbf{s} \in \mathbb{R}^k$. Se a distribuição de \mathbf{X} possui f.d.p, então $\boldsymbol{\Sigma}$ é uma matriz positiva definida e a f.d.p associada à distribuição de \mathbf{X} é dada por

$$f_{\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{x}) = |\boldsymbol{\Sigma}|^{-1/2}h((\mathbf{x} - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})), \mathbf{x} \in \mathbb{R}^n,$$

para alguma função h tal que $h(u) \geq 0$ para $u \geq 0$. Neste caso, a notação será $\mathbf{X} \sim El_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; h)$.

Note que, se $\boldsymbol{\mu} = \mathbf{0}$ e $\boldsymbol{\Sigma} = \mathbf{I}_n$, isto é, $\mathbf{X} \sim El_n(\mathbf{0}, \mathbf{I}_n; h)$, a distribuição de \mathbf{X} é dita ser esférica. A Tabela 3.1 apresenta algumas subclasses de distribuições esféricas n -dimensionais e suas respectivas funções de densidade $h(\mathbf{x}^t \mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n$. Em tal tabela $F_{U|\nu}(u)$ representa a f.d.a de uma variável aleatória U .

Tabela 3.1: Subclasses de distribuições esféricas n -variadas

Tipo	$h(\mathbf{x}^t \mathbf{x})$ ($\mathbf{x} \in \mathbb{R}^n$)
Normal padrão	$h(\mathbf{x}^t \mathbf{x}) \propto \exp \{-\mathbf{x}^t \mathbf{x}/2\}$
t generalizada padrão	$h(\mathbf{x}^t \mathbf{x}) \propto (1 + \mathbf{x}^t \mathbf{x}/s)^{-(n+m)/2}$, $s, m > 0$
Cauchy generalizada	$h(\mathbf{x}^t \mathbf{x}) \propto (1 + \mathbf{x}^t \mathbf{x}/s)^{-(n+1)/2}$, $s > 0$
Logística	$h(\mathbf{x}^t \mathbf{x}) \propto \exp(-\mathbf{x}^t \mathbf{x})[1 + \exp\{-\mathbf{x}^t \mathbf{x}\}]^{-2}$
Normal Independente	$h(\mathbf{x}^t \mathbf{x}) \propto \int_0^\infty u^{n/2} \exp\{-\mathbf{x}^t \mathbf{x}u/2\} dF_{U \nu}(u)$
Pearson II	$h(\mathbf{x}^t \mathbf{x}) \propto (1 - \mathbf{x}^t \mathbf{x})^m$, $m > 0$, $0 \leq u < 1$

Um resultado que será bastante útil ao longo deste trabalho é dado no Teorema 3.1 a seguir onde apresentamos a distribuição de probabilidade de combinações lineares de um vetor aleatório com distribuição conjuntamente elíptica. Este resultado e sua demonstração podem ser encontrados em Fang *et al.* (1990).

TEOREMA 3.1. *Seja $\mathbf{Y} = \boldsymbol{\xi} + \mathbf{A}\mathbf{X}$, em que $\boldsymbol{\xi} \in \mathbb{R}^m$ e \mathbf{A} é uma matriz $m \times n$. Se $\mathbf{X} \sim El_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \varphi)$, então $\mathbf{Y} \sim El_m(\boldsymbol{\xi} + \mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^t; \varphi)$.*

Outro resultado importante sobre distribuições multivariadas, cuja prova também pode ser encontrada em Fang *et al.* (1990), diz respeito as suas distribuições marginais. Para as distribuições elípticas, abordamos esta questão no próximo teorema.

TEOREMA 3.2. *Se $\mathbf{X} \sim El_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \varphi)$, e*

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad e \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

definem uma partição de \mathbf{X} , $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$, respectivamente, em que \mathbf{X}_1 e $\boldsymbol{\mu}_1$ são vetores $m \times 1$ e $\boldsymbol{\Sigma}_{11}$ é uma matriz $m \times m$, $1 \leq m < n$, então $\mathbf{X}_1 \sim El_m(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}; \varphi)$.

Na Figura 16 apresentamos as curvas de nível da f.d.p de casos particulares da distribuição Normal bivariada. O primeiro gráfico é um exemplo de uma distribuição esférica, isto é, representa a distribuição de $\mathbf{X} \sim N_2(\mathbf{0}, \mathbf{I}_2)$, enquanto o segundo gráfico ilustra uma distribuição elíptica que, neste caso, será a distribuição $N_2(\mathbf{0}, \boldsymbol{\Sigma})$ em que $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0,5 \\ 0,5 & 1 \end{pmatrix}$.

Uma subclasse das distribuições elípticas bastante conhecida e que vem sido muito considerada na construção de modelos estatísticos mais flexíveis é a classe de distribuições chamadas de distribuições Normais Independentes (NI) (ver Rogers e Tukey, 1972; Andrews e Mallows, 1974; Lange e Sinsheimer, 1993, entre outros). Esta classe é formada pelas distribuições elípticas originadas de misturas na escala da distribuição Normal. Esta

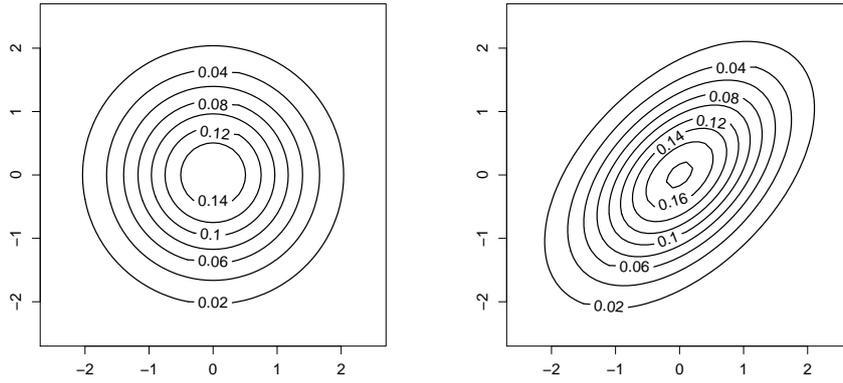


Figura 16: Curvas de nível da f.d.p das distribuições Normais bivariadas esférica (esquerda) e elíptica (direita).

mistura é realizada assumindo que a escala original da distribuição Normal é modificada pela multiplicação de uma variável aleatória positiva U , com f.d.p $f_{U|\boldsymbol{\nu}}(u)$ e f.d.a $F_{U|\boldsymbol{\nu}}(u)$ indexadas pelo vetor de parâmetros $\boldsymbol{\nu}$. Esta subclasse de distribuições elípticas assumirá diferentes formas ao mudarmos a distribuição de probabilidades da variável misturadora U . Fazem parte deste família, entre outras, as distribuições t -Student, *Slash*, Normal Contaminada e Normal. As distribuições t -Student, *Slash* e Normal Contaminada são amplamente utilizadas na literatura para modelar dados com presença de valores atípicos, pois são distribuições que possuem caudas mais pesadas do que a distribuição Normal (ver Souza e Migon, 2010; Lange e Sinsheimer, 1993, entre outros).

Formalmente, definimos a classe de distribuições Normal Independente da seguinte forma. Um vetor aleatório \mathbf{X} tem distribuição na classe distribuições NI, com parâmetros de locação $\boldsymbol{\mu}$, de escala $\boldsymbol{\Sigma}$ e de forma $\boldsymbol{\nu}$, $\mathbf{X} \sim \text{NI}(\boldsymbol{\mu}, \boldsymbol{\Sigma}^2; F_{U|\boldsymbol{\nu}})$, se sua distribuição tem f.d.p dada por

$$f_{\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}}(\mathbf{x}) = \int_0^\infty \frac{u^{n/2}}{(2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} \exp \left\{ -\frac{u}{2} (\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\} dF_{U|\boldsymbol{\nu}}(u). \quad (3.3)$$

A média e a variância de distribuições na classe NI são dadas, respectivamente, por

$$E(\mathbf{X}) = \boldsymbol{\mu} \quad \text{e} \quad V(\mathbf{X}) = E(U^{-1}) \boldsymbol{\Sigma}.$$

Um resultado importante relacionado à família de distribuições NI é sua representação estocástica. Este resultado pode tanto ser útil na implementação computacional dos modelos para fazermos inferências quanto na geração de amostras de distribuições desta família. Lange e Sinsheimer (1993) definem a família de distribuições NI através do resultado a seguir.

PROPOSIÇÃO 3.1. *Se o vetor \mathbf{X} possui distribuição Normal Independente n -variada com parâmetros de locação $\boldsymbol{\mu}$ e escala $\boldsymbol{\Sigma}$ temos que*

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + U^{-1/2} \boldsymbol{\Sigma}^{1/2} \mathbf{T},$$

em que $\mathbf{T} \sim N_n(0, \mathbf{I}_n)$ e U é uma variável positiva com f.d.p $f_{U|\boldsymbol{\nu}}(u)$.

Diferentes escolhas de distribuições para U levam a diferentes distribuições para X , por exemplo,

- (i) Se $U \sim Gama(\nu/2, \nu/2)$ então X tem distribuição t -Student multivariada com parâmetros de locação $\boldsymbol{\mu} \in \mathbb{R}^n$ e de escala $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ e com graus de liberdade $\nu > 0$, a qual será denotada por $X \sim T_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ e cuja f.d.p é dada por

$$f_{\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu}(\mathbf{x}) = \frac{\Gamma((\nu + n)/2)}{|\boldsymbol{\Sigma}|^{1/2} \pi^{n/2} \Gamma(\nu/2) \nu^{n/2}} \left[1 + \frac{(\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})}{\nu} \right]^{-(\nu+n)/2},$$

onde a matriz de covariância de \mathbf{X} é dada por

$$V[\mathbf{X}] = \frac{\nu}{\nu - 2} \boldsymbol{\Sigma}, \text{ para } \nu > 2;$$

- (ii) Se $U \sim Beta(\nu, 1)$ então X tem distribuição *Slash* multivariada com parâmetros de locação $\boldsymbol{\mu} \in \mathbb{R}^n$, de escala $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ e de forma $\nu > 0$, a qual será denotada por $X \sim SL_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ e cuja a f.d.p dada por

$$f_{\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu}(\mathbf{x}) = \nu \int_0^1 u^{\nu-1} \phi_n(\mathbf{x} | \boldsymbol{\mu}, u^{-1} \boldsymbol{\Sigma}) du,$$

onde a matriz de covariância de \mathbf{X} é dada por

$$V[\mathbf{X}] = \frac{\nu}{\nu - 1} \boldsymbol{\Sigma}, \text{ para } \nu > 1;$$

- (iii) Se $f_{U|\nu_1, \nu_2}(u) = \nu_1 \mathbf{1}_{\{u=\nu_2\}} + (1 - \nu_1) \mathbf{1}_{\{u=1\}}$, com $0 \leq \nu_1 \leq 1$ e $0 < \nu_2 \leq 1$, então X tem distribuição Normal Contaminada multivariada com parâmetros de locação $\boldsymbol{\mu} \in \mathbb{R}^n$, de escala $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ e onde ν_1 e ν_2 são parâmetros de forma, denotada por $X \sim NC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu_1, \nu_2)$ e cuja a f.d.p é

$$f_{\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu_1, \nu_2}(\mathbf{x}) = \nu_1 \phi_n(\mathbf{x} | \boldsymbol{\mu}, \nu_2^{-1} \boldsymbol{\Sigma}) + (1 - \nu_1) \phi_n(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

onde a matriz de covariância de \mathbf{X} é dada por

$$V(\mathbf{X}) = \left(\frac{\nu_1}{\nu_2} + 1 - \nu_1 \right) \boldsymbol{\Sigma}.$$

A distribuição Normal também pertence a classe de distribuições NI e é obtida ao considerarmos uma distribuição degenerada em 1 para U .

Com o objetivo de ilustrarmos o comportamento das distribuições pertencentes a classe NI apresentamos gráficos da f.d.p das distribuições citadas anteriormente, no caso univariado e para alguns casos particulares. A escolha pelos gráficos no caso univariado, apesar das distribuições serem multivariadas, foi feita com a intenção de simplificar a comparação entre as distribuições. Na Figura 17 apresentamos gráficos das f.d.p das distribuições t -Student e *Slash* para alguns valores diferentes dos parâmetros.

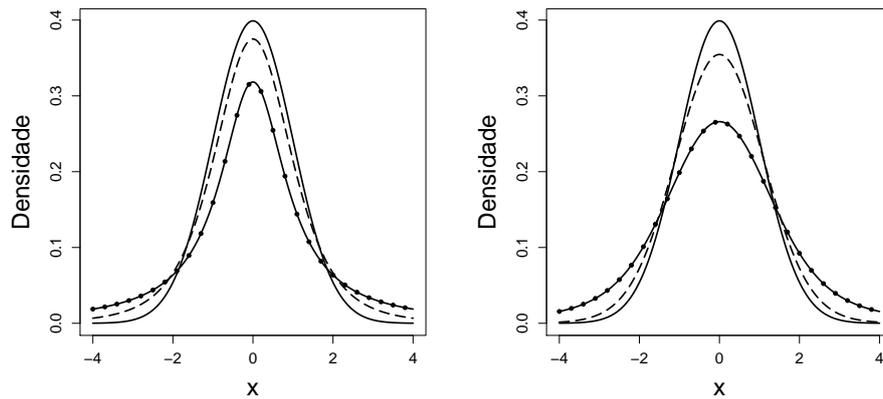


Figura 17: Densidade das distribuições t -Student (esquerda) e *Slash* (direita). No gráfico à esquerda as distribuições Normal padrão (linha sólida), $T_1(0, 1, 1)$ (●) e $T_1(0, 1, 4)$ (linha tracejada). No gráfico à direita as distribuições Normal padrão (linha sólida), $SL_1(0, 1, 1)$ (●) e $SL_1(0, 1, 4)$ (linha tracejada).

Da Figura 17 vemos que as distribuições t -Student e *Slash* possuem caudas mais pesadas que distribuição Normal. Mais especificamente, quanto menor o valor para ν , que denota o número de graus de liberdade na distribuição t -Student e é o parâmetro de forma na distribuição *Slash*, mais pesadas são as caudas destas distribuições. Em ambos os casos, a distribuição Normal é o caso limite quando ν tende a infinito. O caso particular da distribuição t -Student em que $\nu = 1$ equivale a distribuição Cauchy. Note que é possível obtermos distribuições t -Student com caudas mais pesadas do que a distribuição Cauchy bastando, para isto, considerarmos valores para ν menores do que 1.

Na Figura 18 apresentamos os gráficos das f.d.p da distribuição Normal Contaminada para diferentes valores dos parâmetros ν_1 e ν_2 .

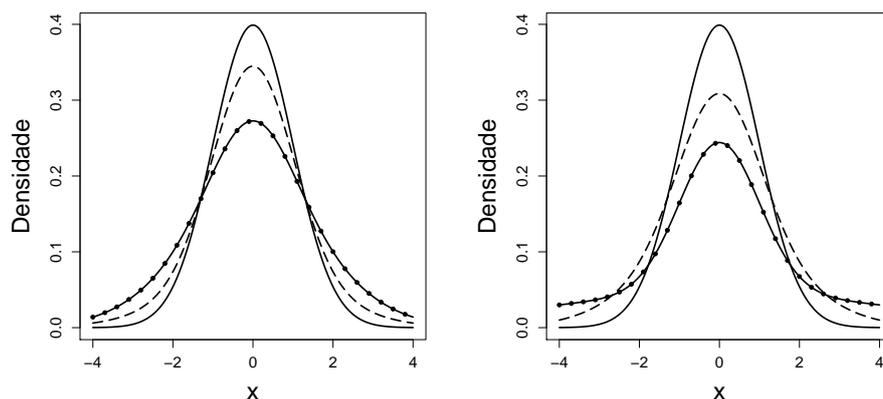


Figura 18: Densidade da distribuição Normal Contaminada. No gráfico à esquerda as distribuições Normal padrão (linha sólida), $NC_1(0; 1; 0,7; 0,3)$ (●) e $NC_1(0; 1; 0,3; 0,3)$ (linha tracejada). No gráfico à direita as distribuições Normal padrão (linha sólida), $NC_1(0; 1; 0,5; 0,05)$ (●) e $NC_1(0; 1; 0,5; 0,3)$ (linha tracejada).

Observamos na Figura 18 que esta distribuição também possui caudas mais pesadas

do que a distribuição Normal. Nesta distribuição, o parâmetro ν_1 pode ser interpretado como a probabilidade de obtermos valores atípicos e o parâmetro ν_2 como um fator de aumento na variabilidade da distribuição dos valores atípicos. Distribuições com caudas mais pesadas podem ser obtidas assumindo ou valores maiores para ν_1 , ou valores menores para ν_2 . A distribuição Normal é obtida nos casos particulares em que $\nu_1 = 0$ ou $\nu_2 = 1$.

Visto que tanto a distribuição t -Student quanto as distribuições *Slash* e Normal Contaminada possuem caudas mais pesadas do que a distribuição Normal, uma pergunta natural é qual entre elas tem caudas mais pesadas. Isso equivale a questionar qual delas é mais flexível para o ajuste de dados com presença de valores atípicos. Não podemos responder esta questão com precisão uma vez que esta comparação dependerá dos valores atribuídos aos parâmetros das distribuições (ver Figuras 17 e 18) e diferentes combinações destes valores podem levar a conclusões conflitantes. Mais detalhes sobre a classe de distribuição obtida da mistura na escala da distribuição Normal podem ser encontrados, por exemplo, em Lange e Sinsheimer (1993).

3.3 Modelo proposto

Suponha que a população esteja dividida em *clusters*, dentre os quais k deles são selecionados independentemente. Para cada *cluster* i selecionado, em que $i = 1, \dots, k$, obtemos uma amostra de n_i indivíduos. Faça $N = \sum_{i=1}^k n_i$. Suponha que cada indivíduo j do *cluster* i seja submetido a n_{ij} experimentos Bernoulli independentes, com probabilidade de sucesso constante para o indivíduo, a qual é denotada por π_{ij} . Seja y_{ij} o número de sucessos observados para tal indivíduo. Assim $y_{ij} \mid \pi_{ij} \sim \text{Binomial}(n_{ij}, \pi_{ij})$. Sejam, $\mathbf{x}_{ij} = (1, x_{ij1}, x_{ij2}, \dots, x_{ijp})^t$ e $\mathbf{z}_{ij} = (1, z_{ij1}, z_{ij2}, \dots, z_{ijq})^t$ vetores coluna de covariáveis de ordem $p+1$ e $q+1$, respectivamente, em que \mathbf{x}_{ij} está associado aos efeitos fixos e \mathbf{z}_{ij} aos aleatórios para o j -ésimo indivíduo do *cluster* i . Assuma que \mathbf{X} e \mathbf{Z} são as matrizes de ordens $N \times (p+1)$ e $N \times (q+1)$ contendo, respectivamente, as informações relacionadas as covariáveis para os efeitos fixos e aleatórios, para todos os N indivíduos observados. Assuma também que $\mathbf{n} = (n_{11}, \dots, n_{1n_1}, \dots, n_{k1}, \dots, n_{kn_k})^t$ e $\mathbf{D} = (\mathbf{X}, \mathbf{Z}, \mathbf{n})$.

Defina $\boldsymbol{\gamma}_i = (\gamma_{i0}, \gamma_{i1}, \dots, \gamma_{iq})^t \in \mathbb{R}^{q+1}$ como sendo o vetor de efeitos aleatórios associados ao i -ésimo *cluster* e $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)^t \in \mathbb{R}^{p+1}$ um vetor de efeitos fixos. Defina também $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^t, \dots, \boldsymbol{\gamma}_k^t)^t$ o vetor coluna de ordem $k(q+1)$ composto pelos efeitos aleatórios associados a todos os *clusters*. Ao utilizarmos a função de ligação logito para modelarmos π_{ij} e o preditor linear $\eta_{ij} = \mathbf{x}_{ij}^t \boldsymbol{\beta} + \mathbf{z}_{ij}^t \boldsymbol{\gamma}_i$, temos que $\pi_{ij} = \exp\{\eta_{ij}\} [1 + \exp\{\eta_{ij}\}]^{-1}$. Como consequência, condicional em $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ e \mathbf{D} , a função de verossimilhança é

$$f(\mathbf{y} \mid \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{D}) = \prod_{i=1}^k \prod_{j=1}^{n_i} \binom{n_{ij}}{y_{ij}} \left[\frac{\exp\{\eta_{ij}\}}{1 + \exp\{\eta_{ij}\}} \right]^{y_{ij}} \left[\frac{1}{1 + \exp\{\eta_{ij}\}} \right]^{n_{ij} - y_{ij}}, \quad (3.4)$$

em que $\mathbf{y} = (y_{11}, \dots, y_{1n_1}, \dots, y_{k1}, \dots, y_{kn_k})^t$ é um vetor com valor observado para variável resposta para todos os indivíduos de todos os *clusters*.

Este modelo é mais geral do que o modelo que apresentamos em (1.1) no Capítulo 1. Para ser mais específico, retornamos à função de verossimilhança em (1.1) quando assumimos $n_{ij} = 1$ em (3.4), ou seja, quando y_{ij} é uma variável binária para todo $i = 1, \dots, k$ e todo $j = 1, \dots, n_i$, e quando $\boldsymbol{\gamma}_i = \gamma_{i0}$ é um escalar.

3.4 Resultados para a razão de chances

Assim como no Capítulo 1, aqui também temos o interesse em comparar dois indivíduos j_1 e j_2 em diferentes *clusters* i_1 e i_2 , respectivamente, através da razão de chances. Considerando o modelo descrito em (3.4) temos que a razão de chances entre estes indivíduos é

$$OR_{12} = \exp \{ \kappa_{12} + W_{12} \}, \quad (3.5)$$

em que $\kappa_{12} = (\mathbf{x}_{i_1 j_1}^t - \mathbf{x}_{i_2 j_2}^t) \boldsymbol{\beta}$ e $W_{12} = \mathbf{z}_{i_1 j_1}^t \boldsymbol{\gamma}_{i_1} - \mathbf{z}_{i_2 j_2}^t \boldsymbol{\gamma}_{i_2}$.

Estendendo os resultados propostos em Larsen *et al.* (2000) especificamos uma distribuição de probabilidade para o vetor $\boldsymbol{\gamma}$ que seja mais geral do que a distribuição Normal. Em um contexto geral, se W_{12} é uma variável aleatória contínua com f.d.p e f.d.a conhecidas, dado $\boldsymbol{\beta}$, a Proposição 3.2 nos fornece a distribuição de OR_{12} e resultados relacionados à *MOR*, condicional nos efeitos fixos.

PROPOSIÇÃO 3.2. *Se W_{12} é uma variável aleatória contínua tal que, dados \mathbf{Z} e os hiperparâmetros $\boldsymbol{\theta}_{12}$, tem f.d.p $f_{W_{12}|\mathbf{Z},\boldsymbol{\theta}_{12}}(\cdot)$ e f.d.a $F_{W_{12}|\mathbf{Z},\boldsymbol{\theta}_{12}}(\cdot)$ então*

- (i) *A quantidade aleatória $OR_{12} = \exp \{ \kappa_{12} + W_{12} \}$, dados $\mathbf{D} = (\mathbf{X}, \mathbf{Z}, \mathbf{n})$, $\boldsymbol{\beta}$ e $\boldsymbol{\theta}_{12}$, tem f.d.p dada por*

$$f_{OR_{12}|\mathbf{D},\boldsymbol{\theta}_{12},\boldsymbol{\beta}}(r) = \frac{1}{r} f_{W_{12}|\mathbf{Z},\boldsymbol{\theta}_{12}}(\ln r - \kappa_{12}), \quad r \in \mathbb{R}_+; \quad (3.6)$$

- (ii) *a mediana da distribuição em (3.6) é tal que*

$$(a) \quad MOR_{12} = \text{med} \{ \exp \{ \kappa_{12} + W_{12} \} \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta} \} = \exp \left\{ \kappa_{12} + F_{W_{12}|\mathbf{Z},\boldsymbol{\theta}_{12}}^{-1}(0,5) \right\};$$

$$(b) \quad MOR_{|12|} = \text{med} \{ \exp \{ |W_{12}| \} \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta} \} = \exp \{ \text{med} \{ |W_{12}| \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta} \} \} \\ = \exp \left\{ F_{|W_{12}|\mathbf{Z},\boldsymbol{\theta}_{12}}^{-1}(0,5) \right\}, \text{ em que } F_{|W_{12}|\mathbf{Z},\boldsymbol{\theta}_{12}}(\cdot) \text{ é a f.d.a da variável aleatória } |W_{12}| \text{ condicional em } \mathbf{Z} \text{ e } \boldsymbol{\theta}_{12};$$

- (iii) *Se $f_{W_{12}|\mathbf{Z},\boldsymbol{\theta}_{12}}(\cdot)$ é uma f.d.p simétrica em torno de μ então*

$$MOR_{12} = \text{med} \{ \exp \{ \kappa_{12} + W_{12} \} \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta} \} = \exp \{ \kappa_{12} + \mu \};$$

- (iv) *Se $f_{W_{12}|\mathbf{Z},\boldsymbol{\theta}_{12}}(\cdot)$ é uma f.d.p simétrica em torno de zero então*

$$(a) \quad MOR_{|12|} = \text{med} \{ \exp \{ |W_{12}| \} \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta} \} = \exp \{ \text{med} \{ |W_{12}| \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta} \} \} \\ = \exp \left\{ F_{|W_{12}|\mathbf{Z},\boldsymbol{\theta}_{12}}^{-1}(0,75) \right\};$$

$$(b) \quad \text{med} \{ \exp \{ -|W_{12}| \} \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta} \} = [\text{med} \{ \exp \{ |W_{12}| \} \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta} \}]^{-1}.$$

DEMONSTRAÇÃO.

- (i) Como $OR_{12} = \exp\{\kappa_{12} + W_{12}\}$, segue que $W_{12} = \ln OR_{12} - \kappa_{12}$. Daí, usando o método do Jacobiano, segue que a f.d.p de OR_{12} no ponto $r \in \mathbb{R}_+$, dados \mathbf{D} , $\boldsymbol{\beta}$ e $\boldsymbol{\theta}_{12}$, é

$$\begin{aligned} f_{OR_{12}|\mathbf{D},\boldsymbol{\theta}_{12},\boldsymbol{\beta}}(r) &= f_{W_{12}|\mathbf{Z},\boldsymbol{\theta}_{12}}(\ln r - \kappa_{12}) \left| \frac{dW_{12}}{dr} \right| \\ &= \frac{1}{r} f_{W_{12}|\mathbf{Z},\boldsymbol{\theta}_{12}}(\ln r - \kappa_{12}). \end{aligned}$$

- (ii) (a) Seja $a = \text{med}\{\exp\{\kappa_{12} + W_{12}\} \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta}\}$. Assim, temos que

$$\begin{aligned} F_{OR_{12}|\mathbf{D},\boldsymbol{\theta}_{12},\boldsymbol{\beta}}(a) &= 0,5 \\ P(\exp\{\kappa_{12} + W_{12}\} \leq a \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta}) &= 0,5 \\ P(W_{12} \leq \ln a - \kappa_{12} \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta}) &= 0,5. \end{aligned}$$

Logo, $a = \exp\left\{\kappa_{12} + F_{W_{12}|\mathbf{Z},\boldsymbol{\theta}_{12}}^{-1}(0,5)\right\}$.

- (b) Seja $a_1 = \text{med}\{\exp\{|W_{12}|\} \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta}\}$. Assim,

$$\begin{aligned} P(\exp\{|W_{12}|\} \leq a_1 \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta}) &= 0,5 \\ P(|W_{12}| \leq \ln a_1 \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta}) &= 0,5 \\ F_{|W_{12}| \mid \mathbf{Z},\boldsymbol{\theta}_{12}}(\ln a_1) &= 0,5. \end{aligned}$$

Logo, $a_1 = \exp\left\{F_{|W_{12}| \mid \mathbf{Z},\boldsymbol{\theta}_{12}}^{-1}(0,5)\right\}$. A outra parte da igualdade, segue do uso direto da Proposição A.2 no Apêndice A.

- (iii) Segue de (ii) que $\text{med}\{\exp\{\kappa_{12} + W_{12}\} \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta}\} = \exp\left\{F_{W_{12}|\mathbf{Z},\boldsymbol{\theta}_{12}}^{-1}(0,5) + \kappa_{12}\right\}$. Como, condicional em \mathbf{Z} e $\boldsymbol{\theta}_{12}$, W_{12} é uma variável aleatória com densidade simétrica em torno de μ , temos que $F_{W_{12}|\mathbf{Z},\boldsymbol{\theta}_{12}}^{-1}(0,5) = \mu$. Desta forma, $\text{med}\{\exp\{\kappa_{12} + W_{12}\} \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta}\} = \exp\{\kappa_{12} + \mu\}$.

- (iv) (a) Seja $a = F_{|W_{12}| \mid \mathbf{Z},\boldsymbol{\theta}_{12}}^{-1}(0,75)$. Como W_{12} , dados \mathbf{Z} e $\boldsymbol{\theta}_{12}$, tem f.d.p simétrica em torno de zero, temos que

$$\begin{aligned} P(|W_{12}| \leq a \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta}) &= 0,5 \\ P(-a \leq W_{12} \leq a \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta}) &= 0,5 \\ F_{W_{12}|\mathbf{Z},\boldsymbol{\theta}_{12}}(a) - F_{W_{12}|\mathbf{Z},\boldsymbol{\theta}_{12}}(-a) &= 0,5 \\ F_{W_{12}|\mathbf{Z},\boldsymbol{\theta}_{12}}(a) &= 0,75. \end{aligned}$$

Logo, $a = F_{|W_{12}| \mid \mathbf{Z},\boldsymbol{\theta}_{12}}^{-1}(0,75)$. Substituindo este resultado no item (iii) segue que $\text{med}\{\exp\{|W_{12}|\} \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta}\} = \exp\left\{F_{|W_{12}| \mid \mathbf{Z},\boldsymbol{\theta}_{12}}^{-1}(0,75)\right\}$.

(b) Seja $h = \text{med}\{\exp\{-|W_{12}|\} \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta}\}$. Assim,

$$\begin{aligned} P(\exp\{-|W_{12}|\} \leq h \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta}) &= 0,5 \\ P(W_{12} \geq -\ln h \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta}) + P(W_{12} \leq \ln h \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta}) &= 0,5 \\ 2P(W_{12} \leq \ln h \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta}) &= 0,5 \\ F_{W_{12} \mid \mathbf{z}, \boldsymbol{\theta}_{12}}(\ln h) &= 0,25. \end{aligned}$$

Desta forma, segue que

$$\begin{aligned} h &= \exp \left\{ F_{W_{12} \mid \mathbf{z}, \boldsymbol{\theta}_{12}}^{-1}(0,25) \right\} \\ h &= \exp \left\{ -F_{W_{12} \mid \mathbf{z}, \boldsymbol{\theta}_{12}}^{-1}(0,75) \right\} \\ h &= [\text{med}\{\exp\{|W_{12}|\} \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta}\}]^{-1}, \end{aligned}$$

portanto, $\text{med}\{\exp\{-|W_{12}|\} \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta}\} = [\text{med}\{\exp\{|W_{12}|\} \mid \mathbf{D}, \boldsymbol{\theta}_{12}, \boldsymbol{\beta}\}]^{-1}$. □

Perceba de (iv) que sempre que a distribuição de W_{12} for simétrica ao redor de zero a mediana da razão de chances será $\exp\{\kappa_{12}\}$. Assim, a comparação dos indivíduos não dependerá dos *clusters* a que pertencem.

Do ponto de vista bayesiano, a construção da distribuição *a priori* sobre $\boldsymbol{\beta}$ não é uma tarefa fácil. Em muitos dos casos, a informação sobre $\boldsymbol{\beta}$ que está disponível é de forma indireta através da razão de chances. Para ajudar a determinar distribuições *a priori* para $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$, pode ser útil conhecermos características da distribuição de OR_{12} dados \mathbf{D} e $\boldsymbol{\theta}_{12}$. Para facilitar a notação, a expressão para OR_{12} em (3.5) será escrita como

$$OR_{12} = \exp\{W_{12}^*\}, \quad (3.7)$$

em que $W_{12}^* = \kappa_{12} + W_{12} = (\mathbf{x}_{i_1 j_1}^t - \mathbf{x}_{i_2 j_2}^t)\boldsymbol{\beta} + \mathbf{z}_{i_1 j_1}^t \boldsymbol{\gamma}_{i_1} - \mathbf{z}_{i_2 j_2}^t \boldsymbol{\gamma}_{i_2}$.

Após especificarmos as distribuições de probabilidade conjunta para os vetores $\boldsymbol{\gamma}$ e $\boldsymbol{\beta}$, podemos obter a f.d.p para a variável aleatória W_{12}^* . Na Proposição 3.3, alguns resultados relacionados à distribuição da razão de chances e sua mediana são obtidos.

PROPOSIÇÃO 3.3. *Se W_{12}^* é uma variável aleatória contínua que, condicional em \mathbf{D} e nos hiperparâmetros $\boldsymbol{\theta}_{12}$, possui f.d.p $f_{W_{12}^* \mid \mathbf{D}, \boldsymbol{\theta}_{12}}(\cdot)$ e f.d.a $F_{W_{12}^* \mid \mathbf{D}, \boldsymbol{\theta}_{12}}(\cdot)$ então*

(i) $OR_{12} = \exp\{W_{12}^*\}$, dados \mathbf{D} e $\boldsymbol{\theta}_{12}$, tem f.d.p dada por

$$f_{OR_{12} \mid \mathbf{D}, \boldsymbol{\theta}_{12}}(r) = \frac{1}{r} f_{W_{12}^* \mid \mathbf{D}, \boldsymbol{\theta}_{12}}(\ln r), \quad r \in \mathbb{R}_+; \quad (3.8)$$

(ii) A mediana da f.d.p em (3.8) é dada por

$$MOR_{12} = \text{med}\{\exp\{W_{12}^*\} \mid \mathbf{D}, \boldsymbol{\theta}_{12}\} = \exp\left\{F_{W_{12}^* \mid \mathbf{D}, \boldsymbol{\theta}_{12}}^{-1}(0,5)\right\};$$

(iii) Se $f_{W_{12}^*|D, \theta_{12}}(\cdot)$ é uma f.d.p simétrica em torno de μ então

$$MOR_{12} = \text{med}\{\exp\{W_{12}^*\} \mid \mathbf{D}, \theta_{12}\} = \exp\{\mu\}.$$

A demonstração desta proposição é similar a demonstração da Proposição 3.2 e, por isso, será omitida do texto.

3.4.1 Razão de chances no modelo logístico elíptico

Considere o modelo em (3.4) e assuma que, condicional na matriz de escala Σ , o vetor de efeitos aleatórios tem distribuição elíptica tal que $\gamma \sim El_n(\mathbf{0}, \Sigma; h)$, cuja f.d.p é dada em (3.1). Do resultado em (3.5) temos que $OR_{12} = \exp\{\kappa_{12} + W_{12}\}$, sendo que $\kappa_{12} = (\mathbf{x}_{i_1 j_1}^t - \mathbf{x}_{i_2 j_2}^t)\beta$ e $W_{12} = \mathbf{z}_{i_1 j_1}^t \gamma_{i_1} - \mathbf{z}_{i_2 j_2}^t \gamma_{i_2}$. A variável aleatória W_{12} pode ser reescrita como $W_{12} = \mathbf{z}_{12}^t \gamma_{12}$, em que $\mathbf{z}_{12} = (\mathbf{z}_{i_1 j_1}^t, -\mathbf{z}_{i_2 j_2}^t)^t$ e $\gamma_{12} = (\gamma_{i_1}^t, \gamma_{i_2}^t)^t$. Decorre do Teorema 3.2 que $\gamma_{12} \sim El_2(\mathbf{0}, \Sigma_{12}; h)$, em que Σ_{12} denota a matriz de escala que define a covariância entre os vetores γ_{i_1} e γ_{i_2} . Consequentemente, aplicando o Teorema 3.1, obtemos que $W_{12} \sim El_1(0, \sigma_{12}^2; h)$, com $\sigma_{12}^2 = \mathbf{z}_{12}^t \Sigma_{12} \mathbf{z}_{12}$. Desta forma, uma vez que a distribuição de W_{12} é simétrica em torno de zero, segue diretamente da Proposição 3.2 o seguinte resultado. Por simplicidade, como em (3.4), $\mathbf{D} = (\mathbf{X}, \mathbf{Z}, \mathbf{n})$.

PROPOSIÇÃO 3.4. Se γ é um vetor aleatório com distribuição $\gamma \sim El_{k(q+1)}(\mathbf{0}, \Sigma; h)$ então, dados \mathbf{D} , β e Σ , segue que

a) a variável aleatória OR_{12} tem distribuição log-elíptica com parâmetros de locação $\kappa_{12} \in \mathbb{R}$ e escala $\sigma_{12}^2 \in \mathbb{R}_+$, denotada por $LEl_1(\kappa_{12}, \sigma_{12}^2; h)$, cuja f.d.p é

$$f_{OR_{12}|D, \beta, \Sigma}(r) = \frac{1}{r \sigma_{12}} h \left(\frac{(\ln r - \kappa_{12})^2}{\sigma_{12}^2} \right), \quad r \in \mathbb{R}_+;$$

b) a razão de chances mediana é tal que

$$(i) \quad MOR_{|12|} = \text{med}\{\exp\{|W_{12}|\} \mid \mathbf{D}, \beta, \Sigma\} = \exp\{\text{med}\{|W_{12}| \mid \mathbf{D}, \beta, \Sigma\}\} \\ = \exp\{F_{El_1 h}^{-1}(0, 75)\}, \text{ em que } F_{El_1 h}(\cdot) \text{ é a f.d.a da distribuição } El_1(0, \sigma_{12}^2; h);$$

$$(ii) \quad \text{med}\{\exp\{-|W_{12}|\} \mid \mathbf{D}, \beta, \Sigma\} = \text{med}\{\exp\{|W_{12}| \mid \mathbf{D}, \beta, \Sigma\}\}^{-1};$$

$$(iii) \quad MOR_{12} = \text{med}\{\exp\{\kappa_{12} + W_{12}\} \mid \mathbf{D}, \beta, \Sigma\} = \exp\{\kappa_{12}\}.$$

O modelo logístico com intercepto aleatório elíptico segue como caso particular do modelo em (3.4) quando assumimos $\mathbf{z}_{12} = (1, -1)^t$ e $\gamma_{12} = (\gamma_{i_1}, \gamma_{i_2})^t$. Além disto, os resultados apresentados em Larsen *et al.* (2000) são casos particulares do Teorema 3.4, caso para o qual a distribuição Normal é a distribuição elíptica utilizada. Outros casos particulares do Teorema 3.4 são apresentados no Corolário 3.1.

COROLÁRIO 3.1. Condicional em \mathbf{D} , β e Σ (e em ν , ν_1 e ν_2 quando for o caso), segue que

(i) Se $\gamma \sim T_{k(q+1)}(\mathbf{0}, \mathbf{\Sigma}, \nu)$ então OR_{12} tem distribuição log-*t*-Student com parâmetros $\kappa_{12} \in \mathbb{R}$, $\sigma_{12}^2 \in \mathbb{R}_+$ e $\nu \in \mathbb{R}_+$, denotada por $LT_1(\kappa_{12}, \sigma_{12}^2, \nu)$, cuja f.d.p é dada por

$$f_{OR_{12}|\mathbf{D}, \mathbf{\beta}, \mathbf{\Sigma}}(r) = \frac{\Gamma((\nu+1)/2)}{r\sigma_{12}\Gamma(\nu/2)(\pi\nu)^{1/2}} \left[1 + \frac{(\ln r - \kappa_{12})^2}{\nu\sigma_{12}} \right]^{-(\nu+1)/2};$$

(ii) Se $\gamma \sim SL_{k(q+1)}(\mathbf{0}, \mathbf{\Sigma}, \nu)$ então OR_{12} tem distribuição log-*Slash* com parâmetros $\kappa_{12} \in \mathbb{R}$, $\sigma_{12}^2 \in \mathbb{R}_+$ e $\nu \in \mathbb{R}_+$, denotada por $LSL_1(\kappa_{12}, \sigma_{12}^2, \nu)$, com f.d.p dada por

$$f_{OR_{12}|\mathbf{D}, \mathbf{\beta}, \mathbf{\Sigma}}(r) = \frac{\nu}{r} \int_0^1 \phi(\ln r | \kappa_{12}, u^{-1}\sigma_{12}^2) du;$$

(iii) Se $\gamma \sim NC_{k(q+1)}(\mathbf{0}, \mathbf{\Sigma}, \nu_1, \nu_2)$ então OR_{12} tem distribuição log-*Normal-contaminada* com parâmetros $\kappa_{12} \in \mathbb{R}$, $\sigma_{12}^2 \in \mathbb{R}_+$, $0 \leq \nu_1 \leq 1$ e $0 < \nu_2 \leq 1$, denotada por $LNC_1(\kappa_{12}, \sigma_{12}^2, \nu_1, \nu_2)$, com f.d.p dada por

$$f_{OR_{12}|\mathbf{D}, \mathbf{\beta}, \mathbf{\Sigma}}(r) = \frac{1}{r} [\nu_1 \phi(\ln r | \kappa_{12}, \nu_2^{-1}\sigma_{12}^2) + (1 - \nu_1) \phi(\ln r | \kappa_{12}, \sigma_{12}^2)].$$

Note que neste capítulo, em geral, a distribuição assumida para γ impõe dependência entre os efeitos aleatórios γ_{i_1} e γ_{i_2} . A exceção a este fato ocorre no caso em que a distribuição elíptica usada é a distribuição Normal com matriz de escala bloco-diagonal. Neste caso obtemos independência entre os efeitos aleatórios para os *clusters*. Nos demais casos, a independência entre γ_{i_1} e γ_{i_2} não decorre da estrutura bloco diagonal para a matriz de escala mas, apesar da relação de dependência, podemos definir $\mathbf{\Sigma}$ de tal forma que γ_{i_1} e γ_{i_2} sejam não correlacionados.

Se considerarmos o caso particular do modelo em que apenas o intercepto é aleatório, isto é, $\gamma = (\gamma_1, \dots, \gamma_k)^t$, e se assumirmos efeitos aleatórios γ_i 's independentes e com distribuição na classe de distribuições elípticas, muitas vezes não conseguimos obter a distribuição de probabilidades de combinações lineares do vetor de efeitos aleatórios γ de forma analítica. Exceções ocorrem para as distribuições Normal e Normal Contaminada. Sendo assim, uma vez que nosso interesse recai sobre a combinação linear $W_{12} = \gamma_{i_1} - \gamma_{i_2}$, não é possível encontrarmos a f.d.p da razão de chances OR_{12} analiticamente. No entanto, podemos facilmente obter uma amostra da distribuição de probabilidades de W_{12} gerando amostras de γ_{i_1} e γ_{i_2} . Através desta amostra gerada podemos estimar quantidades relacionadas a f.d.p de W_{12} . Então, neste caso particular, podemos usar a Proposição 3.5, apresentada a seguir, para calcular MOR_{12} e $MOR_{|12|}$.

PROPOSIÇÃO 3.5. *Em um modelo com intercepto aleatório, se γ é um vetor aleatório tal que $\gamma_i \stackrel{iid}{\sim} El_1(0, \sigma^2; h)$ então, dados \mathbf{D} , $\mathbf{\beta}$ e σ^2 , segue que*

$$MOR_{12} = \exp\{\kappa_{12}\} \quad e \quad MOR_{|12|} = \exp\left\{F_{W_{12}|\mathbf{Z}, \boldsymbol{\theta}_{12}}^{-1}(0, 75)\right\},$$

em que $F_{W_{12}|\mathbf{Z}, \boldsymbol{\theta}_{12}}(\cdot)$ é a f.d.a da variável aleatória W_{12} dado \mathbf{Z} e $\boldsymbol{\theta}_{12} = (\sigma^2, \nu)^t$.

A demonstração da Proposição 3.5 segue diretamente da Proposição A.1 de onde segue que $W_{12} = \gamma_{i_1} - \gamma_{i_2}$ é uma variável aleatória simétrica em torno de zero e dos itens (iii) e (iv) da Proposição 3.2.

Particularmente, se assumimos que $\gamma_i \stackrel{iid}{\sim} NC(0, \sigma^2, \nu_1, \nu_2)$, podemos mostrar que a variável aleatória W_{12} tem f.d.p e f.d.a dadas por

$$\begin{aligned} f_{W_{12}|\sigma^2, \nu_1, \nu_2}(w) &= \nu_1^2 \phi_1 \left(w \mid 0, \frac{2\sigma^2}{\nu_2} \right) + 2\nu_1(1 - \nu_1) \phi_1 \left(w \mid 0, \frac{(1 + \nu_2)\sigma^2}{\nu_2} \right) \\ &\quad + (1 - \nu_1)^2 \phi_1 \left(w \mid 0, 2\sigma^2 \right) \text{ e} \\ F_{W_{12}|\sigma^2, \nu_1, \nu_2}(w) &= \nu_1^2 \Phi_1 \left(w \mid 0, \frac{2\sigma^2}{\nu_2} \right) + 2\nu_1(1 - \nu_1) \Phi_1 \left(w \mid 0, \frac{(1 + \nu_2)\sigma^2}{\nu_2} \right) \\ &\quad + (1 - \nu_1)^2 \Phi_1 \left(w \mid 0, 2\sigma^2 \right), \end{aligned}$$

respectivamente. Neste caso, a quantidade $F_{W_{12}|\mathbf{z}, \boldsymbol{\theta}_{12}}^{-1}(0, 75)$ na Proposição 3.5 pode ser calculada resolvendo numericamente a igualdade $F_{W_{12}|\sigma^2, \nu_1, \nu_2}(w) = 0,75$.

Outro resultado importante desta seção, utilizado apenas sob o enfoque bayesiano, diz respeito a obtermos a distribuição *a priori* de OR_{12} sem condicionarmos nos efeitos fixos $\boldsymbol{\beta}$. Para isto, é necessário adotarmos uma distribuição *a priori* conjunta para $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$.

Defina o vetor coluna $\zeta = (\boldsymbol{\beta}^t, \boldsymbol{\gamma}^t)^t$ de dimensão $m_1 = (p + 1) + k(q + 1)$. Assuma que $\zeta \sim El_{m_1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}; h)$, em que

$$\boldsymbol{\mu}^t = (\boldsymbol{\mu}_\beta^t, \boldsymbol{\mu}_{\gamma_1}^t, \dots, \boldsymbol{\mu}_{\gamma_k}^t) \quad \text{e} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_\beta & \boldsymbol{\Sigma}_{\beta 1} & \cdots & \boldsymbol{\Sigma}_{\beta k} \\ \boldsymbol{\Sigma}_{\beta 1}^t & \boldsymbol{\Sigma}_{11} & \cdots & \boldsymbol{\Sigma}_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{\beta k}^t & \boldsymbol{\Sigma}_{1k}^t & \cdots & \boldsymbol{\Sigma}_{kk} \end{pmatrix}, \quad (3.9)$$

$\boldsymbol{\mu}_\beta \in \mathbb{R}^{p+1}$ e $\boldsymbol{\mu}_{\gamma_i} \in \mathbb{R}^{q+1}$ são vetores de locação relacionados a $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}_i$, $i = 1, \dots, k$, respectivamente, $\boldsymbol{\Sigma}$ é uma matriz de escala simétrica positiva definida, $\boldsymbol{\Sigma}_\beta$ é uma matriz quadrada de ordem $(p + 1)$ trazendo informação sobre covariância entre os elementos de $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_{\beta i}$ é uma matriz $(p + 1) \times (q + 1)$ com informação sobre covariância entre $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}_i$ e $\boldsymbol{\Sigma}_{i_1 i_2}$ é uma matriz quadrada de ordem $(q + 1)$ com informação sobre covariância entre $\boldsymbol{\gamma}_{i_1}$ e $\boldsymbol{\gamma}_{i_2}$. É importante ressaltar que, por exemplo, $\boldsymbol{\Sigma}_\beta$ não é exatamente a matriz de covariâncias assumida *a priori* para $\boldsymbol{\beta}$ pois, em geral, a matriz de covariância nas distribuições elípticas também é função dos parâmetros de forma. Isto ocorre, por exemplo, nas distribuições *t*-Student e *Slash*. O mesmo segue para $\boldsymbol{\Sigma}_{\beta i}$ e $\boldsymbol{\Sigma}_{i_1 i_2}$.

De (3.7) segue que a razão de chances é

$$OR_{12} = \exp \{W_{12}^*\}, \quad (3.10)$$

em que $W_{12}^* = \mathbf{x}_{12}^{*t} \zeta_{12}$, $\mathbf{x}_{12}^* = ((\mathbf{x}_{i_1 j_1} - \mathbf{x}_{i_2 j_2})^t, \mathbf{z}_{i_1 j_1}^t, -\mathbf{z}_{i_2 j_2}^t)^t$ e $\zeta_{12} = (\boldsymbol{\beta}^t, \boldsymbol{\gamma}_{i_1}^t, \boldsymbol{\gamma}_{i_2}^t)^t$.

Como W_{12}^* é uma transformação linear de ζ_{12} se $\zeta_{12} \sim El_{m_{12}}(\boldsymbol{\mu}_{12}, \boldsymbol{\Sigma}_{12}^*; h)$, onde $m_{12} = p + 2q + 3$,

$$\boldsymbol{\mu}_{12} = (\boldsymbol{\mu}_\beta^t, \boldsymbol{\mu}_{\gamma_{i_1}}^t, \boldsymbol{\mu}_{\gamma_{i_2}}^t)^t \quad \text{e} \quad \boldsymbol{\Sigma}_{12}^* = \begin{pmatrix} \boldsymbol{\Sigma}_\beta & \boldsymbol{\Sigma}_{\beta i_1} & \boldsymbol{\Sigma}_{\beta i_2} \\ \boldsymbol{\Sigma}_{\beta i_1}^t & \boldsymbol{\Sigma}_{i_1 i_1} & \boldsymbol{\Sigma}_{i_1 i_2} \\ \boldsymbol{\Sigma}_{\beta i_2}^t & \boldsymbol{\Sigma}_{i_1 i_2}^t & \boldsymbol{\Sigma}_{i_2 i_2} \end{pmatrix},$$

então, como consequência, segue do Teorema 3.1 que

$$W_{12}^* \sim El_1(\mu_{12}^*, \sigma_{12}^{*2}; h), \quad (3.11)$$

em que $\mu_{12}^* = \mathbf{x}_{12}^{*t} \boldsymbol{\mu}_{12}$, $\sigma_{12}^{*2} = \mathbf{x}_{12}^{*t} \boldsymbol{\Sigma}_{12}^* \mathbf{x}_{12}^*$ e \mathbf{x}_{12}^* é definido como em (3.10). Como consequência e considerando a Proposição 3.3 pode-se diretamente obter a distribuição de OR_{12} , condicional em $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ e \mathbf{D} , neste caso. Este resultado é apresentado na Proposição 3.6.

PROPOSIÇÃO 3.6. *Se $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ são variáveis aleatórias tais que $(\boldsymbol{\beta}^t, \boldsymbol{\gamma}^t)^t \sim El_{m_1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}; h)$, com \mathbf{D} , $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ definidos como em (3.9), então, dados $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ e \mathbf{D} , segue que*

a) *a variável aleatória OR_{12} tem distribuição log-elíptica com parâmetros de locação μ_{12}^* e escala σ_{12}^{*2} , denotada por $LEl_1(\mu_{12}^*, \sigma_{12}^{*2}; h)$, cuja f.d.p é dada por*

$$f_{OR_{12}|\mathbf{D}, \boldsymbol{\mu}, \boldsymbol{\Sigma}}(r) = \frac{1}{r \sigma_{12}^*} h \left(\frac{(\ln r - \mu_{12}^*)^2}{\sigma_{12}^{*2}} \right), \quad r \in \mathbb{R}_+, \quad (3.12)$$

em que μ_{12}^* e σ_{12}^{*2} são definidos em (3.11);

b) *A mediana da distribuição em (3.12) é dada por*

$$MOR_{12} = \text{med}\{\exp\{W_{12}^*\} \mid \mathbf{D}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\} = \exp\{\mu_{12}^*\}.$$

Na maior parte das vezes, artigos disponíveis na literatura estatística que abordam assuntos sobre modelos mistos assumem independência entre $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ e, usualmente assume-se distribuição normal para $\boldsymbol{\gamma}$. No enfoque bayesiano, usualmente também se considera normalidade para $\boldsymbol{\beta}$. Este e alguns outros casos particulares da Proposição 3.6 são apresentados no seguinte corolário.

COROLÁRIO 3.2. *Condicional em \mathbf{D} , $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ (e em ν , ν_1 e ν_2 quando for o caso), segue que*

- (i) *Se $(\boldsymbol{\beta}^t, \boldsymbol{\gamma}^t)^t \sim N_{m_1}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, então $OR_{12} \sim \text{log-Normal}(\mu_{12}^*, \sigma_{12}^{*2})$;*
- (ii) *Se $(\boldsymbol{\beta}^t, \boldsymbol{\gamma}^t)^t \sim T_{m_1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, então $OR_{12} \sim LT_1(\mu_{12}^*, \sigma_{12}^{*2}, \nu)$;*
- (iii) *Se $(\boldsymbol{\beta}^t, \boldsymbol{\gamma}^t)^t \sim SL_{m_1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, então $OR_{12} \sim LSL_1(\mu_{12}^*, \sigma_{12}^{*2}, \nu)$;*
- (iv) *Se $(\boldsymbol{\beta}^t, \boldsymbol{\gamma}^t)^t \sim NC_{m_1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu_1, \nu_2)$, então $OR_{12} \sim LNC_1(\mu_{12}^*, \sigma_{12}^{*2}, \nu_1, \nu_2)$*
- (v) *Se $\boldsymbol{\beta}$ e $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^t, \dots, \boldsymbol{\gamma}_k^t)^t$ são independentes tal que $\boldsymbol{\beta} \sim N_{p+1}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$ e $\boldsymbol{\gamma} \sim N_{k(q+1)}(0, \boldsymbol{\Sigma}_\gamma)$, então $OR_{12} \sim \text{log-Normal}(M, V)$, em que*

$$\begin{aligned} M &= (\mathbf{X}_{i_1 j_1} - \mathbf{X}_{i_2 j_2})^t \boldsymbol{\mu}_\beta \quad e \\ V &= (\mathbf{X}_{i_1 j_1} - \mathbf{X}_{i_2 j_2})^t \boldsymbol{\Sigma}_\beta (\mathbf{X}_{i_1 j_1} - \mathbf{X}_{i_2 j_2}) + (\mathbf{Z}_{i_1 j_1}^t, -\mathbf{Z}_{i_2 j_2}^t) \boldsymbol{\Sigma}_\gamma (\mathbf{Z}_{i_1 j_1}^t, -\mathbf{Z}_{i_2 j_2}^t)^t; \end{aligned}$$

(vi) Se β e γ são independentes tal que $\beta \sim N_{p+1}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$ e $\gamma \stackrel{iid}{\sim} N_{k(q+1)}(0, \boldsymbol{\Sigma}_\gamma)$, com $\boldsymbol{\Sigma}_\gamma = \text{bloco-diagonal}\{\boldsymbol{\Sigma}_{\gamma_1}, \dots, \boldsymbol{\Sigma}_{\gamma_k}\}$ tal que $\boldsymbol{\Sigma}_{\gamma_r} = \boldsymbol{\Sigma}_{\gamma_s} \forall r, s = 1, \dots, k$, então $OR_{12} \sim \text{log-Normal}(M, V)$, em que

$$\begin{aligned} M &= (\mathbf{X}_{i_1j_1} - \mathbf{X}_{i_2j_2})^t \boldsymbol{\mu}_\beta \quad e \\ V &= (\mathbf{X}_{i_1j_1} - \mathbf{X}_{i_2j_2})^t \boldsymbol{\Sigma}_\beta (\mathbf{X}_{i_1j_1} - \mathbf{X}_{i_2j_2}) + (\mathbf{Z}_{i_1j_1} - \mathbf{Z}_{i_2j_2})^t \boldsymbol{\Sigma}_{\gamma_i} (\mathbf{Z}_{i_1j_1} - \mathbf{Z}_{i_2j_2}). \end{aligned}$$

Para obtermos os resultados referentes à distribuição de OR condicional em β , e sobre as medidas MOR propostas para interpretarmos os efeitos fixos e para quantificarmos a heterogeneidade entre os *clusters*, é necessário conhecermos a distribuição da variável aleatória W_{12} . Em outras palavras, é necessário conhecermos alguns resultados sobre combinações lineares da distribuição assumida para o vetor de efeitos aleatórios γ . Nem sempre tais resultados são conhecidos mas, nestes casos, ainda podemos interpretar os efeitos fixos e quantificar a heterogeneidade presente nos *clusters* no enfoque clássico usando aproximações computacionais para tais resultados. Isto é, após estimarmos β e os hiperparâmetros da distribuição de γ , podemos simular valores da distribuição assumida para γ e obtermos numericamente, tanto a distribuição para a razão de chances quanto para outras quantidades relacionadas a ela, como por exemplo, a mediana.

3.5 Métodos de máxima verossimilhança no modelo logístico misto NI

Nesta seção, discutimos os procedimentos de inferência, considerando apenas o caso particular do modelo apresentado em (3.4) em que apenas o intercepto é aleatório, isto é, consideramos o modelo assumido em (1.1). Para os efeitos aleatórios assumiremos que $\gamma_i \stackrel{iid}{\sim} El_1(0, \sigma^2; h)$, $i = 1, \dots, k$, em que h é a função geradora de densidade de alguma das distribuições pertencentes a subclasse de distribuições NI. Estas distribuições possuem f.d.p como a apresentada em (3.3). Ao longo deste texto, este modelo será genericamente chamado de modelo Logístico-NI.

Utilizando a Proposição 3.1, o efeito aleatório γ_i pode ser estocasticamente representado por

$$\gamma_i \stackrel{d}{=} U_i^{-1/2} \sigma T, \quad (3.13)$$

em que $T \sim N(0, 1)$ e U_i é uma variável positiva com f.d.p $f_{U_i|\nu}(u_i)$. Assim, segue que $\gamma_i|U_i = u_i \sim N(0, u_i^{-1}\sigma^2)$. Diferentes escolhas de distribuições para U_i levam a diferentes distribuições para γ_i e, conseqüentemente, a diferentes modelos. Por exemplo,

- (i) Modelo Logístico-N: se $f_{U_i}(u_i) = \mathbf{1}_{\{u_i=1\}}$, então $\gamma_i \sim N(0, \sigma^2)$;
- (ii) Modelo Logístico-T: se $U_i \sim \text{Gama}(\nu/2, \nu/2)$, então $\gamma_i \sim T_1(0, \sigma^2, \nu)$;
- (iii) Modelo Logístico-SL: se $U_i \sim \text{Beta}(\nu, 1)$, então $\gamma_i \sim SL_1(0, \sigma^2, \nu)$;
- (iv) Modelo Logístico-NC: se $f_{U_i|\nu_1, \nu_2}(u_i) = \nu_1 \mathbf{1}_{\{u_i=\nu_2\}} + (1 - \nu_1) \mathbf{1}_{\{u_i=1\}}$, com $0 \leq \nu_1 \leq 1$ e $0 < \nu_2 \leq 1$, então $\gamma_i \sim NC_1(0, \sigma^2, \nu_1, \nu_2)$.

Para realizarmos inferência sob o ponto de vista clássico, mais precisamente para obtermos os estimadores de máxima verossimilhança (EMV) para β e os parâmetros da distribuição de γ_i , inicialmente é necessário encontrarmos a função de verossimilhança para os dados completos e, em seguida, a função de verossimilhança marginal para os dados observados. Por dados completos entendemos a variável resposta, os efeitos aleatórios e as demais variáveis latentes presentes no modelo.

Seja $\mathbf{U} = (U_1, \dots, U_k)^t$ o vetor de variáveis misturadoras latentes, $\mathbf{U}_c^t = (\gamma^t, \mathbf{U}^t)$ o vetor com todas as variáveis latentes do modelo, $\mathbf{y}_c^t = (\mathbf{y}^t, \gamma^t, \mathbf{U}^t)$ o vetor de dados completos e $\theta^t = (\beta^t, \sigma^2, \nu)$ o vetor de parâmetros. A função de verossimilhança para os dados completos \mathbf{y}_c é

$$\begin{aligned} L_c(\theta; \mathbf{y}_c) &= f_{\mathbf{y}|D, \beta, \gamma}(\mathbf{y}) f_{\gamma|\sigma^2, U}(\gamma) f_{U|\nu}(\mathbf{u}) \\ &= \prod_{i=1}^k \prod_{j=1}^{n_i} \binom{n_{ij}}{y_{ij}} \left[\frac{\exp\{\eta_{ij}\}}{1 + \exp\{\eta_{ij}\}} \right]^{y_{ij}} \left[\frac{1}{1 + \exp\{\eta_{ij}\}} \right]^{n_{ij} - y_{ij}} \\ &\quad \times \left[\prod_{i=1}^k \left(\frac{u_i}{2\pi\sigma^2} \right)^{1/2} \exp\left\{ -\frac{u_i \gamma_i^2}{2\sigma^2} \right\} \right] \left[\prod_{i=1}^k f_{U_i|\nu}(u_i) \right], \end{aligned} \quad (3.14)$$

em que $\eta_{ij} = \mathbf{x}_{ij}^t \beta + \gamma_i$.

Para continuarmos o processo de inferência é necessário obtermos a função de verossimilhança marginal para os dados observados e maximizarmos com respeito aos parâmetros de interesse. Para isto é necessário integrarmos a função em (3.14) com respeito a γ e \mathbf{U} . No entanto, tal integral multidimensional não possui forma analítica conhecida. O algoritmo EM (do acrônimo *Expectation-Maximization*), proposto por Dempster *et al.* (1977), é uma alternativa para obtermos os estimadores de máxima verossimilhança (EMV) neste tipo de situação. No que segue apresentamos o algoritmo EM em sua forma geral e algumas de suas variações para obtermos os estimadores de máxima verossimilhança para o modelo Logístico-NI.

3.5.1 Algoritmo EM Monte Carlo no modelo Logístico-NI

O algoritmo EM é construído considerando-se a função de verossimilhança completa visando obtermos os EMV para a função de verossimilhança marginal. Cada iteração do algoritmo EM é composta por dois passos denotados, respectivamente, por passo E (*expectation*) e M (*maximization*). Para a iteração $m + 1$ do algoritmo, no passo E calculamos o valor esperado do logaritmo da função de verossimilhança para os dados completos com respeito a f.d.p do vetor de variáveis latentes \mathbf{U}_c , condicional em \mathbf{y} , \mathbf{D} e no vetor de estimativas $\theta^{(m)}$. Esta esperança é expressa por

$$Q(\theta | \theta^{(m)}) = E_{\theta^{(m)}}[l_c(\theta; \mathbf{y}_c) | \mathbf{y}] = \int l_c(\theta; \mathbf{y}_c) f_{\mathbf{U}_c|\theta^{(m)}, \mathbf{y}, \mathbf{D}}(\mathbf{u}_c) d\mathbf{u}_c, \quad (3.15)$$

em que $l_c(\theta; \mathbf{y}_c)$ denota a função de log-verossimilhança para \mathbf{y}_c , $\theta^{(m)}$ denota o argumento que maximizou a função $Q(\theta | \theta^{(m-1)})$ no passo m e $f_{\mathbf{U}_c|\theta^{(m)}, \mathbf{y}, \mathbf{D}}(\mathbf{u}_c)$ representa a f.d.p do vetor aleatório \mathbf{U}_c condicional em \mathbf{y} , \mathbf{D} e $\theta = \theta^{(m)}$. O passo M do algoritmo EM corresponde à maximização de $Q(\theta | \theta^{(m)})$ com respeito a θ .

No caso do modelo Logístico-NI a função de log-verossimilhança para \mathbf{y}_c decorre de (3.14) e é dada por

$$l_c(\boldsymbol{\theta}; \mathbf{y}_c) \propto \left[\sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij} \eta_{ij} - n_{ij} \ln \{1 + \exp \{\eta_{ij}\}\} \right] - \frac{k}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^k u_i \gamma_i^2 + \sum_{i=1}^k \ln f_{U_i|\nu}(u_i),$$

em que o termo $\sum_{i=1}^k \ln f_{U_i|\nu}(u_i)$ assume uma forma diferente para cada distribuição da subclasse NI. Para as distribuições consideradas neste trabalho temos que $\sum_{i=1}^k \ln f_{U_i|\nu}(u_i)$ é igual a

$$\begin{cases} 0 & \text{se } \gamma_i \sim N(0, \sigma^2); \\ k \left[\frac{\nu}{2} \ln \{\nu/2\} - \ln \Gamma(\nu/2) \right] + (\frac{\nu}{2} - 1) \ln u. - \frac{\nu}{2} u. & \text{se } \gamma_i \sim T(0, \sigma^2, \nu); \\ k \ln \nu + (\nu - 1) \ln u. & \text{se } \gamma_i \sim SL(0, \sigma^2, \nu); \\ \frac{1}{1-\nu_2} [(k - u.) \ln \nu_1 + (u. - k\nu_2) \ln \{1 - \nu_1\}] & \text{se } \gamma_i \sim NC(0, \sigma^2, \nu_1, \nu_2), \end{cases} \quad (3.16)$$

em que $u. = \sum_{i=1}^k u_i$ e $\ln u. = \sum_{i=1}^k \ln u_i$.

Note que se os efeitos aleatórios tiverem distribuição Normal Contaminada, a expressão em (3.16) é obtida porque podemos escrever

$$f_{U_i|\nu_1, \nu_2}(u_i) = \nu_1^{(1-u_i)/(1-\nu_2)} (1 - \nu_1)^{(u_i-\nu_2)/(1-\nu_2)}.$$

Em alguns casos não é possível obtermos as expressões analíticas fechadas para o valor esperado dado em (3.15). Isto, por exemplo, é o que ocorre com os casos apresentados aqui. Quando isto ocorre, uma possível solução é utilizarmos um estimador Monte Carlo para estimarmos $Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)})$. Se na iteração $m+1$ do algoritmo obtivermos uma amostra $\mathbf{u}_c^{(m+1,1)}, \dots, \mathbf{u}_c^{(m+1,L)}$ da distribuição $f_{U_c|\boldsymbol{\theta}^{(m)}, \mathbf{y}, \mathcal{D}}(\mathbf{u}_c)$, podemos estimar a esperança em (3.15) através da aproximação

$$Q_L(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)}) = \frac{1}{L} \sum_{l=1}^L l_c(\boldsymbol{\theta}; \mathbf{y}, \mathbf{u}_c^{(m+1,l)}). \quad (3.17)$$

Considerando-se esta aproximação, o algoritmo passa a ser denominado por algoritmo EM Monte Carlo (EMMC) e foi proposto por Wei e Tanner (1990). Como mencionado em Levine e Casella (2001), pela Lei dos Grandes Números, o estimador em (3.17) fornece uma boa aproximação para a esperança teórica dada em (3.15) se o tamanho da amostra Monte Carlo é grande.

Um aspecto importante da implementação do algoritmo EMMC é a escolha do tamanho da amostra Monte Carlo L . Quando $L = 1$ o algoritmo EMMC é equivalente ao algoritmo EM estocástico proposto por Celeux e Diebolt (1985). No entanto, maiores valores para L aumentam a precisão nas estimativas fornecidas pelo algoritmo. Wei e Tanner (1990) afirmam que é ineficiente iniciarmos o algoritmo EMMC assumindo um

valor inicial para L grande, uma vez que as estimativas para os parâmetros de interesse nos passos iniciais podem estar distantes dos valores verdadeiros. Eles sugerem que L deve ser aumentado com o número de iterações. Em função da aproximação Monte Carlo considerada no passo E e do erro Monte Carlo introduzido decorrente disto, no algoritmo EMMC não há garantias de que a cada iteração as estimativas fiquem mais próximas das estimativas de máxima verossimilhança. Por isso, nas iterações finais é necessário considerarmos valores mais altos de L para que o erro Monte Carlo não seja dominante no comportamento das estimativas dos parâmetros. Booth e Hobert (1999) dizem que na escolha do valor L vive-se o dilema de decidirmos entre a acurácia da aproximação da função Q e a velocidade de convergência do algoritmo. Alguns trabalhos como, por exemplo, McCulloch (1994), McCulloch (1997) e Chan e Kuk (1997), sugerem o uso de métodos *ad hoc* para aumentar L . Em McCulloch (1994), aumenta-se o valor de L linearmente com o número de iterações. McCulloch (1997) assume $L = 50$ para iterações de 1 a 19, $L = 200$ para iterações de 20 a 39 e $L = 5000$ a partir da iteração 40. Em Chan e Kuk (1997) usa-se $L = 1000$ para cada uma de suas 300 iterações.

Outros trabalhos consideram estratégias, baseadas na avaliação do erro Monte Carlo, para aumentar automaticamente o número L de réplicas Monte Carlo ao longo do algoritmo. Considerando que as amostras das variáveis latentes são geradas de forma independente, Booth e Hobert (1999) propõem um método para aumentar L a partir da construção de um elipsoide de confiança para $\boldsymbol{\theta}^{(m+1)}$. Levine e Casella (2001) generalizam o trabalho de Booth e Hobert (1999) considerando amostras dependentes, geradas por rotinas MCMC, das variáveis latentes, mas assumem que os componentes do vetor $\boldsymbol{\theta}^{(m+1)}$ são independentes. Levine e Fan (2004), também considerando amostras dependentes, implementam uma estratégia para aumentar L , diferente da estratégia sugerida por Levine e Casella (2001). Estes autores constroem através de um elipsoide de confiança para $\boldsymbol{\theta}^{(m+1)}$ em que estimam a correlação entre os componentes de $\boldsymbol{\theta}^{(m+1)}$. Chen *et al.* (2002) consideram a estratégia proposta em Booth e Hobert (1999) para modelos generalizados mistos utilizando uma abordagem semi-paramétrica para aproximar a densidade dos efeitos aleatórios. Caffo *et al.* (2005) propõe um método denominado “*Ascent-based MCEM*” para aumentar o número de réplicas Monte Carlo em que baseiam-se na propriedade de incremento da função de log-verossimilhança do algoritmo EM. Este método é comparado ao proposto por Booth e Hobert (1999). Neath (2006), entre outras coisas, implementa e compara os métodos automáticos propostos por Booth e Hobert (1999) e Caffo *et al.* (2005) para aumentar o número L de réplicas Monte Carlo. Neath (2006) conclui que o método *Ascent-based MCEM* é mais fácil de entender e produz uma alocação mais eficiente dos recursos Monte Carlo entre as iterações do EMMC do que o método proposto por Booth e Hobert (1999). Além disso, o *Ascent-based MCEM* é menos sensível a escolha de parâmetros pré-fixados.

Definido o valor de L , em uma certa iteração $m+1$ do algoritmo, os passos do algoritmo EMMC são:

- P1) Gerar uma amostra de tamanho L de \mathbf{U}_c da distribuição $f_{\mathbf{U}_c|\boldsymbol{\theta}^{(m)}, \mathbf{y}, \mathbf{D}}(\mathbf{u}_c)$;
- P2) Calcular $Q_L(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)})$;
- P3) Maximizar $Q_L(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)})$ para obter $\boldsymbol{\theta}^{(m+1)}$.

Estes passos são repetidos até que o algoritmo atinja a convergência. Existem diversos

critérios para avaliarmos a convergência de algoritmos do tipo EM. No algoritmo EM usual uma regra muito usada é pararmos o algoritmo quando a mudança relativa nos valores dos parâmetros for muito pequena, isto é, o algoritmo pode ser parado quando

$$\max_r \left(\frac{|\theta_r^{(m+1)} - \theta_r^{(m)}|}{|\theta_r^{(m)}| + \delta_1} \right) < \delta_2, \quad (3.18)$$

em que δ_1 e δ_2 são constantes pré-determinadas, $\max_r A_r$ representa o maior valor A_r com o subscrito r variando para todos os componentes do vetor de parâmetros $\boldsymbol{\theta}$. Em geral, são utilizados valores pequenos como, por exemplo, $\delta_1 = 0,001$ e $\delta_2 = 0,0001$. O critério em (3.18) também pode ser considerado no algoritmo EMMC. Entretanto, para este algoritmo estes valores para δ_1 e δ_2 não são recomendados uma vez que com presença do Erro Monte Carlo, seria necessário amostra Monte Carlo de tamanho L muito grande para atendermos tal precisão. Em Booth e Hobert (1999) são considerados valores para δ_2 entre 0,002 e 0,005. Além disso, tal artigo afirma que o critério de parada em (3.18) pode ser atingido prematuramente em uma iteração do algoritmo EMMC simplesmente por existir um grande erro Monte Carlo associado com a estimação dos parâmetros. Por isso, Booth e Hobert (1999) finaliza o algoritmo EMMC apenas quando a regra em (3.18) é satisfeita em três iterações consecutivas.

Note que o critério de parada em (3.18) pode não ser atingido quando o algoritmo fornecer um valor muito próximo de 0 para a estimativa de algum parâmetro. Isto pode ser um problema quando o valor de algum parâmetro é próximo de zero. Por esta razão, Booth e Hobert (1999) sugerem utilizar um segundo critério de convergência em conjunto com o critério (3.18) baseado na mudança da estimativa dos parâmetro relativa ao seu erro padrão, ou seja, considerar o critério dado por

$$\max_r \left(\frac{|\theta_r^{(m+1)} - \theta_r^{(m)}|}{\sqrt{V(\hat{\theta}_r)} + \delta_1} \right) < \delta_2, \quad (3.19)$$

para valores pequenos de δ_1 e δ_2 , em que $V(\hat{\theta}_i)$ representa a variância das estimativas fornecidas para a i -ésima componente do vetor $\boldsymbol{\theta}$.

Note que para usarmos o critério de parada em (3.19) é necessário calcularmos a variância $Var(\hat{\theta}_i)$ em cada iteração do algoritmo, o que pode tornar o processo de estimação mais lento.

Em algoritmos determinísticos, outra opção de critério de parada que é utilizado é a avaliação das mudanças na log-verossimilhança observada. No algoritmo EMMC esse tipo de regra não é muito conveniente pois não é possível encontrarmos a forma analítica para a log-verossimilhança observada. Então faz-se necessário utilizarmos uma aproximação Monte Carlo para esta função. Além disso, como diz Neath (2006), este tipo de critério é perigoso pois uma pequena mudança da $l_{obs}(\boldsymbol{\theta}, \mathbf{y})$ não indica necessariamente que $\boldsymbol{\theta}^{(m)}$ está perto do estimador de máxima verossimilhança. A mudança na log-verossimilhança pode ser pequena porque estimativas dos parâmetros estão perto da convergência ou porque no movimento de atualização dos parâmetros atingiu-se uma região muito suave da superfície de verossimilhança, não necessariamente próxima do seu máximo. A segunda possibilidade é devida ao fato de o algoritmo EMMC não possuir a propriedade de ascensão da log-verossimilhança que existe no EM determinístico.

Outra questão delicada no algoritmo EMMC é a geração das amostras da distribuição condicional completa $f_{\mathbf{U}_c|\boldsymbol{\theta}^{(m)},\mathbf{y},\mathbf{D}}(\mathbf{u}_c)$ de $\mathbf{U}_c = (\boldsymbol{\gamma}^t, \mathbf{U}^t)^t$. Existem várias possibilidades para gerarmos amostras desta distribuição. Alguns procedimentos sugeridos para este fim são a amostragem por rejeição, a amostragem por importância e os métodos MCMC. Jank (2005) discute as vantagens e desvantagens de cada uma destas alternativas. Uma desvantagem da amostragem por rejeição é que esta funciona bem apenas para problemas em que é necessário gerarmos amostras de distribuições que têm baixa dimensão. Além disso, este método depende da escolha de uma boa distribuição de importância ou candidata para que a taxa de aceitação do algoritmo seja alta. A amostragem por importância apresenta como principais vantagens o fato de gerarmos amostras independentes e utilizarmos todos os valores simulados, entretanto, este método depende fortemente da amostra escolhida. Os métodos MCMC são muito úteis quando é necessário gerarmos amostras de distribuições multivariadas com alta dimensão, no entanto, as amostras geradas não são independentes, o que faz com que a variância das estimativas aumente. Conseqüentemente, isto leva ao aumento da variância das estimativas podendo acarretar problemas de convergência do algoritmo. Booth e Hobert (1999) discutem os resultados obtidos quando as amostras dos efeitos aleatórios são geradas utilizando-se o algoritmo de Metropolis-Hastings, o método da amostragem por rejeição e amostragem por importância. Quintana *et al.* (1999) e Levine e Casella (2001) utilizam amostragem por importância, Chen *et al.* (2002) consideram um método denominado por “Dupla” amostragem por rejeição e Levine e Fan (2004) utilizam métodos MCMC para gerar dos efeitos aleatórios.

Neste trabalho, devido as dificuldades computacionais, utilizamos o amostrador de Gibbs para gerarmos amostras da distribuição $f_{\mathbf{U}_c|\boldsymbol{\theta}^{(m)},\mathbf{y},\mathbf{D}}(\mathbf{u}_c)$. Para garantirmos a independência da amostra gerada, L cadeias independentes foram consideradas. Considerando a representação dada em (3.13), as distribuições condicionais completas para \mathbf{U}_c nos casos discutidos no início da seção são dadas a seguir. Com o intuito de simplificar a notação, omitimos a dependência em $\boldsymbol{\theta} = \boldsymbol{\theta}^{(m)}$ na apresentação das distribuições condicionais completas (d.c.c) obtidas. Tais distribuições são dadas por

$$\begin{aligned}
f_{\boldsymbol{\gamma}|\boldsymbol{\theta},\mathbf{y},\mathbf{U},\mathbf{D}}(\boldsymbol{\gamma}) &\propto f_{\mathbf{y}|\mathbf{D},\boldsymbol{\beta},\boldsymbol{\gamma}}(\mathbf{y})f_{\boldsymbol{\gamma}|\sigma^2,\mathbf{U}}(\boldsymbol{\gamma}) \\
&\propto \left[\prod_{i=1}^k \prod_{j=1}^{n_i} (\pi_{ij})^{y_{ij}} (1 - \pi_{ij})^{n_{ij} - y_{ij}} \right] \exp \left\{ -\frac{\sum_{i=1}^n u_i \gamma_i^2}{2\sigma^2} \right\}, \\
f_{\mathbf{U}|\boldsymbol{\theta},\mathbf{y},\boldsymbol{\gamma},\mathbf{D}}(\mathbf{u}) &\propto f_{\boldsymbol{\gamma}|\sigma^2,\mathbf{U}}(\boldsymbol{\gamma})f_{\mathbf{U}|\boldsymbol{\nu}}(\mathbf{u}) \\
&\propto \prod_{i=1}^k u_i^{\frac{1}{2}} \exp \left\{ -\frac{U_i \gamma_i^2}{2\sigma^2} \right\} f_{U_i|\boldsymbol{\nu}}(u_i), \tag{3.20}
\end{aligned}$$

em que $\pi_{ij} = \exp \{ \eta_{ij} \} [1 + \exp \{ \eta_{ij} \}]^{-1}$ e $\eta_{ij} = \mathbf{x}_{ij}^t \boldsymbol{\beta} + \gamma_i$.

A d.c.c em (3.20) assumirá uma expressão diferente em cada modelo no qual assumimos uma distribuição da classe NI para os efeitos aleatórios. No modelo Logístico-N

temos $f_{U|\boldsymbol{\theta}, \mathbf{y}, \boldsymbol{\gamma}, \mathbf{D}}(\mathbf{u}) = \prod_{i=1}^k \mathbf{1}_{\{u_i=1\}}$. Se assumimos o modelo Logístico-T, então

$$f_{U|\boldsymbol{\theta}, \mathbf{y}, \boldsymbol{\gamma}, \mathbf{D}}(\mathbf{u}) \propto \prod_{i=1}^k u_i^{\frac{\nu-1}{2}} \exp \left\{ -\frac{u_i}{2} \left(\frac{\gamma_i^2}{\sigma^2} + \nu \right) \right\},$$

a partir de onde conclui-se que $U_i|\mathbf{y}, \boldsymbol{\gamma} \sim \text{Gama} \left(\frac{\nu+1}{2}, \frac{1}{2} \left(\frac{\gamma_i^2}{\sigma^2} + \nu \right) \right)$. Ao adotarmos o modelo Logístico-SL, temos que

$$f_{U|\boldsymbol{\theta}, \mathbf{y}, \boldsymbol{\gamma}, \mathbf{D}}(\mathbf{u}) \propto \prod_{i=1}^k u_i^{\nu-\frac{1}{2}} \exp \left\{ -\frac{u_i \gamma_i^2}{2\sigma^2} \right\} \mathbf{1}_{\{0 < u_i < 1\}},$$

assim $U_i|\mathbf{y}, \boldsymbol{\gamma} \sim \text{Gama} \left(\nu + \frac{1}{2}, \frac{\gamma_i^2}{2\sigma^2} \right) \mathbf{1}_{\{0 < u_i < 1\}}$. Para gerarmos de uma distribuição Gama truncada a direita, como a distribuição condicional completa de U_i no modelo Logístico-SL, Philippe (1997) propõe gerar utilizando mistura finita de distribuições Beta. Neste caso, devido ao fato da distribuição truncada assumir valores onde, em geral, a distribuição Gama tem pouca massa de probabilidade, a alternativa de utilizarmos o método da inversa da distribuição acumulada leva a erros numéricos. Se consideramos o modelo Logístico-NC, então

$$f_{U|\boldsymbol{\theta}, \mathbf{y}, \boldsymbol{\gamma}, \mathbf{D}}(\mathbf{u}) \propto \prod_{i=1}^k u_i^{\frac{1}{2}} \exp \left\{ -\frac{u_i \gamma_i^2}{2\sigma^2} \right\} \left[\nu_1 \mathbf{1}_{\{u_i=\nu_2\}} + (1 - \nu_1) \mathbf{1}_{\{u_i=1\}} \right],$$

de forma que, condicional em $\boldsymbol{\theta}$, \mathbf{y} e $\boldsymbol{\gamma}$, u_i assume o valor ν_2 com probabilidade $B_1/(B_1 + B_2)$ e assume o valor 1 com probabilidade $B_2/(B_1 + B_2)$, em que $B_1 = \nu_1 \nu_2^{1/2} \exp \left\{ -\frac{\nu_2 \gamma_i^2}{2\sigma^2} \right\}$ e $B_2 = (1 - \nu_1) \exp \left\{ -\frac{\gamma_i^2}{2\sigma^2} \right\}$.

Note que a distribuição $f_{\boldsymbol{\gamma}|U, \boldsymbol{\theta}, \mathbf{y}, \mathbf{D}}(\boldsymbol{\gamma})$ não possui forma analítica conhecida. Para gerarmos desta distribuição utilizamos o algoritmo Metropolis-Hastings (ver, por exemplo, Gilks *et al.*, 1996). Analisaremos três alternativas distintas de Metropolis-Hastings que diferem quanto a distribuição candidata. A primeira alternativa de algoritmo, a qual denotaremos por Metropolis-Hastings de McCulloch, foi proposta por McCulloch (1997), onde foi utilizada especificamente no modelo logístico com efeitos aleatórios Normais. Aqui, esta proposta foi adaptada para os modelos em que $\boldsymbol{\gamma}$ possui distribuição na subclasse NI. Esta alternativa consiste em considerarmos a distribuição de importância ou candidata como sendo $h_{\gamma_i}(\gamma_i) = f_{\gamma_i|U_i, \sigma^2}(\gamma_i)$ e isto faz com que a probabilidade de aceitação da proposta $\boldsymbol{\gamma}^*$, dados os valores atuais de $\boldsymbol{\gamma}$, no algoritmo Metropolis-Hastings seja simplificada para

$$A_i(\boldsymbol{\gamma}^*, \boldsymbol{\gamma}) = \min \left\{ 1, \frac{f_{\mathbf{y}|\boldsymbol{\gamma}^*, \mathbf{D}, \boldsymbol{\beta}}(\mathbf{y})}{f_{\mathbf{y}|\boldsymbol{\gamma}, \mathbf{D}, \boldsymbol{\beta}}(\mathbf{y})} \right\}.$$

Contudo, esta alternativa apresenta duas desvantagens: A primeira é que não é possível controlarmos a taxa de aceitação do Metropolis-Hastings e a segunda desvantagem é que cada elemento do vetor $\boldsymbol{\gamma}$ é gerado separadamente, o que pode tornar mais lento o programa computacional. Por outro lado, se tentarmos atualizar mais de um elemento de $\boldsymbol{\gamma}$ por vez, a taxa de aceitação do algoritmo torna-se muito baixa. Esta alternativa de algoritmo (Algoritmo 1) é descrita a seguir.

Algoritmo 1 *Metropolis-Hastings de McCulloch*

Seja $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_k)$ valores prévios da distribuição condicional $f_{\boldsymbol{\gamma}|\mathbf{U},\boldsymbol{\theta},\mathbf{y},\mathbf{D}}(\boldsymbol{\gamma})$.

1. Inicialize $i = 1$;
 2. Gere γ_i^* da distribuição candidata $h_{\gamma_i}(\gamma_i) = f_{\gamma_i|U_i,\sigma^2}(\gamma_i)$;
 3. Faça $\boldsymbol{\gamma}^* = (\gamma_1, \dots, \gamma_i^*, \dots, \gamma_k)$;
 4. Calcule $A_i(\boldsymbol{\gamma}^*, \boldsymbol{\gamma}) = \min \left\{ 1, \frac{f_{\mathbf{y}|\boldsymbol{\gamma}^*,\mathbf{D},\boldsymbol{\beta}}(\mathbf{y})}{f_{\mathbf{y}|\boldsymbol{\gamma},\mathbf{D},\boldsymbol{\beta}}(\mathbf{y})} \right\}$;
 5. Aceite $\boldsymbol{\gamma}^*$ como novo valor com probabilidade $A_i(\boldsymbol{\gamma}^*, \boldsymbol{\gamma})$, caso contrário mantenha o valor prévio.
 6. Faça $i = i + 1$;
 7. Repita de 2-6 enquanto $i \leq k$.
-

Perceba que, uma vez que as observações para a variável resposta são independentes condicionalmente nos efeitos aleatórios, a probabilidade de aceitação no Algoritmo 1 pode ser simplificada para

$$A_i(\gamma_i^*, \gamma_i) = \min \left\{ 1, \frac{\prod_{j=1}^{n_i} f_{y_{ij}|\gamma_i^*,\mathbf{D},\boldsymbol{\beta}}(y_{ij})}{\prod_{j=1}^{n_i} f_{y_{ij}|\gamma_i,\mathbf{D},\boldsymbol{\beta}}(y_{ij})} \right\}. \quad (3.21)$$

Em termos de eficiência na implementação computacional do algoritmo, utilizar a probabilidade de aceitação dada em (3.21) traz grande benefício pois diminui bastante o número de contas a serem feitas, o que reduz consideravelmente o tempo de execução do algoritmo. O algoritmo MCMC com o algoritmo Metropolis-Hastings de McCulloch para gerarmos da distribuição condicional completa dos efeitos aleatórios denotaremos por EMMC-M.

Visando melhorar os problemas do Algoritmo 1 previamente mencionados, propomos uma alternativa (Algoritmo 2) que consiste em utilizar como distribuição candidata a distribuição Normal com média igual ao valor no passo anterior e com variância igual τ^2 . Quando a distribuição candidata é formulada desta forma o algoritmo é denotado por Metropolis com Passeio aleatório (ver, por exemplo, Gilks *et al.*, 1996). Aqui denotaremos este algoritmo por Metropolis-Hastings Passeio Aleatório e denotaremos o algoritmo EMMC utilizando esta proposta de algoritmo Metropolis por EMMC-PA. O valor de τ^2 é escolhido de forma a manter a taxa de aceitação do algoritmo Metropolis-Hastings próxima do ideal. Segundo Gamerman e Lopes (2006) e referências mencionadas por eles, o valor ideal para a taxa de aceitação está entre 20% e 50% para o caso univariado e próximo a 24% para problemas de grandes dimensões. Uma vantagem desta alternativa é que podemos gerar valores de $\boldsymbol{\gamma}$ dividindo-o em blocos de tamanho l_1 , o que torna o algoritmo mais veloz e pode melhorar a convergência do algoritmo.

Uma terceira alternativa de algoritmo Metropolis-Hastings considerada neste problema, aqui denominada algoritmo Metropolis-Hastings de Quintana, foi proposta por Quintana *et al.* (1999). Ao algoritmo EMMC com esta proposta denominamos por EMMC-Q. A ideia desta proposta é obtermos uma distribuição candidata mais próxima da distribuição alvo para aumentar a taxa de aceitação do algoritmo. Em Quintana *et*

Algoritmo 2 *Metropolis-Hastings Passeio Aleatório*

Seja $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_k)$ valores prévios da distribuição condicional $f_{\boldsymbol{\gamma}|\mathbf{U}, \boldsymbol{\theta}, \mathbf{y}, \mathbf{D}}(\boldsymbol{\gamma})$.

1. Inicialize $b = 1$, defina o tamanho l_1 dos blocos a serem gerados e calcule $i_1 = (b - 1)l_1 + 1$ e $i_2 = bl_1$;
2. Gere $\boldsymbol{\gamma}_{[i_1:i_2]}^* = (\gamma_{i_1}^*, \dots, \gamma_{i_2}^*)$ da distribuição candidata $N(\boldsymbol{\gamma}_{[i_1:i_2]}, \mathbf{I}_{l_1} \tau^2)$ em que $\boldsymbol{\gamma}_{[i_1:i_2]} = (\gamma_{i_1}, \dots, \gamma_{i_2})$;
3. Faça $\boldsymbol{\gamma}^* = (\gamma_1, \dots, \boldsymbol{\gamma}_{[i_1:i_2]}^*, \dots, \gamma_k)$;
4. Calcule $A_b(\boldsymbol{\gamma}^*, \boldsymbol{\gamma}) = \min \left\{ 1, \frac{f_{\mathbf{y}|\boldsymbol{\gamma}^*, \mathbf{D}, \boldsymbol{\beta}}(\mathbf{y}) f_{\boldsymbol{\gamma}|\mathbf{U}, \sigma^2}(\boldsymbol{\gamma}^*)}{f_{\mathbf{y}|\boldsymbol{\gamma}, \mathbf{D}, \boldsymbol{\beta}}(\mathbf{y}) f_{\boldsymbol{\gamma}|\mathbf{U}, \sigma^2}(\boldsymbol{\gamma})} \right\}$;
5. Aceite $\boldsymbol{\gamma}^*$ como novo valor com probabilidade $A_b(\boldsymbol{\gamma}^*, \boldsymbol{\gamma})$, caso contrário mantenha o valor prévio.
6. Faça $b = b + 1$;
7. Repita de 2-6 enquanto $b \leq (k/l_1)$.

Obs.: Se k/l_1 não for inteiro o último bloco terá um tamanho menor.

al. (1999) a distribuição candidata é inspirada na teoria assintótica para distribuição *a posteriori* de $\boldsymbol{\gamma}$ ao considerarmos $\boldsymbol{\beta}$ conhecido. Em tal artigo afirma-se que uma amostragem razoável da distribuição de γ_i é obtida ao gerarmos valores de uma f.d.p candidata que satisfaça

$$h_{\gamma_i}(\gamma_i) \propto \exp \left\{ -\frac{cI_i}{2}(\gamma_i - \hat{\gamma}_i)^2 + \log \{f_{\gamma_i|U_i, \sigma^2}(\gamma_i)\} \right\},$$

em que I_i é a informação de Fisher observada, com respeito a γ_i , avaliada em $\hat{\gamma}_i$ e c é uma constante positiva. De acordo com Quintana *et al.* (1999), a constante c é introduzida para permitir-nos amostrar de uma maior diversidade de distribuições e melhorar a estabilidade numérica. Em um exemplo com o modelo Logístico-N, o autor relata não observar mudanças significativas na convergência do algoritmo para valores de c entre $[1; 2,5]$, exceto quanto a taxa de aceitação da implementação Monte Carlo do passo E.

Em nosso caso, como $\gamma_i|U_i \sim N(0, u_i^{-1}\sigma^2)$, a distribuição candidata torna-se

$$N((cI_i + u_i/\sigma^2)^{-1}cI_i\hat{\gamma}_i, (cI_i + u_i/\sigma^2)^{-1}), \quad (3.22)$$

em que $\hat{\gamma}_i$ é a solução da equação $\frac{\partial}{\partial \gamma_i} \sum_{j=1}^{n_i} y_{ij} \eta_{ij} - n_{ij} \ln \{1 + \exp \{\eta_{ij}\}\} = 0$ e $I_i = \sum_{j=1}^{n_i} n_{ij} \hat{\pi}_{ij} (1 - \hat{\pi}_{ij})$, com $\hat{\pi}_{ij} = \exp \{\mathbf{x}_{ij}^t \boldsymbol{\beta} + \hat{\gamma}_i\} [1 + \exp \{\mathbf{x}_{ij}^t \boldsymbol{\beta} + \hat{\gamma}_i\}]^{-1}$. A quantidade $(cI_i + u_i/\sigma^2)^{-1}cI_i$ está sempre contida no intervalo entre 0 e 1. Por isso, a média da distribuição candidata está entre 0 e $\hat{\gamma}_i$. Note que a escolha de valores maiores de c diminuem a variância da distribuição candidata e aproximam a média desta distribuição de $\hat{\gamma}_i$, pois torna a quantidade $(cI_i + u_i/\sigma^2)^{-1}cI_i$ mais próxima do valor 1. A partir da distribuição candidata em (3.22), a probabilidade de aceitação de um valor proposto γ_i^* do algoritmo Metropolis-Hastings é dada por

$$A_i(\gamma_i^*, \gamma_i) = \min \left\{ 1, \left(\frac{\prod_{j=1}^{n_i} f_{y_{ij}|\gamma_i^*, \mathbf{D}, \boldsymbol{\beta}}(y_{ij})}{\prod_{j=1}^{n_i} f_{y_{ij}|\gamma_i, \mathbf{D}, \boldsymbol{\beta}}(y_{ij})} \right) g(\gamma_i^*, \gamma_i, c) \right\}, \quad (3.23)$$

onde $g(\gamma_i^*, \gamma_i, c) = \exp \left\{ \frac{cI_i}{2} [(\gamma_i^* - \hat{\gamma}_i)^2 - (\gamma_i - \hat{\gamma}_i)^2] \right\}$.

A esquematização deste algoritmo (Algoritmo 3) é descrita a seguir.

Algoritmo 3 *Metropolis-Hastings de Quintana*

Seja $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_k)$ valores prévios da distribuição condicional $f_{\boldsymbol{\gamma} | \mathbf{U}, \boldsymbol{\theta}, \mathbf{y}, \mathbf{D}}(\boldsymbol{\gamma})$.

1. Inicialize $i = 1$;
 2. Calcule $\hat{\gamma}_i$ e I_i ;
 3. Gere γ_i^* da distribuição candidata $N((cI_i + u_i/\sigma^2)^{-1}cI_i\hat{\gamma}_i, (cI_i + u_i/\sigma^2)^{-1})$;
 4. Faça $\boldsymbol{\gamma}^* = (\gamma_1, \dots, \gamma_i^*, \dots, \gamma_k)$;
 5. Calcule $A_i(\gamma_i^*, \gamma_i) = \min \left\{ 1, \left(\frac{\prod_{j=1}^{n_i} f_{y_{ij} | \gamma_i^*, \mathbf{D}, \boldsymbol{\beta}}(y_{ij})}{\prod_{j=1}^{n_i} f_{y_{ij} | \gamma_i, \mathbf{D}, \boldsymbol{\beta}}(y_{ij})} \right) g(\gamma_i^*, \gamma_i, c) \right\}$;
 6. Aceite $\boldsymbol{\gamma}^*$ como novo valor com probabilidade $A_i(\boldsymbol{\gamma}^*, \boldsymbol{\gamma})$, caso contrário mantenha o valor prévio.
 7. Faça $i = i + 1$;
 8. Repita de 2-7 enquanto $i \leq k$.
-

A etapa de maximização do algoritmo EMMC fornece as estimativas $\boldsymbol{\theta}$ e pode ser dividida em três partes uma vez que sob o modelo que estamos considerando em (3.14), o valor esperado em (3.15) torna-se

$$Q_L(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)}) = Q_L(\boldsymbol{\beta} | \boldsymbol{\beta}^{(m)}) + Q_L(\sigma^2 | \sigma^{2(m)}) + Q_L(\boldsymbol{\nu} | \boldsymbol{\nu}^{(m)}).$$

Assim, na iteração $m + 1$ do algoritmo EMMC, a maximização de $Q_L(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)})$ pode ser feita maximizando-se com respeito a $\boldsymbol{\beta}$, σ^2 e $\boldsymbol{\nu}$ separadamente. Para obtermos as estimativas de $\boldsymbol{\beta}$, σ^2 e $\boldsymbol{\nu}$ devemos maximizar, respectivamente, os termos a seguir:

$$Q_L(\boldsymbol{\beta} | \boldsymbol{\beta}^{(m)}) = \frac{1}{L} \sum_{l=1}^L \left[\sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij} \eta_{ij}^{(m+1,l)} - n_{ij} \ln \{1 + \exp \{ \eta_{ij}^{(m+1,l)} \} \} \right], \quad (3.24)$$

$$Q_L(\sigma^2 | \sigma^{2(m)}) = -\frac{k}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \frac{1}{L} \sum_{l=1}^L \sum_{i=1}^k u_i^{(m+1,l)} \left(\gamma_i^{(m+1,l)} \right)^2, \quad (3.25)$$

$$Q_L(\boldsymbol{\nu} | \boldsymbol{\nu}^{(m)}) = \frac{1}{L} \sum_{l=1}^L \sum_{i=1}^k \ln f_{U_i | \boldsymbol{\nu}}(u_i^{(m+1,l)}), \quad (3.26)$$

em que $\eta_{ij}^{(m+1,l)} = \mathbf{x}_{ij}^t \boldsymbol{\beta} + \gamma_i^{(m+1,l)}$.

Após gerarmos amostras de \mathbf{U}_c , a maximização de (3.24) com respeito a $\boldsymbol{\beta}$, em cada passo, é realizada através de métodos numéricos. O estimador de σ^2 é obtido analiticamente ao maximizarmos (3.25) e é dado por

$$(\sigma^2)^{(m+1)} = \frac{1}{kL} \sum_{l=1}^L \sum_{i=1}^k u_i^{(m+1,l)} \left(\gamma_i^{(m+1,l)} \right)^2. \quad (3.27)$$

No modelo Logístico-T, para estimarmos $\boldsymbol{\nu}$, a maximização de (3.26) também é feita numericamente. Visto que a distribuição t -Student possui variância para $\nu > 2$ e também que a f.d.p não se altera muito para valores altos de ν , a maximização pode ser realizada colocando restrições para ν , por exemplo, considerando ν entre 2,1 e 30.

Já no modelo Logístico-SL existe solução analítica para o estimador de ν e o ponto de máximo de (3.26) no passo $m + 1$ é

$$\nu^{(m+1)} = -k \left[\frac{1}{L} \sum_{l=1}^L \sum_{i=1}^k \ln u_i^{(m+1,l)} \right]^{-1}. \quad (3.28)$$

A distribuição *Slash* possui variância apenas se $\nu > 0$. O estimador obtido em (3.28) não garante que tal condição seja atendida e, por isso, se for desejada a existência da variância da distribuição dos efeitos aleatórios, é necessário adicionarmos esta restrição no algoritmo.

Ao assumirmos a distribuição Normal Contaminada (modelo Logístico-NC) para representar o comportamento do efeitos aleatórios temos que $\boldsymbol{\nu} = (\nu_1, \nu_2)$. Em Souza e Migon (2010) os parâmetros ν_1 e ν_2 não são estimados, eles são definido de forma *ad-hoc* pelo pesquisador. Para estimá-los, considerando (3.16) temos que a expressão em (3.26) torna-se

$$Q_L(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(m)}) = \frac{1}{1 - \nu_2} [(k - \bar{u}^{(m+1)}) \ln \nu_1 + (\bar{u}^{(m+1)} - k\nu_2) \ln \{1 - \nu_1\}], \quad (3.29)$$

em que $\bar{u}^{(m+1)} = \frac{1}{L} \sum_{l=1}^L \sum_{i=1}^k u_i^{(m+1,l)}$.

Com o intuito de avaliarmos o comportamento da função $Q_L(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(m)})$ em (3.29) com respeito a ν_1 e ν_2 apresentamos a Figura 19 a seguir, em que assumimos $k = 100$ e $\bar{u}^{(m+1)} = 70$. A Figura 19 apresenta o gráfico da função $Q_L(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(m)})$ sob dois ângulos distintos. Perceba que $\bar{u}^{(m+1)} = 70$ é o valor esperado em uma situação em que tenhamos $\nu_1 = 0,4$ e $\nu_2 = 0,25$ como valores reais, pois seria esperado que 60 vezes u_i fosse igual a 1 e em 40 vezes igual a 0,25.

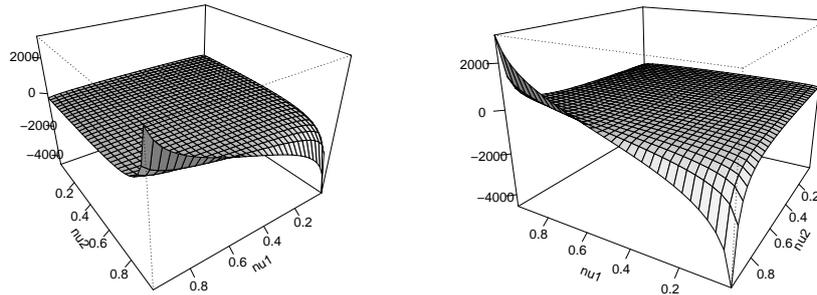


Figura 19: Gráfico da função $Q_L(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(m)})$ para $\bar{u}^{(m+1)} = 70$ visto por dois ângulos distintos.

Observamos na Figura 19 que o máximo da função $Q_L(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(m)})$ ocorre quando $\nu_1 = 1$ e $\nu_2 \rightarrow 1$. Este comportamento é similar ao obtido para qualquer valor de $\bar{u}^{(m+1)}$. Em

virtude disto, vemos que não é possível estimarmos os dois parâmetros simultaneamente. Note também que, em (3.29), apenas $\bar{u}^{(m+1)}$ carrega informação da amostra para a estimação de ν_1 e ν_2 . Isto explica o motivo de não ser possível estimarmos os dois parâmetros conjuntamente a partir da maximização da expressão em (3.29). Sendo assim, analisamos a opção de fixarmos um dos parâmetros e estimarmos o outro. Inicialmente verificamos a possibilidade de considerarmos o parâmetro ν_1 fixo. Para tal derivamos $Q_L(\boldsymbol{\nu} | \boldsymbol{\nu}^{(m)})$ com respeito a ν_2 e obtemos

$$\frac{\partial}{\partial \nu_2} Q_L(\boldsymbol{\nu} | \boldsymbol{\nu}^{(m)}) = \frac{k - \bar{u}^{(m+1)}}{(1 - \nu_2)^2} \ln \left\{ \frac{\nu_1}{1 - \nu_1} \right\}. \quad (3.30)$$

Note que se $\nu_1 = 0,5$ a expressão em (3.29) é simplificada para $Q_L(\boldsymbol{\nu} | \boldsymbol{\nu}^{(m)}) = k \log \{0,5\}$ e esta expressão é constante com respeito a ν_2 . Note também que, como $k - \bar{u}^{(m+1)} \geq 0$ e $(1 - \nu_2)^2 > 0$, se $\nu_1 > 0,5$ então $\frac{\partial}{\partial \nu_2} Q_L(\boldsymbol{\nu} | \boldsymbol{\nu}^{(m)}) > 0$, o que significa que a função $Q_L(\boldsymbol{\nu} | \boldsymbol{\nu}^{(m)})$ é estritamente crescente e seu máximo ocorre quando $\nu_2 \rightarrow 1$. Por outro lado, se $\nu_1 < 0,5$ segue que $\frac{\partial}{\partial \nu_2} Q_L(\boldsymbol{\nu} | \boldsymbol{\nu}^{(m)}) < 0$ e, portanto, a função $Q_L(\boldsymbol{\nu} | \boldsymbol{\nu}^{(m)})$ é estritamente decrescente. Assim sendo, se $\nu_1 < 0,5$ a estimativa de máxima verossimilhança para ν_2 tende a zero. Concluímos que para qualquer valor de ν_1 fixado temos uma péssima estimativa para ν_2 e, portanto, descartamos esta possibilidade.

Analisamos agora a outra opção, a de fixarmos um valor para ν_2 e estimarmos ν_1 . Derivando a expressão em (3.29) com respeito a ν_1 uma e duas vezes obtemos, respectivamente, que

$$\begin{aligned} \frac{\partial}{\partial \nu_1} Q_L(\boldsymbol{\nu} | \boldsymbol{\nu}^{(m)}) &= \frac{1}{1 - \nu_2} \left(\frac{k - \bar{u}^{(m+1)}}{\nu_1} - \frac{\bar{u}^{(m+1)} - k\nu_2}{1 - \nu_1} \right); \\ \frac{\partial^2}{\partial \nu_1^2} Q_L(\boldsymbol{\nu} | \boldsymbol{\nu}^{(m)}) &= \frac{1}{1 - \nu_2} \left(-\frac{k - \bar{u}^{(m+1)}}{\nu_1^2} - \frac{\bar{u}^{(m+1)} - k\nu_2}{(1 - \nu_1)^2} \right). \end{aligned}$$

A partir disto, obtemos que o estimador para ν_1 no passo $(m + 1)$, quando fixamos ν_2 , é dado por

$$\nu_1^{(m+1)} = \frac{k - \bar{u}^{(m+1)}}{k(1 - \nu_2)}.$$

Em resumo, o algoritmo EMMC para estimarmos o modelo logístico misto NI é esquematicamente representado a seguir.

No início do Algoritmo 4 é necessário definirmos os valores iniciais tanto para os parâmetros quanto para as variáveis latentes presentes no modelo. Para as variáveis latentes, optamos por inicializar os efeitos aleatórios $\boldsymbol{\gamma}^{(0,l)}$, $l = 1, \dots, L$, com zeros e as variáveis misturadas $\boldsymbol{u}^{(0,l)}$ com valores um. Para os parâmetros, inicializamos os efeitos fixos $\boldsymbol{\beta}$ com as estimativas obtidas no modelo de regressão logística sem efeitos aleatórios. Já os valores $\sigma^{2(0)}$ e $\nu^{(0)}$ são definidos arbitrariamente. Perceba também, que o número L de réplicas Monte Carlo deve aumentar ao longo das iterações do algoritmo EMMC. Desta forma é necessário inicializarmos novas cadeias independentes toda vez que aumentamos o valor de L . Quando isto ocorre, inicializamos as L novas cadeias sorteando, com reposição e mesma probabilidade, os valores assumidos pelas variáveis latentes na iteração anterior do algoritmo.

Algoritmo 4 *EMMC para modelo logístico misto*

1. Inicialize $\boldsymbol{\theta}^{(0)}$ e $\mathbf{u}_c^{(0,1)}, \dots, \mathbf{u}_c^{(0,L)}$;
 2. Faça $m = 1$;
 3. Gere L amostras da distribuição $f_{\mathbf{U}_c | \boldsymbol{\theta}^{(m-1)}, \mathbf{y}, \mathbf{D}}(\mathbf{u}_c)$ via L cadeias independentes de *Gibbs Sampler*. Para cada cadeia l , para $l = 1, \dots, L$, faça:
 - i. Defina valores iniciais $\mathbf{u}_c^{(m-1,l)}$;
 - ii. Gere $\mathbf{u}^{(m,l)}$ de $f_{\mathbf{U} | \boldsymbol{\gamma}^{(m-1,l)}, \boldsymbol{\theta}^{(m-1)}, \mathbf{y}, \mathbf{D}}(\mathbf{u})$;
 - iii. Gere $\boldsymbol{\gamma}^{(m,l)}$ de $f_{\boldsymbol{\gamma} | \mathbf{U}^{(m,l)}, \boldsymbol{\theta}^{(m-1)}, \mathbf{y}, \mathbf{D}}(\boldsymbol{\gamma})$ via Metropolis-Hastings (Algoritmos 1, 2 ou 3);
 4. Calcule $Q_L(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)})$;
 5. Maximize $Q_L(\boldsymbol{\theta} | \boldsymbol{\theta}^{(m)})$ para obter $\boldsymbol{\theta}^{(m)}$;
 6. Faça $m = m + 1$;
 7. Repita os itens 3-6 até atingir a convergência.
-

3.5.2 Matriz de informação de Fisher observada

O cálculo exato da matriz de informação de Fisher no modelo Logístico-NI não é uma tarefa fácil. No que segue apresentamos o cálculo da matriz de informação de Fisher a qual será útil na estimação da variância assintótica dos estimadores de máxima verossimilhança. A matriz de informação de *Fisher* observada é dada por

$$IF(\boldsymbol{\theta}) = -\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t} l_{obs}(\boldsymbol{\theta}; \mathbf{y}),$$

em que $l_{obs} = \ln \left\{ f_{\mathbf{y} | \mathbf{D}, \boldsymbol{\beta}^{(m)}}(\mathbf{y}) \right\}$. O cálculo direto de $-\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t} l_{obs}(\boldsymbol{\theta}; \mathbf{y})$ não é conveniente por envolver integração sob $f_{\mathbf{U}_c | \sigma^2, \boldsymbol{\nu}}(\mathbf{u}_c)$. Louis (1982) mostra que a matriz de informação de Fisher observada pode ser escrita como

$$IF(\boldsymbol{\theta}) = E \left[-\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t} l_c(\boldsymbol{\theta}; \mathbf{y}_c) | \mathbf{y}; \boldsymbol{\theta} \right] - V \left[\frac{\partial}{\partial \boldsymbol{\theta}} l_c(\boldsymbol{\theta}; \mathbf{y}_c) | \mathbf{y}; \boldsymbol{\theta} \right], \quad (3.31)$$

em que as esperanças são calculadas com respeito a $f_{\mathbf{U}_c | \mathbf{y}, \mathbf{D}, \boldsymbol{\theta}}(\mathbf{u}_c)$.

Usando estimadores Monte Carlo, em uma iteração $m + 1$ do algoritmo EMMC, a expressão em (3.31) pode ser aproximada por

$$IF(\boldsymbol{\theta}) \approx -\frac{1}{L} \sum_{l=1}^L \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t} l_c(\boldsymbol{\theta}; \mathbf{y}_c^{(m+1,l)}) - \frac{1}{L} \sum_{l=1}^L \left[\frac{\partial}{\partial \boldsymbol{\theta}} l_c(\boldsymbol{\theta}; \mathbf{y}_c^{(m+1,l)}) \right] \left[\frac{\partial}{\partial \boldsymbol{\theta}} l_c(\boldsymbol{\theta}; \mathbf{y}_c^{(m+1,l)}) \right]^t + \left[\frac{1}{L} \sum_{l=1}^L \frac{\partial}{\partial \boldsymbol{\theta}} l_c(\boldsymbol{\theta}; \mathbf{y}_c^{(m+1,l)}) \right] \left[\frac{1}{L} \sum_{l=1}^L \frac{\partial}{\partial \boldsymbol{\theta}} l_c(\boldsymbol{\theta}; \mathbf{y}_c^{(m+1,l)}) \right]^t, \quad (3.32)$$

em que $\mathbf{y}_c^{(l)} = (\mathbf{y}^t, (\mathbf{u}_c^{(m+1,l)})^t)^t$ e $\mathbf{u}_c^{(m+1,l)}$ denotam amostras da f.d.p $f_{\mathbf{U}_c | \boldsymbol{\theta}, \mathbf{y}, \mathbf{D}}(\mathbf{u}_c)$ na iteração $m + 1$.

Visto que $\boldsymbol{\theta} = (\boldsymbol{\beta}^t, \sigma^2, \nu)$, encontramos as derivadas parciais de primeira e segunda ordem com respeito a $\boldsymbol{\beta}$, σ^2 e ν . Qualquer que seja a distribuição dos efeitos aleatórios na família de distribuições NI as derivadas parciais com respeito $\boldsymbol{\beta}$ e σ^2 são idênticas, ou seja, são as mesmas em todos os modelos. As derivadas de primeira e segunda ordem são dadas por

$$\frac{\partial}{\partial \boldsymbol{\theta}} l_c(\boldsymbol{\theta}; \mathbf{y}_c) = \begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\beta}} l_c(\boldsymbol{\theta}; \mathbf{y}_c) \\ \frac{\partial}{\partial \sigma^2} l_c(\boldsymbol{\theta}; \mathbf{y}_c) \\ \frac{\partial}{\partial \nu} l_c(\boldsymbol{\theta}; \mathbf{y}_c) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - n_{ij} \pi_{ij}) \mathbf{x}_{ij} \\ \frac{\sum_{i=1}^k u_i \gamma_i^2 - k \sigma^2}{2(\sigma^2)^2} \\ \frac{\partial}{\partial \nu} l_c(\boldsymbol{\theta}; \mathbf{y}_c) \end{pmatrix}, \quad (3.33)$$

$$\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t} l_c(\boldsymbol{\theta}; \mathbf{y}_c) = \begin{pmatrix} -\sum_{i=1}^k \sum_{j=1}^{n_i} n_{ij} \pi_{ij} (1 - \pi_{ij}) \mathbf{x}_{ij} \mathbf{x}_{ij}^t & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{k \sigma^2 - 2 \sum_{i=1}^k u_i \gamma_i^2}{2(\sigma^2)^3} & 0 \\ \mathbf{0} & 0 & \frac{\partial^2}{\partial \nu^2} l_c(\boldsymbol{\theta}; \mathbf{y}_c) \end{pmatrix}. \quad (3.34)$$

No modelo Logístico-N temos que $\boldsymbol{\theta} = (\boldsymbol{\beta}^t, \sigma^2)$. Portanto, as matrizes em (3.33) e (3.34) têm dimensão reduzida. Além disso, as derivadas $\frac{\partial}{\partial \nu} l_c(\boldsymbol{\theta}; \mathbf{y}_c)$ e $\frac{\partial^2}{\partial \nu^2} l_c(\boldsymbol{\theta}; \mathbf{y}_c)$ em (3.33) e (3.34) são diferentes nos demais modelos Logísticos-NI. Nos modelos aqui considerados estas derivadas são:

(i) No modelo Logístico-T:

$$\begin{aligned} \frac{\partial}{\partial \nu} l_c(\boldsymbol{\theta}; \mathbf{y}_c) &= \frac{1}{2} \left[k \ln \{\nu/2\} + k - k \psi^{(0)}(\nu/2) + \sum_{i=1}^k (\ln u_i - u_i) \right], \\ \frac{\partial^2}{\partial \nu^2} l_c(\boldsymbol{\theta}; \mathbf{y}_c) &= \frac{k}{2\nu} - \frac{k \psi^{(1)}(\nu/2)}{4}, \end{aligned}$$

em que $\psi(x) = \frac{\partial^{x+1}}{\partial \nu^{x+1}} \ln \Gamma(x)$;

(ii) No modelo Logístico-SL:

$$\frac{\partial}{\partial \nu} l_c(\boldsymbol{\theta}; \mathbf{y}_c) = \frac{k}{\nu} + \sum_{i=1}^k \ln u_i \quad \text{e} \quad \frac{\partial^2}{\partial \nu^2} l_c(\boldsymbol{\theta}; \mathbf{y}_c) = -\frac{k}{\nu^2};$$

(iii) No modelo Logístico-NC:

$$\begin{aligned} \frac{\partial}{\partial \nu_1} l_c(\boldsymbol{\theta}; \mathbf{y}_c) &= \frac{k - \sum_{i=1}^k u_i}{\nu_1(1 - \nu_2)} - \frac{\sum_{i=1}^k u_i - k \nu_2}{(1 - \nu_1)(1 - \nu_2)}, \\ \frac{\partial^2}{\partial \nu_1^2} l_c(\boldsymbol{\theta}; \mathbf{y}_c) &= -\frac{k - \sum_{i=1}^k u_i}{\nu_1^2(1 - \nu_2)} - \frac{\sum_{i=1}^k u_i - k \nu_2}{(1 - \nu_1)^2(1 - \nu_2)}. \end{aligned}$$

Chen *et al.* (2002) argumenta que na prática, por causa do erro Monte Carlo, a matriz obtida em (3.32) nem sempre é positiva definida e, por isso, ele propõe aproximar a informação de Fisher observada diretamente via Monte Carlo. Aqui vamos considerar a expressão em (3.32) para obtermos a variância assintótica dos estimadores.

Na Seção 3.6 temos como meta comparar o desempenho dos algoritmos 1, 2 e 3 usados no passo 3 do algoritmo 4.

3.6 Estudo em dados simulados: Implementação do algoritmo EMMC

Três propostas de algoritmos são sugeridos na literatura para gerarmos da distribuição condicional completa dos efeitos aleatórios. Nesta subseção apresentamos um estudo, considerando dados simulados, com o objetivo de avaliarmos a implementação do algoritmo EMMC, com cada um destes sub-algoritmos, para ajustarmos o modelo logístico com efeitos aleatórios na classe de distribuições NI. Um dos objetivos principais também é avaliarmos o comportamento das estimativas fornecidas pelo algoritmo EMMC para diversos valores do número k de *clusters*, do número n_i de observações dentro de cada *cluster* e do número n_{ij} de experimentos Bernoulli para cada indivíduo j dentro dos *clusters*.

Para gerarmos os bancos de dados nos inspiramos em um exemplo de ilustração denotado por “Modelo Logístico-Normal” apresentado em McCulloch (1997). Este exemplo, que também foi estudado em Booth e Hobert (1999), Levine e Casella (2001) e Caffo *et al.* (2005), considera dados gerados sob o modelo logístico com intercepto aleatório e observações separadas em *clusters*. O exemplo de aplicação analisado em Quintana *et al.* (1999) é muito similar ao que será apresentado nesta seção. Em McCulloch (1997) o banco de dados é gerado considerando uma variável resposta Bernoulli, poucos efeitos aleatórios (apenas 10) e efeitos aleatórios normalmente distribuídos. Neste trabalho, abordamos situações um pouco mais gerais do que as abordadas em trabalhos anteriores e consideramos distribuições na classe NI para os efeitos aleatórios. No processo de geração dos bancos de dados simulados assumimos

$$Y_{ij}|\pi_{ij} \sim \text{Binomial}(n_{ij}, \pi_{ij}), \text{ para } i = 1, 2, \dots, k \text{ e } j = 1, 2, \dots, n_i;$$
$$\ln \{ \pi_{ij} / (1 - \pi_{ij}) \} = \beta_0 + \beta_1 x_{ij} + \gamma_i, \text{ com } x_{ij} = j/n_i \text{ e } \gamma_i \stackrel{iid}{\sim} \text{NI}(0, \sigma^2; F_{U|\nu}).$$

Grande parte dos trabalhos disponíveis na literatura Estatística sobre modelo logístico misto em que o algoritmo EMMC é usado, por exemplo Chen *et al.* (2002), apresentam exemplos de aplicação em que o modelo ajustado não assume o intercepto β_0 . Isto também é visto nos dados sobre acasalamentos de salamandras (McCullagh e Nelder, 1989) analisados por Booth e Hobert (1999), Chen *et al.* (2002) e Levine e Fan (2004). Quintana *et al.* (1999) afirmam que a sequência de estimativas obtidas no EMMC para o parâmetro de intercepto β_0 é menos estável que os outros coeficientes de efeitos fixos. No modelo de intercepto aleatório é intuitivo pensar que o erro Monte Carlo presente no EMMC tem maior impacto na estimação de β_0 , pois é visto como uma espécie de valor médio para o intercepto aleatório. Por isto, entre outras coisas, neste estudo pretendemos também avaliar o efeito de estimarmos β_0 ou assumirmos que β_0 é conhecido.

Para avaliarmos o comportamento do algoritmo implementado, consideramos diferentes valores para k , n_i e n_{ij} . Nos vários cenários abordados, todos os conjuntos de dados são simulados assumindo $\beta_0 = 0$, $\beta_1 = 2$ e $V(\gamma_i) = 1$. Inicialmente, na Subseção 3.6.1, comparamos os algoritmos Metropolis-Hastings descritos na Seção 3.5 em bancos de dados gerados assumindo efeitos aleatórios com distribuição Normal e t -Student. Nas subseções seguintes analisaremos os cenários em que os efeitos aleatórios têm distribuição com caudas mais pesadas do que a distribuição Normal, em particular, as distribuições t -Student, Slash e Normal Contaminada.

Nesta seção, no ajuste dos modelos Logístico-N e Logístico-T rodamos 1100 iterações do algoritmo EMMC e especificamos o número de réplicas Monte Carlo L tal que $L = 10$ para as iterações 1-400, $L = 50$ para as iterações 401-700, $L = 200$ para as iterações 701-900, $L = 500$ para as iterações 901-1000 e $L = 1000$ para as iterações 1001-1100. Por apresentar convergência mais lenta, no modelo Logístico-SL rodamos 1800 iterações em que especificamos $L = 10$ para as iterações 1-400, $L = 50$ para as iterações 401-800, $L = 200$ para as iterações 801-1300, $L = 500$ para as iterações 1301-1550 e $L = 1000$ para as iterações 1551-1800. No modelo Logístico-NC rodamos 1300 iterações, considerando a mesma especificação do modelo Logístico-N até a iteração 900 e, para as iterações seguintes consideramos, $L = 500$ para as iterações 901-1100 e $L = 1000$ para as iterações 1101-1300.

3.6.1 Cenário 1: Comparação dos algoritmos para gerar de γ

No primeiro cenário, o nosso objetivo é avaliar se existem diferenças quanto a convergência das estimativas fornecidas pelo algoritmo EMMC ao utilizarmos as três diferentes versões de algoritmo Metropolis-Hastings, descritas anteriormente na Subseção 3.5.1, para gerarmos das distribuição condicional dos efeitos aleatórios. Para fazermos tal comparação, geramos 8 bancos de dados considerando efeitos aleatórios com distribuição Normal e t -Student. Os bancos de dados são gerados considerando $n_i = 10$ e as demais especificações na Tabela 3.2 a seguir.

Tabela 3.2: Organização dos bancos de dados no Cenário 1

k	n_{ij}	$\gamma_i \stackrel{iid}{\sim} N(0, 1)$	$\gamma_i \stackrel{iid}{\sim} T(0; 0,6; 5)$
50	1	Dados 1	Dados 5
	5	Dados 2	Dados 6
100	1	Dados 3	Dados 7
	5	Dados 4	Dados 8

Os histogramas dos efeitos aleatórios gerados e a f.d.p teórica assumida para os efeitos aleatórios em cada banco de dados podem ser vistos nas figuras apresentadas no Apêndice E.

Para cada banco de dados, ajustamos o modelo que considera a distribuição correta dos efeitos aleatórios e estimamos β_0 , β_1 , σ^2 e ν (quando se trata da t -Student). Para ajustarmos o modelo, utilizamos o algoritmo EMMC com os algoritmos Metropolis-Hastings de McCulloch (EMMC-M), com Passeio Aleatório (EMMC-PA) e de Quintana (EMMC-Q) para gerarmos valores da distribuição condicional dos efeitos aleatórios. No EMMC-PA utilizamos blocos de tamanho $l_1 = 5$ e no algoritmo EMMC-Q consideramos $c = 1$, escolha sugerida pelos autores no caso Normal. Nos algoritmos para o caso do modelo Logístico-T, restringimos os valores de ν entre 2,1 e 30.

Como o algoritmo EMMC aproxima a esperança necessária no passo E do algoritmo EM por uma média Monte Carlo, as estimativas dos parâmetros obtidas neste algoritmo apresentam erro Monte Carlo, dificultando a convergência pontual. Aqui, calculamos a média das estimativas obtidas em uma janela final de iterações para obtermos a estimativa de máxima verossimilhança dos parâmetros. A Tabela 3.3 apresenta a média e o desvio padrão das estimativas obtidas nas últimas 100 iterações do algoritmo EMMC.

Tabela 3.3: Média (desvio padrão) das estimativas fornecidas nas últimas 100 iterações dos algoritmos EMMC-M, EMMC-PA e EMMC-Q

Parâmetro	Real	EMMC-M	EMMC-PA	EMMC-Q
Dados 1				
β_0	0	-0,341 (0,004)	-0,342 (0,003)	-0,342 (0,003)
β_1	2	2,720 (0,003)	2,719 (0,002)	2,719 (0,003)
σ^2	1	0,959 (0,009)	0,955 (0,007)	0,958 (0,008)
Dados 2				
β_0	0	-0,031 (0,003)	-0,039 (0,002)	-0,033 (0,003)
β_0	2	1,949 (0,001)	1,949 (0,001)	1,949 (0,001)
σ^2	1	0,913 (0,003)	0,913 (0,004)	0,911 (0,003)
Dados 3				
β_0	0	0,200 (0,003)	0,198 (0,002)	0,199 (0,003)
β_1	2	1,708 (0,001)	1,708 (0,001)	1,708 (0,002)
σ^2	1	1,171 (0,007)	1,168 (0,004)	1,167 (0,008)
Dados 4				
β_0	0	0,035 (0,002)	0,035 (0,001)	0,035 (0,003)
β_1	2	1,990 (0,001)	1,990 (0,000)	1,989 (0,001)
σ^2	1	1,049 (0,003)	1,048 (0,002)	1,045 (0,003)
Dados 5				
β_0	0	0,313 (0,004)	0,320 (0,004)	0,314 (0,004)
β_1	2	1,554 (0,002)	1,557 (0,003)	1,556 (0,002)
σ^2	0,6	1,255 (0,012)	1,275 (0,013)	1,267 (0,009)
ν	5	28,830 (1,024)	29,466 (0,587)	29,061 (0,792)
$V(\gamma_i)$	1	1,348 (0,011)	1,368 (0,013)	1,361 (0,011)
Dados 6				
β_0	0	0,073 (0,002)	0,080 (0,002)	0,078 (0,003)
β_1	2	1,907 (0,001)	1,907 (0,001)	1,908 (0,001)
σ^2	0,6	0,744 (0,005)	0,737 (0,004)	0,735 (0,006)
ν	5	9,986 (0,233)	9,649 (0,144)	9,338 (0,230)
$V(\gamma_i)$	1	0,931 (0,003)	0,930 (0,006)	0,935 (0,005)
Dados 7				
β_0	0	0,027 (0,002)	0,028 (0,001)	0,027 (0,003)
β_1	2	2,048 (0,001)	2,050 (0,001)	2,048 (0,002)
σ^2	0,6	0,643 (0,007)	0,646 (0,003)	0,635 (0,009)
ν	5	4,749 (0,132)	4,770 (0,052)	4,639 (0,114)
$V(\gamma_i)$	1	1,111 (0,012)	1,113 (0,010)	1,117 (0,009)
Dados 8				
β_0	0	0,020 (0,002)	0,024 (0,001)	0,023 (0,002)
β_1	2	1,936 (0,001)	1,936 (0,001)	1,936 (0,001)
σ^2	0,6	0,709 (0,003)	0,686 (0,005)	0,701 (0,004)
ν	5	6,741 (0,061)	5,971 (0,110)	6,417 (0,070)
$V(\gamma_i)$	1	1,009 (0,004)	1,031 (0,004)	1,019 (0,004)

Analisando a Tabela 3.3, vemos que os três algoritmos conduzem a estimativas pontuais muito próximas para todos os parâmetros e o algoritmo EMMC-PA tende a produzir menor variabilidade nas estimativas. Diferenças maiores são observadas nas estimativas pontuais para o número de graus de liberdade ν nos Dados 5 a 8. Além disso, como é esperado, percebemos que as estimativas tendem a ficar mais próximas dos valores reais e a terem menor erro Monte Carlo (menor desvio padrão das estimativas das últimas 100 iterações do EMMC) nos bancos de dados com maior número de observações (em que $n_{ij} = 5$). Com respeito ao tempo de simulação, não foi verificado ganho ao utilizarmos o

algoritmo EMMC-PA. Para rodar o mesmo número de iterações, os três algoritmos gastam aproximadamente o mesmo tempo. Quanto a taxa de aceitação, como já havia sido afirmado em Quintana *et al.* (1999), o algoritmo EMMC-Q apresenta taxa de aceitação muito maior do que a observada para o algoritmo EMMC-M.

Para avaliarmos melhor a convergência do algoritmo EMMC, analisamos graficamente as estimativas dos parâmetros de interesse obtidas em cada iteração do algoritmo EMMC. As Figuras 20 a 23 apresentam os gráficos das estimativas dos parâmetros em todas as iterações do algoritmo EMMC para os Dados 1, 2, 5 e 7 considerando os três algoritmos Metropolis-Hastings citados. Nestes gráficos, a linha tracejada horizontal representa a média das estimativas obtidas nas últimas 100 iterações do EMMC e as linhas tracejadas verticais representam as iterações em que o número de réplicas Monte Carlo L é aumentado. Para os demais bancos de dados os gráficos das estimativas são apresentados no Apêndice E.

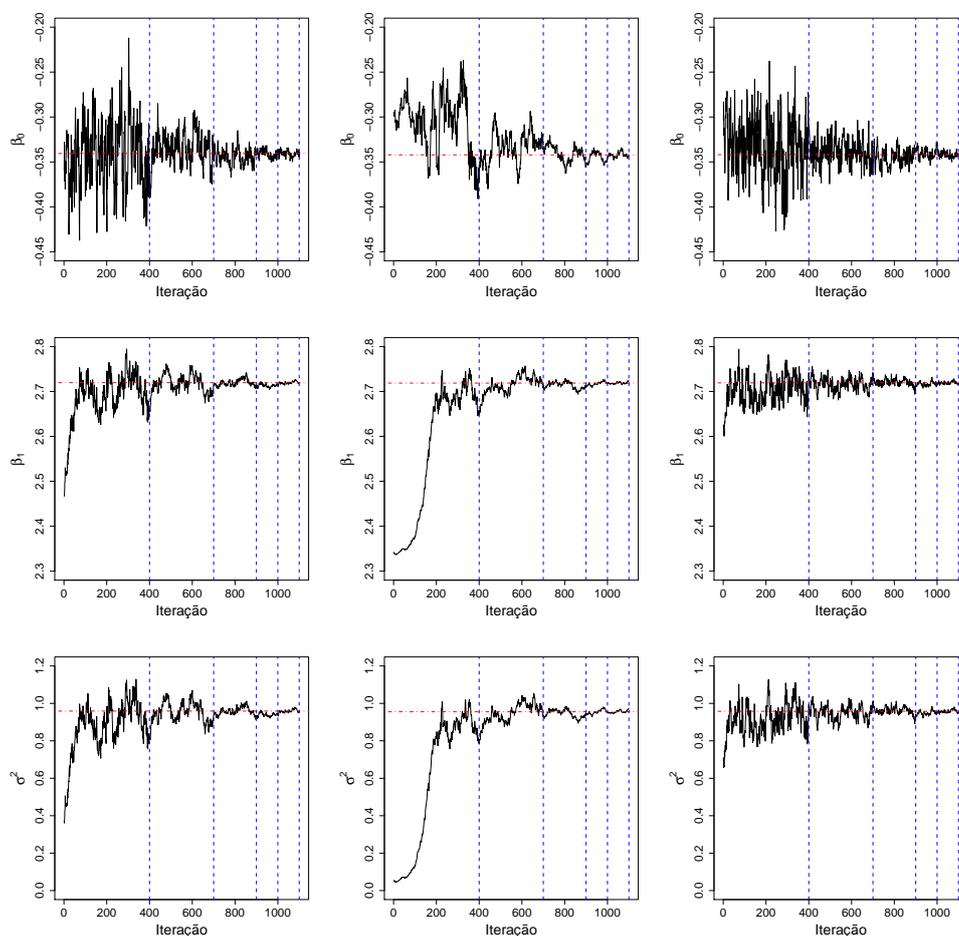


Figura 20: Gráficos das estimativas dos parâmetros *versus* iterações, e média das estimativas nas últimas 100 iterações (linha tracejada), obtidas com os algoritmos EMMC-M (esquerda), EMMC-PA (centro) e EMMC-Q (direita) para Dados 1 no Cenário 1.

Das Figuras 20 a 23 concluímos, como já dito em trabalhos anteriores, que maiores valores para o número L de réplicas Monte Carlo no EMMC levam a uma menor variabilidade nas estimativas, ou seja, diminui-se o erro Monte Carlo das estimativas. Como já

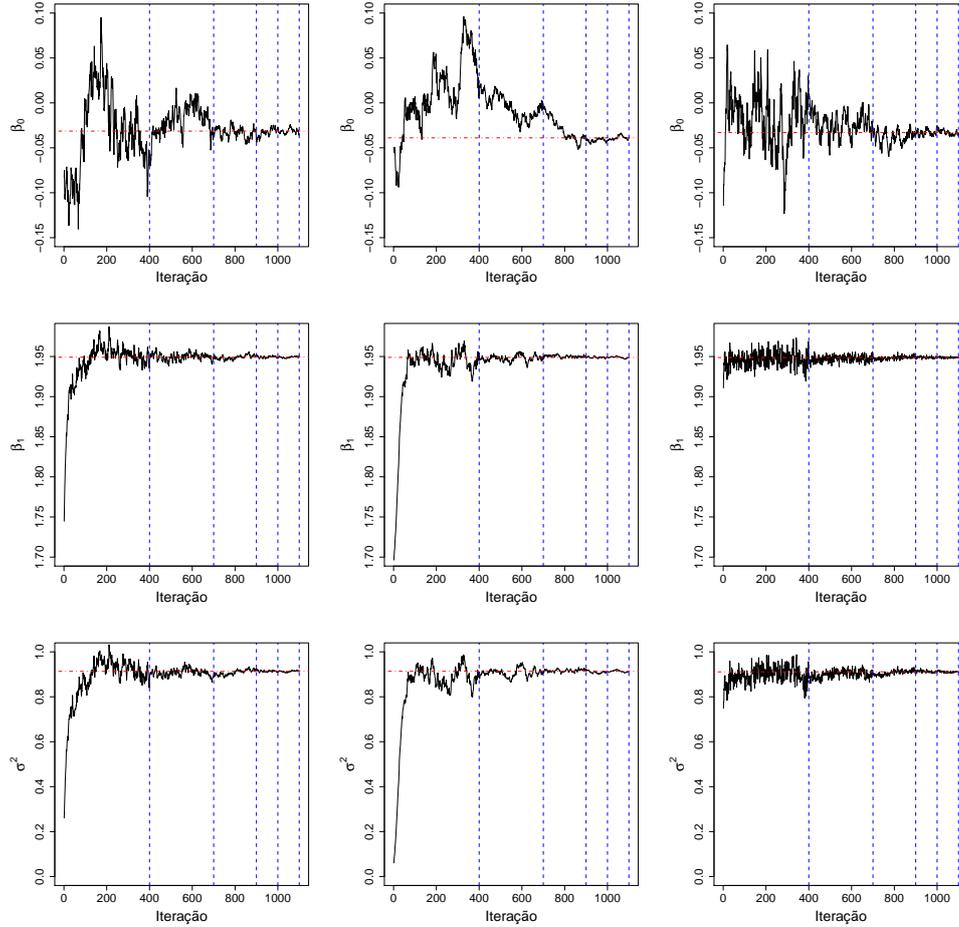


Figura 21: Gráficos das estimativas dos parâmetros *versus* iterações, e média das estimativas nas últimas 100 iterações (linha tracejada), obtidas com os algoritmos EMMC-M (esquerda), EMMC-PA (centro) e EMMC-Q (direita) para Dados 2 no Cenário 1.

háviamos notado na Tabela 3.3, as estimativas obtidas com algoritmo EMMC convergem para praticamente os mesmos valores independentemente da alternativa de Metropolis-Hastings utilizada para gerarmos dos efeitos aleatórios. As estimativas obtidas com o EMMC-Q, principalmente para β_1 e σ^2 , convergem mais rapidamente do que aquelas obtidas usando os outros dois algoritmos. O EMMC-Q aparentemente fornece estimativas mais estáveis, com menor auto-correlação entre as iterações, do que os demais algoritmos, principalmente, no Modelo Logístico-N. As estimativas de β_0 parecem ser mais influenciadas pelo tipo de Metropolis usado para gerarmos da distribuição dos efeitos aleatórios. Notamos que no EMMC-PA as estimativas de β_0 se estabilizam em torno do valor de convergência após um número maior de iterações, mas, neste caso, as estimativas obtidas após a convergência tem menor variabilidade em torno da média das últimas estimativas.

No Modelo Logístico-T, Figuras 22 e 23, observamos que o parâmetro ν é mais difícil de ser estimado. As estimativas deste parâmetro demoram mais a convergir e apresentam maior instabilidade quando valores altos para ν são estimados, veja, por exemplo, as estimativas de ν obtidas para Dados 5. Apesar da instabilidade na estimação de ν , a variância dos efeitos aleatórios $V(\gamma_i)$ apresenta bom comportamento tanto na veloci-

dade de convergência, quanto na variabilidade das estimativas ao longo das iterações do EMMC.

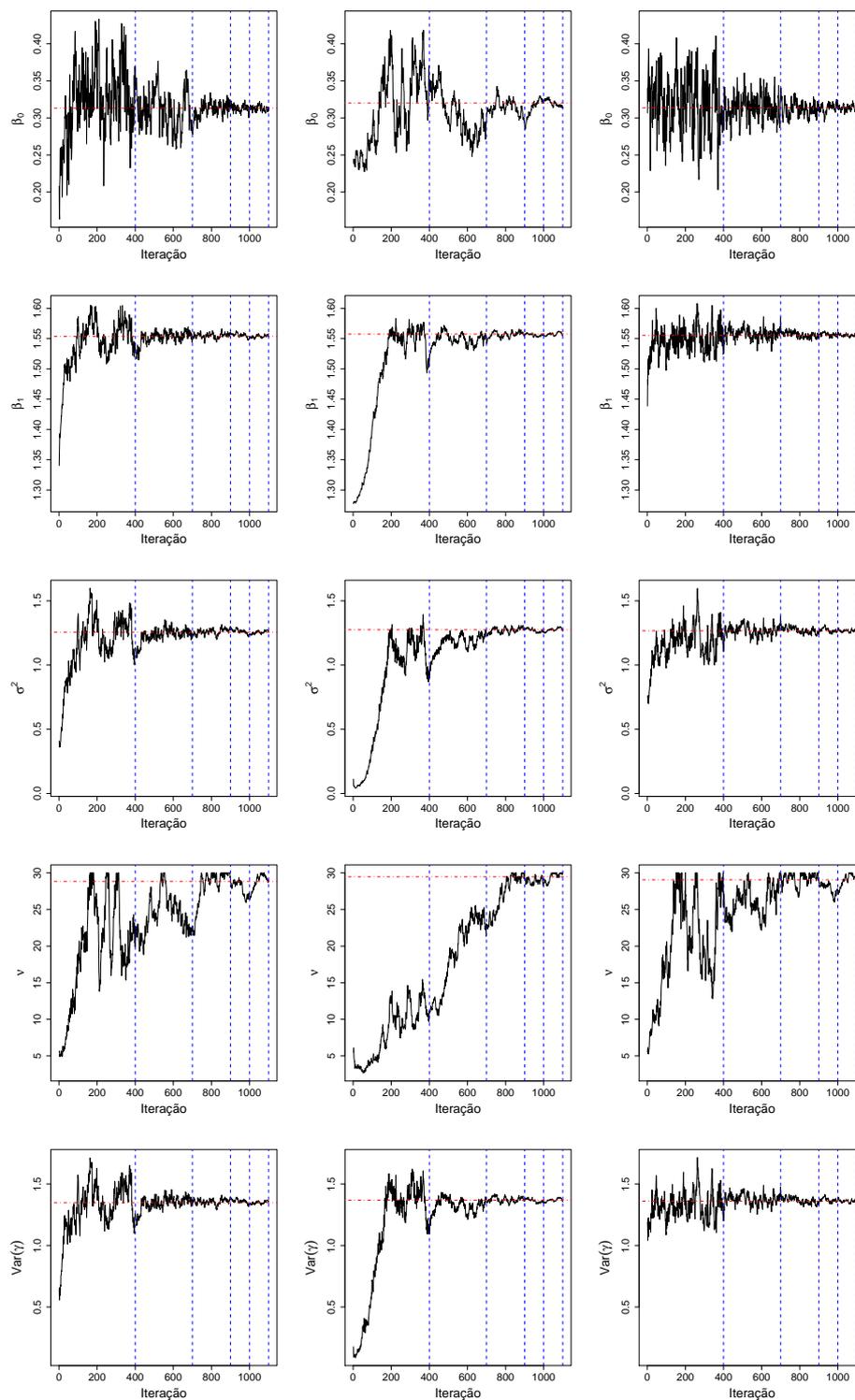


Figura 22: Gráficos das estimativas dos parâmetros *versus* iterações, e média das estimativas nas últimas 100 iterações (linha tracejada), obtidas com os algoritmos EMMC-M (esquerda), EMMC-PA (centro) e EMMC-Q (direita) para Dados 5 no Cenário 1.

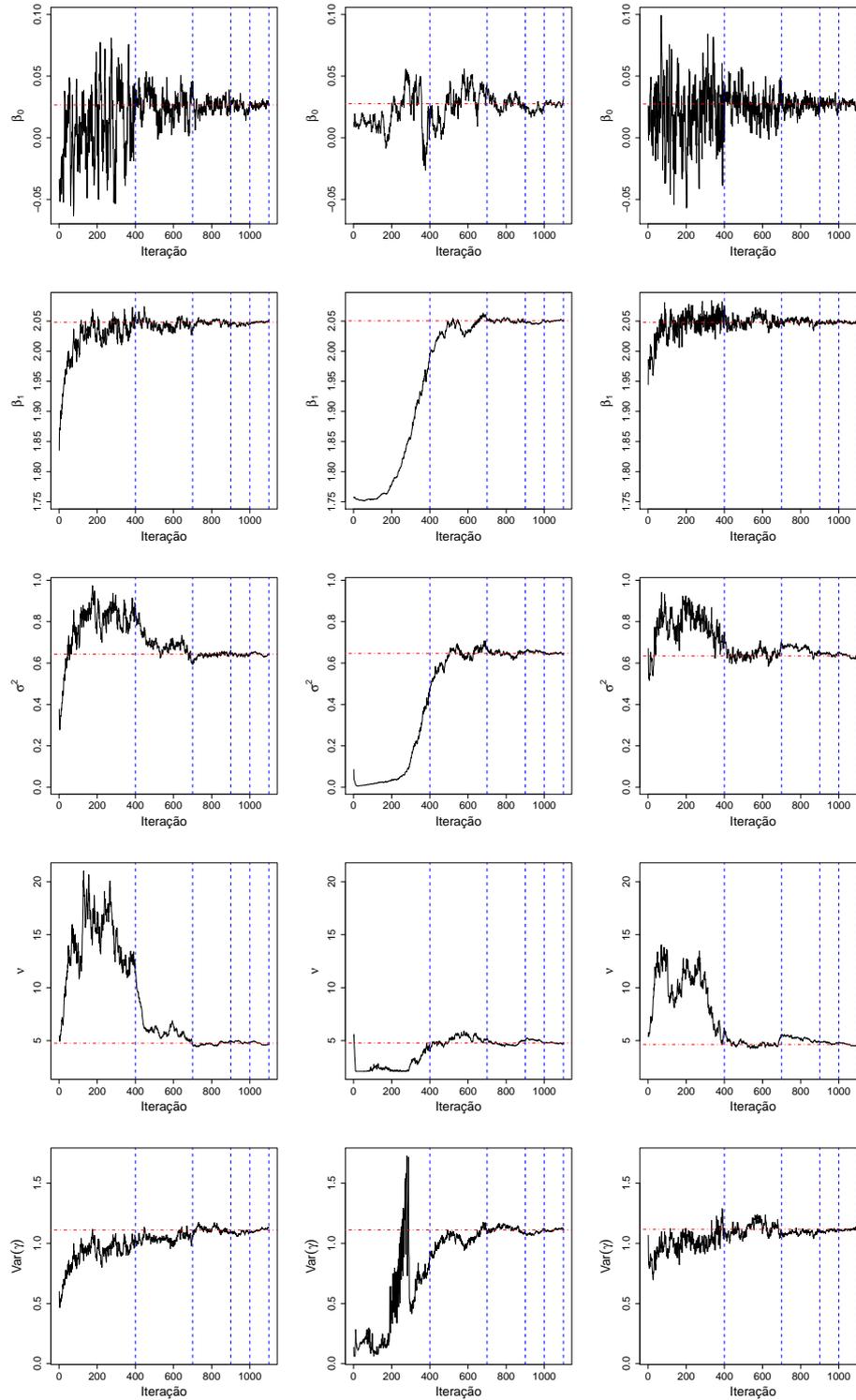


Figura 23: Gráficos das estimativas dos parâmetros *versus* iterações, e média das estimativas nas últimas 100 iterações (linha tracejada), obtidas com os algoritmos EMMC-M (esquerda), EMMC-PA (centro) e EMMC-Q (direita) para Dados 7 no Cenário 1.

3.6.2 Cenário 2: Modelo Logístico-T

Neste segundo cenário, temos como objetivo avaliar a implementação do algoritmo EMMC no modelo Logístico-T em diversos cenários. Desejamos avaliar como o número de *clusters*, o número de observações e o peso da cauda da distribuição dos efeitos aleatórios influenciam no comportamento do algoritmo proposto. Pretendemos também avaliar o impacto de estimarmos o número ν de graus de liberdade ou assumirmos como conhecido. Para isto, geramos bancos de dados explorando diferentes configurações de valores para k , n_i , n_{ij} e ν . Visamos explorar situações em que a distribuição dos efeitos aleatórios tenham caudas mais pesadas do que a distribuição Normal. Para tal, mantendo $V(\gamma_i) = 1$, assumimos $\nu = 4$ e 8 . Na Figura 24 apresentamos os gráficos da f.d.p das distribuições $N(0, 1)$, $T(0; 0,5; 4)$ e $T(0; 0,75; 8)$.

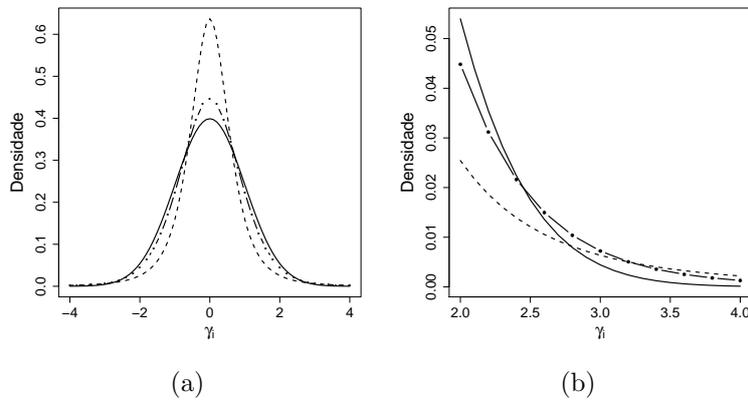


Figura 24: Comparação entre as f.d.p das distribuições Normal padrão (linha sólida), $T(0; 0,5; 4)$ (linha tracejada) e $T(0; 0,75; 8)$ (linha tracejada com ●) em (a) e um *zoom* da cauda a direita em (b).

Os bancos de dados são gerados considerando 7 configurações de valores para as quantidades σ^2 , ν , k , n_i e n_{ij} . Geramos 4 bancos de dados distintos para cada uma destas especificações, totalizando 28 bancos de dados. O objetivo de gerar mais de um banco de dados com a mesma configuração é tornar mais confiáveis as conclusões tiradas a respeito da convergência do algoritmo.

Em todos os bancos de dados, por ter apresentado melhor desempenho no estudo apresentado na Subseção 3.6.1, ajustamos o modelo Logístico-T considerando apenas o algoritmo EMMC-Q. Assim como na Subseção 3.6.1, consideramos a constante $c = 1$.

A seguir iniciamos a análise do comportamento do algoritmo em cenários em que os efeitos aleatórios são gerados de uma distribuição t -Student com caudas muito pesadas ($\nu = 4$) e os quais são considerados diferentes valores para o número n_{ij} de experimentos Bernoulli ao qual é submetido o indivíduo j do *cluster* i .

3.6.2.1 Avaliando o efeito de n_{ij} em cenários com mesmo número de *clusters*

Neste estudo, temos por objetivo explorar o efeito de n_{ij} no que diz respeito ao comportamento do algoritmo e, de forma geral, avaliarmos o comportamento nos casos

em que ν é um valor baixo. Os bancos de dados foram gerados considerando as seguintes especificações:

Configuração 1 (Dados 1 a 4): $\gamma_i \stackrel{iid}{\sim} T(0; 0,5; 4)$, $k = 100$, $n_i = 10$ e $n_{ij} = 1$;

Configuração 2 (Dados 5 a 8): $\gamma_i \stackrel{iid}{\sim} T(0; 0,5; 4)$, $k = 100$, $n_i = 10$ e $n_{ij} = 5$;

Configuração 3 (Dados 9 a 12): $\gamma_i \stackrel{iid}{\sim} T(0; 0,5; 4)$, $k = 100$, $n_i = 10$ e $n_{ij} = 10$.

Inicialmente, em cada banco de dados, ajustamos o modelo Logístico-T e estimamos os graus de liberdade. Para avaliarmos a qualidade das estimativas fornecidas para o grau de liberdade ν da distribuição t -Student dos efeitos aleatórios, apresentamos na Figura 25 os histogramas e f.d.p teórica dos efeitos aleatórios gerados nos bancos de dados com as configurações de 1 a 3.

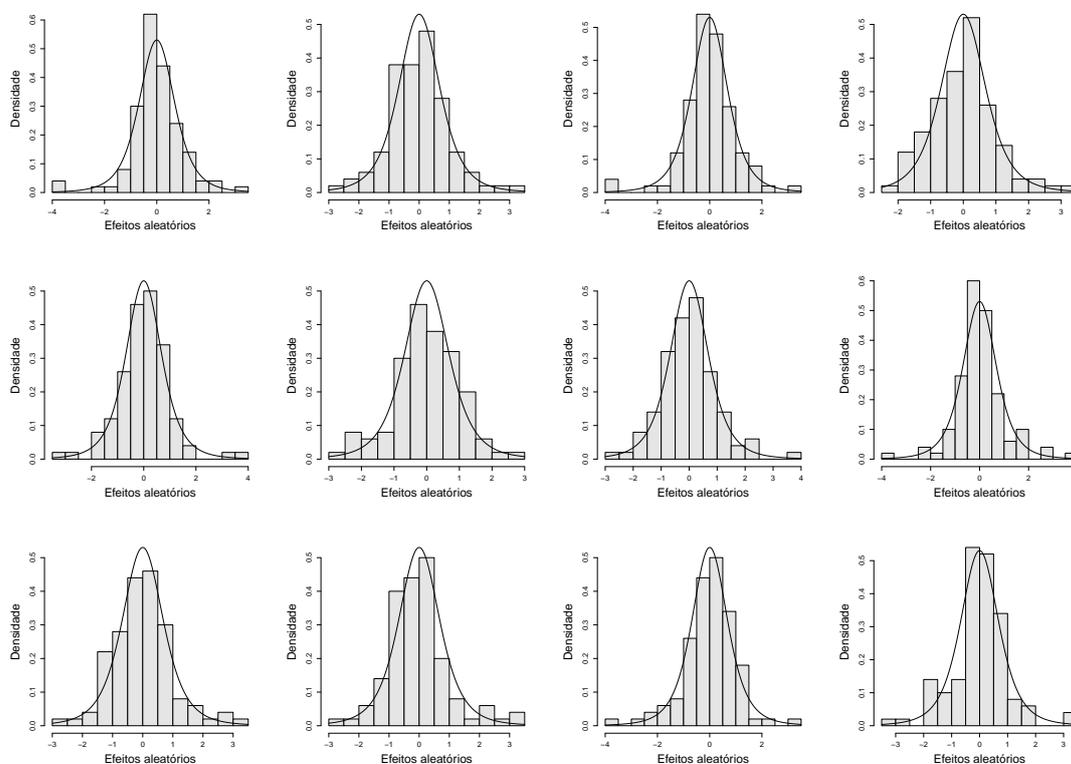


Figura 25: Histograma e f.d.p teórica dos efeitos aleatórios gerados no Cenário 2 para Dados 1 a 12 (da esquerda para direita).

Da Figura 25 percebemos que, em todos os bancos de dados, os histogramas dos valores gerados para os efeitos aleatórios possuem forma razoavelmente similar ao gráfico da f.d.p teórica da distribuição assumida, ou seja, os efeitos aleatórios gerados parecem ser representativos para a suposição de que $\gamma_i \stackrel{iid}{\sim} T(0; 0,5; 4)$.

Na Figura 26 apresentamos os gráficos das estimativas dos parâmetros obtidas nas iterações do algoritmo EMMC-Q para Dados 1 a 4.

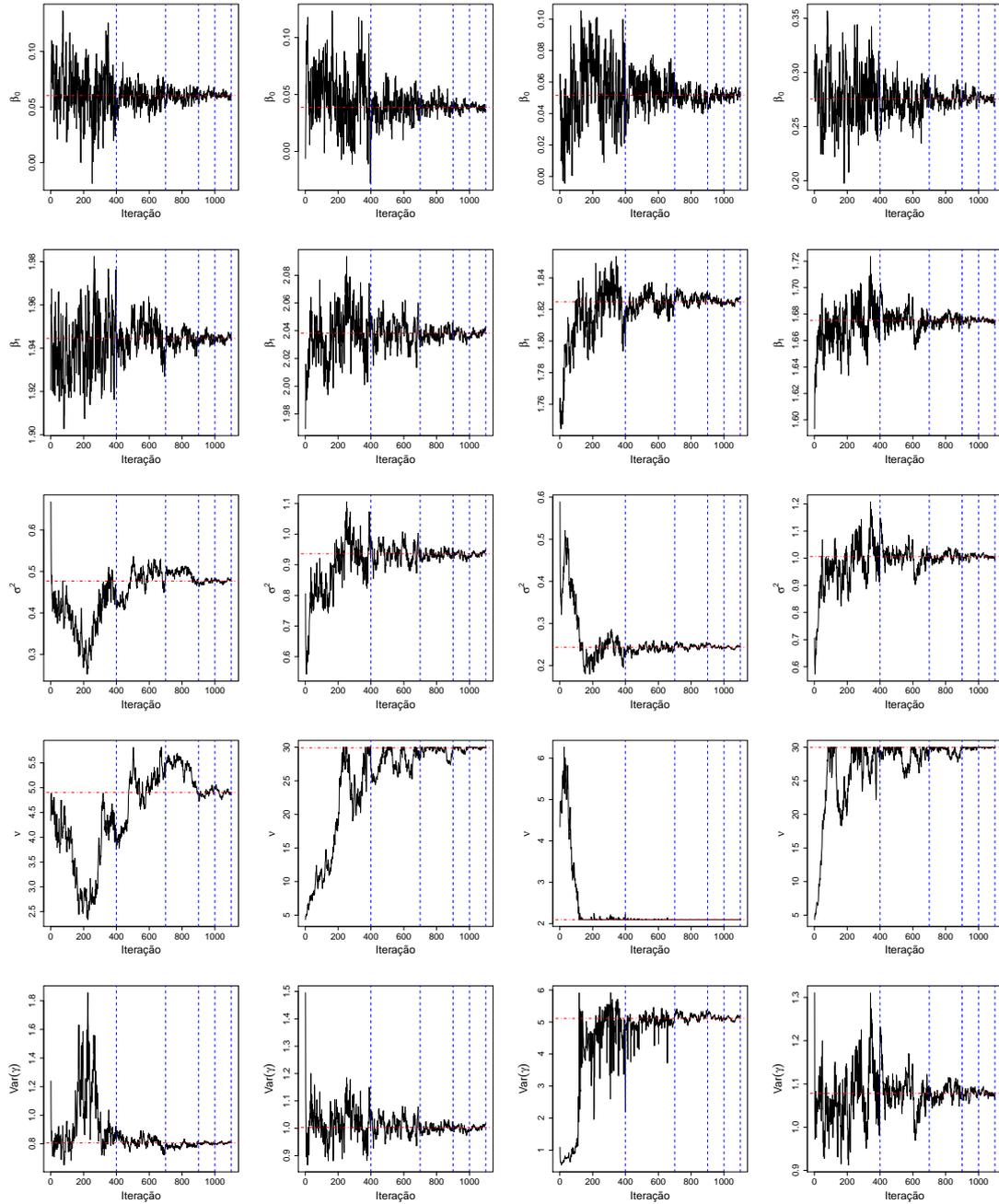


Figura 26: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-Q e média das estimativas nas últimas 100 iterações (linha tracejada) para Dados 1 a 4 (da esquerda para a direita) no Cenário 2.

Notamos na Figura 26 que, as estimativas dos parâmetros de efeitos fixos β_0 e β_1 têm um bom comportamento e apresentam estabilidade em torno dos valores de convergência das estimativas. Naturalmente, quanto maior o valor L menor é o tamanho do erro Monte Carlo e as estimativas se tornam mais estáveis, como pode ser visto nas iterações finais do EMMC. As estimativas para o parâmetro de escala σ^2 também apresentam um bom

comportamento quanto a convergência, mas mostram que são muito influenciadas pelas estimativas do grau de liberdade ν . Nesta configuração onde os bancos de dados são gerados assumindo-se $n_{ij} = 1$, a estimação de ν se torna mais complicada. Vemos que para Dados 2 e 4, o modelo não foi capaz de detectar as caudas pesadas da distribuição dos efeitos aleatórios, estimando ν como sendo 30. Para Dados 1 o valor de ν foi estimado em 5, valor próximo ao valor real de geração, a saber, $\nu = 4$. No entanto, para Dados 3, ν foi estimado em um valor próximo de 2, indicando caudas mais pesadas do que aquela considerada na distribuição teórica. Uma explicação para este comportamento pode ser encontrada ao analisarmos os histogramas dos efeitos aleatórios na Figura 25, em que vemos que os histogramas para Dados 1 e 3 apresentam caudas um pouco mais pesadas do que os Dados 2 e 4. Quanto à estimação da variância da distribuição dos efeitos aleatórios $V(\gamma_i)$, obtivemos estimativas próximas do valor real que é 1, exceto para Dados 3, em que estimativas de ν próximas de 2 levam a estimativas de $V(\gamma_i)$ muito maiores do que o valor real.

A seguir analisamos o comportamento do algoritmo EMMC-Q em bancos de dados com um maior número de observações dentro dos *clusters* (e, conseqüentemente, mais observações) e mesmo número k de *clusters*, como ocorre nos Dados 5 a 8 (em que $n_{ij} = 5$) e nos Dados 9 a 12 (em que $n_{ij} = 10$). Para os Dados 5 a 8 os gráficos das estimativas *versus* iteração são apresentados na Figura 27 e, para os Dados 9 a 12, na Figura 28. Analisando as Figuras 27 e 28 e comparando-as com a Figura 26, vemos que a estimação de ν tende a ser melhor com o aumento de n_{ij} . Nos Dados 5 a 8, vemos que ν é estimado em valores relativamente próximos ao valor real e as estimativas apresentam comportamento bastante estável em torno de um valor de convergência (exceto Dados 7). O mesmo ocorre para σ^2 e $V(\gamma_i)$. Para os efeitos fixos β_0 e β_1 , a convergência do algoritmo torna-se ainda melhor do que o observado para Dados 1 a 4. Na Figura 28 vemos que quando $n_{ij} = 10$, a convergência do algoritmo para todos os parâmetros melhora ainda mais, apresentando grande estabilidade e pouca variabilidade em torno dos valores para os quais convergem as estimativas do EMMC-Q. Em resumo, como é esperado, quanto maior n_{ij} melhor se torna o comportamento do algoritmo EMMC-Q no modelo Logístico-T.

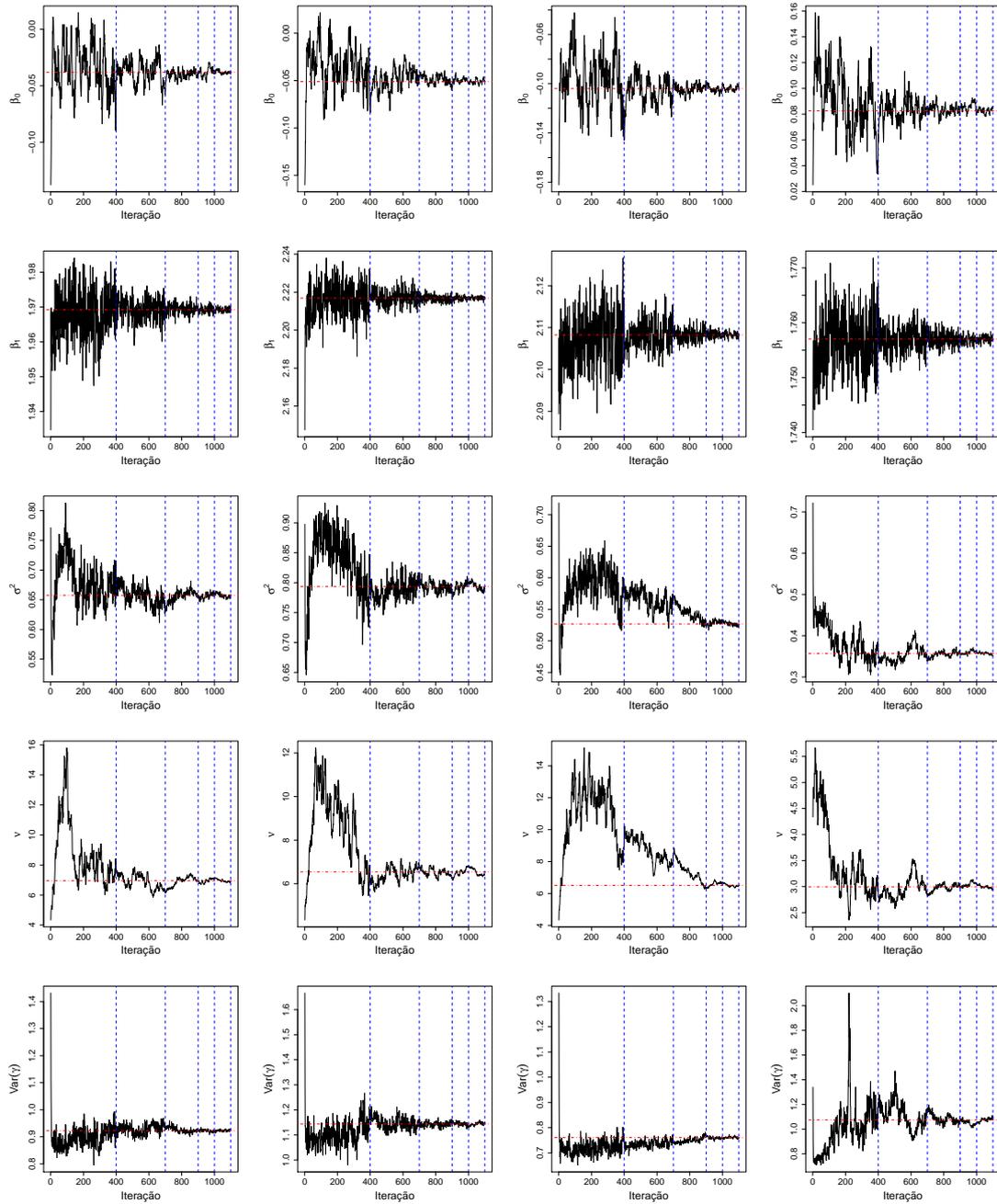


Figura 27: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-Q e média das estimativas nas últimas 100 iterações (linha tracejada) para Dados 5 a 8 (da esquerda para a direita) no Cenário 2.

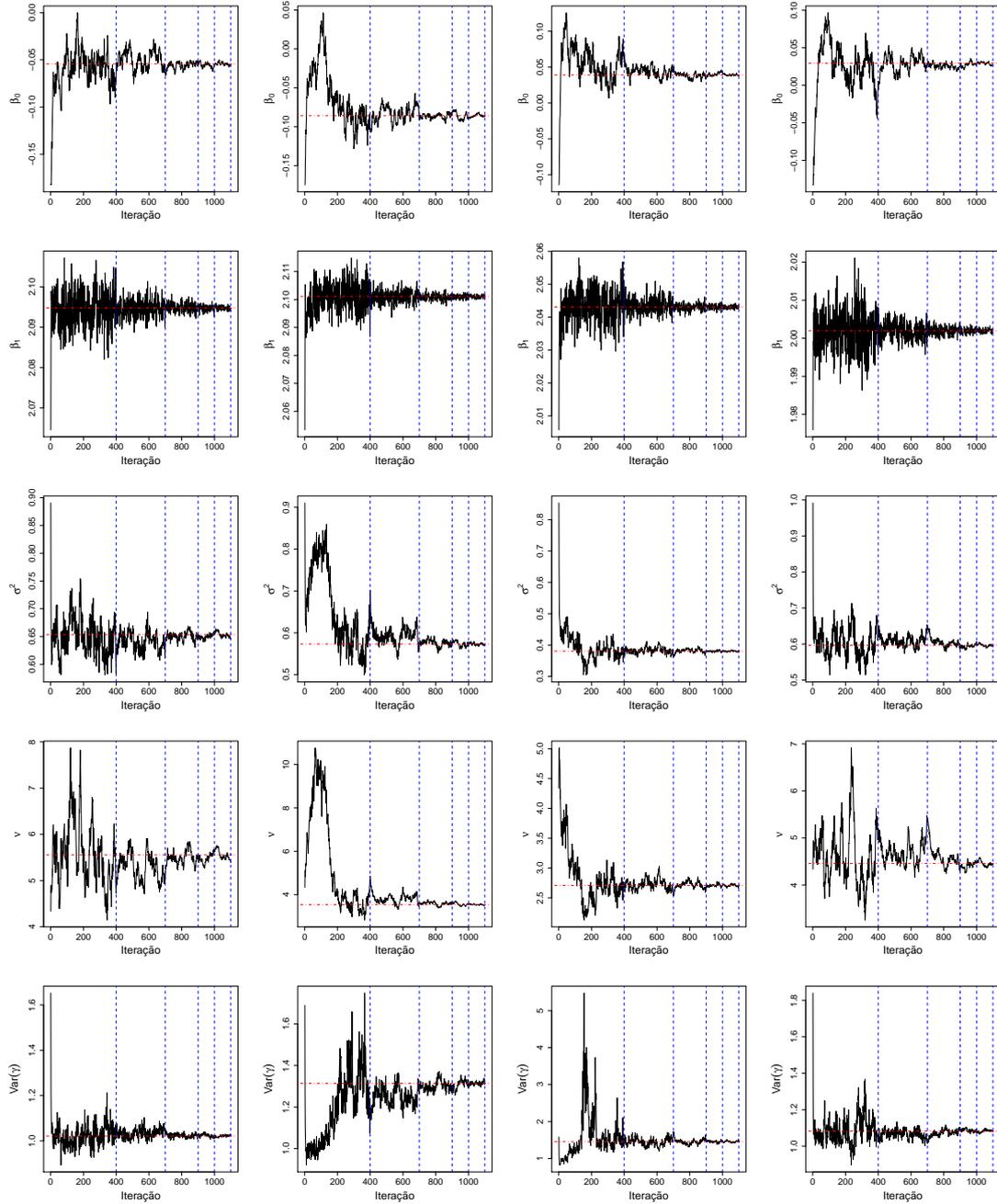


Figura 28: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-Q e média das estimativas nas últimas 100 iterações (linha tracejada) para Dados 9 a 12 (da esquerda para a direita) no Cenário 2.

3.6.2.2 Avaliando o efeito de fixarmos ν

Após esta análise, nos questionamos sobre o comportamento do EMMC-Q se considerarmos como conhecido o grau de liberdade da distribuição *t*-Student assumida para representar o comportamento dos efeitos aleatórios. Para avaliarmos este aspecto, ajustamos novamente o modelo Logístico-T aos Dados 1 a 4, mas agora considerando ν como

um parâmetro conhecido e igual a 4 (valor real). Na Figura 29 apresentamos os gráficos das estimativas dos parâmetros do modelo no algoritmo EMMC.

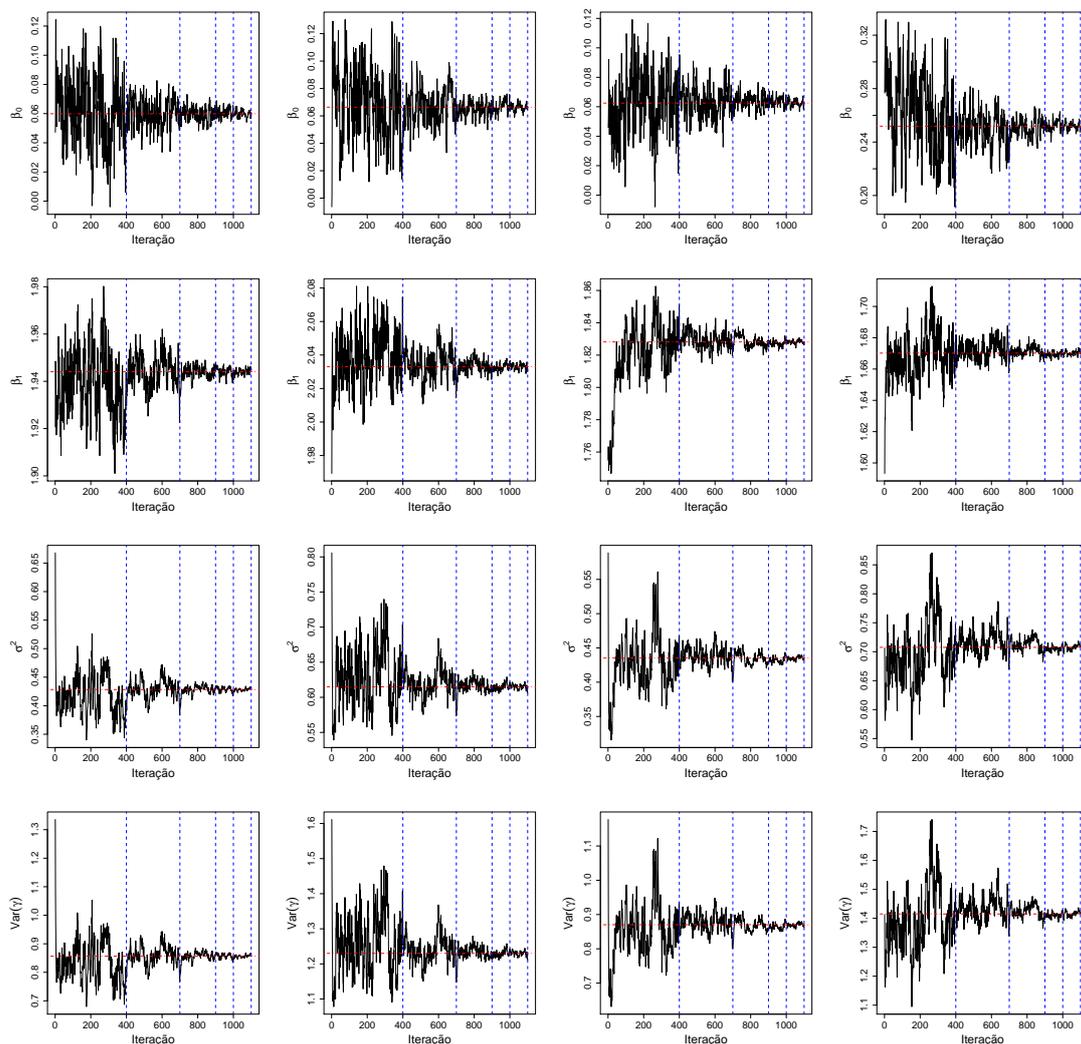


Figura 29: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-Q e média das estimativas nas últimas 100 iterações (linha tracejada) para Dados 1 a 4 (da esquerda para a direita) no Cenário 2 ao considerarmos $\nu = 4$.

Vemos na Figura 29 que ao assumirmos ν conhecido e igual ao valor real, todos os parâmetros apresentam bom comportamento de convergência nos 4 bancos de dados analisados. Ao compararmos com os resultados da Figura 26, onde ν era estimado, vemos que não existe grande mudança quanto as estimativas finais dos parâmetros β associados aos efeitos fixos.

3.6.2.3 Avaliando o efeito da existência de β_0 no modelo

Como dito anteriormente, em muitos dos exemplos apresentados nos artigos que estudam a implementação do algoritmo EMMC em dados binários ou binomiais os modelos ajustados não consideram o intercepto β_0 e não apresentam justificativa para a ausência

de tal parâmetro. Agora pretendemos avaliar se estimar tal parâmetro traz alguma consequência nas estimativas dos demais parâmetros. Com o objetivo de avaliarmos o impacto de estimarmos ou não β_0 , para Dados 5 a 8, reajustamos o modelo retirando β_0 do preditor linear. Na Figura 30 apresentamos os gráficos das estimativas dos parâmetros do modelo no algoritmo EMMC ao assumirmos $\beta_0 = 0$.

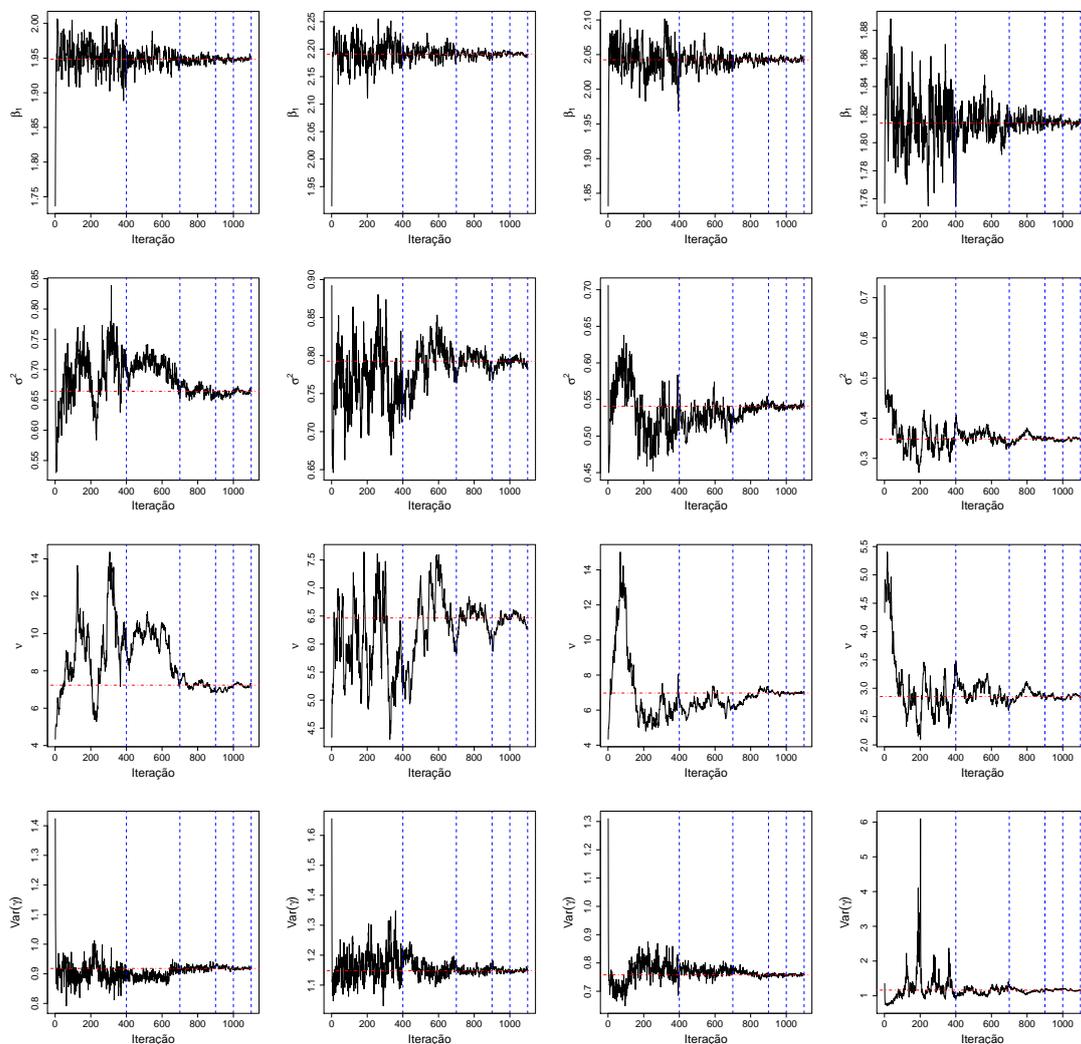


Figura 30: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-Q e média das estimativas nas últimas 100 iterações (linha tracejada) para Dados 5 a 8 (da esquerda para a direita) no Cenário 2 ao considerarmos $\beta_0 = 0$.

Ao compararmos as Figuras 27 e 30 não vemos grande diferença quanto a convergência das estimativas fornecidas pelo algoritmo EMMC-Q ao fixarmos o valor de β_0 . Analisando o comportamento da sequência de estimativas para ν para Dados 7, temos que as estimativas se tornam um pouco mais estáveis se não estimarmos β_0 . Para Dados 6 também existe uma variação pequena nas estimativas no início do algoritmo, tanto é que as estimativas de ν ficaram entre 4,5 e 7,5 enquanto anteriormente variavam entre 4 e 12. De forma geral, não é possível apontarmos um grande ganho na qualidade das

estimativas dos outros parâmetros do modelo ao assumirmos o valor de β_0 como conhecido. No entanto, como vimos em outros cenários, parece haver uma dificuldade maior na convergência das estimativas para β_0 .

3.6.2.4 Avaliando os efeitos de k , n_i e do aumento de ν

Prosseguimos a análise considerando bancos de dados em que os efeitos aleatórios são gerados a partir de uma distribuição t -Student cujo grau de liberdade tem valor intermediário, isto é, uma distribuição t -Student que não tenha cauda muito pesada e que tão pouco seja muito próxima da distribuição Normal. Além de avaliarmos se o algoritmo EMMC-Q para o modelo Logístico-T funciona bem neste caso, queremos estudar o comportamento do algoritmo para diferentes números de *clusters* k e diferentes números de observações n_i dentro de cada *cluster* i . Para isto, geramos dados assumindo valores de $k = 50$ e 100 e assumindo $n_i = 5$ e 10 . Os bancos de dados são gerados seguindo as seguintes configurações:

Configuração 4 (Dados 13 a 16): $\gamma_i \stackrel{iid}{\sim} T(0; 0,75; 8)$, $k = 50$, $n_i = 5$ e $n_{ij} = 5$;

Configuração 5 (Dados 17 a 20): $\gamma_i \stackrel{iid}{\sim} T(0; 0,75; 8)$, $k = 50$, $n_i = 10$ e $n_{ij} = 5$;

Configuração 6 (Dados 21 a 24): $\gamma_i \stackrel{iid}{\sim} T(0; 0,75; 8)$, $k = 100$, $n_i = 5$ e $n_{ij} = 5$;

Configuração 7 (Dados 25 a 28): $\gamma_i \stackrel{iid}{\sim} T(0; 0,75; 8)$, $k = 100$, $n_i = 10$ e $n_{ij} = 5$.

Os histogramas dos efeitos aleatórios gerados nos casos descritos acima com respectiva f.d.p teórica são apresentados na Figura 86 disponível no Apêndice E.

Nas Figuras 31 a 34 apresentamos os gráficos das estimativas dos parâmetros do modelo Logístico-T *versus* iteração para Dados 13 a 16. Vemos que, em todos os bancos de dados, o algoritmo EMMC-Q apresentou bom comportamento para os efeitos fixos β_0 e β_1 . Quando comparamos o impacto do número de observações dentro de cada *cluster*, observamos para os bancos de dados com $n_i = 5$, em geral, o algoritmo não consegue estimar valores baixos para ν e, por isso, não consegue captar que a distribuição dos efeitos aleatórios tem caudas razoavelmente pesadas. Considerando k fixo, o comportamento do algoritmo é melhor em termos de convergência das estimativas nos bancos de dados que possuem $n_i = 10$, principalmente, no que se refere a estimação de ν .

Para avaliarmos a influência do número k de *clusters*, algumas comparações diferentes são possíveis. A princípio podemos comparar o comportamento das sequências de estimativas obtidas para os bancos de dados com mesmo valor de n_i , mas com diferentes valores de k . Podemos comparar, por exemplo, os gráficos apresentados na Figura 32 ($k = 50$ e $n_i = 10$) com os gráficos da Figura 34 ($k = 100$ e $n_i = 10$). Nesta comparação, concluímos que parece não existir diferenças quanto ao comportamento da sequência de estimativas obtidas para os parâmetros, ou seja, o algoritmo se comporta de forma semelhante no que diz respeito a convergência das estimativas dos parâmetros. Um maior valor de k significa que existem mais variáveis latentes no modelo e, sendo assim, seria natural esperarmos uma piora nas estimativa fornecidas pelo modelo, o que não ocorre. Uma explicação possível para este efeito é que, ao compararmos bancos de dados com mesmo n_i , estamos comparando bancos de dados com tamanho N amostral diferente. Por exemplo, na comparação realizada anteriormente, Dados 17 a 20 (Figura 32) são compostos por observações de uma variável resposta binomial de 500 indivíduos enquanto que

para Dados 25 a 28 (Figura 34) são compostos por observações de 100 indivíduos. Uma outra comparação possível para avaliarmos o impacto de k é compararmos as estimativas fornecidas considerando Dados 17 a 20 com as estimativas para Dados 21 a 24, quando todos os bancos de dados possuem tamanho amostral de $N = 500$ observações. Neste caso, notamos que aumentar o número de *clusters* k mantendo fixo o tamanho da amostra tende a piorar o comportamento das estimativas dos parâmetros da distribuição dos efeitos aleatórios.

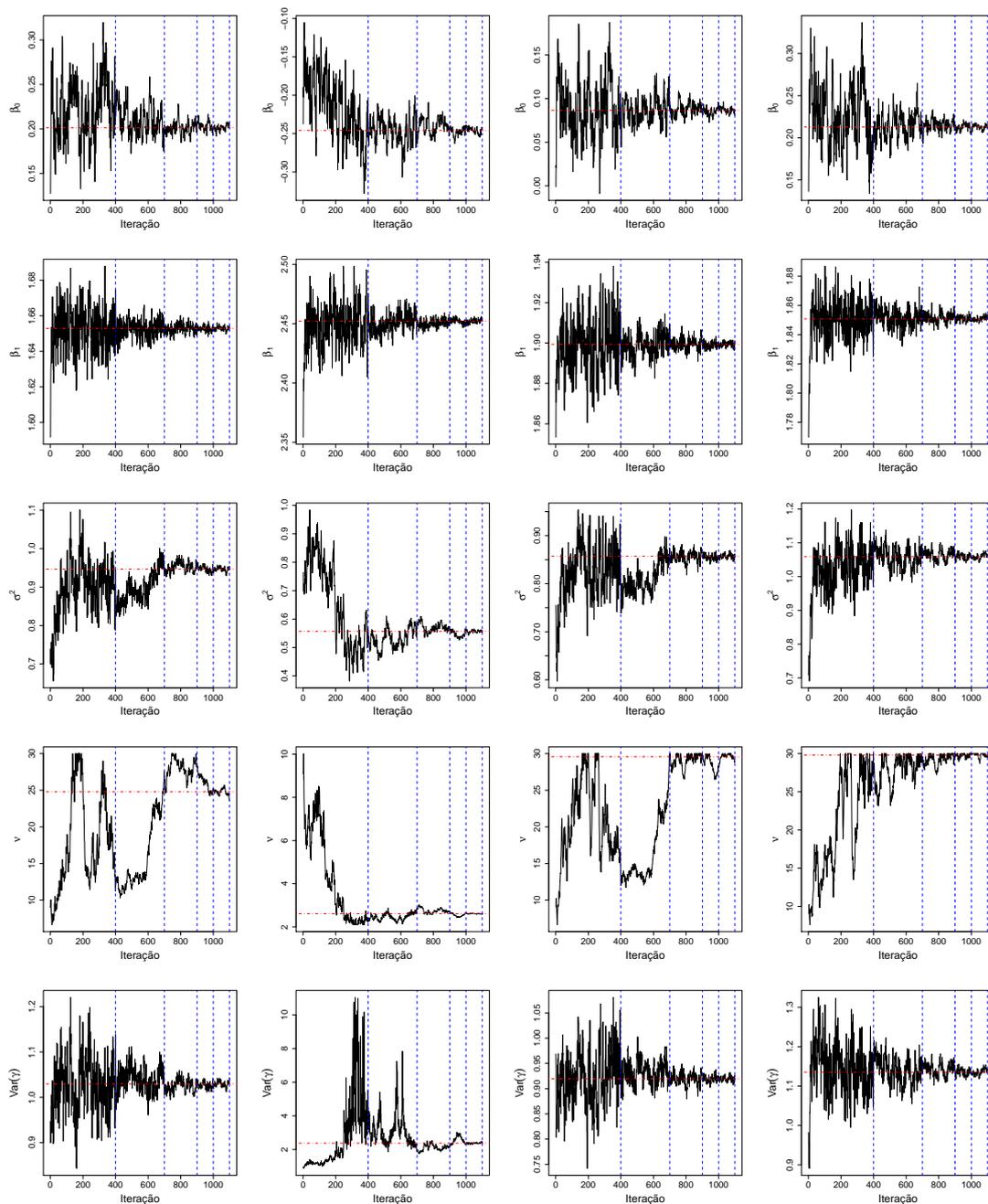


Figura 31: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-Q e média das estimativas nas últimas 100 iterações (linha tracejada) para Dados 13 a 16 (da esquerda para a direita) no Cenário 2.

Por último, comparamos os gráficos das estimativas dos parâmetros para Dados 5 a 8 (Figura 27) e Dados 25 a 28 (Figura 34) para avaliarmos se o valor real de ν influencia na convergência do algoritmo. Observando estes gráficos não é possível notar diferenças do algoritmo quanto ao comportamento de convergência das estimativas, ou seja, o peso da cauda da distribuição dos efeitos aleatórios parece não influenciar na velocidade da convergência e no erro Monte Carlo das estimativas.

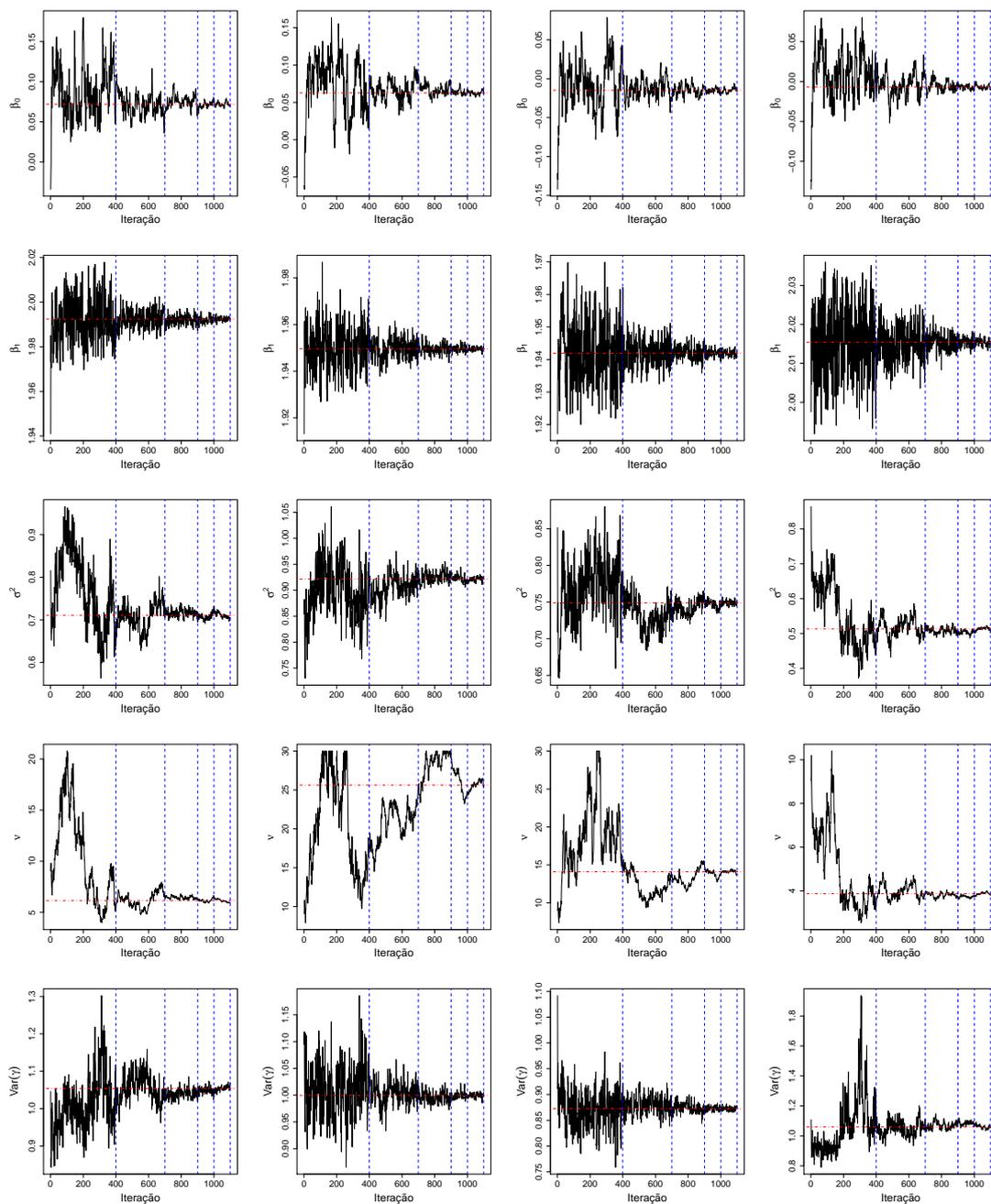


Figura 32: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-Q e média das estimativas nas últimas 100 iterações (linha tracejada) para Dados 17 a 20 (da esquerda para a direita) no Cenário 2.

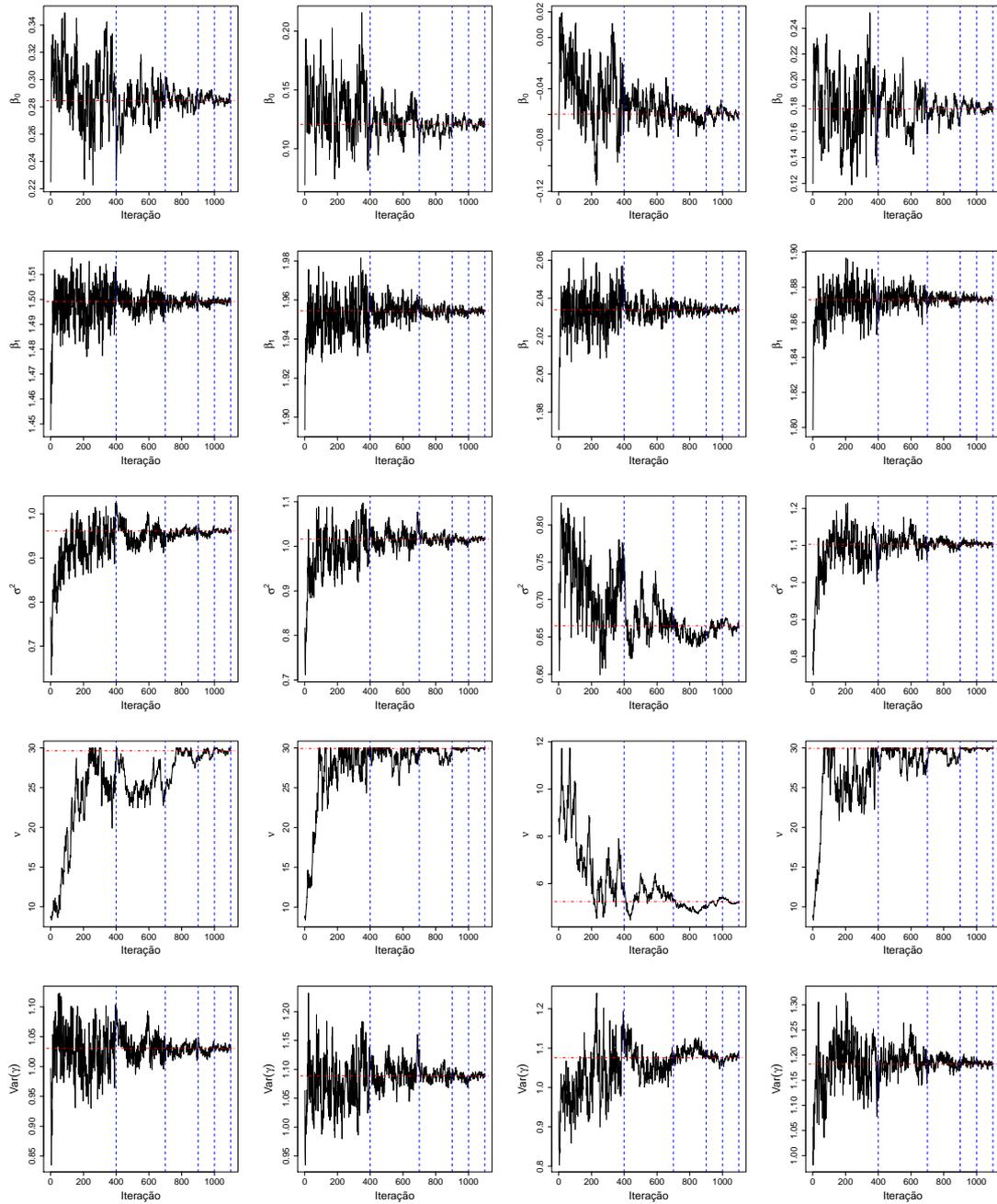


Figura 33: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-Q e média das estimativas nas últimas 100 iterações (linha tracejada) para Dados 21 a 24 (da esquerda para a direita) no Cenário 2.

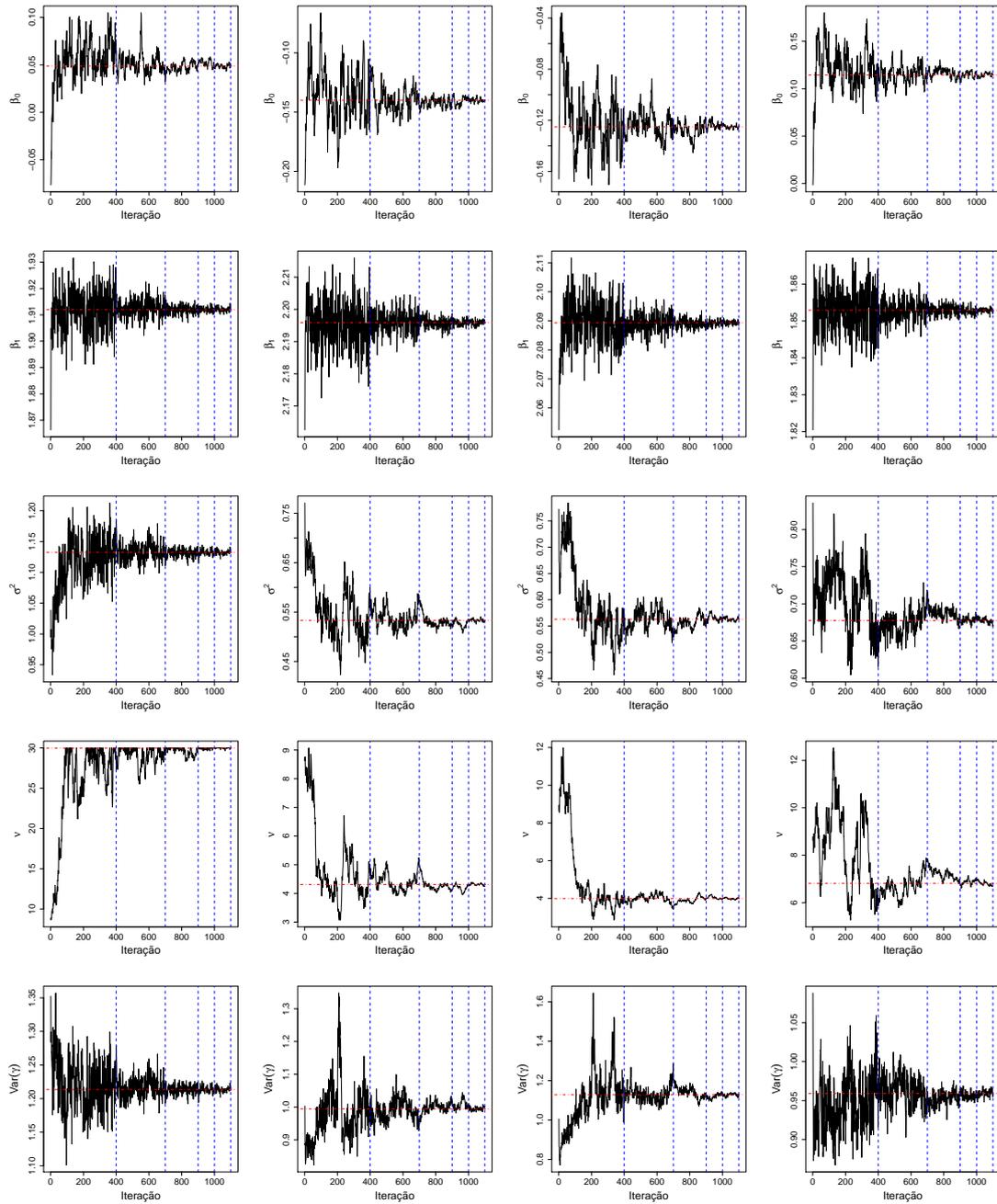


Figura 34: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC e média das estimativas nas últimas 100 iterações (linha tracejada) para Dados 25 a 28 (da esquerda para a direita) no Cenário 2.

3.6.3 Cenário 3: Modelo Logístico-SL

Nosso objetivo aqui é avaliar o comportamento das estimativas obtidas no algoritmo EMMC ao assumirmos que os efeitos aleatórios possuem distribuição de probabilidade Slash. Pretendemos fazer esta avaliação considerando diferentes valores de ν visando avaliar se o valor de tal parâmetro influencia, de alguma forma, na qualidade das estimativas fornecidas pelo algoritmo.

No modelo Logístico-T, obtivemos bons resultados ao utilizarmos o algoritmo EMMC-Q especificando $c = 1$, como sugerido por Quintana *et al.* (1999) para o caso Normal. Inicialmente, nossa ideia era utilizar apenas o algoritmo EMMC-Q, como ocorreu na Subseção 3.6.2. No entanto, como veremos adiante, não obtivemos bons resultados com EMMC-Q ($c = 1$) ao ajustarmos o modelo Logístico-SL. Para efeito de comparação, nesta seção, ajustaremos o modelo Logístico-SL aos bancos de dados simulados utilizando os algoritmos EMMC-M e EMMC-Q. Com o objetivo de avaliarmos a influência da especificação da constante c no algoritmo EMMC-Q, consideraremos $c = 0, 1, 0, 3, 0, 5$ e 1 .

Nesta seção, consideramos 12 bancos de dados em que $k = 100$, $n_i = 10$, $n_{ij} = 5$ e $V(\gamma_i) = 1$, que diferem quanto a distribuição assumida para os efeitos aleatórios e seguem as seguintes especificações para a distribuição dos efeitos aleatórios:

Configuração 1 (Dados 1 a 4): $\gamma_i \stackrel{iid}{\sim} SL(0; 0,333; 1,5)$;

Configuração 2 (Dados 5 a 8): $\gamma_i \stackrel{iid}{\sim} SL(0; 0,667; 3)$;

Configuração 3 (Dados 9 a 12): $\gamma_i \stackrel{iid}{\sim} SL(0; 0,875; 8)$.

Na Figura 35 apresentamos uma comparação entre as f.d.p da distribuição Slash com os valores considerados para os parâmetros e a f.d.p da distribuição $N(0, 1)$. Notamos que a distribuição Slash com $\nu = 1,5$ ou 3 possui caudas mais pesadas do que a distribuição Normal padrão, principalmente para $\nu = 1,5$. Se $\nu = 8$ a distribuição Slash possui f.d.p muito similar a distribuição Normal e, conseqüentemente, as curvas estão sobrepostas nos gráficos da Figura 35.

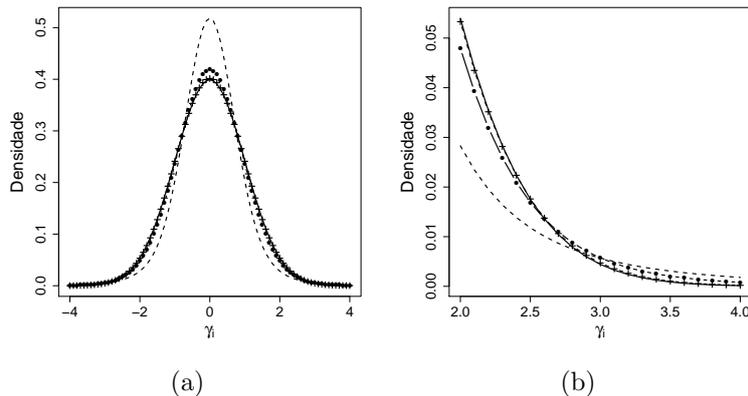


Figura 35: Comparação entre as f.d.p das distribuições Normal padrão (linha sólida), $SL(0; 0,333; 1,5)$ (linha tracejada), $SL(0; 0,667; 3)$ (linha tracejada com \bullet) e $SL(0; 0,875; 8)$ (linha tracejada com $+$) em (a) e uma visualização mais próxima da cauda a direita em (b).

Na Figura 36 apresentamos os histogramas e f.d.p teóricas dos efeitos aleatórios gerados no Cenário 3. Vemos que, em geral, os histogramas estão bem ajustados as f.d.p teóricas. Como já seria esperado, os histogramas para Dados 1 a 4 possuem mais valores atípicos, atestando nestes casos a distribuição dos efeitos aleatórios têm caudas mais pesadas do que nos demais bancos de dados.

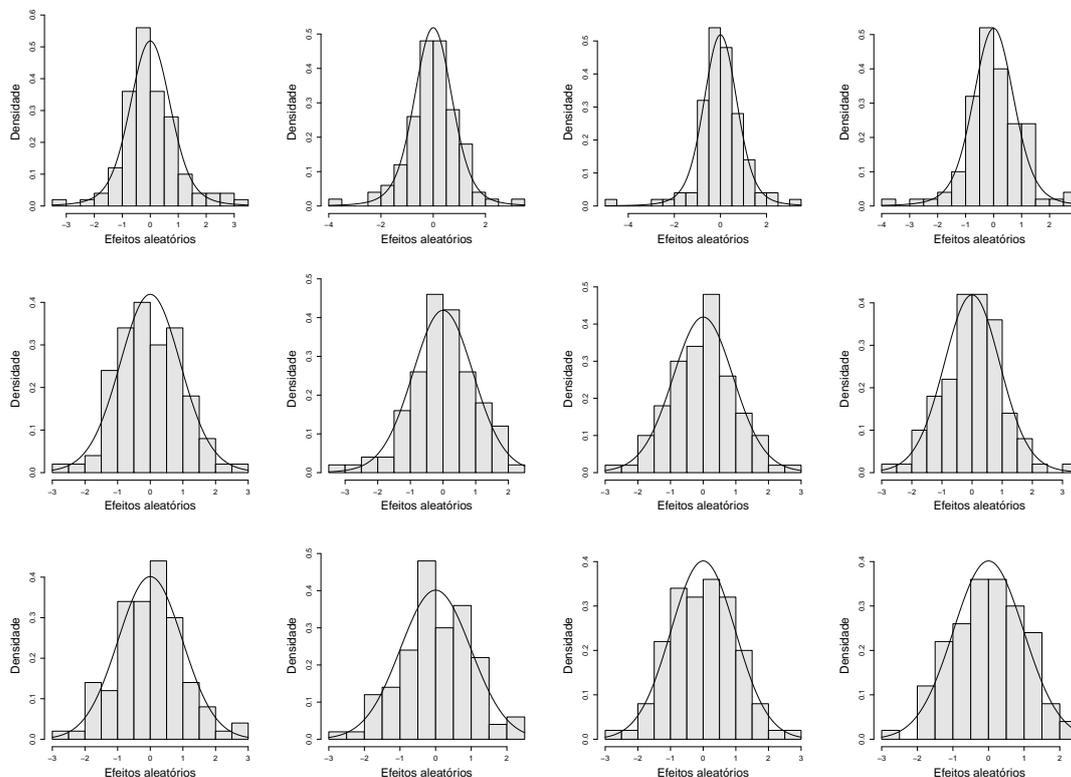


Figura 36: Histograma e f.d.p teórica dos efeitos aleatórios gerados no Cenário 3 para Dados 1 a 12 (da esquerda para direita).

Na implementação do modelo Logístico-SL, para gerarmos valores das variáveis latentes U_i , utilizamos misturas de 30 distribuições Betas. Optamos por um número alto de componentes de mistura para que a probabilidade de aceitação do algoritmo proposto por Philippe (1997) seja muito próxima de 1 e possamos, assim, eliminar a etapa de aceitação do algoritmo. Para garantirmos a existência e estabilidade nas estimativas para $V(\gamma_i)$, nos algoritmos maximizamos ν com a restrição de que $\nu > 1,1$.

Inicialmente, ajustamos o modelo Logístico-SL aos bancos de dados simulados considerando apenas o algoritmo EMMC-M e estes resultados são apresentados a seguir.

3.6.3.1 Avaliando o efeito do aumento de ν : Algoritmo EMMC-M

Ao ajustarmos o modelo Logístico-SL aos bancos de dados simulados com o algoritmo EMMC-M, as taxas de aceitação na geração de valores para γ_i ficaram entre 26% e 37%, sendo que na maior parte das vezes estiveram mais próximas de 30%. Na Figura 37

apresentamos os Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-M para os Dados 1 a 4.

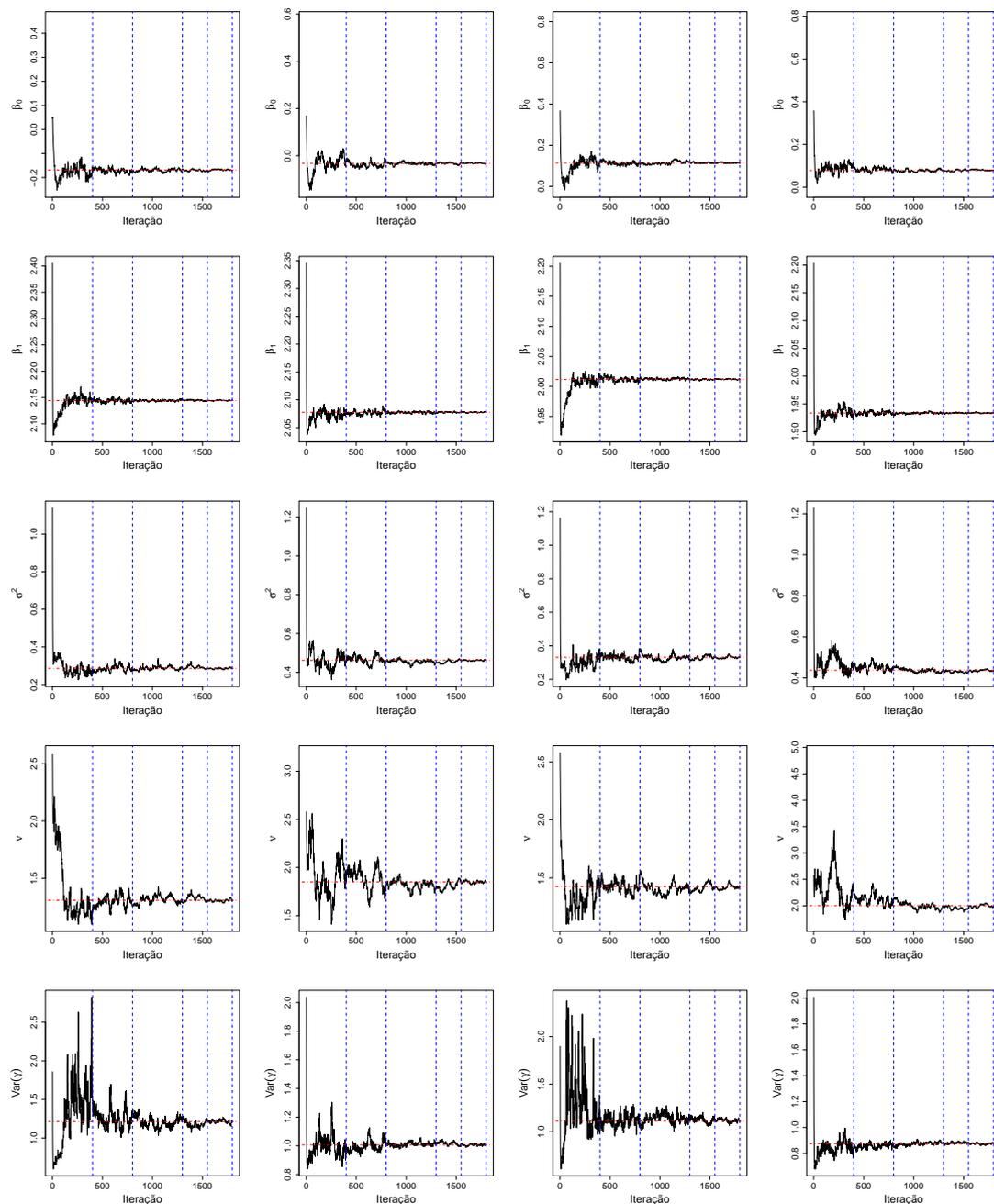


Figura 37: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-M e média das estimativas nas últimas 100 iterações (linha tracejada) para Dados 1 a 4 (da esquerda para a direita) no Cenário 3.

Observamos que, em geral, as estimativas dos parâmetros apresentam bom comportamento quanto a convergência, principalmente, aquelas para os efeitos fixos. O modelo forneceu estimativas baixas para o parâmetro ν , próximos ao valor real de 1,5, demonstrando que foi capaz de captar que os efeitos aleatórios tem cauda pesada.

Nas Figuras 38 e 39 apresentamos os gráficos das estimativas dos parâmetros em todas as iterações do algoritmo EMMC-M para os Dados 5 a 12.

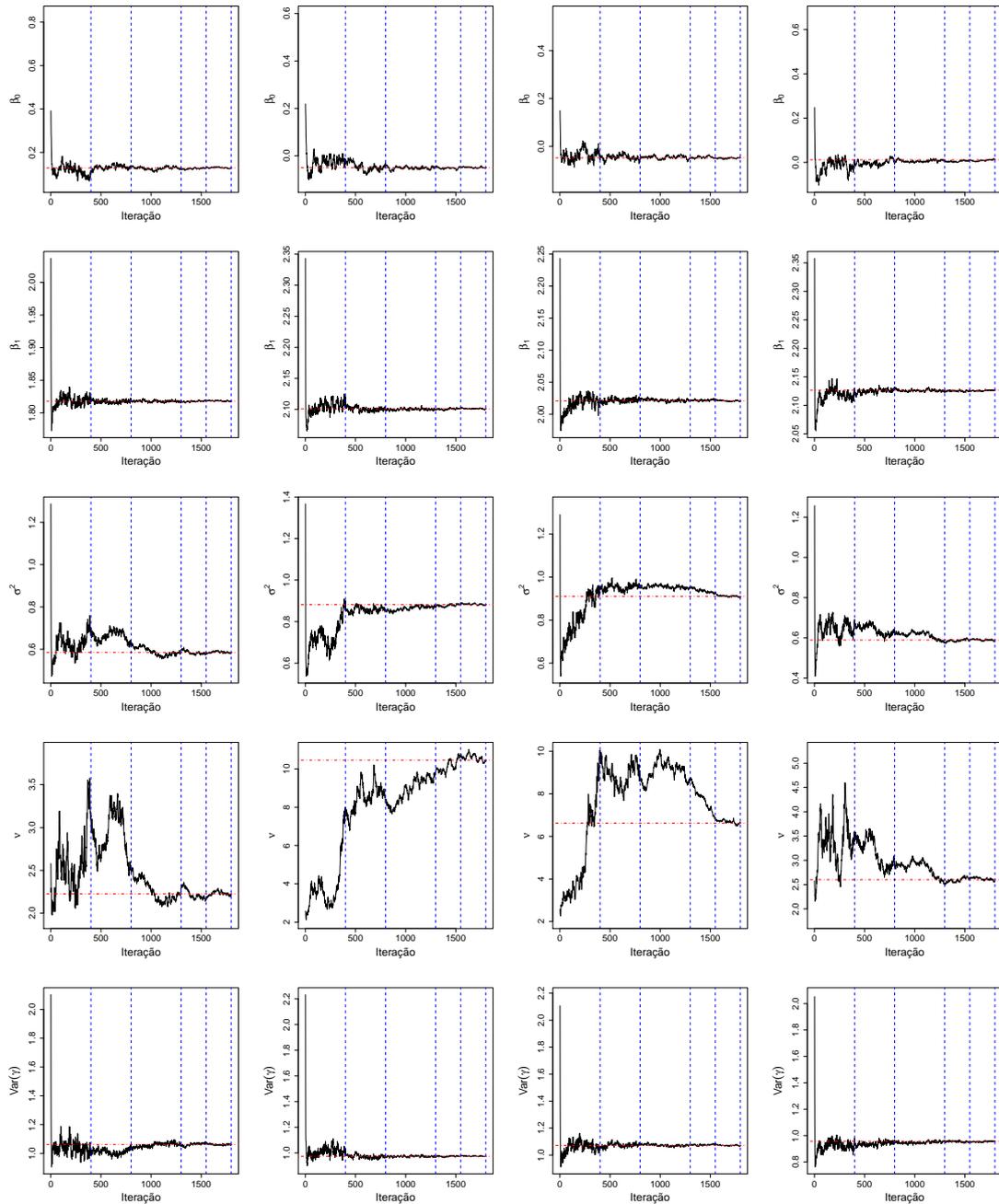


Figura 38: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-M e média das estimativas nas últimas 100 iterações (linha tracejada) para Dados 5 a 8 (da esquerda para a direita) no Cenário 3.

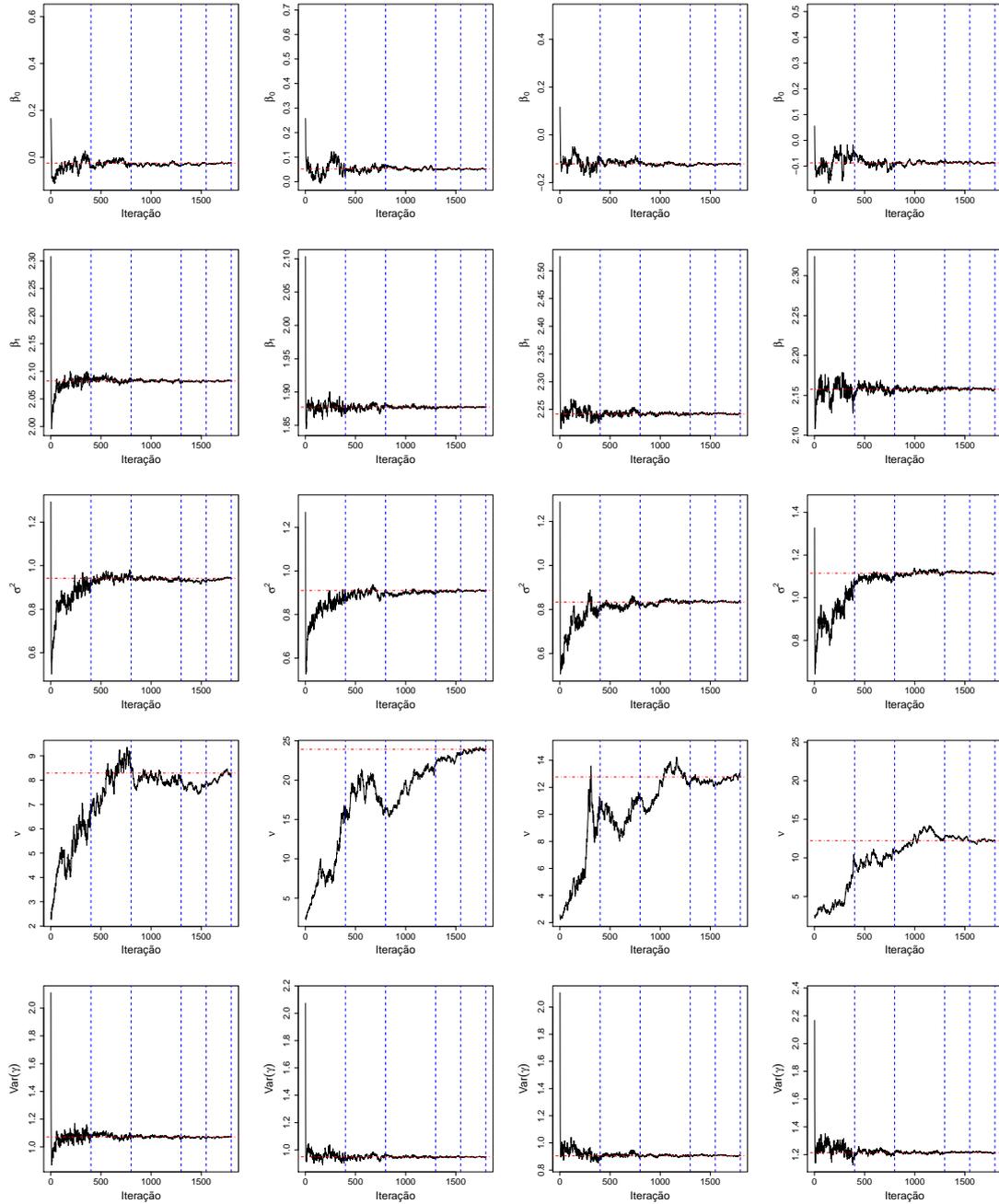


Figura 39: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-M e média das estimativas nas últimas 100 iterações (linha tracejada) para Dados 9 a 12 (da esquerda para a direita) no Cenário 3.

Observando as Figuras 38 e 39, vemos comportamento semelhante ao já apresentado na Figura 37 e também similar ao do modelo Logístico-T estudado na Subseção 3.6.2. As estimativas de β_0 , β_1 e $V(\gamma_i)$ apresentam boa convergência e têm valores próximos aos valores reais. O parâmetro que apresenta mais problemas de convergência é o parâmetro ν e estes problemas aparecem, principalmente, quando valores grandes de ν são estimados. Uma possível explicação é o fato de que para valores altos de ν a f.d.p da distribuição

Slash é muito similar à da distribuição Normal. As estimativas de σ^2 apresentam comportamento estável, mas algumas vezes estas estimativas se distanciam dos valores reais de forma a compensar estimativas ruins de ν e deixar as estimativas de $V(\gamma_i)$ próximas do valor real.

3.6.3.2 Avaliando o efeito do aumento de ν : Algoritmo EMMC-Q

Após usarmos o algoritmo EMMC-M, ajustamos agora o modelo Logístico-SL utilizando o algoritmo EMMC-Q para gerarmos da distribuição condicional completa de γ_i . O objetivo é analisarmos se o algoritmo EMMC-Q, por apresentar maior taxa de aceitação do que o EMMC-M, tem convergência mais rápida. Para avaliarmos o impacto da especificação da constante c , ajustamos o modelo considerando para c os valores 0,1, 0,3, 0,5 e 1. Nas Figuras 40, 41 e 42 apresentamos os gráficos com as estimativas dos parâmetros para Dados 4, 5 e 9, respectivamente. Os gráficos das estimativas para os demais bancos de dados são disponibilizados no Apêndice E. A partir destas figuras, percebemos que, ao contrário do que foi sugerido por Quintana *et al.* (1999) para o modelo Logístico-N, no modelo Logístico-SL o valor estabelecido para a constante c na distribuição candidata em (3.22) influencia não somente na estabilidade numérica das estimativas, mas também os valores para os quais as estimativas dos parâmetros convergem. Quintana *et al.* (1999) menciona que em seu exemplo valores de c entre 1 e 2,5 alteraram apenas a taxa de aceitação do algoritmo e não tiveram impacto na convergência das estimativas. Aqui, vemos que isto não ocorreu quando consideramos $c = 1$. Para $c = 1$, além de problemas na estabilidade das estimativas, os valores estimados para os parâmetros foram muito distantes dos valores reais. Note que o algoritmo Metropolis de Quintana utilizado tem distribuição candidata dada por $N((cI_i + u_i/\sigma^2)^{-1}cI_i\hat{\gamma}_i, (cI_i + u_i/\sigma^2)^{-1})$. A constante c tem influência tanto na média quanto na variância da distribuição candidata. Para estes bancos de dados simulados vemos que ao assumirmos $c = 1$, o fator $(cI_i + u_i/\sigma^2)^{-1}cI_i$ assume valores próximos de 1 (por volta de 0,95) e a variância da distribuição candidata assume valores muito pequenos. Isto é devido ao fato de no modelo Logístico-SL a variável latente U_i assumir valores apenas entre 0 e 1. Em resumo, a aparente causa dos problemas de convergência das estimativas do EMMC é que os valores gerados da distribuição candidata ficam muito concentrados próximos a $\hat{\gamma}_i$. De forma geral, apenas a especificação de $c = 0,1$ no algoritmo EMMC-Q levou a estimativas similares as encontradas ao utilizarmos o algoritmo EMMC-M. Com os demais valores de c , o algoritmo EMMC-Q tende a estimar valores pequenos para ν em todos os bancos de dados. Vemos na Figura 89, por exemplo, que ao especificarmos $c = 0,3$ as estimativas de ν ficam em torno de 2,86 enquanto as estimativas obtidas usando o algoritmo EMMC-M ficam próximas de 11. Observando apenas os gráficos das estimativas para os efeitos fixos, percebemos estimativas semelhantes ao especificarmos $c = 0,1$ e 0,3, mas grandes mudanças para os demais valores considerados para c .

Embora o algoritmo EMMC-Q tenha maior taxa de aceitação (por volta de 45%) do que o algoritmo EMMC-M, não houve melhora na velocidade de convergência das estimativas dos parâmetros nos bancos de dados estudados. Juntando isto ao fato do algoritmo EMMC-Q ter se mostrado muito sensível a especificação da constante c , concluímos não ser uma boa alternativa utilizarmos o algoritmo EMMC-Q para ajustarmos o modelo Logístico-SL.

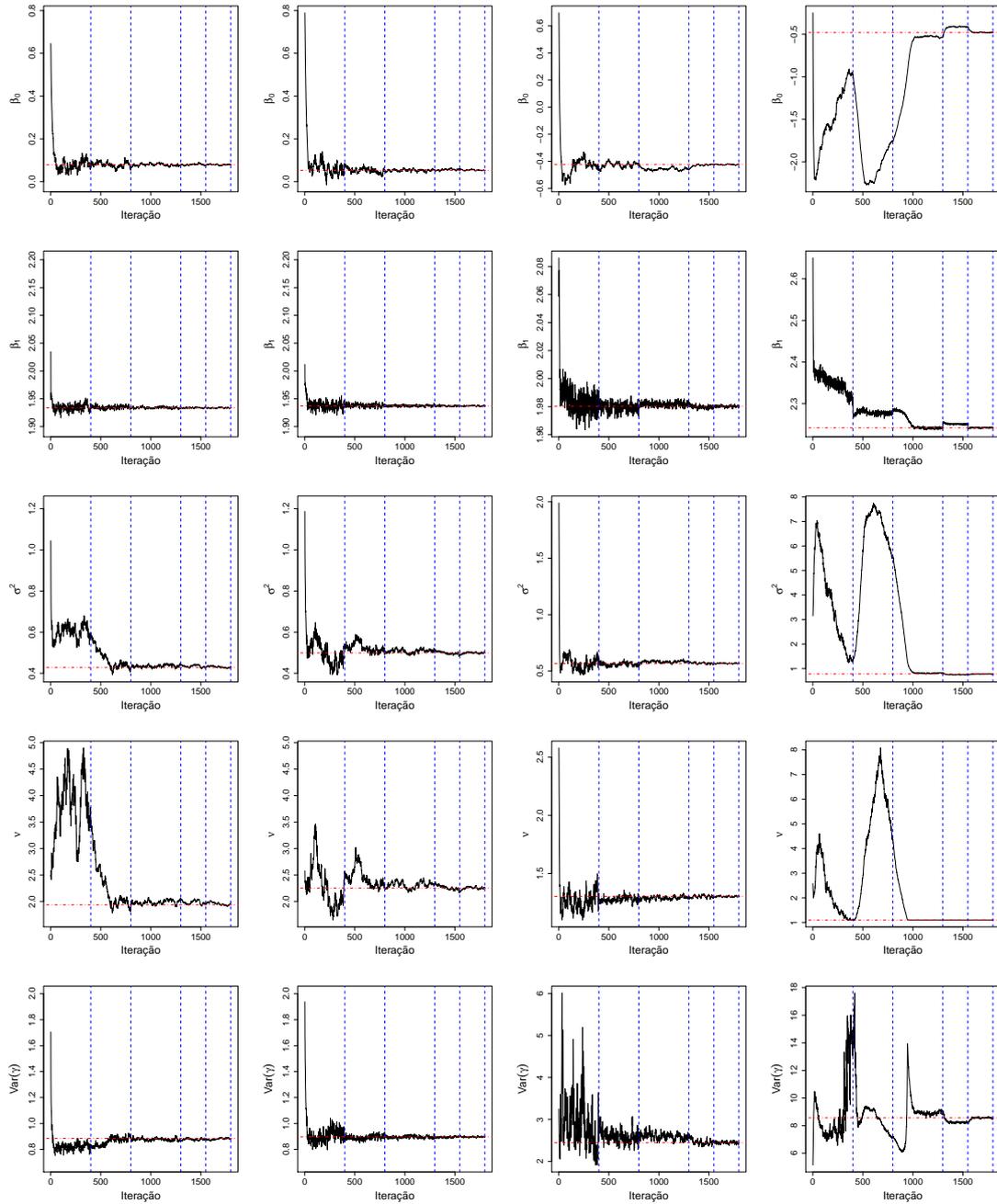


Figura 40: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-Q com constante $c = 0,1$ (primeira coluna), $0,3$ (segunda coluna), $0,5$ (terceira coluna) e 1 (quarta coluna) e média das estimativas nas últimas 100 iterações (linha tracejada) com Dados 4 no Cenário 3.

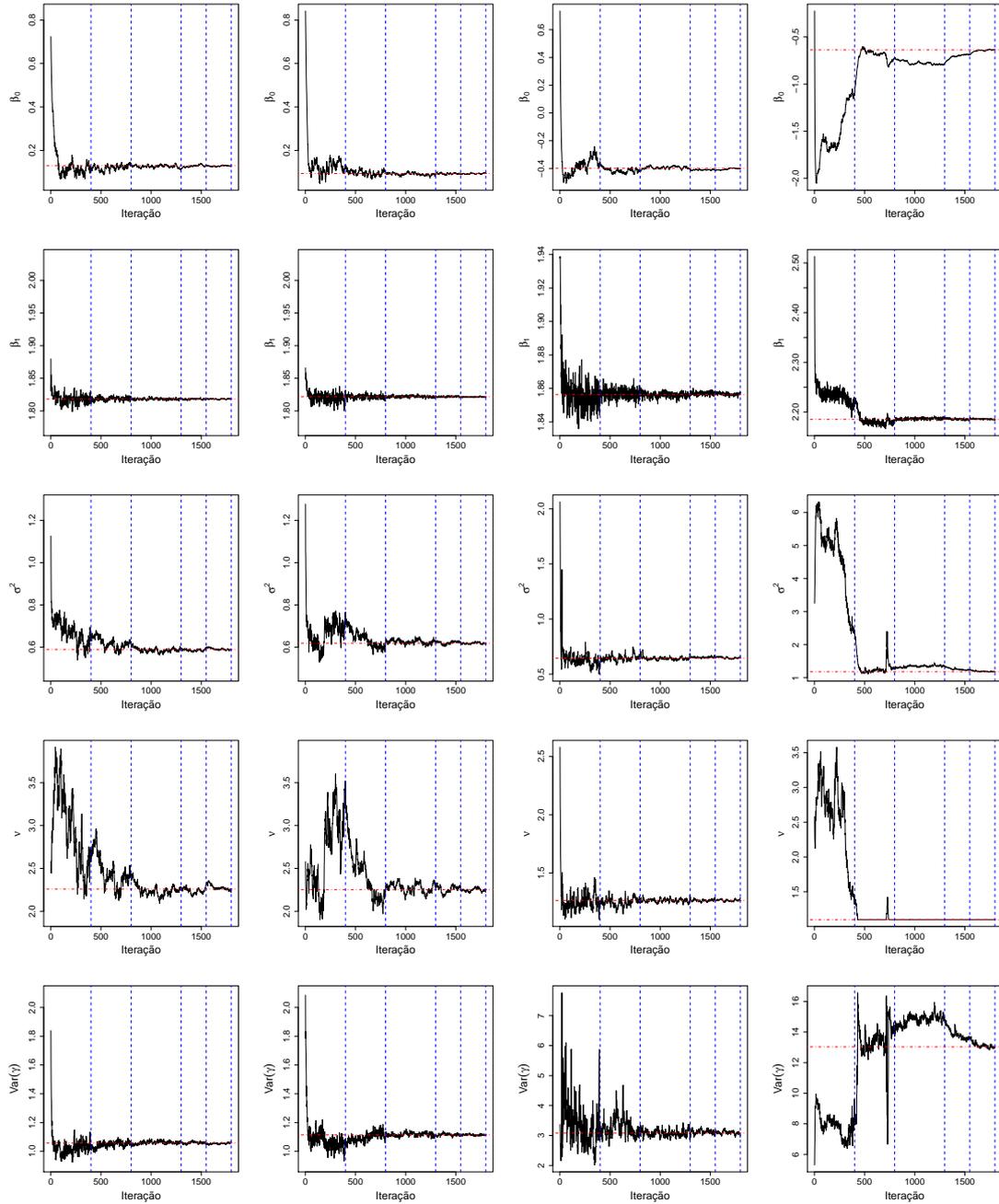


Figura 41: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-Q com constante $c = 0,1$ (primeira coluna), $0,3$ (segunda coluna), $0,5$ (terceira coluna) e 1 (quarta coluna) e média das estimativas nas últimas 100 iterações (linha tracejada) com Dados 5 no Cenário 3.

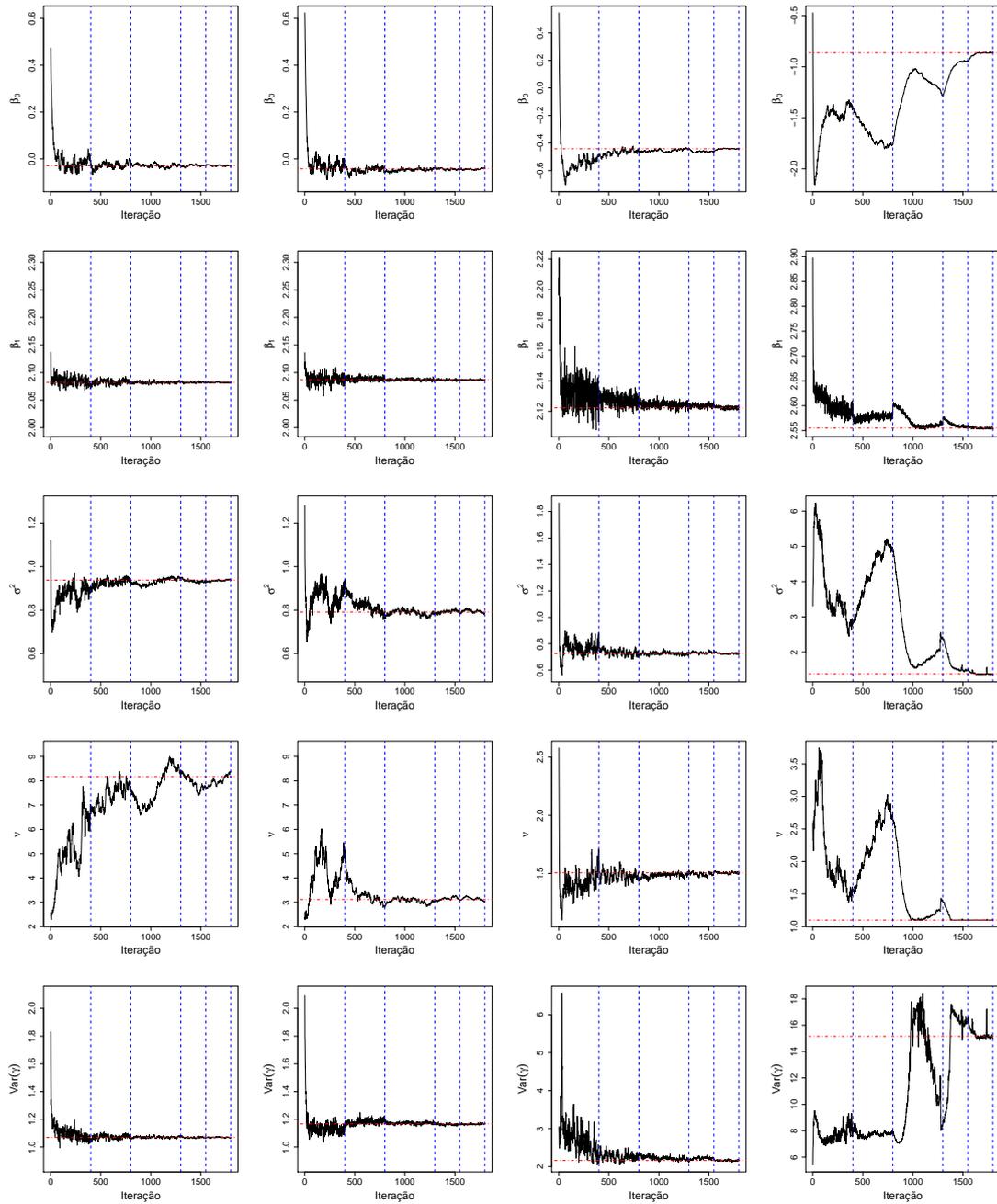


Figura 42: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-Q com constante $c = 0,1$ (primeira coluna), $0,3$ (segunda coluna), $0,5$ (terceira coluna) e 1 (quarta coluna) e média das estimativas nas últimas 100 iterações (linha tracejada) com Dados 9 no Cenário 3.

3.6.4 Cenário 4: Modelo Logístico-NC

Na Subseção 3.6.4 temos por objetivo avaliar o comportamento do algoritmo EMMC ao ajustarmos o modelo Logístico-NC, em diferentes situações quanto ao formato das caudas da distribuição dos efeitos aleatórios na geração dos bancos de dados. Para ajustarmos este modelo, utilizamos apenas o algoritmo EMMC-M. Esta escolha ocorreu em virtude de simplificarmos a análise e é devida aos resultados ruins obtidos pelos algoritmos EMMC-Q, na Subseção 3.6.3, e EMMC-PA, na Subseção 3.6.1.

Na geração dos bancos de dados, assumimos que os efeitos aleatórios têm distribuição Normal contaminada com variância igual a 1. Consideramos sempre o parâmetro ν_2 igual a 0,2. Isto significa que a distribuição dos efeitos aleatórios é uma mistura de duas distribuições Normais com média 0 e tal que uma das componentes da mistura tem variância igual 5 vezes a variância da outra componente. Além disso, para o parâmetro ν_1 utilizamos os valores 0,1, 0,2 e 0,4, o que significa que, respectivamente, aproximadamente 10%, 20% e 40% dos valores dos efeitos aleatórios são gerados da distribuição normal com maior variância. Escolhemos o valor de σ^2 de tal forma que $V(\gamma_i) = 1$. A ideia é considerarmos situações em que a distribuição dos efeitos aleatórios tenha caudas com diferentes formatos. Na Figura 43 apresentamos uma comparação da f.d.p das distribuição Normal Contaminada com a distribuição Normal.

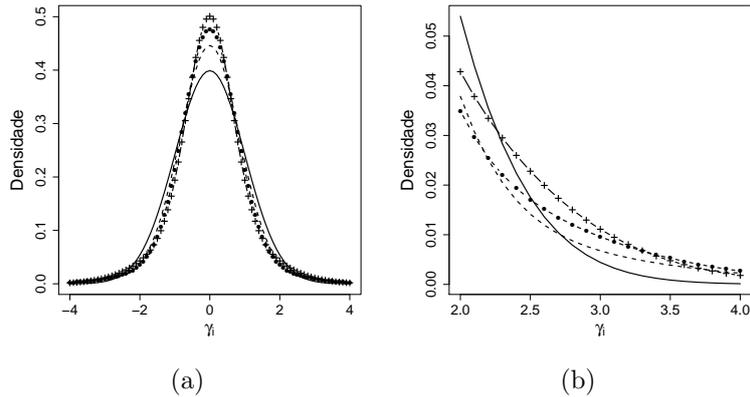


Figura 43: Comparação entre as f.d.p das distribuições Normal padrão (linha sólida), $NC(0; 0,714; 0,10; 0,20)$ (linha tracejada), $NC(0; 0,555; 0,20; 0,20)$ (linha tracejada com \bullet) e $NC(0; 0,385; 0,40; 0,20)$ (linha tracejada com $+$) em (a) e um *zoom* da cauda a direita em (b).

Note na Figura 43 que a distribuição Normal contaminada possui cauda mais pesada que a distribuição Normal nas três configurações de parâmetros consideradas. Veja também que, pelo fato da variância dos efeitos aleatórios estar fixa, a distribuição com $\nu_1 = 0,4$ não possui a f.d.p com cauda mais pesada entre as distribuições apresentadas. Esta distribuição atribui maior massa de probabilidade para valores entre, aproximadamente, 2,4 e 3,2, mas a distribuição NC com $\nu_1 = 0,2$ é a que atribui maior massa de probabilidade para valores maiores do que 3,5.

Geramos os bancos de dados utilizando os mesmos valores para os efeitos fixos das seções anteriores ($\beta_0 = 0$ e $\beta_1 = 2$), assumindo $k = 100$, $n_i = 10$ e $n_{ij} = 5$ e tal que:

Configuração 1 (Dados 1 a 4): $\gamma_i \stackrel{iid}{\sim} NC(0; 0,714; 0,1; 0,2)$;
 Configuração 2 (Dados 5 a 8): $\gamma_i \stackrel{iid}{\sim} NC(0; 0,555; 0,2; 0,2)$;
 Configuração 3 (Dados 9 a 12): $\gamma_i \stackrel{iid}{\sim} NC(0; 0,385; 0,4; 0,2)$.

Apresentamos na Figura 44 os gráficos dos histogramas e f.d.p teórica dos efeitos aleatórios nos 12 bancos de dados considerados nesta seção. Observamos que, em geral, os efeitos aleatórios gerados se ajustam bem a f.d.p teórica do modelo Logístico-NC.

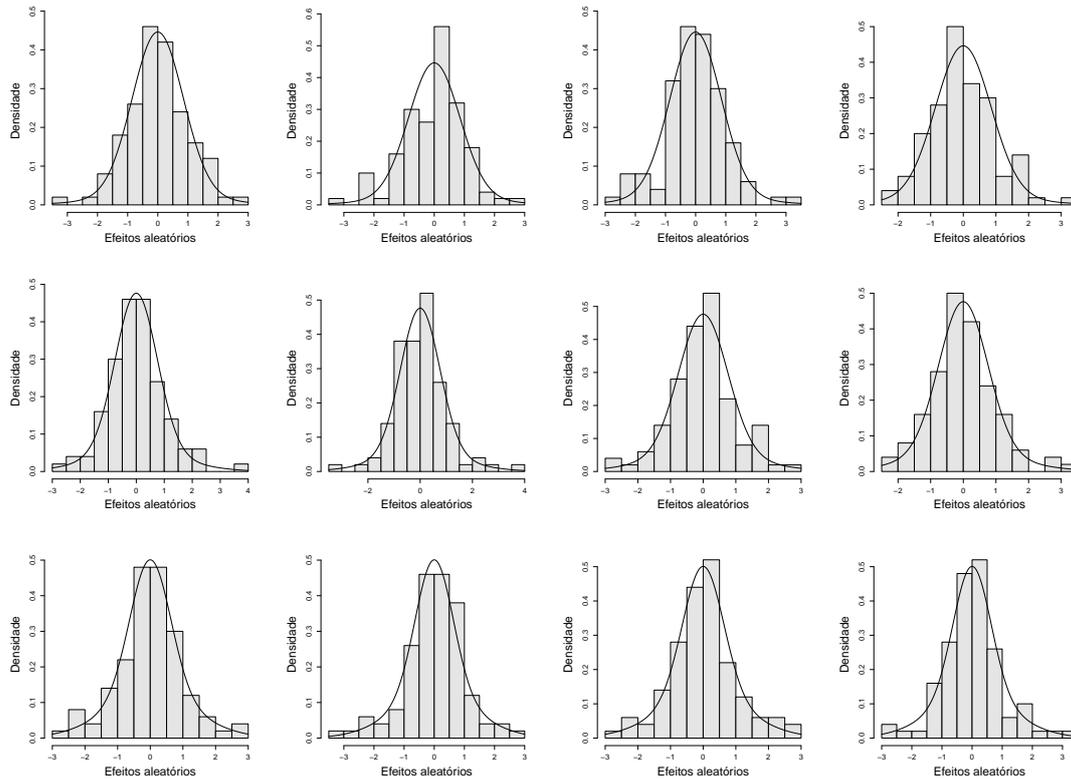


Figura 44: Histograma e f.d.p teórica dos efeitos aleatórios gerados no Cenário 4 para Dados 1 a 12 (da esquerda para direita).

Ajustamos o modelo Logístico-NC aos bancos de dados simulados assumindo que conhecemos o valor real do parâmetros ν_2 , isto é, assumimos $\nu_2 = 0,2$.

Nas Figuras 45, 46 e 47 apresentamos os gráficos das estimativas dos parâmetros do modelo *versus* as iterações do algoritmo EMMC-M para todos os bancos de dados simulados neste cenário. Em todos os casos, vemos bom comportamento de convergência para as estimativas dos parâmetros de efeitos fixos β_0 e β_1 e também para a variância dos efeitos aleatórios $V(\gamma_i)$. Para estas quantidades, além das estimativas serem estáveis ao longo das iterações, os valores de convergência são relativamente próximos aos valores reais especificados na geração dos bancos de dados. Ao considerarmos ν_2 fixo, os parâmetros ν_1 e σ^2 controlam a variância e as caudas da distribuição estimada para os efeitos aleatórios. Observamos que as estimativas de ν_1 e σ^2 convergem, mas as estimativas de ν_1 tendem a serem superestimadas. Isso pode ser observado, por exemplo, para Dados 1 a 4 (Figura 45) em que o valor real é $\nu_1 = 0,1$, mas as estimativas finais obtidas

são aproximadamente 0,18, 0,63, 0,58 e 0,68, respectivamente. Um estudo Monte Carlo mais aprofundado é necessário para ver se essa tendência se confirma.

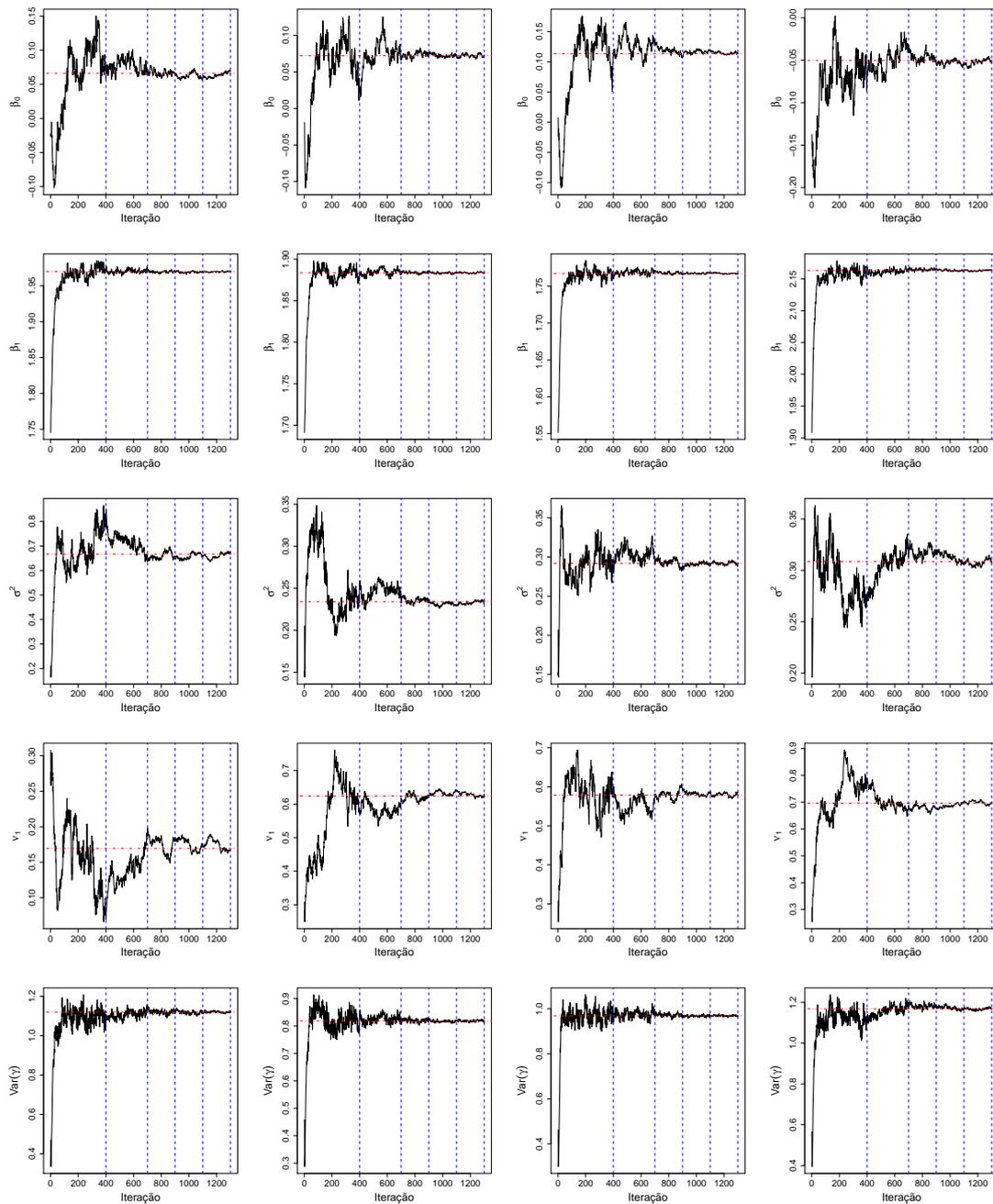


Figura 45: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-M e média das estimativas nas últimas 100 iterações (linha tracejada) com Dados 1 a 4 (da esquerda para a direita) no Cenário 4.

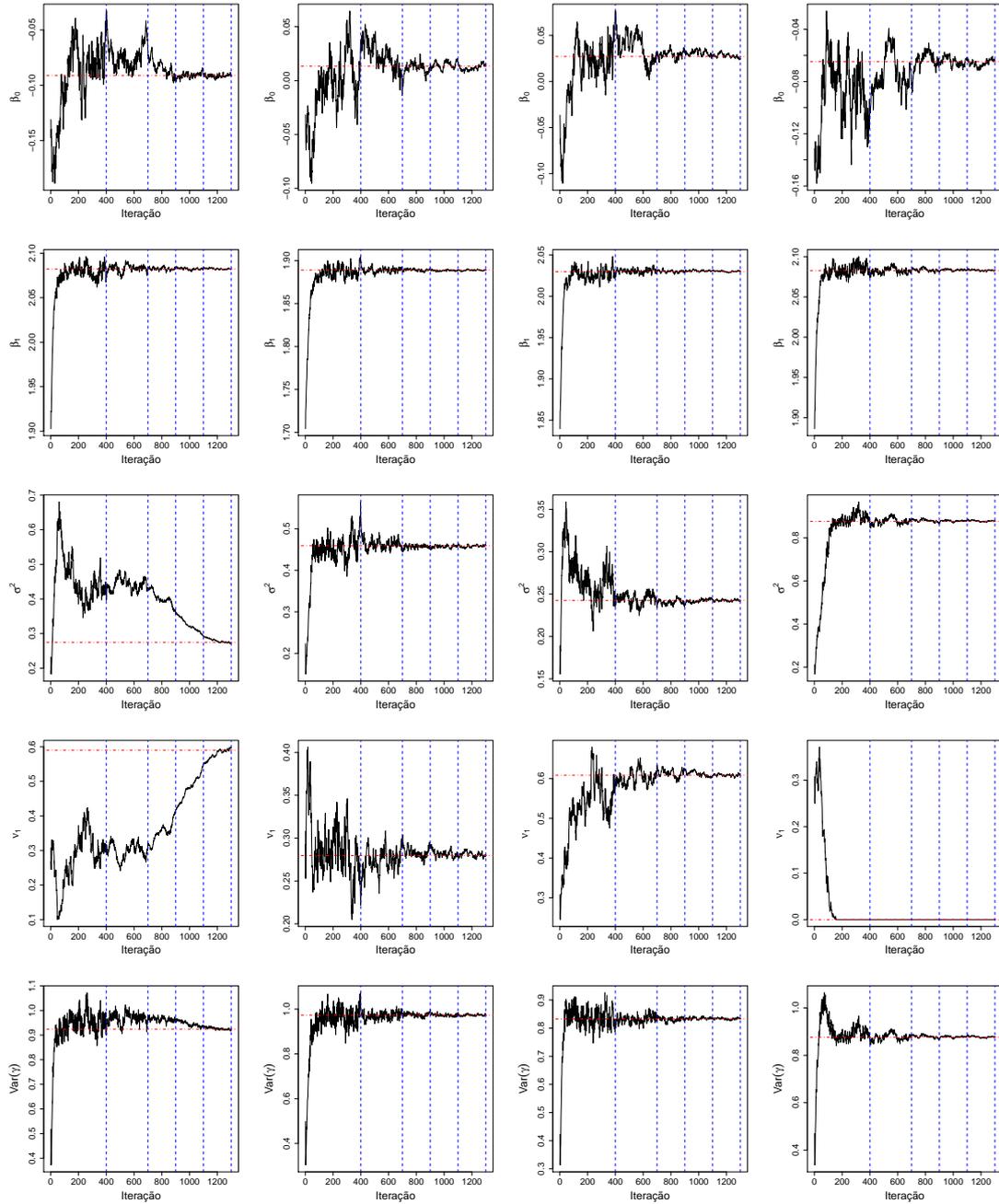


Figura 46: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-M e média das estimativas nas últimas 100 iterações (linha tracejada) com Dados 5 a 8 (da esquerda para a direita) no Cenário 4.

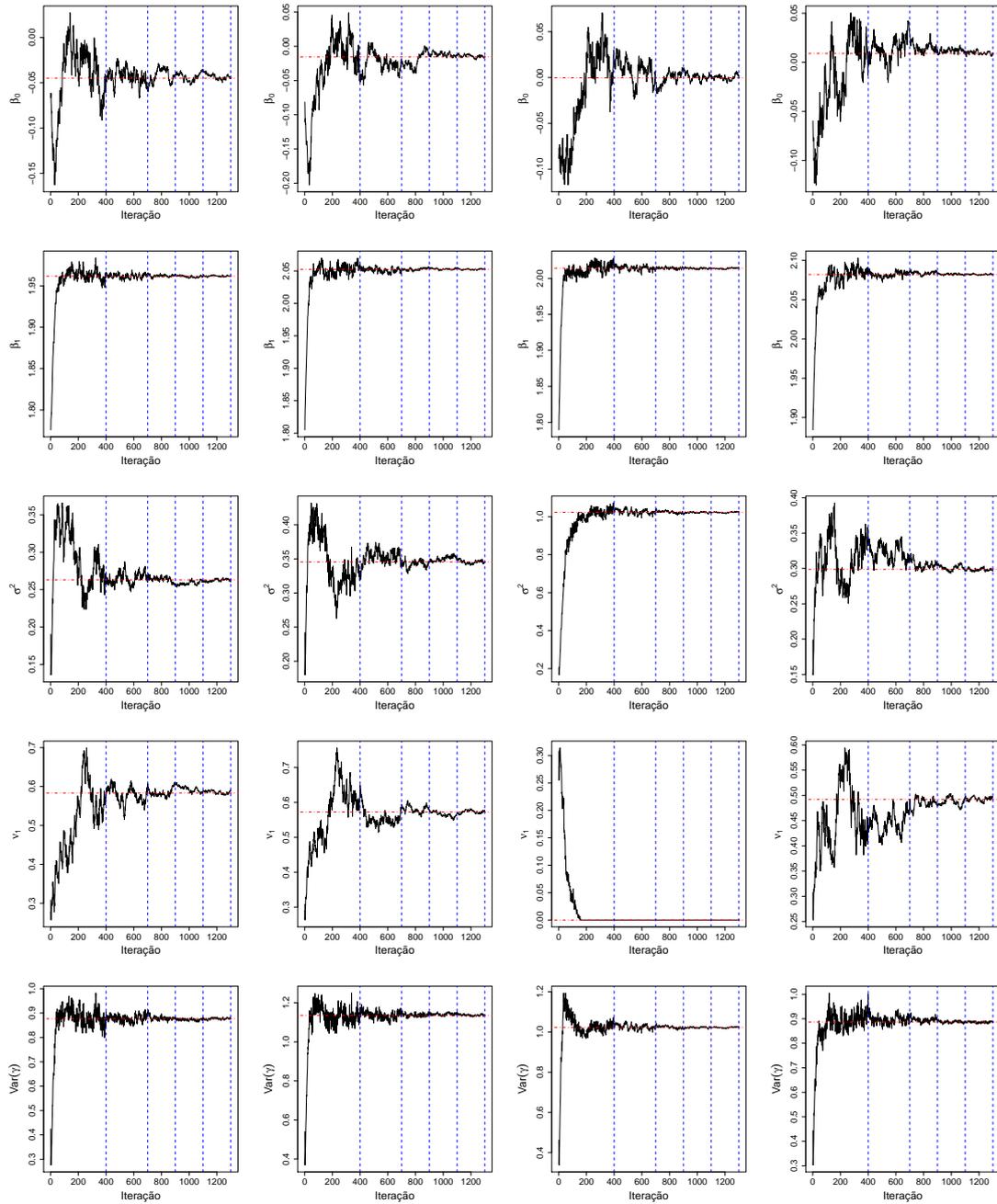


Figura 47: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-M e média das estimativas nas últimas 100 iterações (linha tracejada) com Dados 9 a 12 (da esquerda para a direita) no Cenário 4.

3.6.5 Discussão sobre a implementação do algoritmo EMMC

Nesta Subseção vamos discutir resumidamente os resultados obtidos na implementação dos modelos Logístico-NI com o algoritmo EMMC. Nas Subseções 3.6.1 e 3.6.3 comparamos os algoritmos Metropolis propostos para gerarmos da distribuição condicional completa do vetor de efeitos aleatórios γ . Observamos que o algoritmo EMMC com as três propostas para gerar da distribuição de γ (EMMC-M, EMMC-PA e EMMC-Q) conduzem a estimativas pontuais iguais para os parâmetros de interesse. No entanto, estes algoritmos diferem quanto a convergência das estimativas fornecidas. Nos modelos Logístico-N e Logístico-T, em geral, as estimativas dos parâmetros fornecidas pelo algoritmo EMMC-Q são mais estáveis e se aproximam da convergência com menor número de iterações do que as estimativas fornecidas pelos demais algoritmos. O algoritmo EMMC-M apresentou resultados similares no que diz respeito a convergência das estimativas fornecidas. O algoritmo EMMC-PA foi o pior na questão de velocidade de convergência, principalmente para o parâmetro ν no modelo Logístico-T e de β_0 nos dois modelos comparados. Nos modelos Logístico-N e Logístico-T ajustamos o algoritmo EMMC-Q considerando $c = 1$, que é um dos valores sugeridos pela literatura para o modelo Logístico-N, sem ter qualquer tipo de problema. Entretanto, a escolha do valor deste parâmetro *ad hoc* no algoritmo EMMC-Q se mostrou uma questão delicada no modelo Logístico-SL. Uma escolha errada para o valor de c influencia não somente na velocidade e qualidade da convergência das estimativas fornecidas para os parâmetros de interesse, como também nos valores finais das estimativas. Como não conseguimos criar uma regra não *ad-hoc* para determinarmos um valor ótimo para c no algoritmo EMMC-Q, concluímos que o algoritmo EMMC-M é a melhor opção dentre os algoritmos comparados.

Ao ajustarmos o modelo Logístico-N, principalmente como os algoritmos EMMC-M e EMMC-Q, as estimativas de todos os parâmetros apresentaram bom comportamento quanto a convergência, existindo pouca variabilidade das estimativas ao longo das iterações finais, com o número L de amostras Monte Carlo especificado no passo de esperança do algoritmo. No modelo Logístico-T, notamos que o parâmetro ν é o mais difícil de estimar, pois demora a atingir um padrão de convergência quando valores altos são estimados. Se este parâmetro é fixado, as estimativas ao longo das iterações para os demais parâmetros apresentam bom comportamento. Notamos também que se fixarmos o parâmetro de intercepto β_0 não há alteração substancial no comportamento na estimação dos outros parâmetros do modelo. No entanto, as estimativas para β_0 tendem a convergir mais lentamente e a serem mais instáveis após a convergência ser atingida.

Como seria natural esperar, bancos de dados com maiores valores de n_i e n_{ij} têm melhor convergência das estimativas. Aumentar o número k de *clusters*, mantendo iguais os valores de n_i e n_{ij} , parece não influenciar no comportamento das estimativas dos parâmetros. Comparando bancos de dados com o mesmo número total de observações e com diferentes quantidades de *clusters*, concluímos que o maior número de *clusters* leva a piora da convergência das estimativas. O modelo Logístico-SL implementado com o algoritmo EMMC-M, de forma similar ao modelo Logístico-T, apresenta problemas de convergência apenas para o parâmetro ν se valores mais altos são estimados. No modelo Logístico-NC, observamos bom comportamento de convergência das estimativas dos parâmetros, exceto, em alguns bancos de dados, para o parâmetro ν_1 . Além disto, em alguns bancos de dados, as estimativas deste parâmetro convergem para o valor zero,

apontando inadequadamente que os efeitos aleatórios são normalmente distribuídos. Em resumo, em todos os modelos implementados, notamos boa convergência para as estimativas dos efeitos fixos e da variância dos efeitos aleatórios.

Pontos ainda a serem melhorados no algoritmo EMMC são a especificação do tamanho L da amostra Monte Carlo dos valores para os efeitos aleatórios no passo de esperança do algoritmo e um critério de parada para o algoritmo. Existem propostas interessantes (Booth e Hobert (1999) e Caffo *et al.* (2005)) para a definição automática do valor L , ao longo das iterações do algoritmo EMMC, e que apresentam bons resultados para o modelo Logístico-N. No entanto, em nossa implementação (não apresentada neste trabalho), estes métodos não apresentam bons resultados nos demais modelos Logísticos-NI.

Na próxima seção realizamos um estudo de simulação Monte Carlo onde avaliamos a qualidade da estimação nos modelos Logístico-NI implementados e os comparamos.

3.7 Estudo Monte Carlo: Avaliação da qualidade de estimação

Nesta seção, realizamos um estudo de simulação Monte Carlo com o objetivo de avaliarmos a qualidade das estimativas para os efeitos fixos, para os parâmetros relacionados a distribuição dos efeitos aleatórios e para a razão de chances mediana nos modelos Logístico-NI. Mostramos anteriormente que, no modelo logístico misto, a razão de chances mediana é uma importante ferramenta para interpretar os efeitos fixos e medir a heterogeneidade entre *clusters*. Nosso foco será apenas na estimação de $MOR_{|12|}$, que é uma medida utilizada para avaliarmos a heterogeneidade entre os *clusters*. Apesar da razão de chances mediana MOR_{12} ser de interesse para interpretar os efeitos fixos, como a distribuição dos efeitos aleatórios é simétrica nos modelos implementados neste capítulo, a qualidade das estimativas de MOR_{12} é consequência direta da qualidade da estimação do vetor de parâmetros de efeitos fixos β . Isto decorre do fato da expressão para MOR_{12} ser função apenas de β , como pode ser visto na Proposição 3.5 da Seção 3.4.1. Sendo assim, para não tornarmos a análise repetitiva, omitiremos uma análise das estimativas para MOR_{12} .

Vários cenários são considerados no estudo e, em cada um deles, 200 amostras são geradas considerando o seguinte modelo:

$$Y_{ij}|\pi_{ij} \sim \text{Binomial}(n_{ij}, \pi_{ij}), \text{ para } i = 1, 2, \dots, k \text{ e } j = 1, 2, \dots, n_i;$$

$$\ln \{ \pi_{ij} / (1 - \pi_{ij}) \} = \beta_0 + \beta_1 x_{ij} + \gamma_i, \text{ com } x_{ij} = j/n_i, \text{ e } \gamma_i \stackrel{iid}{\sim} \text{NI}(0, \sigma^2; F_{U|\nu}),$$

em que $\beta_0 = 0$ e $\beta_1 = 2$.

A escolha do número de réplicas ocorre em função do grande tempo computacional gasto no ajuste de cada modelo.

Os cenários analisados se diferenciam pelos valores especificados para k ($k = 50$ e 100), n_i ($n_i = 5$ e 10), n_{ij} ($n_{ij} = 2$ e 4) e, principalmente, pela distribuição assumida para os efeitos aleatórios na geração dos dados. Para a distribuição dos efeitos aleatórios consideramos as distribuições Normal, t -Student, Slash e Normal Contaminada.

Em cada cenário, os modelos implementados na Seção 3.5 serão ajustados e comparados com respeito a qualidade das estimativas dos parâmetros do modelo e para a razão

de chances mediana. Neste estudo de simulação, ajustamos os modelos Logístico-N, Logístico-T, Logístico-SL e Logístico-NC utilizando o algoritmo EMMC-M. No Modelo Logístico-T, para garantir a existência da variância $V(\gamma_i)$ da distribuição dos efeitos aleatórios e evitar que valores muito altos sejam estimados quando ν é estimado como sendo muito próximo de 2, estimamos ν restrito ao intervalo de valores entre 2,5 e 30. Com o mesmo objetivo, no Modelo Logístico-SL, colocamos a restrição de que ν assumira valores entre 1,25 e 15. Como já mencionado na Subseção 3.4.1, para calcularmos $MOR_{|12|}$, em alguns casos, é necessário gerarmos amostras da distribuição de W_{12} . Neste capítulo, consideramos amostras de tamanho 200.000 da distribuição de W_{12} para calcularmos a $MOR_{|12|}$ nos modelos Logístico-T e Logístico-SL.

Na Seção 3.6, notamos que o algoritmo EMMC-M tende a ter melhor comportamento e que as estimativas fornecidas pelos modelos Logístico-NI não convergem com mesma velocidade em todos os casos. O modelo Logístico-N, em geral, apresenta convergência mais rápida para as estimativas dos parâmetros do que o modelo Logístico-T. Da mesma forma, o modelo Logístico-T frequentemente atinge a convergência em um número menor de iterações do que o modelo Logístico-SL. Tanto no modelo Logístico-T, quanto no modelo Logístico-SL, a convergência também se torna mais rápida quando assumimos o parâmetro ν como fixo e conhecido. Em função disto, implementamos apenas o algoritmo EMMC-M e, em cada modelo, determinamos de forma diferente o número L de réplicas Monte Carlo ao longo das iterações do algoritmo. A Tabela 3.4 apresenta o número total de iterações considerado em cada modelo ajustado e o número L de réplicas Monte Carlo considerado em cada uma das iterações do algoritmo EMMC-M.

Tabela 3.4: Número L de réplicas Monte Carlo, ao longo das iterações, do algoritmo EMMC-M em cada modelo

Modelo	Número L de réplicas Monte Carlo					Total de iterações
	10	50	200	500	1000	
Logístico-N	1-300	301-600	601-800	801-900	901-1000	1000
Logístico-T	1-400	401-700	701-1000	1001-1300	1301-1400	1400
Logístico-T (ν fixo)	1-300	301-600	601-800	801-900	901-1000	1000
Logístico-SL	1-400	401-700	701-1200	1201-1600	1601-1700	1700
Logístico-SL (ν fixo)	1-300	301-600	601-800	801-900	901-1000	1000
Logístico-NC	1-400	401-700	701-900	901-1200	1201-1300	1300

Em todos os casos, consideramos como estimativas finais dos parâmetros a média das estimativas fornecidas nas últimas 100 iterações do algoritmo.

Além das medidas usuais (viés e EQM) que refletem a qualidade das estimativas dos parâmetros de interesse, os modelos também são comparados através da qualidade de ajuste fornecida pelo AIC (*Akaike information criterion*). Existem algumas versões de AIC para compararmos modelos mistos. Greven e Kneib (2009) consideram as versões marginal e condicional do AIC e as aplicam a modelos lineares mistos. Vaida e Blanchard (2005), Yu e Yau (2012) e Saefken *et al.* (2014) discutem sobre a estimação do AIC condicional em modelos generalizados mistos. Vaida e Blanchard (2005) afirmam que o AIC marginal é indicado quando o foco da pesquisa está na estimação dos parâmetros ditos populacionais (parâmetros de efeitos fixos) enquanto o AIC condicional é indicado quando o foco da pesquisa está em “parâmetros” específicos a *clusters*, isto é, o foco está na estimação dos efeitos aleatórios. Neste trabalho, seguindo Vaida e Blanchard

(2005), utilizamos o AIC marginal e obtemos uma aproximação para a função de verossimilhança marginal através de integração Monte Carlo e condicionando nas estimativas dos parâmetros da distribuição dos efeitos aleatórios.

Este estudo de simulação está organizado como descrito a seguir. Na Subseção 3.7.1 analisaremos os resultados dos cenários em que os bancos de dados são gerados sob o modelo Logístico-N, nas Subseções 3.7.2, 3.7.3 e 3.7.4 consideramos, respectivamente, os modelos Logístico-T, Logístico-SL e Logístico-NC na geração dos bancos dados. Por fim, na Subseção 3.7.5, apresentamos nossas conclusões finais sobre o estudo de simulação Monte Carlo.

3.7.1 Dados simulados do modelo Logístico-N

Nesta Subseção, os dados são gerados sob o modelo Logístico-N. Nosso objetivo é comparar a implementação do modelo Logístico-N via algoritmo EMMC com a implementação do mesmo modelo via métodos numéricos de integração. Comparamos as implementações do modelo com respeito a aspectos relacionados a qualidade das estimativas pontuais e intervalares dos parâmetros de interesse. O modelo Logístico-N via métodos de integração numérica está implementado no pacote estatístico “lme4” do *software* gratuito R. Ajustamos o modelo utilizando quadratura de Gauss-Hermite adaptativa, avaliando em 9 pontos, para obtermos a função de log-verossimilhança marginal.

Na Tabela 3.5, apresentamos as particularidades dos cenários que analisaremos, os quais são denotados por A1 a A8, quando $V(\gamma_i) = 0,5$, e por B1 a B8 quando $V(\gamma_i) = 1$. Além da variância dos efeitos aleatórios, os cenários também diferem quanto ao número k de *clusters*, ao número n_i de observações dentro de cada *clusters* e ao número n_{ij} de experimentos Bernoulli para cada observação.

Tabela 3.5: Organização dos cenários gerados sob modelo Logístico-N

	k	$n_i = 5$		$n_i = 10$	
		$n_{ij} = 2$	$n_{ij} = 4$	$n_{ij} = 2$	$n_{ij} = 4$
$\gamma_i \stackrel{iid}{\sim} N(0; 0,5)$	50	A1	A2	A3	A4
	100	A5	A6	A7	A8
$\gamma_i \stackrel{iid}{\sim} N(0, 1)$	50	B1	B2	B3	B4
	100	B5	B6	B7	B8

Os valores de k , n_i e n_{ij} escolhidos permitem realizar comparações que atendem a diversos objetivos. Entre eles, podemos avaliar o impacto do número de *clusters* na qualidade das estimativas dos parâmetros de interesse e o efeito de aumentarmos o tamanho da amostra através de n_i ou n_{ij} . Para isto, basta compararmos os cenários aumentando uma das quantidades e mantendo as demais fixas. No entanto, para avaliarmos se uma das quantidades k , n_i ou n_{ij} têm maior influência na qualidade das estimativas fornecidas pelos modelos, isto é, se existe uma melhor estratégia quanto a escolha de qual quantidade seria melhor aumentarmos, podemos comparar cenários com igual número total $k \times n_i \times n_{ij}$ de observações de experimentos Bernoulli. A comparação pode ser realizada de três maneiras distintas:

- Cenários com igual número k de *clusters*, mas com diferentes valores de n_i e n_{ij}

como, por exemplo, a comparação entre os cenários A2 e A3 ou entre os cenários A6 e A7;

- Cenários com igual número n_{ij} de experimentos Bernoulli para cada observação dentro dos *clusters*, mas com diferentes valores de k e n_i como ocorre, por exemplo, na comparação entre os cenários A3 e A5 ou entre os cenários A4 e A6;
- Cenários com igual número n_i de observações dentro de cada *cluster*, mas com diferentes valores de k e n_{ij} . Essa é uma comparação possível entre os cenários A2 e A5 ou cenários A4 e A7, por exemplo.

Em cada cenário, comparamos as implementações do modelo Logístico-N através do algoritmo EMMC-M e de integração numérica no que diz respeito a qualidade da estimação dos efeitos fixos β_0 e β_1 e, também, da variância dos efeitos aleatórios, que aqui é igual ao parâmetro σ^2 . Como medidas de qualidade utilizamos o viés e o erro quadrático médio (EQM) da estimativa pontual dos parâmetros de interesse e, também, a probabilidade de cobertura do intervalo de confiança (IC). Consideramos duas alternativas para obtermos os intervalos de 95% de confiança assintóticos para os efeitos fixos e para σ^2 , que diferem somente quanto ao erro padrão obtido para as estimativas. Na primeira, a qual denotamos a probabilidade de cobertura do intervalo de confiança por PC_F , construímos o intervalo de confiança assintótico usando o erro padrão das estimativas obtido a partir da matriz de informação de Fisher observada. Na segunda alternativa, utilizamos o desvio padrão das estimativas obtidas no estudo Monte Carlo na construção do intervalo de confiança. Neste caso, denotamos a probabilidade de cobertura do intervalo de confiança por PC_{MC} . Não foi possível obtermos o intervalo de confiança para a variância dos efeitos aleatórios ao utilizarmos a quadratura de Gauss-Hermite e a matriz de informação de Fisher observada, uma vez que o pacote estatístico “lme4”, do *software* R, não fornece estimativa para o erro padrão das estimativas de σ^2 .

Na Tabela 3.6, apresentamos viés, EQM e probabilidade de cobertura dos intervalos de 95% confiança para β_0 , β_1 e σ^2 , ao ajustarmos o modelo Logístico-N aos bancos de dados dos Cenários A1 a A8, utilizando o algoritmo EMMC-M e quadratura de Gauss-Hermite adaptativa. Nesta tabela, percebemos que as duas forma de implementação do modelo Logístico-N, EMMC-M e Quadratura de Gauss-Hermite, fornecem resultados iguais ou muito próximos para todas as medidas consideradas e em todos os cenários analisados. Isto confirma a boa implmentação do modelo Logístico-N com o algoritmo EMMC-M.

Ao analisarmos o comportamento do viés das estimativas, vemos que, em algumas vezes, o viés das estimativas de um dos parâmetros aumenta ao aumentarmos o número de observações do bancos de dados, tanto quando aumentamos n_i quanto n_{ij} . Este fato ocorre, por exemplo, para β_0 nos Cenários A1 e A2, para β_1 nos Cenários A3 e A4 e para σ^2 nos Cenários A1 e A3. Talvez a explicação para isto seja que o número de 200 réplicas de bancos de dados não seja o suficientemente grande para o estudo Monte Carlo retratar perfeitamente média da distribuição dos estimadores de máxima verossimilhança neste modelo. Notamos também que, se fixarmos n_i e n_{ij} , os cenários com maior número de *clusters* k apresentam menor EQM para as estimativas de todos os parâmetros. Conclusão idêntica é tirada ao compararmos cenários, com mesmo k , em que apenas n_i ou n_{ij} muda. Isto é esperado porque, se aumentamos uma das quantidades, k , n_i ou n_{ij} , e mantivermos as outras fixas, obtemos bancos de dados com um número maior de observações.

Tabela 3.6: Viés, EQM e probabilidade de cobertura do IC (95%) para os efeitos fixos e variância dos efeitos aleatórios, nos Cenários A1 a A8, sob o modelo Logístico-N

		Implementação							
Cenário		EMMC				Gauss-Hermite			
		Viés	EQM	PC_{MC}	PC_F	Viés	EQM	PC_{MC}	PC_F
A1	β_0	0,004	0,084	0,945	0,920	0,004	0,085	0,945	0,920
	β_1	0,007	0,210	0,935	0,905	0,010	0,212	0,935	0,905
	σ^2	-0,001	0,061	0,925	0,930	0,007	0,057	0,935	—
A2	β_0	0,012	0,038	0,940	0,955	0,012	0,038	0,940	0,955
	β_1	0,022	0,082	0,945	0,945	0,023	0,083	0,945	0,945
	σ^2	-0,008	0,031	0,955	0,945	-0,006	0,031	0,955	—
A3	β_0	0,011	0,036	0,945	0,930	0,011	0,036	0,945	0,930
	β_1	-0,009	0,074	0,940	0,950	-0,009	0,074	0,940	0,950
	σ^2	-0,022	0,028	0,950	0,915	-0,021	0,029	0,950	—
A4	β_0	-0,008	0,022	0,945	0,930	-0,008	0,022	0,945	0,930
	β_1	0,017	0,035	0,980	0,980	0,018	0,036	0,980	0,980
	σ^2	-0,010	0,020	0,960	0,920	-0,009	0,020	0,960	—
A5	β_0	0,015	0,040	0,950	0,950	0,015	0,040	0,950	0,950
	β_1	-0,020	0,083	0,935	0,935	-0,020	0,083	0,935	0,935
	σ^2	-0,001	0,031	0,955	0,950	-0,001	0,032	0,955	—
A6	β_0	-0,003	0,017	0,955	0,970	-0,003	0,017	0,955	0,970
	β_1	-0,003	0,041	0,950	0,945	-0,003	0,042	0,950	0,945
	σ^2	-0,025	0,015	0,950	0,915	-0,025	0,015	0,945	—
A7	β_0	-0,015	0,017	0,935	0,945	-0,015	0,017	0,935	0,945
	β_1	0,016	0,035	0,960	0,975	0,016	0,035	0,960	0,975
	σ^2	-0,006	0,014	0,960	0,925	-0,006	0,014	0,955	—
A8	β_0	0,003	0,010	0,955	0,965	0,003	0,010	0,955	0,970
	β_1	0,008	0,021	0,955	0,945	0,008	0,021	0,955	0,945
	σ^2	0,001	0,009	0,955	0,960	0,001	0,009	0,955	—

Em cenários cujos os dados são gerados assumindo um número total de observações Bernoulli igual, se o mesmo número k de *clusters* é considerado, vemos que os cenários com maior n_i e menor n_{ij} fornecem EQM igual ou ligeiramente menor para todos os parâmetros estimados. Ao compararmos cenários com valores iguais para n_{ij} , observamos um padrão de comportamento no EQM apenas para o parâmetro β_1 . Neste caso, observamos que em cenários com maior k e menor n_i o EQM para β_1 foi maior. Por fim, comparando os cenários com valores de n_i iguais, o EQM se comportou de forma diferente quando mudamos de cenários com $n_i = 5$ para cenários com $n_i = 10$.

Ao compararmos a probabilidade de cobertura dos IC construídos com o valor nominal do IC, isto é, com 95%, vemos que as duas abordagens apresentam desempenho similar. O IC com erro padrão das estimativas obtido nas amostras Monte Carlo apresenta leve vantagem na estimação da σ^2 . Notamos também que não há melhoria na probabilidade de cobertura do IC, para nenhum parâmetro, ao aumentarmos o tamanho da amostra. Os histogramas das estimativas de σ^2 nos Cenários A1 a A8, mostrados na Figura 48, revelam uma grande assimetria. Isto pode explicar o fraco desempenho do intervalo de confiança em alguns cenários, uma vez que, nas duas abordagens adotadas, os intervalos

são construídos com base no teorema Central do limite para as estimativas de máxima verossimilhança, ou seja, assumindo normalidade para as estimativas.

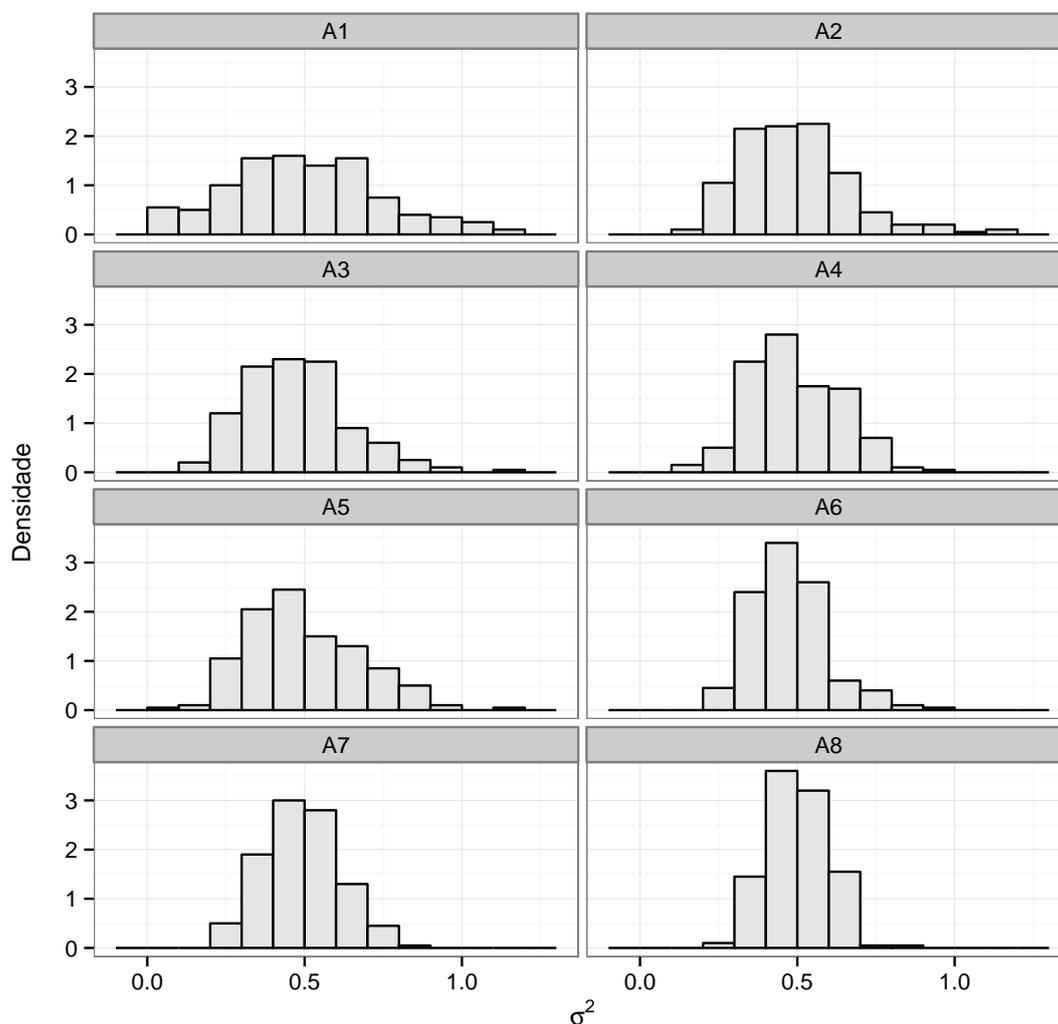


Figura 48: Histograma das estimativas de σ^2 nos Cenários A1 a A8.

Na Tabela 3.7, apresentamos o viés, o EMQ e a probabilidade de cobertura dos intervalos 95% de confiança para as estimativas dos parâmetros ao ajustarmos o modelo Logístico-N nos Cenários B1 a B8. Da Tabela 3.7, observamos que as conclusões para os Cenários B1 a B8 com respeito a k , n_i e n_{ij} são similares àquelas obtidas para os Cenários A1 a A8. As duas implementações do modelo têm desempenho igual na estimação dos parâmetros e ao aumentarmos qualquer uma destas quantidades, mantendo fixas as demais, obtemos uma diminuição do EQM das estimativas de todos os parâmetros.

Comparando cenários que têm o mesmo total de observações Bernoulli e, inicialmente, mesmo número k de *clusters*, vemos que os cenários com maior n_i e menor n_{ij} tendem a fornecer EQM menor para todos os parâmetros estimados. Ao compararmos cenários em que n_{ij} é igual, observamos que similar ao que observamos na Tabela 3.6 para o parâmetro β_1 , o EQM foi maior em cenários com maior k e menor n_i . Para β_0 e σ^2 , os cenários

Tabela 3.7: Viés, EQM e probabilidade de cobertura do IC (95%) para os efeitos fixos e variância dos efeitos aleatórios, nos Cenários B1 a B8, sob o modelo Logístico-N

		Implementação							
Cenário		EMMC				Gauss-Hermite			
		Viés	EQM	PC_{MC}	PC_F	Viés	EQM	PC_{MC}	PC_F
B1	β_0	-0,010	0,083	0,940	0,935	-0,010	0,083	0,940	0,940
	β_1	0,062	0,179	0,940	0,940	0,062	0,180	0,940	0,940
	σ^2	0,034	0,157	0,945	0,950	0,035	0,158	0,945	—
B2	β_0	-0,015	0,045	0,955	0,955	-0,015	0,045	0,955	0,955
	β_1	0,020	0,088	0,955	0,950	0,020	0,088	0,955	0,950
	σ^2	0,018	0,114	0,940	0,905	0,017	0,114	0,940	—
B3	β_0	0,009	0,046	0,960	0,945	0,009	0,046	0,960	0,945
	β_1	0,013	0,080	0,940	0,945	0,013	0,080	0,940	0,945
	σ^2	-0,020	0,090	0,935	0,925	-0,019	0,091	0,935	—
B4	β_0	0,001	0,033	0,955	0,950	0,001	0,033	0,950	0,950
	β_1	0,007	0,038	0,955	0,955	0,008	0,038	0,950	0,955
	σ^2	-0,015	0,052	0,955	0,940	-0,014	0,052	0,955	—
B5	β_0	-0,015	0,041	0,950	0,960	-0,015	0,041	0,955	0,960
	β_1	0,023	0,090	0,965	0,955	0,023	0,090	0,965	0,955
	σ^2	-0,025	0,071	0,945	0,920	-0,025	0,071	0,945	—
B6	β_0	-0,007	0,025	0,945	0,940	-0,007	0,025	0,945	0,945
	β_1	0,000	0,039	0,960	0,960	0,000	0,039	0,960	0,960
	σ^2	0,004	0,044	0,960	0,940	0,005	0,044	0,950	—
B7	β_0	-0,017	0,020	0,965	0,975	-0,017	0,020	0,960	0,975
	β_1	0,026	0,043	0,955	0,945	0,026	0,044	0,955	0,945
	σ^2	-0,005	0,043	0,955	0,915	-0,004	0,043	0,955	—
B8	β_0	0,010	0,019	0,945	0,945	0,010	0,019	0,945	0,940
	β_1	-0,004	0,022	0,935	0,920	-0,004	0,022	0,935	0,920
	σ^2	0,008	0,028	0,945	0,945	0,008	0,028	0,940	—

com maior k e menor n_i apresentaram menor EQM. Conclusão similar ocorreu quando comparamos os cenários com igual valor de n_i . Neste caso, os cenários com maior k e menor n_{ij} tiveram menor EQM para β_0 e σ^2 , mas maior EQM para β_1 .

Comparando os Cenários A1-A8 e B1-B8, observamos que o EQM das estimativas dos parâmetros de interesse é, em geral, maior nos Cenários B1 a B8, em que a variância real dos efeitos aleatórios é maior. Isto ocorre, principalmente, para as estimativas de β_0 e σ^2 . Estas conclusões são corroboradas pelos *boxplots* (Figura 49) das estimativas de β_0 , β_1 e σ^2 fornecidas pela implementação do modelo Logístico-N com o algoritmo EMMC-M. Nesta figura notamos que existe maior variabilidade nas estimativas, principalmente para σ^2 , nos Cenários em que o valor real de σ^2 é 1, isto é, nos Cenários B1 a B8.

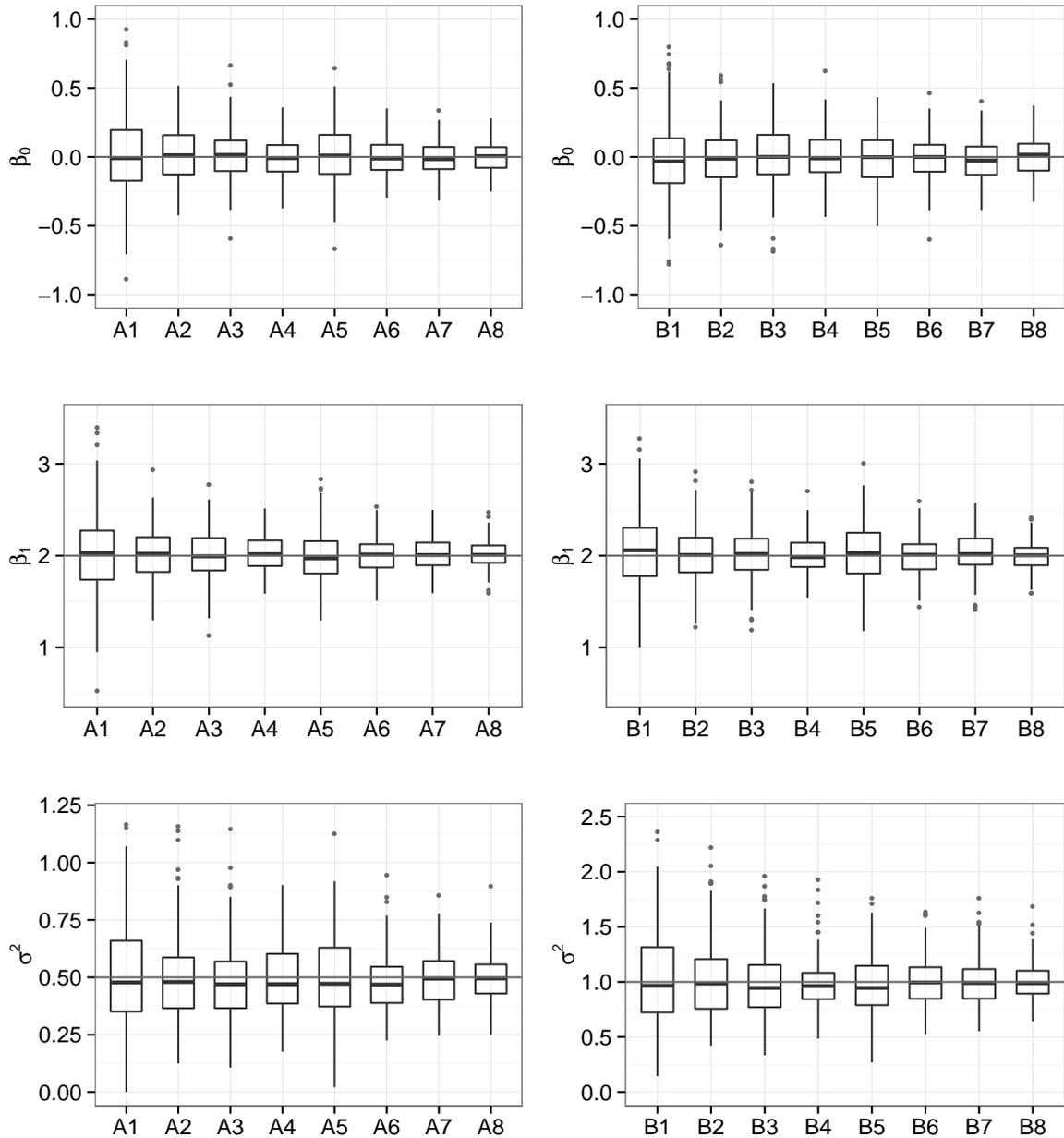


Figura 49: *Boxplots* das estimativas fornecidas para os efeitos fixos e para a variância dos efeitos aleatórios, e valores reais (linha horizontal), nos Cenários A1 a A8 e B1 a B8.

3.7.2 Dados simulados do modelo Logístico-T

Nesta subseção temos como meta compararmos as estimativas fornecidas pelos modelos Logístico-N (Modelo M1), Logístico-T estimando o grau de liberdade ν (Modelo M2), Logístico-T com ν fixado no valor real (Modelo M3) e Logístico-SL (Modelo M4), quando os dados são gerados assumindo uma distribuição de probabilidade t -Student para os efeitos aleatórios. Para investigarmos o efeito do peso das caudas nas estimativas, geramos os bancos de dados de tal forma que $V(\gamma_i) = 1$ e assumindo $\nu = 3, 6$ e 20 .

As mesmas especificações utilizadas na Subseção 3.7.1 são consideradas para o número k de *clusters*, n_i de observações dentro de cada *cluster* e n_{ij} de experimentos Bernoulli para cada observação dentro dos *clusters*. Na Tabela 3.8 apresentamos os cenários que analisaremos.

Tabela 3.8: Organização dos cenários gerados sob modelo Logístico-T

	k	$n_i = 5$		$n_i = 10$	
		$n_{ij} = 2$	$n_{ij} = 4$	$n_{ij} = 2$	$n_{ij} = 4$
$\gamma_i \stackrel{iid}{\sim} T(0; 0,333; 3)$	50	C1	C2	C3	C4
	100	C5	C6	C7	C8
$\gamma_i \stackrel{iid}{\sim} T(0; 0,667; 6)$	50	D1	D2	D3	D4
	100	D5	D6	D7	D8
$\gamma_i \stackrel{iid}{\sim} T(0; 0,9; 20)$	50	E1	E2	E3	E4
	100	E5	E6	E7	E8

3.7.2.1 Qualidade das estimativas dos parâmetros

Iniciamos nossa análise considerando cenários em que os dados são gerados assumindo uma distribuição com caudas muito pesadas para os efeitos aleatórios ($\nu = 3$). Nestes cenários, por causa da característica da distribuição Normal, esperamos que o modelo Logístico-N tenha um desempenho ruim se comparado aos demais modelos. Na Tabela 3.9, apresentamos os resultados para viés e EQM obtidos para as estimativas dos parâmetros de efeitos fixos e da distribuição dos efeitos aleatórios nos Cenários C1 a C8. Apresentamos o viés e o EQM para σ^2 apenas nos modelos M2 e M3 pois somente nestes casos tal parâmetro possui a mesma interpretação do que no modelo considerado na geração dos dados. O mesmo ocorre para ν , em que os resultados são apresentados apenas para o modelo M2, visto que no modelo M3 o parâmetro ν é assumido conhecido.

Analisando a Tabela 3.9, vemos que os modelos ajustados não apresentam muita diferença no que diz respeito a estimação dos efeitos fixos. Em todos os cenários analisados, os modelos fornecem valores semelhantes para o viés e o EQM destes parâmetros. O EQM para σ^2 é menor no Modelo M3 do que o Modelo M2 em todos os cenários. Isto é esperado, uma vez que os dois modelos assumem corretamente a distribuição *t*-Student para os efeitos aleatórios, mas o Modelo M3 traz consigo a informação adicional quanto ao valor correto para ν . Avaliando a estimação de $V(\gamma_i)$, vemos que sob o Modelo Logístico-N (M1), as estimativas apresentam viés negativo alto em todos os cenários, mas têm EQM baixo. Sob o modelo M1, a $V(\gamma_i)$ tem menor EQM nos Cenários C1, C3 e C5, enquanto que sob o Modelo M3, o EQM de $V(\gamma_i)$ é menor nos demais cenários. O Modelo M2 é o modelo que fornece as estimativas para $V(\gamma_i)$ com maior EQM. Acreditamos que isto se deve ao fato de, no Modelo M2, muitas vezes o parâmetro ν ser estimado em valores muito pequenos (próximo a 2,5), o que pode levar a instabilidade no cálculo de $V(\gamma_i)$. De forma geral, os Modelos M2 e M3 são os que fornecem as estimativas menos viesadas para $V(\gamma_i)$.

Tabela 3.9: Viés e EQM dos parâmetros de todos os modelos nos Cenários C1 a C8

Cenário		Modelo							
		M1		M2		M3		M4	
		Viés	EQM	Viés	EQM	Viés	EQM	Viés	EQM
C1	β_0	0,021	0,075	0,024	0,075	0,017	0,072	0,027	0,075
	β_1	-0,027	0,196	-0,022	0,197	-0,024	0,197	-0,016	0,197
	σ^2	-	-	0,145	0,087	0,029	0,038	-	-
	ν	-	-	11,341	283,077	-	-	-	-
	$V(\gamma_i)$	-0,259	0,199	0,057	0,632	0,087	0,345	-0,098	0,343
C2	β_0	-0,011	0,043	0,001	0,041	-0,006	0,040	0,003	0,042
	β_1	0,016	0,095	0,022	0,096	0,021	0,096	0,027	0,097
	σ^2	-	-	0,079	0,050	-0,008	0,020	-	-
	ν	-	-	9,044	218,650	-	-	-	-
	$V(\gamma_i)$	-0,325	0,190	-0,074	0,299	-0,024	0,179	-0,132	0,299
C3	β_0	0,009	0,037	0,014	0,036	0,008	0,035	0,020	0,036
	β_1	-0,003	0,081	0,001	0,081	0,000	0,081	0,006	0,081
	σ^2	-	-	0,128	0,068	0,030	0,024	-	-
	ν	-	-	9,065	216,027	-	-	-	-
	$V(\gamma_i)$	-0,275	0,156	0,041	0,394	0,088	0,215	-0,083	0,251
C4	β_0	-0,012	0,027	-0,001	0,025	-0,004	0,024	0,002	0,025
	β_1	0,042	0,043	0,048	0,044	0,047	0,044	0,051	0,045
	σ^2	-	-	0,053	0,025	-0,005	0,012	-	-
	ν	-	-	7,017	170,661	-	-	-	-
	$V(\gamma_i)$	-0,333	0,167	-0,004	0,304	-0,015	0,110	-0,101	0,225
C5	β_0	-0,004	0,037	-0,003	0,036	-0,009	0,036	-0,001	0,037
	β_1	0,019	0,086	0,023	0,087	0,022	0,087	0,030	0,089
	σ^2	-	-	0,111	0,058	0,004	0,020	-	-
	ν	-	-	8,712	200,172	-	-	-	-
	$V(\gamma_i)$	-0,302	0,159	-0,034	0,292	0,010	0,184	-0,179	0,174
C6	β_0	0,012	0,022	0,016	0,022	0,011	0,022	0,020	0,022
	β_1	-0,022	0,044	-0,016	0,044	-0,017	0,044	-0,012	0,044
	σ^2	-	-	0,053	0,025	-0,013	0,010	-	-
	ν	-	-	5,883	135,183	-	-	-	-
	$V(\gamma_i)$	-0,350	0,156	-0,068	0,195	-0,039	0,086	-0,202	0,132
C7	β_0	0,012	0,019	0,012	0,018	0,008	0,018	0,018	0,019
	β_1	-0,003	0,040	0,001	0,040	0,000	0,040	0,006	0,040
	σ^2	-	-	0,069	0,026	0,002	0,008	-	-
	ν	-	-	5,751	132,442	-	-	-	-
	$V(\gamma_i)$	-0,326	0,133	-0,020	0,180	0,004	0,071	-0,160	0,106
C8	β_0	-0,006	0,013	0,009	0,013	0,007	0,012	0,013	0,013
	β_1	-0,019	0,023	-0,013	0,022	-0,013	0,022	-0,009	0,022
	σ^2	-	-	0,035	0,018	-0,005	0,007	-	-
	ν	-	-	3,878	86,586	-	-	-	-
	$V(\gamma_i)$	-0,329	0,136	0,032	0,206	-0,015	0,059	-0,095	0,122

É possível notar que, em todos os modelos, mantendo as demais características fixas, os cenários com maiores valores de n_i ou n_{ij} apresentam menor valor de EQM para as estimativas de todos os parâmetros, exceto para $V(\gamma_i)$, no Modelo M2 para os Cenários C7

e C8, por exemplo. Quando comparamos cenários com diferentes números k de *clusters* e com valores iguais de n_i e n_{ij} , observamos que os cenários com maior k tem menor EQM para todos os parâmetros de interesse.

Se compararmos cenários em que os bancos de dados têm o mesmo número total $k \times n_i \times n_{ij}$ de observações de experimentos Bernoulli e mesmo número de *clusters*, como é o caso dos Cenários C2 e C3 ou dos Cenários C6 e C7, notamos que os cenários com maior n_i e menor n_{ij} tendem a apresentar menor EMQ para β_0 , β_1 e ν e menor viés para β_1 em todos os modelos ajustados. Conclusão similar sobre o EQM pode ser feita para $V(\gamma_i)$ nos Modelos M1 e M4 apenas. Em geral, as estimativas para $V(\gamma_i)$ foram menos viciadas nos cenários com maior n_i e menor n_{ij} . Fazendo uma comparação entre cenários com mesmo total de observações de experimentos Bernoulli, mas divididos em números k de *clusters* diferentes, notamos em cenários com igual especificação para n_{ij} (Cenários C3 e C5 e, Cenários C4 e C6) e sob os Modelos M2 e M3, os cenários com maior k tendem a fornecer menor EQM para β_0 , σ^2 , ν e $V(\gamma_i)$ e maior EMQ para β_1 . Comparando cenários com valor igual para n_i , isto é, os Cenários C2 e C5 ou os Cenários C4 e C7, observamos que os cenários com maior k (C5 e C7) apresentam menor EQM para β_0 , β_1 , ν e $V(\gamma_i)$, sob todos os modelos.

Na Figura 50 apresentamos os *boxplots* das estimativas para ν sob os Modelos M2 e M3, nos Cenários C1 a C8.

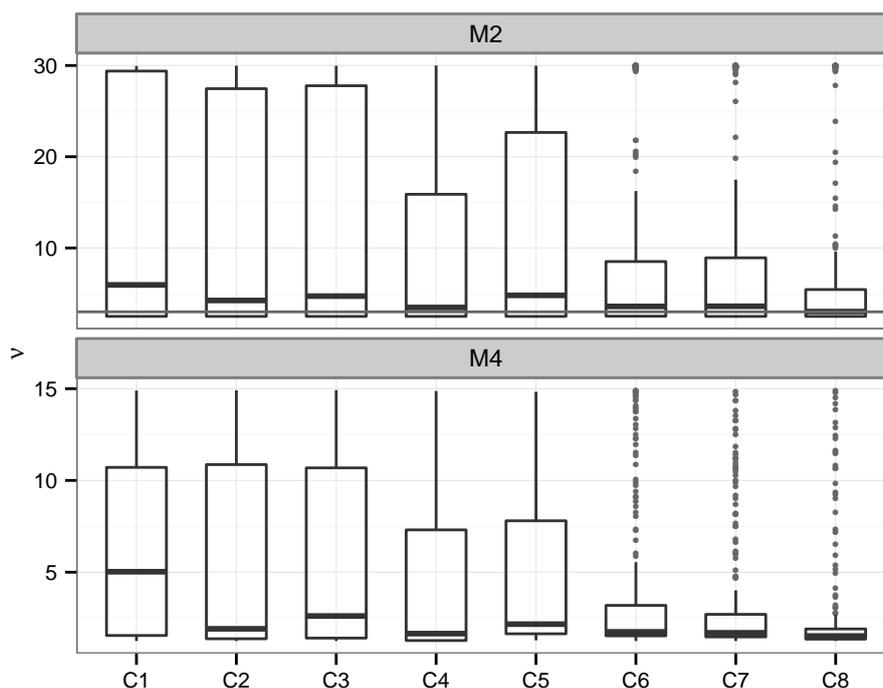


Figura 50: *Boxplots* das estimativas fornecidas para ν e valor real (linha horizontal), sob os Modelos M2 e M4, nos Cenários C1 a C8.

Da Figura 50 observamos que, sob o Modelo M2, as estimativas para ν se aproximam do valor real 3 nos cenários com maior número de observações. Além disso, notamos que a estimação deste parâmetro é bem melhor nos cenários em que $k = 100$. No entanto, muitas vezes, estimativas altas para ν são obtidas em todos os cenários. Para o

Modelo M4, notamos que nos cenários mais informativos, as estimativas de ν (que tem mesma interpretação no Modelo M2) se concentram em valores menores, atestando que a distribuição dos efeitos aleatórios tem caudas pesadas. Para aprofundar a análise das estimativas fornecidas para o parâmetro ν apresentamos, na Figura 51, os histogramas das estimativas para ν , nos Cenários C1 a C8, fornecidas pelo Modelo Logístico-T (Modelo M2). Notamos que a distribuição para as estimativas de ν é bimodal, uma das modas se concentrando em valores próximos do valor real ($\nu = 3$) e outra concentrada no valor 30. Provavelmente estas estimativas concentradas no valor 30 se espalhariam por valores maiores se não tivéssemos colocado restrições na estimação do parâmetro ν . Isto significa que em muitos bancos de dados o valor estimado para ν indica normalidade para os efeitos aleatórios.

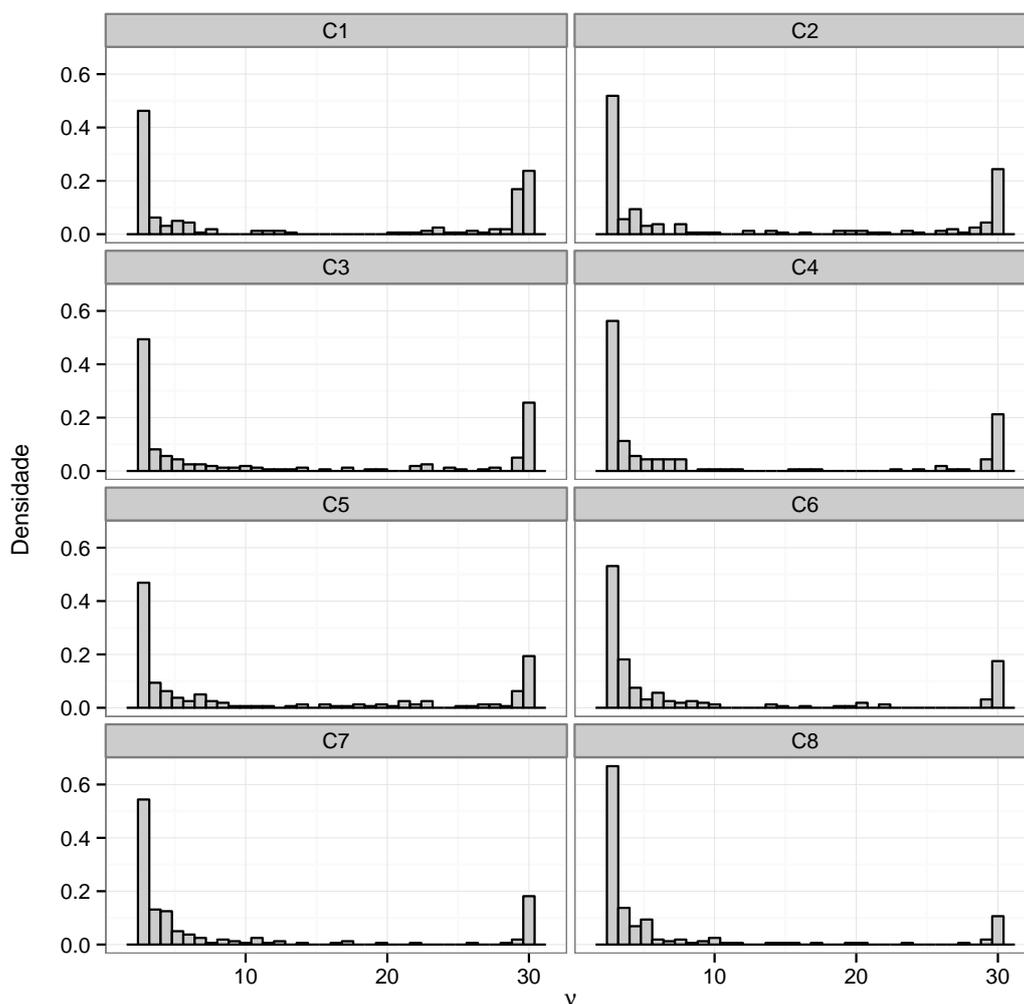


Figura 51: Histogramas das estimativas fornecidas para ν , sob o Modelo M2, nos Cenários C1 a C8.

A seguir, na Figura 52, exibimos os *boxplots* das estimativas da variância dos efeitos aleatórios, sob todos os modelos, nos Cenários C1 a C8.

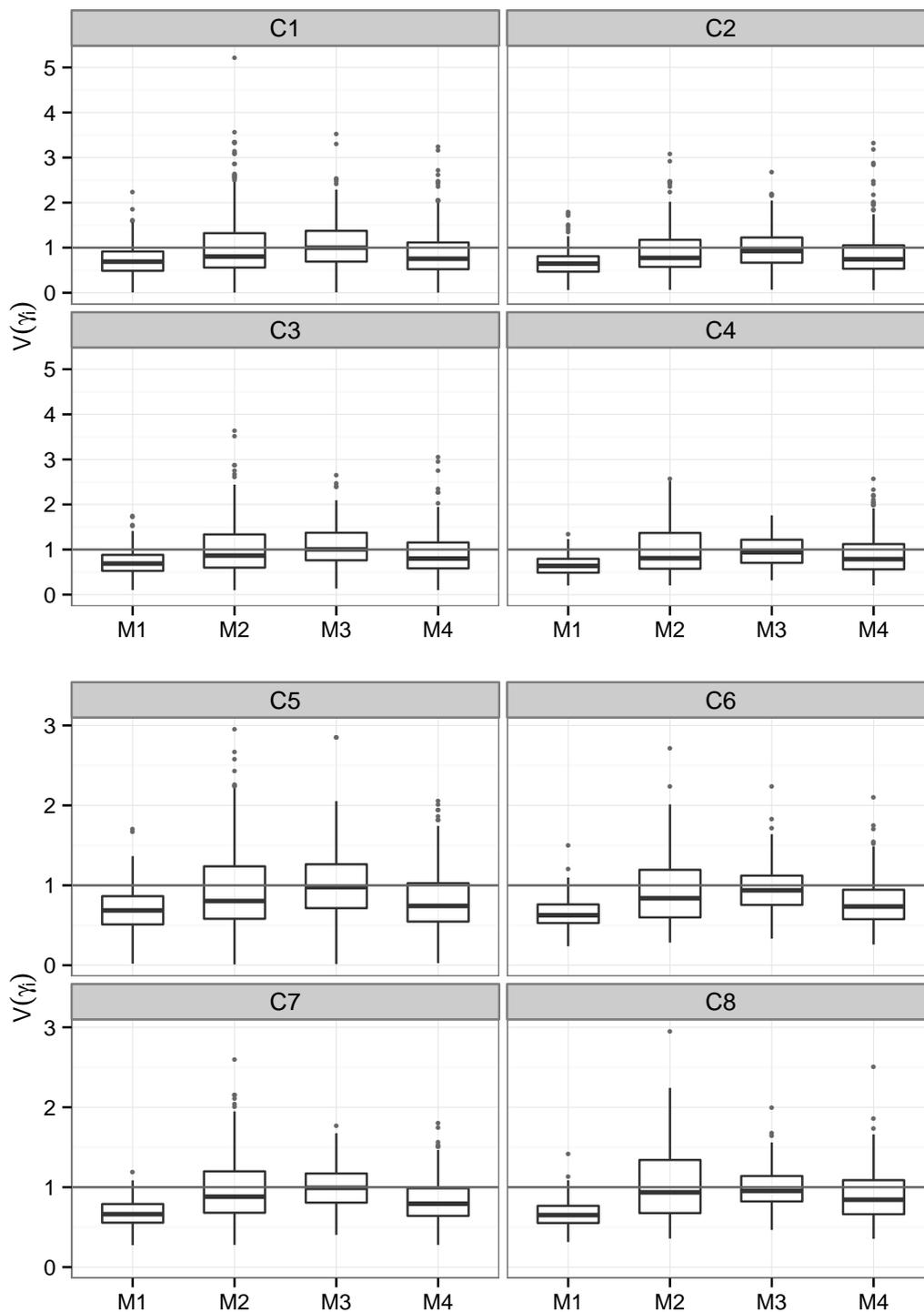


Figura 52: *Boxplots* das estimativas fornecidas para a variância dos efeitos aleatórios e valor real (linha horizontal), sob todos os modelos, nos Cenários C1 a C8.

Avaliando a Figura 52, observamos que a má-especificação da distribuição para os efeitos aleatórios, Modelos M1 e M4, conduz a estimativas muito viciadas para $V(\gamma_i)$. Além disso, sob o Modelo M1 (Logístico-N), em geral, as estimativas da $V(\gamma_i)$ têm menor variabilidade. Apesar da distribuição das estimativas de $V(\gamma_i)$ obtidas sob o Modelo M2 estar centrada próxima ao valor real de $V(\gamma_i)$ nos cenários com maior número de observações, a variância desta distribuição é muito grande, indicando uma falta de consistência do estimador. As melhores estimativas para $V(\gamma_i)$ são obtidas sob o Modelo M3. Neste caso, as estimativas estão centradas no valor real e têm pouca variabilidade. Todos os modelos fornecem estimativas com menor variabilidade nos cenários com mais *clusters* ($k = 100$) do que em cenários em que $k = 50$.

Na Tabela 3.10 são apresentados os vieses e EQM das estimativas dos efeitos fixos e dos parâmetros da distribuição dos efeitos aleatórios nos Cenários D1 a D8. O objetivo agora é avaliar a qualidade das estimativas fornecidas pelos modelos em cenários em que ν assume um valor intermediário, isto é, quando a distribuição dos efeitos aleatórios tem caudas mais pesadas do que a distribuição Normal, mas não tão pesadas quanto a assumida no caso analisado anteriormente. Quando comparamos os modelos ajustados, notamos que para os efeitos fixos, as estimativas obtidas em todos os modelos são muito próximas, principalmente, no que refere ao EQM. Observamos que, sob o Modelo M3, as estimativas para σ^2 têm viés e EQM menores do que as obtidas sob o Modelo M2. Ao analisarmos as estimativas para $V(\gamma_i)$, notamos que as estimativas obtidas sob o Modelo M1 apresentam viés negativo em todos os cenários e sob este modelo, o valor de EQM é o menor em todos os cenários, exceto no Cenário D8. Os valores de EQM sob o Modelo M3 são bem próximos aos obtidos sob o Modelo M1. Além disso, sob o Modelo M3 obtemos o menor viés nas estimativas nos Cenários D1 a D5. Sob o Modelo M4, as estimativas de $V(\gamma_i)$ têm menor viés nos Cenários D6 a D8, enquanto sob o Modelo M2 as estimativas apresentam com maior EQM e alto valor de viés em todos dos cenários.

Em nossa análise vemos que, em todos os modelos, um maior valor para qualquer uma das características k , n_i ou n_{ij} e demais variáveis iguais, conduz a estimativas com menor EQM para todos os parâmetros. Algumas poucas exceções a isto ocorreram para as estimativas de ν no Modelo M2 e para as estimativas de $V(\gamma_i)$ nos Modelos M2 e M4.

Ao compararmos apenas os cenários com a mesma quantidade total de observações Bernoulli, vemos que as conclusões obtidas da Tabela 3.9 para os Cenários C1 a C8 não são válidas para os cenários gerados com $\nu = 6$ (Cenários D1 a D8). Na comparação entre cenários com mesmo número k de *clusters*, notamos que os cenários com maior n_i e menor n_{ij} tiveram maior EQM para β_1 em todos os modelos. Para os demais parâmetros os resultados foram inconclusivos, visto que houve um comportamento nos cenários em que $k = 50$ e outro comportamento para cenários em que $k = 100$. Comparando os cenários com igual especificação de n_{ij} vemos que, sob todos os modelos, os cenários com mais *clusters* e menor n_i tiveram menor EQM para β_0 , β_1 , σ^2 e $V(\gamma_i)$. Por fim, fazendo a comparação entre cenários com valor igual para n_i , temos que os cenários com maior valor de k e menor valor de n_{ij} apresentam menor EQM apenas para β_0 e $V(\gamma_i)$.

Quando comparamos os resultados obtidas nas Tabelas 3.9 e 3.10, percebemos que os modelos forneceram estimativas com menor EQM para $V(\gamma_i)$ e maior EQM para β_0 , σ^2 e ν nos cenários em que os bancos de dados foram gerados com maior valor de ν .

Tabela 3.10: Viés e EQM dos parâmetros de todos os modelos nos Cenários D1 a D8

Cenário		Modelo							
		M1		M2		M3		M4	
		Viés	EQM	Viés	EQM	Viés	EQM	Viés	EQM
D1	β_0	0,006	0,086	0,006	0,086	0,003	0,085	0,008	0,086
	β_1	0,023	0,200	0,026	0,201	0,025	0,200	0,029	0,202
	σ^2	—	—	0,056	0,146	0,013	0,085	—	—
	ν	—	—	12,282	295,258	—	—	—	—
	$V(\gamma_i)$	-0,066	0,150	0,149	0,406	0,020	0,192	0,028	0,225
D2	β_0	0,000	0,048	0,002	0,047	0,000	0,046	0,005	0,047
	β_1	-0,007	0,083	-0,006	0,083	-0,006	0,083	-0,004	0,083
	σ^2	—	—	0,035	0,084	0,006	0,056	—	—
	ν	—	—	11,466	281,555	—	—	—	—
	$V(\gamma_i)$	-0,082	0,106	0,153	0,446	0,010	0,125	0,039	0,209
D3	β_0	-0,008	0,043	-0,006	0,043	-0,008	0,043	-0,003	0,043
	β_1	0,021	0,085	0,022	0,085	0,022	0,085	0,024	0,086
	σ^2	—	—	0,022	0,076	0,003	0,055	—	—
	ν	—	—	10,516	259,908	—	—	—	—
	$V(\gamma_i)$	-0,087	0,103	0,151	0,413	0,006	0,123	0,037	0,202
D4	β_0	0,007	0,032	0,014	0,030	0,015	0,030	0,017	0,031
	β_1	-0,005	0,040	-0,003	0,040	-0,003	0,040	-0,001	0,040
	σ^2	—	—	-0,001	0,068	0,001	0,048	—	—
	ν	—	—	8,510	216,944	—	—	—	—
	$V(\gamma_i)$	-0,079	0,097	0,183	0,491	0,002	0,107	0,107	0,279
D5	β_0	-0,006	0,037	-0,008	0,036	-0,010	0,037	-0,005	0,037
	β_1	0,013	0,079	0,015	0,079	0,014	0,079	0,017	0,079
	σ^2	—	—	0,050	0,085	0,001	0,040	—	—
	ν	—	—	11,311	265,949	—	—	—	—
	$V(\gamma_i)$	-0,082	0,073	0,110	0,260	0,003	0,090	-0,016	0,101
D6	β_0	0,018	0,024	0,019	0,024	0,017	0,024	0,022	0,024
	β_1	-0,022	0,038	-0,020	0,038	-0,020	0,038	-0,019	0,038
	σ^2	—	—	0,019	0,049	-0,008	0,022	—	—
	ν	—	—	8,440	198,861	—	—	—	—
	$V(\gamma_i)$	-0,097	0,049	0,070	0,142	-0,012	0,050	-0,004	0,075
D7	β_0	0,025	0,024	0,026	0,025	0,025	0,025	0,029	0,025
	β_1	-0,032	0,049	-0,031	0,049	-0,031	0,049	-0,029	0,049
	σ^2	—	—	0,032	0,049	-0,003	0,028	—	—
	ν	—	—	10,059	242,847	—	—	—	—
	$V(\gamma_i)$	-0,094	0,058	0,082	0,203	-0,003	0,062	0,002	0,099
D8	β_0	-0,007	0,017	-0,002	0,017	-0,003	0,016	-0,001	0,017
	β_1	0,007	0,021	0,009	0,021	0,009	0,021	0,010	0,021
	σ^2	—	—	0,027	0,039	-0,009	0,017	—	—
	ν	—	—	8,822	208,426	—	—	—	—
	$V(\gamma_i)$	-0,104	0,041	0,040	0,127	-0,014	0,038	0,001	0,071

Apresentamos na Figura 53 os histogramas das estimativas de ν quando ajustamos o Modelo M2, nos Cenários D1 a D8. Os *boxplots* das estimativas de ν sob os modelos M2 e M4, assim como os histogramas das estimativas de ν no Modelo M4 para todos os cenários analisados nesta subseção, podem ser encontradas no Apêndice F.

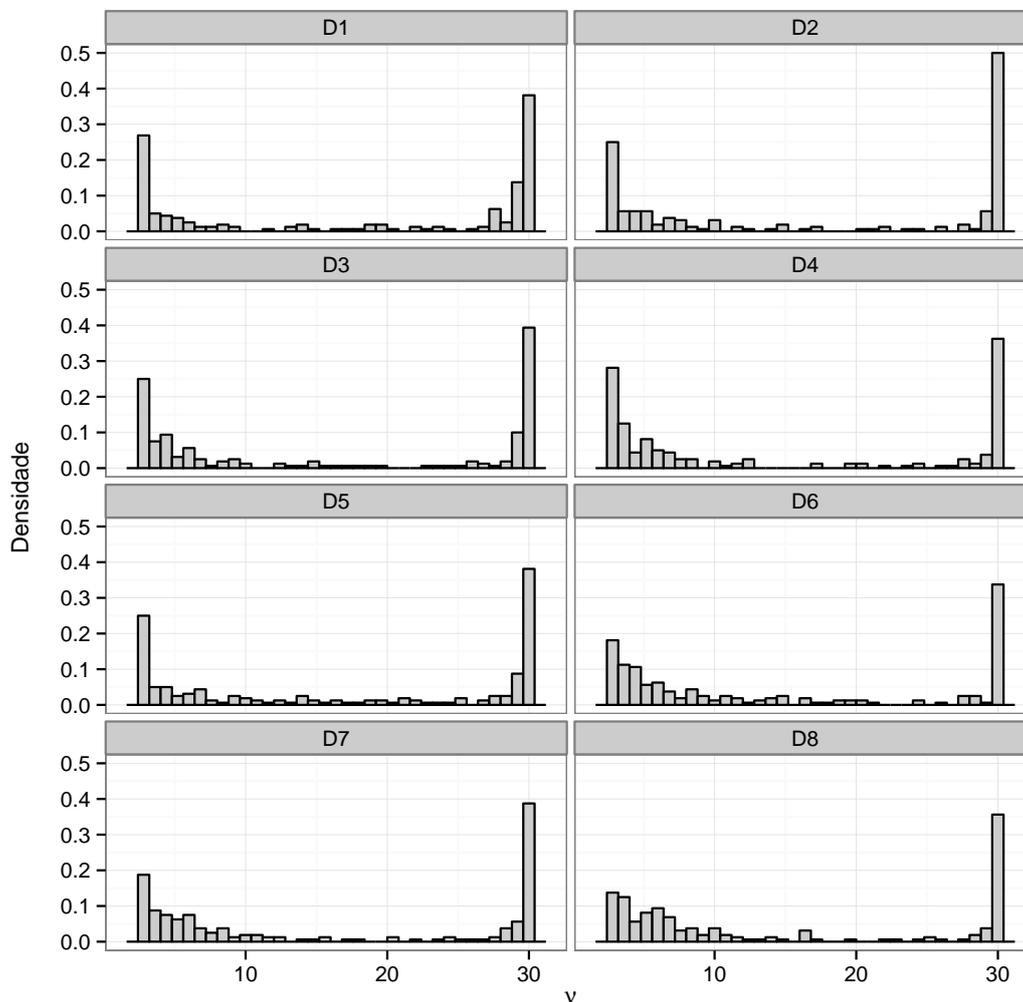


Figura 53: Histogramas das estimativas fornecidas para ν , sob o Modelo M2, nos Cenários D1 a D8.

Da Figura 53, percebemos que a estimação do parâmetro ν no Modelo M2 é ruim em todos dos cenários. A maior parte das estimativas se concentra em valores muito baixos ou muito altos se comparados ao valor real $\nu = 6$. Aparentemente, no Cenário D8 existe uma concentração um pouco maior de estimativas próximas de $\nu = 6$.

Observamos na Figura 54 os *boxplots* das estimativas para a variância da distribuição dos efeitos aleatórios $V(\gamma_i)$, fornecidas por todos os modelos ajustados, nos Cenários D1 a D8. Assim como ocorreu para os Cenários C1 a C8, o Modelo M3 (Logístico-T com ν conhecido) faz bom papel na estimação de $V(\gamma_i)$, fornecendo estimativas centradas no valor real e com pouca variabilidade. O modelo Logístico-N fornece estimativas com viés negativo e com pouca variabilidade. O Modelo M2 fornece estimativas centradas no valor correto de $V(\gamma_i)$, no entanto, estas estimativas têm grande variabilidade. Isto ocorre

provavelmente devido a pobre estimação dos graus de liberdade ν .

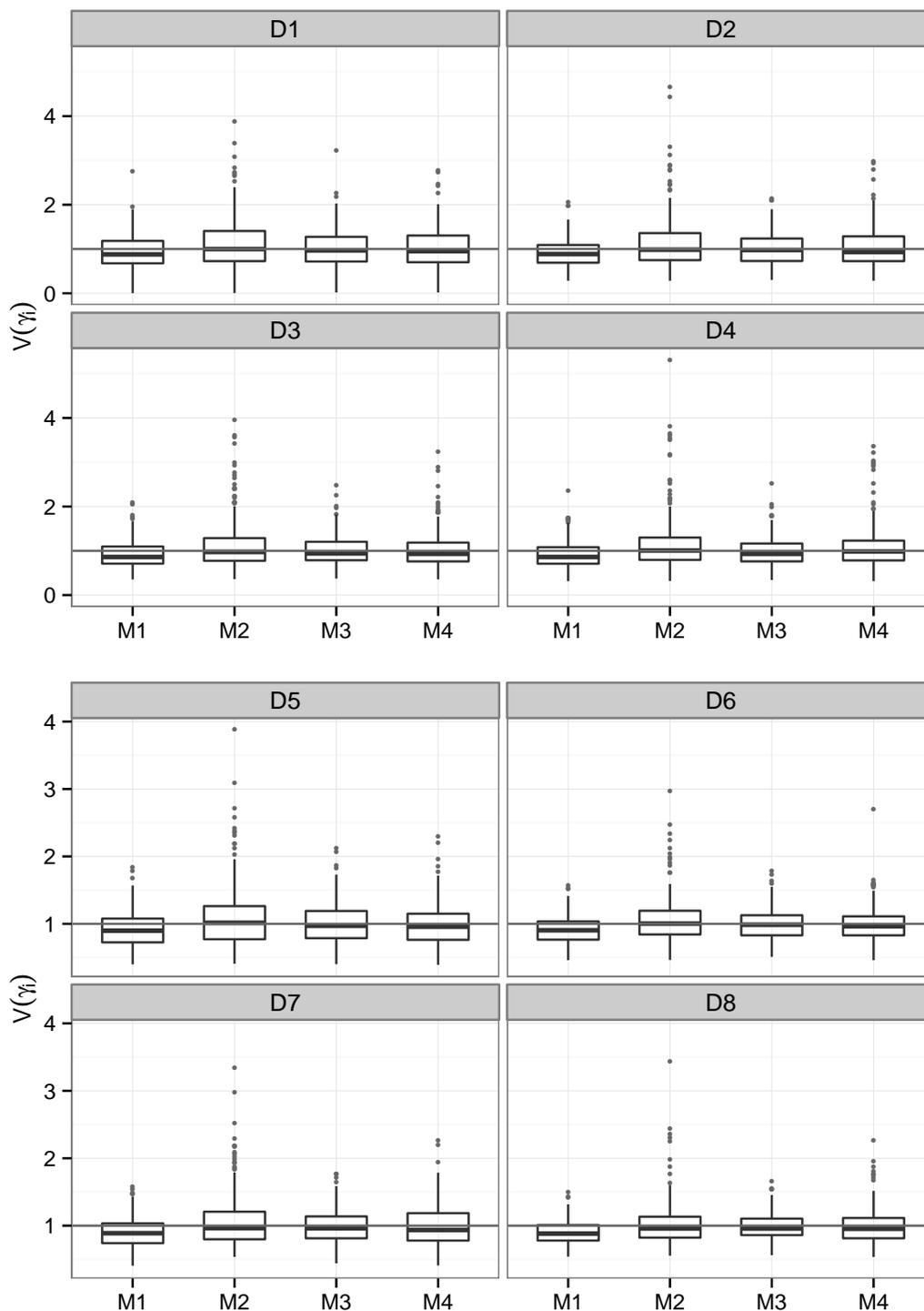


Figura 54: *Boxplots* das estimativas fornecidas para a variância dos efeitos aleatórios e valor real (linha horizontal), sob todos os modelos, nos Cenários D1 a D8.

Na Tabela 3.11, apresentamos o viés e o EQM dos estimadores dos parâmetros de efeitos fixos e da distribuição dos efeitos aleatórios nos Cenários E1 a E8.

Tabela 3.11: Viés e EQM dos parâmetros de todos os modelos nos Cenários E1 a E8

Cenário		Modelo							
		M1		M2		M3		M4	
		Viés	EQM	Viés	EQM	Viés	EQM	Viés	EQM
E1	β_0	-0,044	0,088	-0,044	0,088	-0,045	0,088	-0,043	0,088
	β_1	0,048	0,161	0,049	0,162	0,048	0,161	0,051	0,163
	σ^2	-	-	-0,081	0,147	0,013	0,132	-	-
	ν	-	-	0,483	132,453	-	-	-	-
	$V(\gamma_i)$	-0,003	0,153	0,183	0,494	0,015	0,163	0,052	0,206
E2	β_0	0,016	0,042	0,016	0,042	0,016	0,042	0,018	0,042
	β_1	-0,002	0,067	-0,002	0,067	-0,002	0,067	-0,001	0,067
	σ^2	-	-	-0,068	0,110	-0,009	0,088	-	-
	ν	-	-	2,332	122,076	-	-	-	-
	$V(\gamma_i)$	-0,030	0,102	0,076	0,227	-0,010	0,109	0,032	0,157
E3	β_0	-0,017	0,042	-0,015	0,042	-0,016	0,042	-0,015	0,042
	β_1	0,011	0,072	0,012	0,072	0,012	0,072	0,012	0,072
	σ^2	-	-	-0,064	0,078	0,013	0,066	-	-
	ν	-	-	0,873	128,675	-	-	-	-
	$V(\gamma_i)$	-0,004	0,079	0,113	0,179	0,015	0,082	0,057	0,115
E4	β_0	-0,009	0,032	-0,010	0,032	-0,007	0,032	-0,008	0,032
	β_1	0,013	0,043	0,013	0,043	0,013	0,043	0,013	0,043
	σ^2	-	-	-0,072	0,068	-0,012	0,049	-	-
	ν	-	-	1,693	126,804	-	-	-	-
	$V(\gamma_i)$	-0,032	0,059	0,070	0,147	-0,013	0,061	0,024	0,088
E5	β_0	0,012	0,041	0,011	0,041	0,012	0,041	0,013	0,041
	β_1	-0,036	0,072	-0,035	0,072	-0,036	0,072	-0,034	0,072
	σ^2	-	-	-0,106	0,082	-0,019	0,057	-	-
	ν	-	-	0,473	129,399	-	-	-	-
	$V(\gamma_i)$	-0,038	0,068	0,121	0,286	-0,021	0,071	0,008	0,108
E6	β_0	-0,011	0,026	-0,011	0,026	-0,011	0,026	-0,010	0,026
	β_1	0,014	0,043	0,015	0,043	0,015	0,043	0,015	0,043
	σ^2	-	-	-0,067	0,057	-0,006	0,043	-	-
	ν	-	-	-0,166	120,177	-	-	-	-
	$V(\gamma_i)$	-0,025	0,051	0,032	0,071	-0,007	0,054	0,005	0,062
E7	β_0	0,011	0,025	0,012	0,024	0,011	0,025	0,013	0,025
	β_1	-0,022	0,037	-0,022	0,037	-0,022	0,037	-0,022	0,037
	σ^2	-	-	-0,074	0,058	-0,017	0,032	-	-
	ν	-	-	0,982	120,013	-	-	-	-
	$V(\gamma_i)$	-0,037	0,038	0,045	0,076	-0,019	0,040	0,004	0,050
E8	β_0	-0,011	0,017	-0,010	0,017	-0,010	0,017	-0,008	0,017
	β_1	0,001	0,022	0,001	0,022	0,001	0,022	0,001	0,022
	σ^2	-	-	-0,056	0,044	-0,001	0,024	-	-
	ν	-	-	0,478	120,228	-	-	-	-
	$V(\gamma_i)$	-0,018	0,028	0,045	0,046	-0,001	0,030	0,029	0,041

Da Tabela 3.11, notamos que mais uma vez todos os modelos forneceram estimativas muito próximas para os efeitos fixos. Para a variância dos efeitos aleatórios, os Modelos M1 e M3 tiveram melhor desempenho, apresentando valores muito próximos para EQM. Quanto ao viés, em geral, o modelo M3 leva pequena vantagem. Para o parâmetro σ^2 , o Modelo M3 fornece estimativas um pouco melhores em viés e EQM do que aquelas obtidas quando o Modelo M2 foi ajustado.

Comparando cenários com mesma quantidade de experimentos Bernoulli, vemos que as conclusões não foram iguais as obtidas nos cenários anteriores. Observamos que sob os Modelos M1, M3 e M4, quando comparamos cenários com mesmo número de *clusters*, os cenários com maior valor de n_i e menor valor de n_{ij} tendem a fornecer estimativas com menor EQM para β_0 , σ^2 e $V(\gamma_i)$. Resultados similares são observados para os mesmos parâmetros quando comparamos cenários com valor igual de n_{ij} , onde notamos que os cenários com maior valor de k e menor valor de n_i têm menor EQM ou quando comparamos cenários com valor igual de n_i , em que os cenários com maior k e menor n_{ij} tendem a ter menor EQM.

Para avaliarmos melhor a qualidade da estimação de ν e $V(\gamma_i)$, nas Figuras 55 e 56 apresentamos, respectivamente, os histogramas das estimativas de ν , fornecidas pelo Modelo M2, e os *boxplots* das estimativas de $V(\gamma_i)$ nos Cenários E1 a E8.

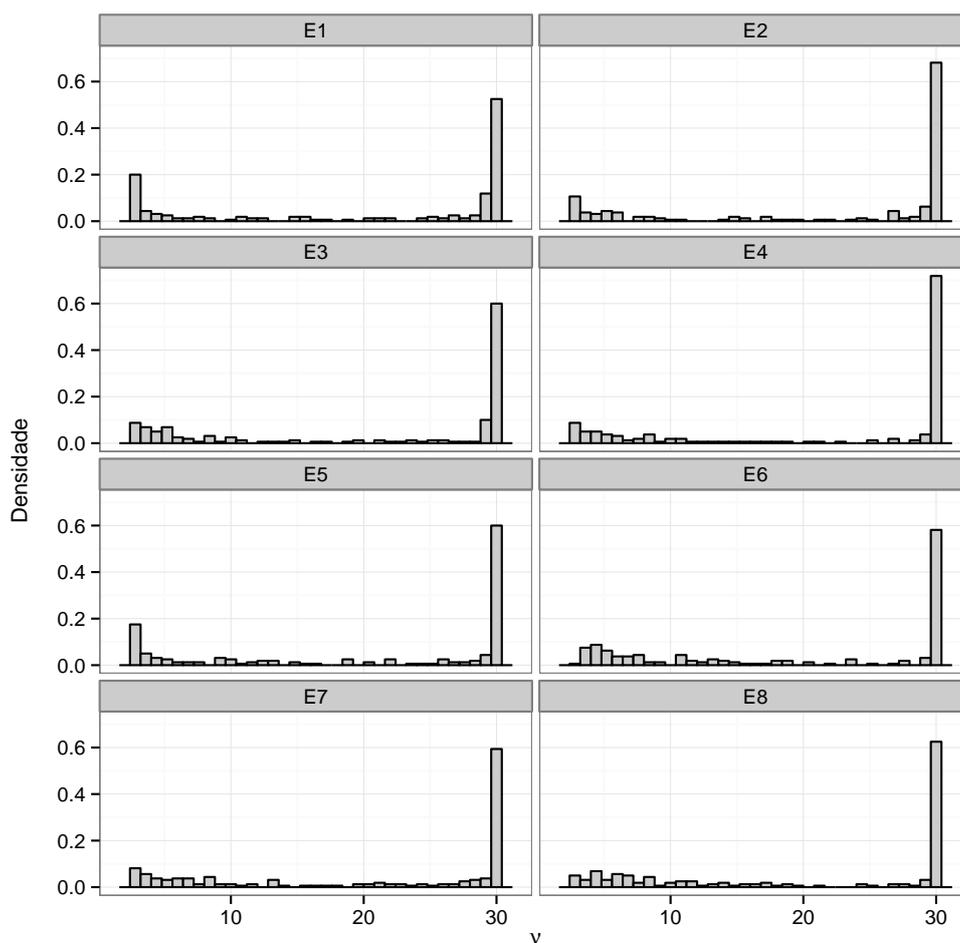


Figura 55: Histogramas das estimativas fornecidas para ν , sob o Modelo M2, nos Cenários E1 a E8.

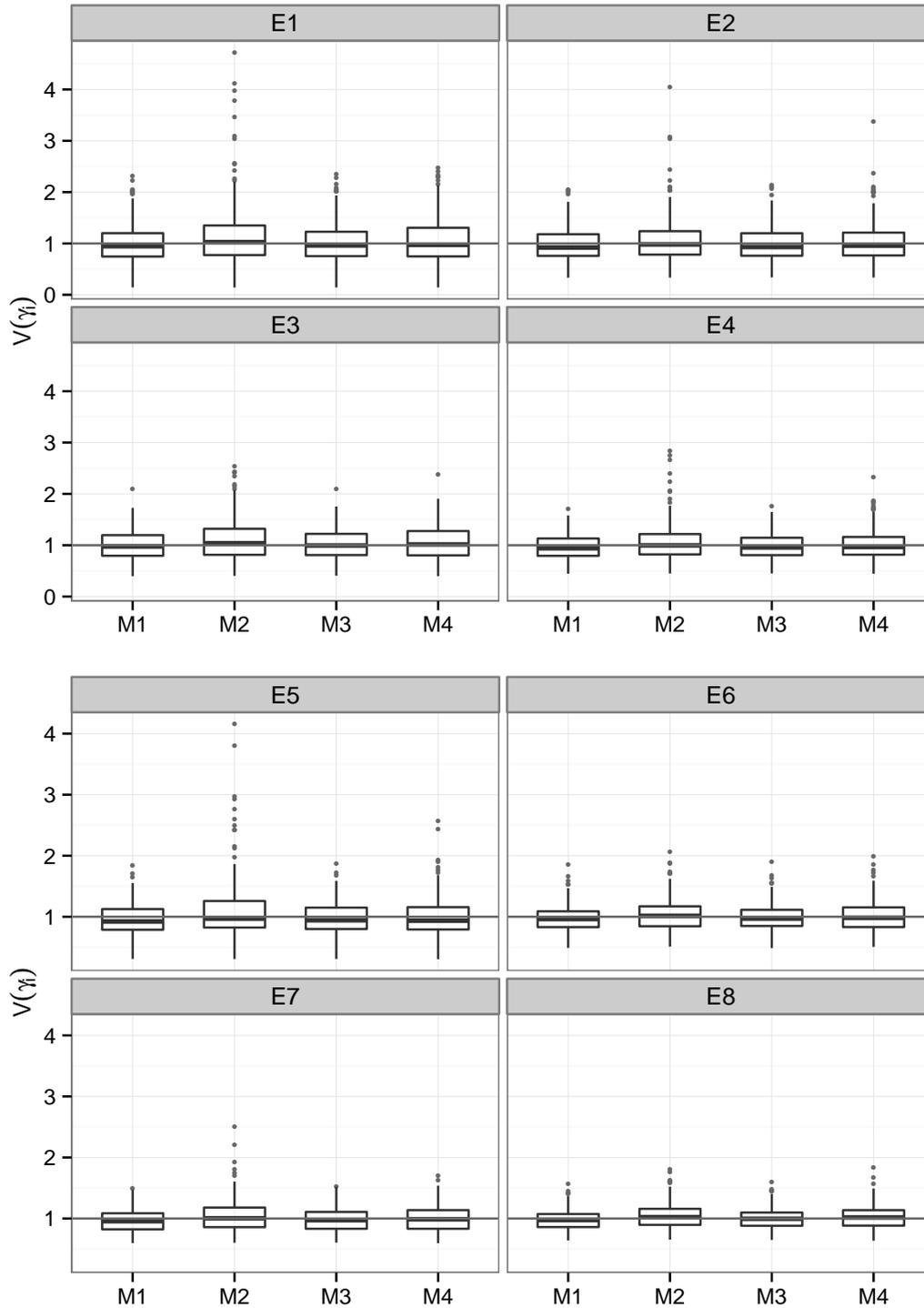


Figura 56: *Boxplots* das estimativas fornecidas para a variância dos efeitos aleatórios e valor real (linha horizontal), sob todos os modelos, nos Cenários E1 a E8.

Da Figura 55, observamos que o número ν de graus de liberdade da distribuição t -Student, utilizada para representar o comportamento dos efeitos aleatórios, no Modelo M2 é estimado, predominante, no valor 30, indicando que a cauda da distribuição é mais leve do que a real ($\nu = 20$). Isto indica que, na maior parte das vezes, o Modelo M2 aponta normalidade para o comportamento dos efeitos aleatórios. Observamos que

a maior concentração de estimativas em valores baixos ocorre nos cenários com menor número de observações dentro de cada *cluster* como, por exemplo, nos Cenários E1 e E5. Na Figura 56, notamos que todos os modelos fornecem estimativas centradas no valor real, isto é, em $V(\gamma_i) = 1$. Além disto, sob os Modelos M2 e M4, as estimativas têm maior variabilidade, enquanto sob os Modelos M1 e M3, a distribuição das estimativas é muito parecida. Vemos, também, que a variabilidade das estimativas diminui ao aumentarmos k . O mesmo ocorre se aumentarmos n_i ou n_{ij} .

3.7.2.2 A qualidade das estimativas da razão de chances mediana

Passamos então para a análise das estimativas fornecida pelos modelos ajustados para a razão de chances mediana $MOR_{|12|}$ nos cenários estudados nesta seção. Com as especificações consideradas para os parâmetros da distribuição *t*-Student dos efeitos aleatórios, temos que o valor real para $MOR_{|12|}$ é aproximadamente 1,999 nos Cenários C1 a C8, 2,393 nos Cenários D1 a D8 e 2,548 nos Cenários E1 a E8. Na Tabela 3.12 apresentamos o Viés e o EQM das estimativas de $MOR_{|12|}$ obtidas sob os Modelos M1 a M4 em todos os cenários.

Tabela 3.12: Viés e EQM para $MOR_{|12|}$ em todos os modelos nos cenários C1 a C8, D1 a D8 e E1 a E8

Cenários	Modelo							
	M1		M2		M3		M4	
	Viés	EQM	Viés	EQM	Viés	EQM	Viés	EQM
C1	0,261	0,289	0,105	0,179	0,055	0,176	0,226	0,261
C2	0,182	0,174	0,011	0,091	-0,018	0,090	0,110	0,129
C3	0,247	0,192	0,092	0,109	0,062	0,105	0,188	0,152
C4	0,174	0,121	0,007	0,053	-0,007	0,055	0,091	0,081
C5	0,213	0,159	0,062	0,099	0,007	0,093	0,179	0,136
C6	0,155	0,080	0,003	0,042	-0,023	0,042	0,094	0,057
C7	0,186	0,079	0,033	0,039	0,009	0,035	0,122	0,053
C8	0,182	0,078	0,004	0,030	-0,004	0,029	0,098	0,046
D1	0,116	0,250	0,005	0,232	0,021	0,220	0,094	0,242
D2	0,098	0,166	-0,003	0,138	0,012	0,140	0,051	0,150
D3	0,092	0,158	-0,009	0,135	0,008	0,139	0,047	0,147
D4	0,103	0,153	-0,021	0,115	0,005	0,120	0,030	0,124
D5	0,098	0,113	0,003	0,109	0,006	0,101	0,075	0,108
D6	0,081	0,068	-0,016	0,063	-0,008	0,056	0,033	0,061
D7	0,084	0,083	-0,003	0,071	0,001	0,070	0,046	0,073
D8	0,073	0,052	-0,012	0,046	-0,009	0,042	0,024	0,047
E1	0,042	0,246	-0,037	0,237	0,018	0,239	0,027	0,237
E2	0,010	0,159	-0,044	0,164	-0,010	0,160	-0,014	0,161
E3	0,041	0,124	-0,019	0,119	0,018	0,117	0,015	0,119
E4	0,007	0,090	-0,044	0,090	-0,014	0,088	-0,019	0,088
E5	0,000	0,104	-0,073	0,107	-0,022	0,104	-0,013	0,100
E6	0,016	0,079	-0,034	0,080	-0,006	0,077	-0,007	0,080
E7	0,001	0,057	-0,051	0,065	-0,022	0,057	-0,022	0,060
E8	0,025	0,043	-0,026	0,048	0,002	0,043	-0,003	0,046

Nas Figuras 57, 58 e 59 apresentamos os *boxplots* das estimativas de MOR_{12} obtidas, sob cada um dos modelos, nos cenários C1 a C8, D1 a D8 e E1 a E8, respectivamente.

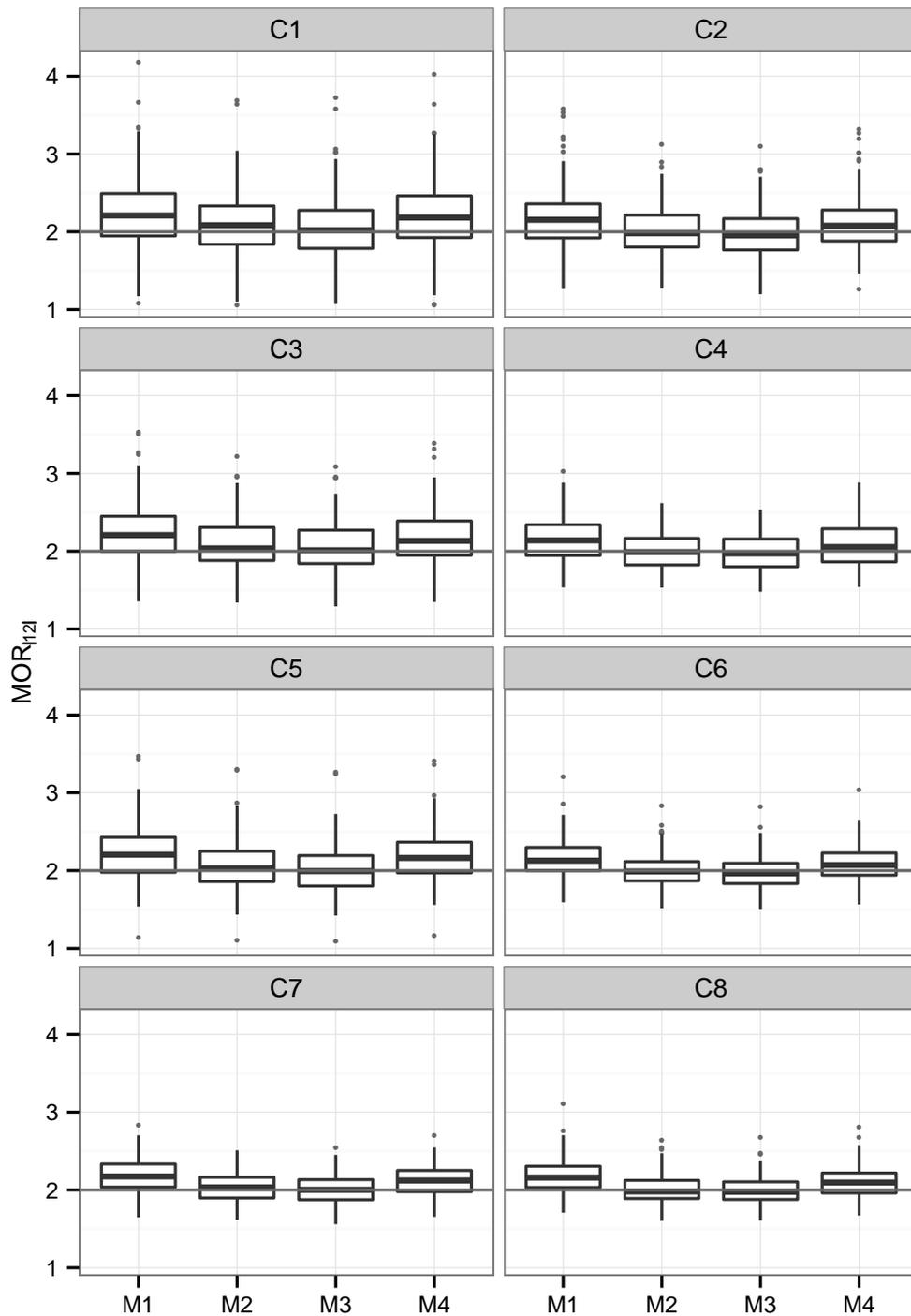


Figura 57: *Boxplots* das estimativas de MOR_{12} e valor real (linha horizontal), sob todos os modelos, nos Cenários C1 a C8.

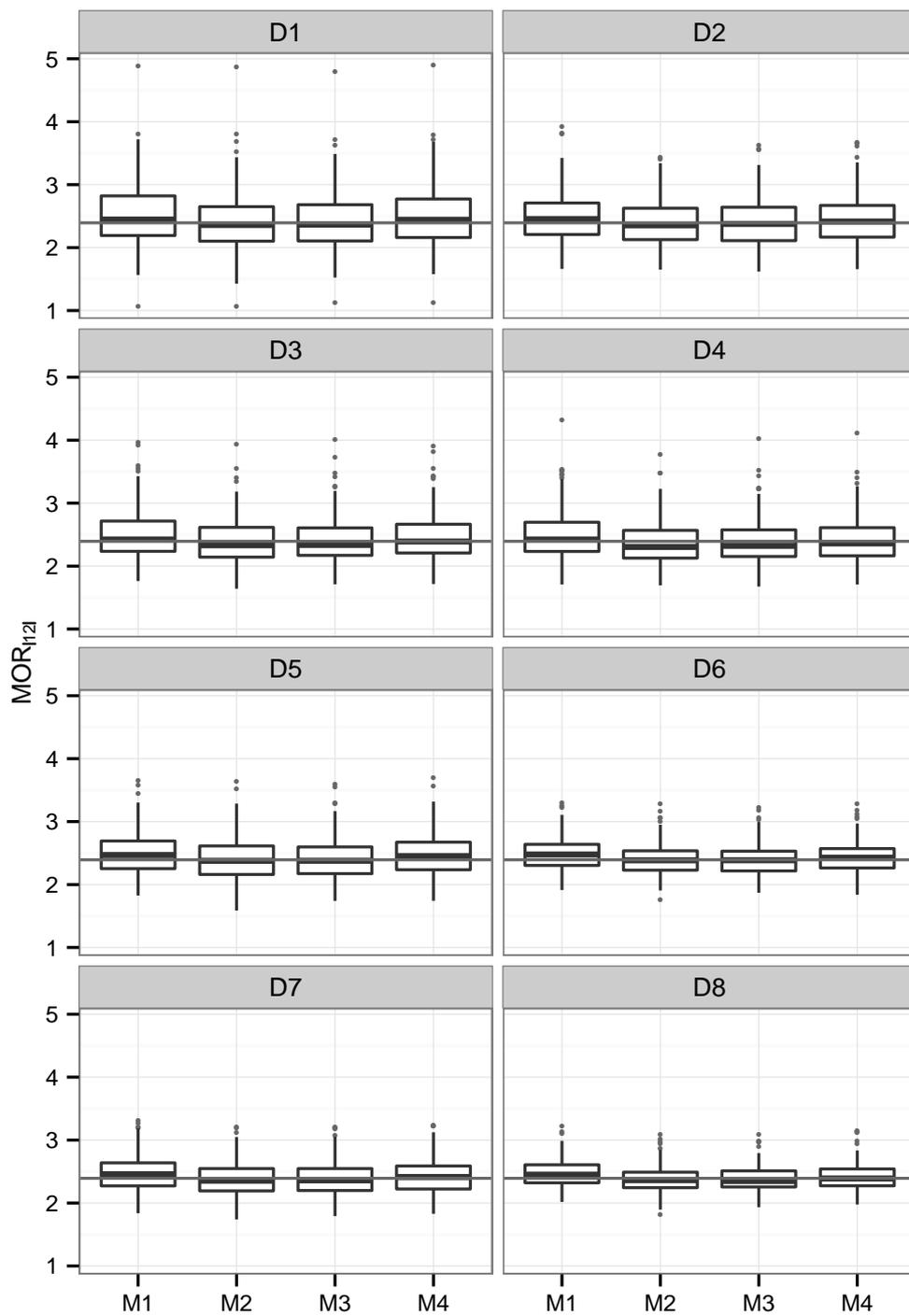


Figura 58: *Boxplots* das estimativas de MOR_{12} e valor real (linha horizontal), sob todos os modelos, nos Cenários D1 a D8.

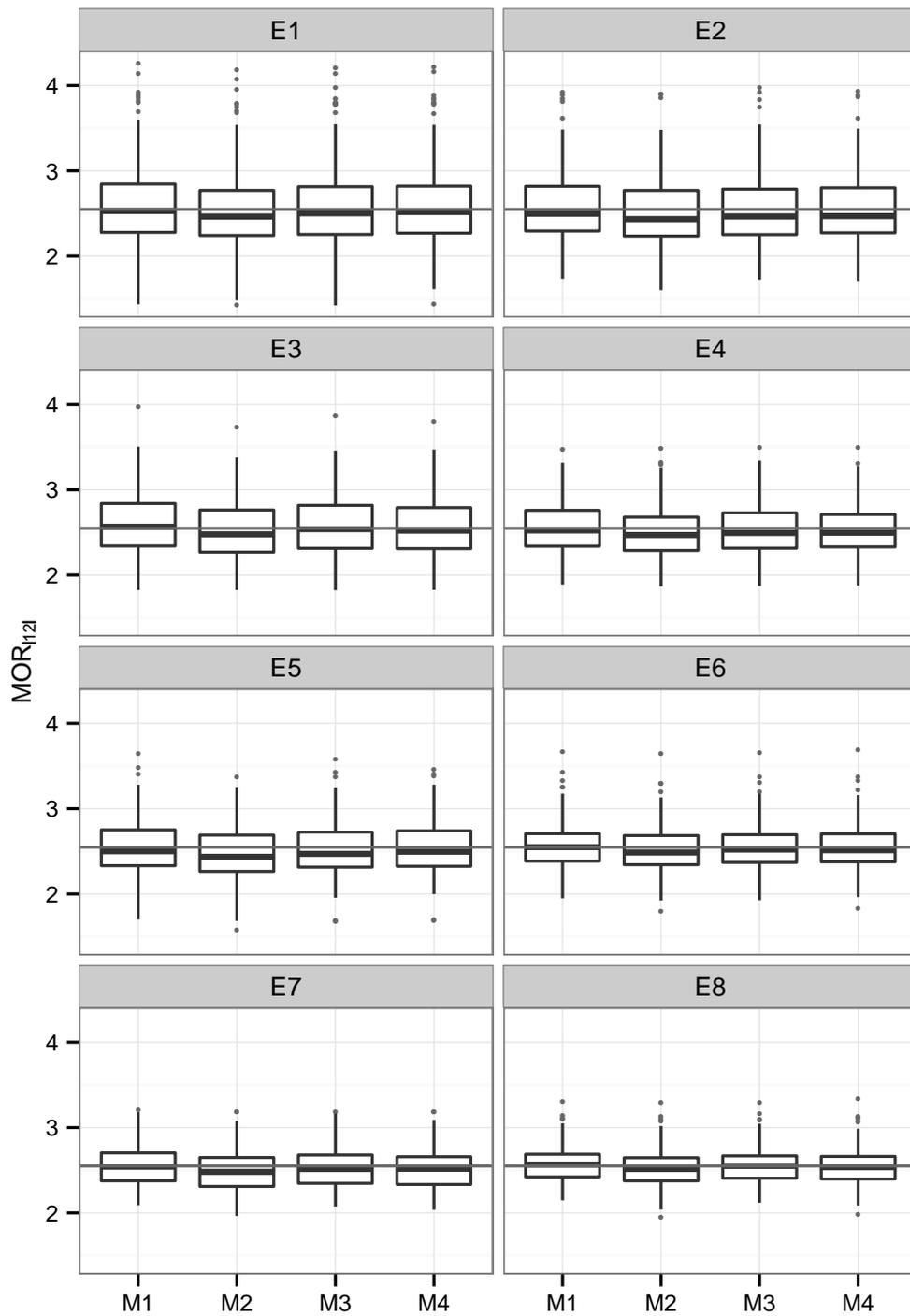


Figura 59: *Boxplots* das estimativas de MOR_{12} e valor real (linha horizontal), sob todos os modelos, nos Cenários E1 a E8.

Da Tabela 3.12, vemos que sob os Modelos M2 e M3, com ligeira vantagem para M3, obtemos as melhores estimativas para $MOR_{|12|}$, com menor viés e menor EQM, nos cenários C1 a C8. Nos Cenários D1 a D8, os Modelos M2 e M3 novamente tendem a fornecer as estimativas com menores viés e EQM. O EQM sob o Modelo M4 tem valores próximos aos menores valores obtidos sob os cenários D5 a D8. Nos Cenários E1 a E2, os resultados para viés e EQM foram muito similares para todos os modelos, mas o Modelo M2, geralmente, tem desempenho um pouco pior que os demais. Independente do valor real de ν , aumentar o valor de k (ou n_i ou n_{ij}), mantendo fixas as demais quantidades, leva a menores valores do EQM para as estimativas de $MOR_{|12|}$.

Quando comparamos cenários que têm o mesmo total de observações de experimentos Bernoulli, para cenários com mesma quantidade de *clusters* não é possível tirar conclusões. Se os cenários comparados tem mesmo valor de n_{ij} , notamos que os cenários com maior k e menor n_i têm menor EQM para $MOR_{|12|}$ em todos os modelos. Contudo, se a comparação envolve cenários com igual valor para n_i , existe uma tendência de que cenários com maior k e menor n_{ij} tendem a ter menor EQM para $MOR_{|12|}$, mas isso não é verdade quando comparamos os cenários C2 e C5 para as estimativas fornecidas pelos Modelos M2 a M4.

Observamos da Figura 57, que nos Cenários C1 a C8, sob os Modelos M2 e M3, as estimativas são mais próximas do valor real do que para os demais modelos. Nos cenários D1 a D8, esse comportamento persiste, mas com menor intensidade, como podemos ver na Figura 58. Já na Figura 59, vemos que para os Cenários E1 a E8, as estimativas fornecidas por todos os modelos são muito similares, mas com pequena desvantagem para o Modelo M2.

3.7.2.3 Comparação de modelos

Na Tabela 3.13, avaliamos qual dos modelos melhor se ajusta aos bancos de dados segundo o AIC. Apresentamos a proporção de vezes em que cada modelo tem o menor valor de AIC, isto é, apresentou o melhor ajuste aos banco de dados gerados. A comparação de modelos é conduzida de duas formas distintas. Na primeira, a qual denotamos por Seleção 1, calculamos a proporção de escolha de cada modelos entre os 4 modelos ajustados. Na segunda comparação (Seleção 2), a proporção de escolha de cada modelo é calculada considerando apenas os modelos M1, M2 e M4, ou seja, excluimos da comparação o Modelo M3, modelo em que assumimos como conhecido o valor de ν . O objetivo desta segunda comparação é avaliarmos, de acordo com o AIC, qual dos modelos tem melhor comportamento quando não dispomos de informação sobre ν . Comparando todos os modelos, concluímos que o Modelo M3 é o melhor modelo seguido pelo Modelo M2 nos Cenários C1 a C8 e D1 a D8, principalmente, quando $k = 100$. Em geral, nos Cenários E1 a E4, o modelo M3 apresenta maior proporção de escolha, seguido pelo modelo M2. O contrário ocorre nos Cenários E5 a E8, em que o Modelo M2 apresenta maior proporção de escolha e o modelo M3 é o segundo melhor. Ao retirarmos o Modelo M3 da seleção, vemos que o Modelo M2 apresenta maior proporção de escolha em quase todos os cenários contemplados pela tabela, exceto no cenários C1.

Tabela 3.13: Proporção de escolha de cada modelo, usando o critério AIC, nos cenários gerados sob o modelo Logístico-T

Cenário	Seleção 1				Seleção 2		
	M1	M2	M3	M4	M1	M2	M4
C1	0,230	0,130	0,525	0,115	0,395	0,345	0,260
C2	0,110	0,210	0,450	0,230	0,230	0,400	0,370
C3	0,095	0,185	0,575	0,145	0,205	0,520	0,275
C4	0,065	0,225	0,515	0,195	0,140	0,520	0,340
C5	0,030	0,220	0,700	0,050	0,125	0,715	0,160
C6	0,005	0,280	0,660	0,055	0,045	0,835	0,120
C7	0,005	0,265	0,670	0,060	0,030	0,815	0,155
C8	0,000	0,350	0,600	0,050	0,015	0,895	0,090
D1	0,165	0,200	0,505	0,130	0,345	0,390	0,265
D2	0,100	0,265	0,505	0,130	0,225	0,520	0,255
D3	0,085	0,230	0,505	0,180	0,200	0,505	0,295
D4	0,070	0,260	0,490	0,180	0,135	0,540	0,325
D5	0,030	0,240	0,665	0,065	0,130	0,740	0,130
D6	0,010	0,345	0,615	0,030	0,070	0,795	0,135
D7	0,010	0,295	0,640	0,055	0,090	0,800	0,110
D8	0,000	0,310	0,650	0,040	0,040	0,865	0,095
E1	0,225	0,220	0,380	0,175	0,345	0,380	0,275
E2	0,170	0,255	0,405	0,170	0,260	0,485	0,255
E3	0,165	0,315	0,335	0,185	0,245	0,465	0,290
E4	0,145	0,300	0,340	0,215	0,195	0,475	0,330
E5	0,070	0,445	0,415	0,070	0,185	0,620	0,195
E6	0,040	0,450	0,445	0,065	0,105	0,765	0,130
E7	0,035	0,515	0,395	0,055	0,085	0,815	0,100
E8	0,035	0,520	0,385	0,060	0,055	0,840	0,105

3.7.3 Dados simulados do modelo Logístico-SL

Nesta subseção, consideramos cenários em que os dados são gerados sob o modelo Logístico-SL. Na geração dos bancos de dados, assumimos duas especificações para os parâmetros da distribuição Slash dos efeitos aleatórios. Assumimos $\gamma_i \stackrel{iid}{\sim} SL(0; 0,333; 1,5)$ para os Cenários E1 a E8 e $\gamma_i \stackrel{iid}{\sim} SL(0; 0,667; 3)$ para os Cenários F1 a F8. Em ambos os casos temos que $E(\gamma_i) = 0$ e $V(\gamma_i) = 1$, mas as distribuições diferem quanto ao peso de suas caudas. Isto pode ser visto na Figura 35 da Subseção 3.6.3. Os cenários também diferem quanto aos valores escolhidos para k , n_i e n_{ij} na geração dos dados. Os cenários analisados estão descritos na Tabela 3.14.

Tabela 3.14: Organização dos cenários gerados sob modelo Logístico-N

	$n_i = 5$			$n_i = 10$	
	k	$n_{ij} = 2$	$n_{ij} = 4$	$n_{ij} = 2$	$n_{ij} = 4$
$\gamma_i \stackrel{iid}{\sim} SL(0; 0,333; 1,5)$	50	F1	F2	F3	F4
	100	F5	F6	F7	F8
$\gamma_i \stackrel{iid}{\sim} SL(0; 0,667; 3)$	50	G1	G2	G3	G4
	100	G5	G6	G7	G8

Em cada cenário, ajustamos os modelos Logístico-N (Modelo M1), Logístico-T estimando ν (Modelo M2), Logístico-SL estimando ν (Modelo M3) e Logístico-SL com ν sendo conhecido (Modelo M4).

3.7.3.1 A qualidade das estimativas dos parâmetros

Apresentamos, na Tabela 3.19, os resultados de viés e EQM para β_0 , β_1 , σ , ν e $V(\gamma_i)$ ao ajustarmos os Modelos M1 a M4 a dados dos Cenários F1 a F8. Estes são os cenários em que a distribuição dos efeitos aleatórios possui caudas mais pesadas e, por isso, é onde esperamos os piores resultados do modelo Logístico-N. Comparando os modelos quanto a estimação dos efeitos fixos, notamos que todos os modelos fornecem estimativas com valores muito próximos para EQM. Os modelos que assumem distribuição Slash para os efeitos aleatórios, Modelos M3 e M4, fornecem estimativas menos viesadas para β_1 do que os demais modelos nos Cenários F3, F4 e F8, que são cenários em que $n_i = 10$. Ao compararmos os Modelos M3 e M4 quanto a estimação de σ^2 , temos que o Modelo M4 fornece estimativas menos viesadas e com menor EQM. Observando os resultados para $V(\gamma_i)$, as estimativas com menor viés são obtidas quando o Modelo M2 é ajustado, seguido pelo Modelo M4, enquanto que os menores valores de EQM são obtidos sob os Modelos M1 (principalmente, quando $k = 50$) e M4 (quando $k = 100$). Como já havíamos notado, sob o modelo Logístico-N (Modelo M1), as estimativas para $V(\gamma_i)$ possuem viés alto e negativo quando os efeitos aleatórios são gerados de distribuições com caudas bem mais pesadas do que a distribuição Normal. O modelo Logístico-T fornece estimativas para $V(\gamma_i)$ com valor alto de EQM quando estimamos o número de graus de liberdade ν .

Tabela 3.15: Viés e EQM dos parâmetros de todos os modelos nos Cenários F1 a F8

Cenário		Modelo							
		M1		M2		M3		M4	
		Viés	EQM	Viés	EQM	Viés	EQM	Viés	EQM
F1	β_0	-0,015	0,086	-0,014	0,085	-0,012	0,085	-0,021	0,086
	β_1	0,046	0,192	0,049	0,192	0,054	0,194	0,055	0,194
	σ^2	-	-	-	-	0,249	0,139	0,042	0,031
	ν	-	-	-	-	5,766	58,370	-	-
	$V(\gamma_i)$	-0,201	0,169	0,023	0,470	-0,078	0,304	0,125	0,283
F2	β_0	0,013	0,041	0,015	0,040	0,018	0,041	0,013	0,040
	β_1	0,011	0,083	0,015	0,084	0,017	0,085	0,017	0,084
	σ^2	-	-	-	-	0,170	0,073	0,003	0,013
	ν	-	-	-	-	5,633	62,119	-	-
	$V(\gamma_i)$	-0,267	0,127	-0,064	0,228	-0,139	0,171	0,008	0,113
F3	β_0	0,017	0,040	0,019	0,039	0,023	0,039	0,016	0,039
	β_1	-0,009	0,073	-0,007	0,073	-0,004	0,074	-0,004	0,074
	σ^2	-	-	-	-	0,175	0,072	0,009	0,017
	ν	-	-	-	-	5,843	64,538	-	-
	$V(\gamma_i)$	-0,254	0,146	0,000	0,395	-0,101	0,273	0,025	0,149
F4	β_0	-0,007	0,028	0,004	0,026	0,007	0,027	0,006	0,026
	β_1	-0,016	0,039	-0,012	0,039	-0,010	0,039	-0,010	0,039
	σ^2	-	-	-	-	0,132	0,056	0,021	0,014
	ν	-	-	-	-	3,623	40,603	-	-
	$V(\gamma_i)$	-0,212	0,112	0,063	0,294	-0,013	0,186	0,062	0,130
F5	β_0	-0,014	0,038	-0,007	0,035	-0,012	0,037	-0,017	0,037
	β_1	0,023	0,086	0,005	0,079	0,032	0,087	0,035	0,088
	σ^2	-	-	-	-	0,173	0,068	0,022	0,016
	ν	-	-	-	-	3,427	28,638	-	-
	$V(\gamma_i)$	-0,236	0,124	0,015	0,320	-0,122	0,166	0,065	0,145
F6	β_0	-0,012	0,021	-0,011	0,021	-0,008	0,021	-0,012	0,020
	β_1	0,008	0,036	0,011	0,036	0,014	0,037	0,015	0,037
	σ^2	-	-	-	-	0,159	0,056	0,020	0,010
	ν	-	-	-	-	3,546	34,176	-	-
	$V(\gamma_i)$	-0,231	0,098	0,019	0,282	-0,102	0,134	0,058	0,090
F7	β_0	-0,015	0,018	-0,012	0,018	-0,008	0,018	-0,013	0,018
	β_1	0,025	0,046	0,028	0,046	0,032	0,046	0,033	0,047
	σ^2	-	-	-	-	0,143	0,049	0,019	0,009
	ν	-	-	-	-	3,181	29,541	-	-
	$V(\gamma_i)$	-0,233	0,091	0,047	0,240	-0,095	0,100	0,057	0,078
F8	β_0	-0,011	0,016	-0,003	0,015	0,000	0,015	-0,003	0,016
	β_1	-0,005	0,026	-0,002	0,026	0,000	0,026	0,000	0,026
	σ^2	-	-	-	-	0,124	0,039	0,018	0,006
	ν	-	-	-	-	2,901	29,280	-	-
	$V(\gamma_i)$	-0,227	0,083	0,012	0,175	-0,061	0,105	0,052	0,058

Na Figura 60, apresentamos os *boxplots* das estimativas de ν , obtidas quando os Modelos M2 e M3 são ajustados nos Cenários F1 a F8.

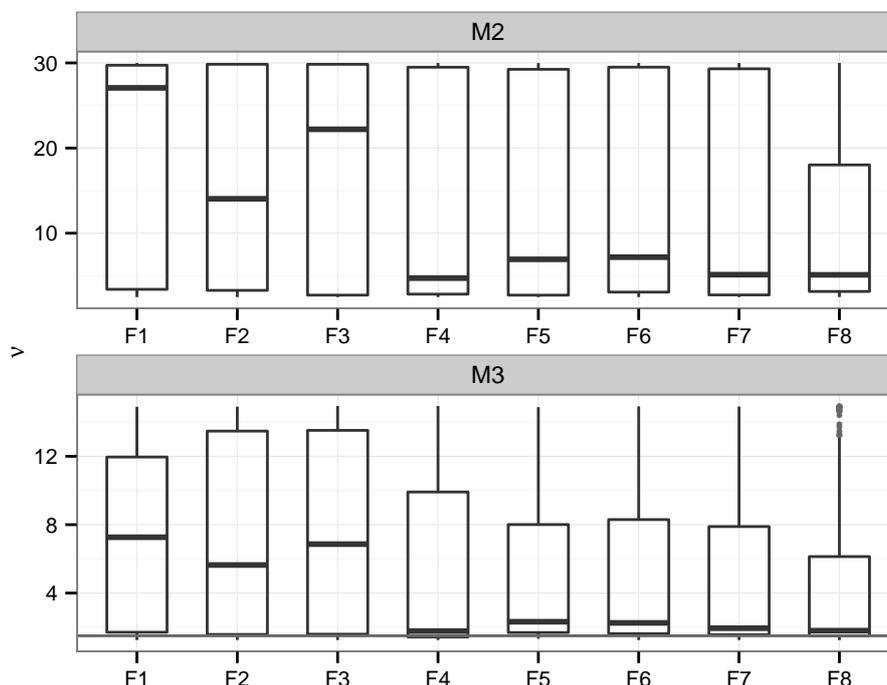


Figura 60: *Boxplots* das estimativas fornecidas para ν e valor real (linha horizontal), sob os Modelos M2 e M4, nos Cenários F1 a F8.

Observamos que, em todos os cenários, sob os dois modelos, o parâmetro ν tende a ser superestimado em muitos bancos de dados. Estimativas mais baixas de ν , que indicam que as caudas da distribuição dos efeitos aleatórios são pesadas, são observadas mais frequentemente nos cenários em que $k = 100$ e no cenário com $k = 50$, $n_i = 10$ e $n_{ij} = 4$, isto é, no cenário mais informativo dentre os que possuem 50 *clusters*.

Na Figura 61, são apresentados os histogramas das estimativas de ν para o Modelo M3, nos Cenários F1 a F8, com o objetivo de mostrar o comportamento da distribuição do estimador de ν no modelo em que a distribuição do efeito aleatório é especificada corretamente. Os histogramas das estimativas de ν no Modelo M2 podem ser encontrados no Apêndice F. Da Figura 61, notamos que as estimativas mais altas de ν são obtidas nos cenários com poucos *clusters* ($k = 50$). Nos cenários em que $k = 100$, há uma melhora na estimação de ν e as estimativas se concentram em valores pequenos, próximos ao valor real $\nu = 1,5$.

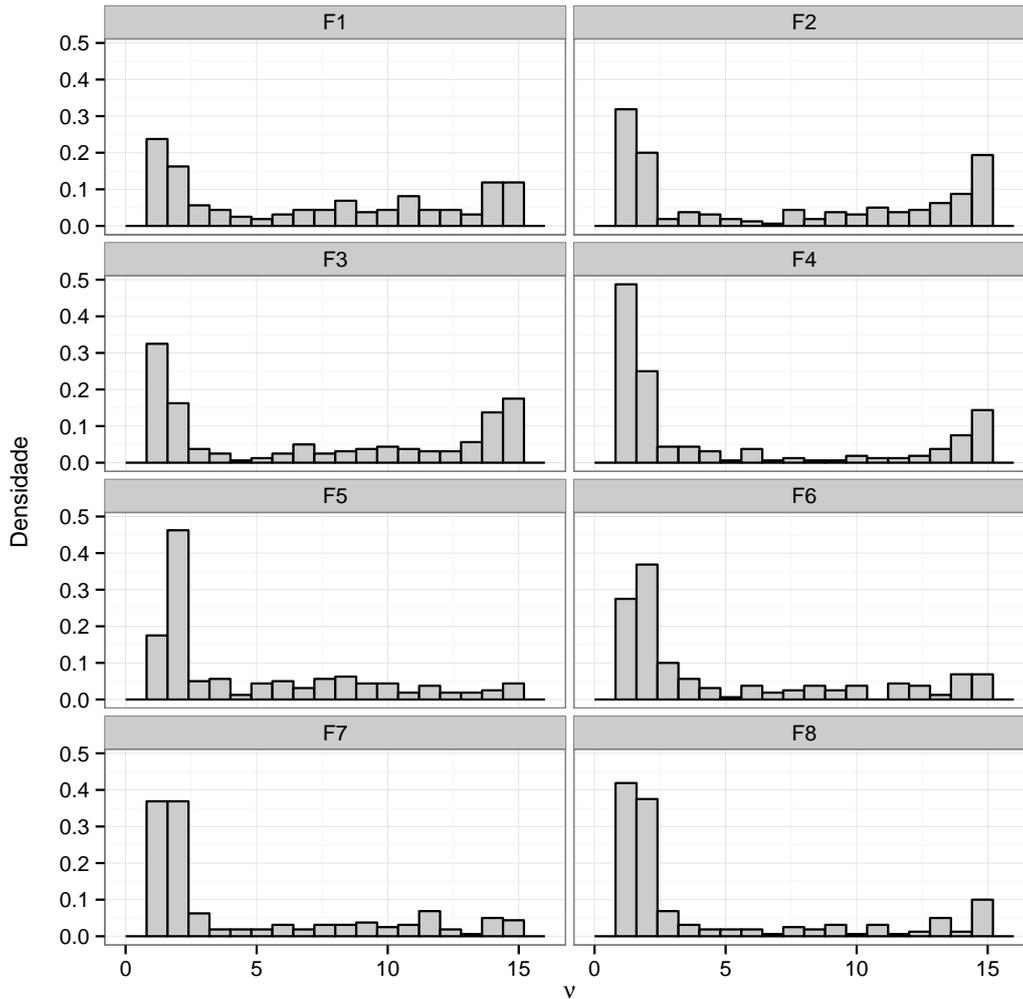


Figura 61: Histogramas das estimativas fornecidas para ν , sob o Modelo M3, nos Cenários F1 a F8.

A Figura 62 apresenta os *boxplots* das estimativas obtidas para a variância dos efeitos aleatórios, sob os Modelos M1 a M4, nos Cenários F1 a F8. O Modelo M4 faz o melhor papel na estimação de $V(\gamma_i)$ pois é o único modelo em que a distribuição das estimativas está centrada no valor real em todos os cenários. Além disto, as estimativas obtidas para este modelo apresentam pouca variabilidade. No Modelo M1, vemos a $V(\gamma_i)$ é subestimada, pois as estimativas estão, em geral, concentradas em valores abaixo do valor real e apresentam pequena variabilidade. Notamos da Tabela 3.15 que o Modelo Logístico-T (M2) foi aquele cujas estimativas apresentaram menor viés. Apesar disto, os *boxplots* revelam que as distribuições das estimativas estão sempre centradas abaixo do valor real e existe muita variabilidade nas estimativas, ocorrendo muitas estimativas que podem ser consideradas *outliers* positivos. Aparentemente, o baixo viés no Modelo M2 é consequência de um efeito de compensação entre as estimativas, em sua maioria, abaixo do valor real e os valores atípicos muito acima do valor real. Sob o Modelo M3, que assume corretamente a distribuição Slash para os efeitos aleatórios e estima ν , temos que a estimação de $V(\gamma_i)$ é ruim, pois as estimativas apresentam grande variabilidade e a distribuição das estimativas está centrada distante do valor real em todos os cenários.

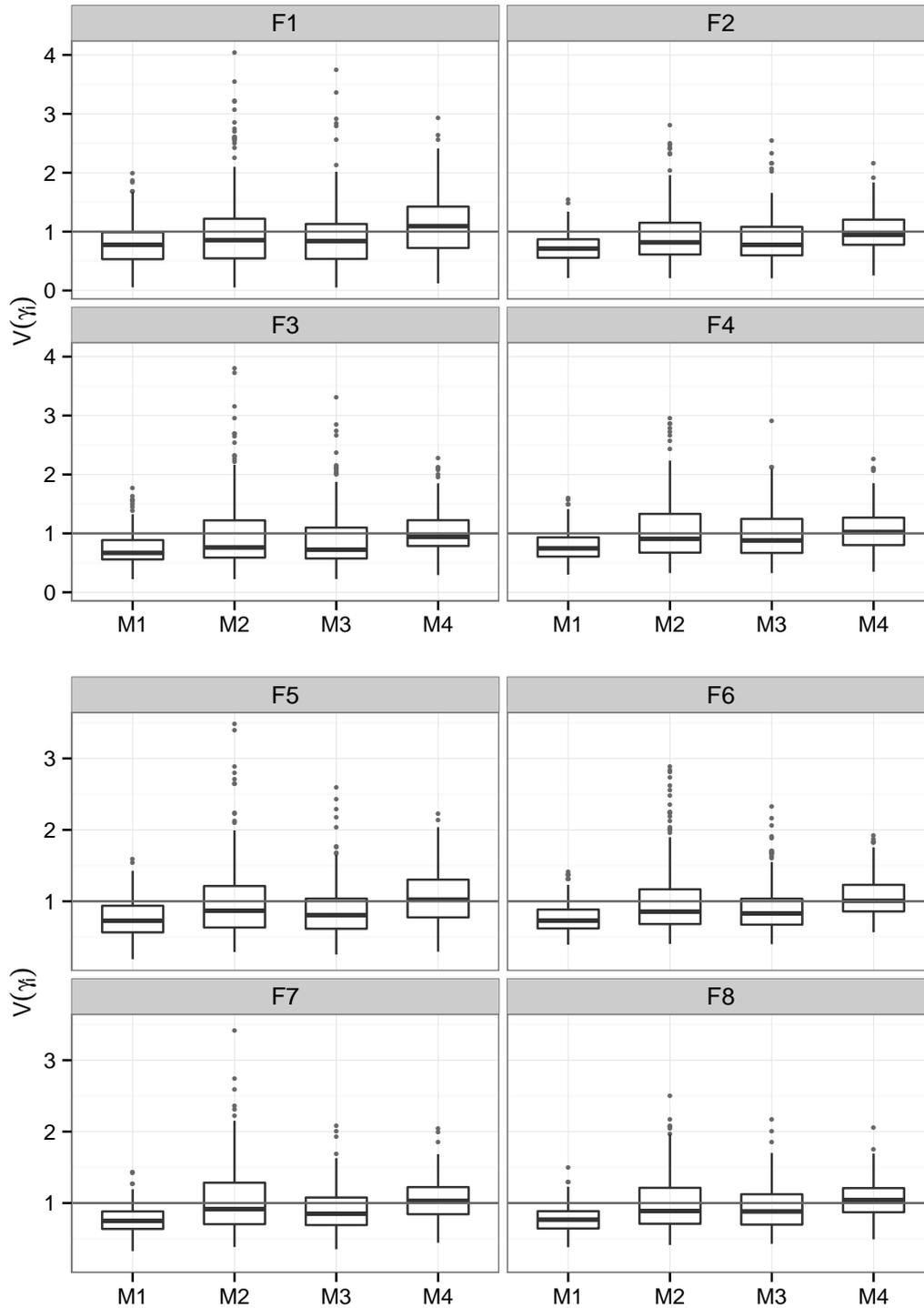


Figura 62: *Boxplots* das estimativas fornecidas para a variância dos efeitos aleatórios e valor real (linha horizontal), sob todos os modelos, nos Cenários F1 a F8.

Analisamos agora as estimativas obtidas nos cenários em os efeitos aleatórios são gerados de uma distribuição Slash com caudas mais pesadas do que a distribuição Normal, mas não muito pesadas, isto é, assumindo $\nu = 3$. Na Tabela 3.17, apresentamos o viés e o EQM das estimativas obtidas para os parâmetros dos Modelos M1 a M4 nos Cenários

G1 a G8.

Tabela 3.16: Viés e EQM dos parâmetros de todos os modelos nos Cenários G1 a G8

Cenário		Modelo							
		M1		M2		M3		M4	
		Viés	EQM	Viés	EQM	Viés	EQM	Viés	EQM
G1	β_0	-0,019	0,094	-0,017	0,093	-0,016	0,093	-0,019	0,093
	β_1	0,025	0,180	0,027	0,180	0,030	0,181	0,026	0,180
	σ^2	-	-	-	-	0,059	0,111	-0,019	0,074
	ν	-	-	-	-	5,433	53,938	-	-
	$V(\gamma_i)$	-0,058	0,157	0,199	0,698	0,049	0,293	-0,028	0,167
G2	β_0	0,002	0,048	0,002	0,048	0,004	0,048	0,003	0,048
	β_1	-0,010	0,083	-0,009	0,083	-0,009	0,083	-0,010	0,083
	σ^2	-	-	-	-	0,050	0,085	-0,022	0,042
	ν	-	-	-	-	5,469	58,704	-	-
	$V(\gamma_i)$	-0,064	0,091	0,077	0,241	0,012	0,127	-0,033	0,096
G3	β_0	-0,016	0,048	-0,017	0,048	-0,014	0,048	-0,016	0,048
	β_1	0,033	0,080	0,033	0,080	0,035	0,080	0,033	0,080
	σ^2	-	-	-	-	0,092	0,090	-0,004	0,038
	ν	-	-	-	-	6,070	64,583	-	-
	$V(\gamma_i)$	-0,041	0,080	0,117	0,271	0,045	0,143	-0,006	0,085
G4	β_0	-0,020	0,030	-0,021	0,032	-0,019	0,031	-0,018	0,030
	β_1	0,008	0,037	0,009	0,037	0,010	0,037	0,009	0,037
	σ^2	-	-	-	-	0,090	0,083	0,007	0,034
	ν	-	-	-	-	6,044	67,777	-	-
	$V(\gamma_i)$	-0,026	0,071	0,099	0,200	0,055	0,121	0,012	0,076
G5	β_0	0,004	0,037	0,003	0,037	0,005	0,037	0,003	0,037
	β_1	0,008	0,074	0,009	0,074	0,010	0,074	0,008	0,074
	σ^2	-	-	-	-	0,113	0,095	0,015	0,039
	ν	-	-	-	-	4,768	45,277	-	-
	$V(\gamma_i)$	-0,010	0,079	0,120	0,190	0,041	0,098	0,022	0,088
G6	β_0	0,011	0,028	0,011	0,028	0,013	0,028	0,012	0,028
	β_1	-0,010	0,043	-0,010	0,043	-0,009	0,043	-0,010	0,043
	σ^2	-	-	-	-	0,074	0,071	-0,003	0,026
	ν	-	-	-	-	4,693	48,437	-	-
	$V(\gamma_i)$	-0,035	0,055	0,060	0,118	0,023	0,078	-0,004	0,058
G7	β_0	0,001	0,024	0,002	0,023	0,003	0,023	0,002	0,024
	β_1	0,004	0,036	0,004	0,036	0,005	0,036	0,004	0,036
	σ^2	-	-	-	-	0,112	0,072	0,016	0,023
	ν	-	-	-	-	5,641	59,210	-	-
	$V(\gamma_i)$	-0,010	0,047	0,106	0,164	0,055	0,079	0,025	0,052
G8	β_0	-0,001	0,018	0,000	0,018	0,001	0,018	0,001	0,018
	β_1	-0,001	0,021	-0,001	0,021	0,000	0,021	-0,001	0,021
	σ^2	-	-	-	-	0,073	0,053	-0,002	0,016
	ν	-	-	-	-	5,019	53,431	-	-
	$V(\gamma_i)$	-0,036	0,035	0,033	0,060	0,024	0,056	-0,003	0,036

Da Tabela 3.16, percebemos que não há diferenças entre os modelos ajustados no que diz respeito à estimação dos efeitos fixos. As estimativas para β_0 e β_1 sob todos os modelos são muito similares, com valores muito próximos de viés e EQM. Isto já havia sido percebido anteriormente na Subseção 3.7.2 e, também, na Tabela 3.15 e esta em acordo com trabalhos anteriormente publicados. Da mesma forma que ocorreu nos Cenários F1 a F8, nos Cenários G1 a G8, as melhores estimativas para σ^2 são obtidas sob o Modelo M4, e não sob o Modelo M3, tanto com relação ao viés quanto em relação ao EQM. Isto é um resultado natural, uma vez que ambos os modelos assumem a distribuição Slash para os efeitos aleatórios, mas o Modelo M4 tem o ganho adicional ao assumir como conhecido o valor real do parâmetro ν . Quando avaliamos as estimativas para a variância da distribuição assumida para os efeitos aleatórios, em geral, sob os Modelos M4 e M3 (nesta ordem) temos os melhores resultados para o viés das estimativas. Além disto, em todos os cenários analisados na Tabela 3.16, sob o Modelo M1, as estimativas da $V(\gamma_i)$ apresentam menor EQM, seguidas por aquelas obtidas sob o Modelo M4. As estimativas da $V(\gamma_i)$ obtidas sob o Modelo M2 têm o maior EQM.

Observamos nas Tabelas 3.15 e 3.16, que quando comparamos cenários em que os bancos de dados têm os valores iguais para n_i e n_{ij} , os cenários com maior número k de *clusters* têm menor EQM para todos os parâmetros e modelos. Comparando apenas cenários com mesmos valores para k e n_i , notamos que sob os Modelos M1, M2 e M4, os cenários com maior valor de n_{ij} têm menor EQM para as estimativas de todos os parâmetros. Sob o Modelo M3, este resultado também foi observado para β_0 , β_1 e σ^2 em todos os cenários avaliados nesta subseção. Ao avaliarmos as estimativas de $V(\gamma_i)$ percebemos que sob o Modelo M3 também existe esta tendência, mas ocorreu uma exceção ao compararmos os Cenários F7 e F8. Se a comparação ocorre entre cenários com valores iguais para k e n_{ij} , vemos que existe a tendência de cenários com maior valor para n_i terem menor EQM para as estimativa de todos os parâmetros sob os Modelos M1, M2 e M4. Sob o Modelo M3, conclusão similar é tirada, exceto para o grau de liberdade ν da distribuição Slash.

Agora, tentamos determinar se, para a estimação dos parâmetros, existe uma melhor estratégia a ser seguida na definição do número k de *clusters*, do número n_i de indivíduos dentro de cada *cluster* e do número n_{ij} de experimentos Bernoulli realizados para cada indivíduo. O objetivo é avaliarmos, por exemplo, se é melhor estratégia considerarmos um número maior de *clusters* ou de indivíduos dentro de cada *cluster*. Para isto, focamos a comparação apenas entre cenários para os quais os bancos de dados gerados possuem a mesma quantidade $k \times n_i \times n_{ij}$ de observações de experimentos Bernoulli e mesma distribuição para os efeitos aleatórios. Inicialmente, considerando fixo o número k de *clusters*, notamos que os cenários com maior valor de n_i e menor valor de n_{ij} fornecem estimativas para β_0 com menor EQM, em todos os modelos e cenários. Esta tendência parece também ocorrer na estimação de outros parâmetros como, por exemplo, na estimação de β_1 em todos os modelos utilizados ou na estimação da $V(\gamma_i)$ nos Modelos M1 e M4. No entanto, esta conclusão não é confirmada em todas as comparações realizadas. Se comparamos apenas os cenários com valores iguais para n_{ij} , vemos que em todos os modelos, geralmente, os cenários com maior valor de k e menor valor de n_i fornecem estimativas com menor EQM para β_0 , ν e $V(\gamma_i)$. Por fim, se comparamos cenários com valor igual de n_i , observamos que em todos os modelos, os cenários com maior valor para k e menor para n_{ij} têm as estimativas com menor EQM para β_0 , ν e $V(\gamma_i)$. Para o parâmetro

σ^2 , as comparações feitas não revelam um padrão de comportamento que possa ajudar na determinação de quais quantidades são mais importantes na qualidade da estimação. Além disso, se as comparações são avaliadas quanto ao viés das estimativas, os resultados obtidos não se mostram conclusivos.

Para compreender melhor o comportamento das estimativas obtidas para os graus de liberdade das distribuições *t*-Student e Slash consideramos os *boxplots* das estimativas de ν , exibidos na Figura 63.

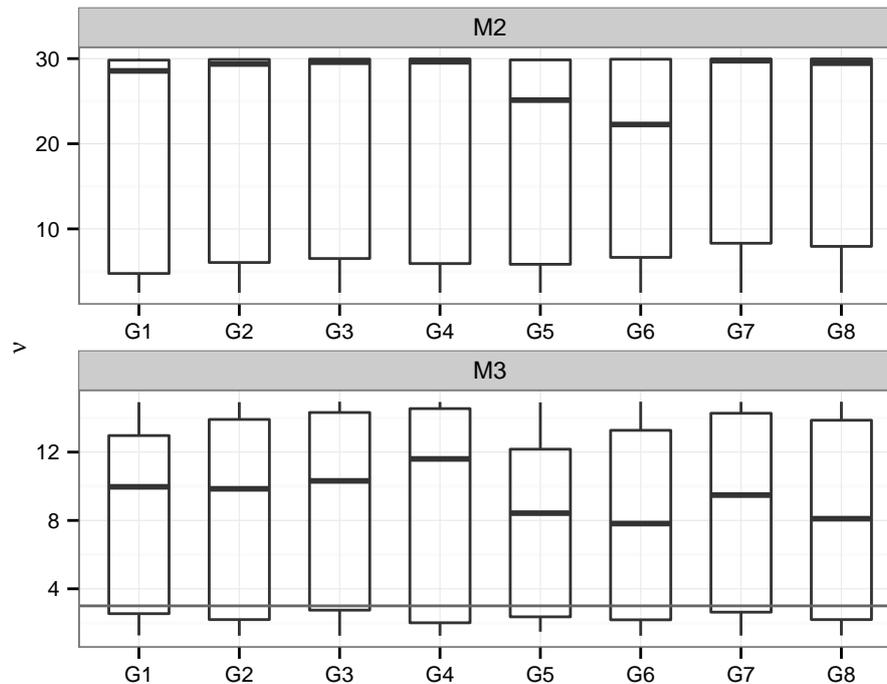


Figura 63: *Boxplots* das estimativas fornecidas para ν e valor real (linha horizontal), sob os Modelos M2 e M4, nos Cenários G1 a G8.

Analisando a Figura 63, percebemos que a estimação de ν é muito ruim tanto sob o Modelo M2 quanto sob o Modelo M3. No Modelo M2, temos que as estimativas de ν tendem a ser muito maiores do que o valor real, como pode ser visto, por exemplo, na mediana da distribuição das estimativas que está acima do valor 20 em todos os cenários avaliados. Além disto, existe variabilidade muito grande nas estimativas para ν no Modelo M2. Note que a cauda inferior do *boxplots* atinge o valor 2,5 e o primeiro quartil está baixo do valor 4. O mesmo ocorre para o Modelo M3, onde vemos que as distribuições das estimativas não se concentraram próximas ao valor real $\nu = 3$ em nenhum cenário.

Os histogramas das estimativas de ν quando ajustamos o Modelo M3 são dados na Figura 64 (ver Apêndice F para o caso do Modelo M2). No texto, nosso foco incide sobre a análise das estimativas fornecidas pelo Modelo M3 pelo fato deste modelo ser especificado corretamente segundo a geração dos bancos de dados. Notamos que nos Cenários G1 a G8, em geral, a distribuição das estimativas de ν é bimodal. Existe uma concentração de estimativas em valores pequenos, que apontam caudas muito pesadas para a distribuição dos efeitos aleatórios e uma segunda moda das estimativas no valor no valor 15, que é o maior valor permitido pelo algoritmo na estimação de ν . Provavelmente

esta segunda moda é causada pela restrição do parâmetro imposta na implementação do modelo. Para valores altos de ν a curva da f.d.p da distribuição Slash é praticamente igual a da distribuição Normal. Nos cenários G5 a G8, em que os bancos de dados são divididos em 100 *clusters*, a primeira moda da distribuição tende a ficar mais próxima do valor real $\nu = 3$.

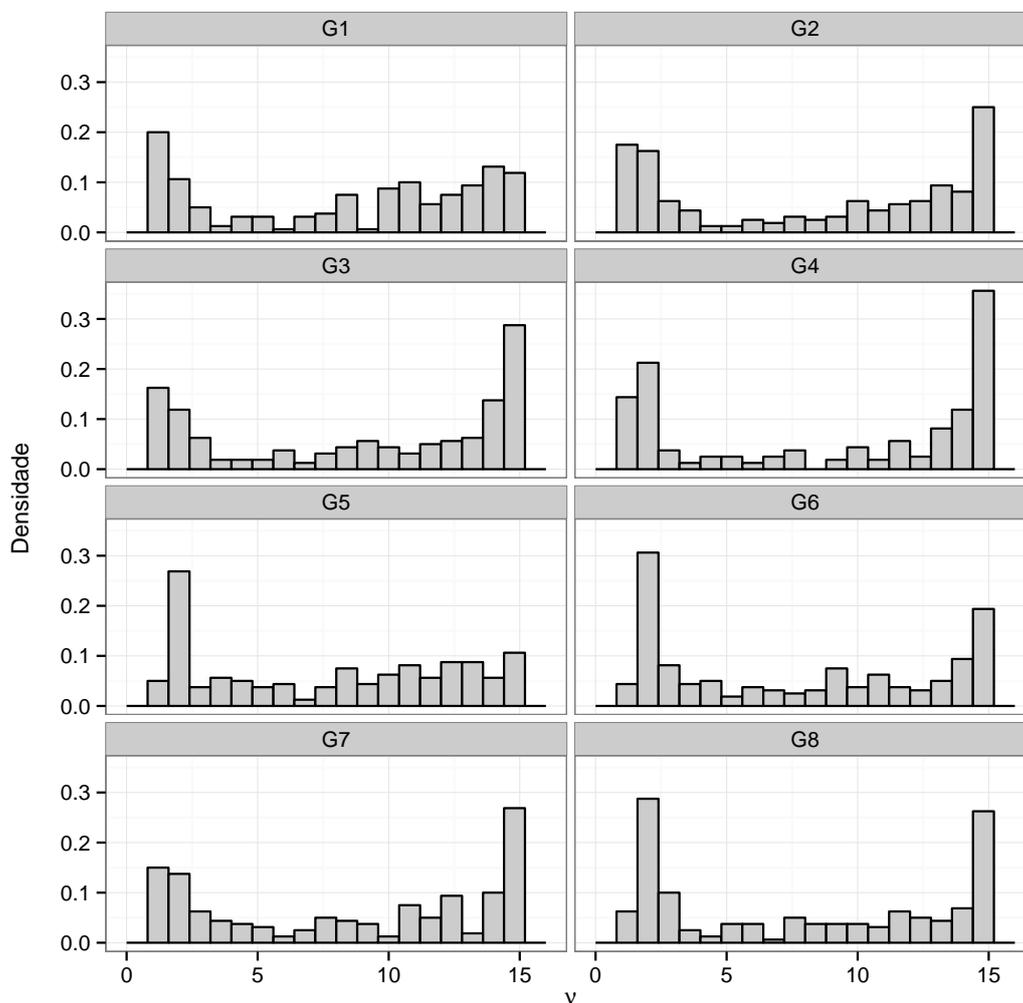


Figura 64: Histogramas das estimativas fornecidas para ν , sob o Modelo M3, nos Cenários G1 a G8.

Na Figura 65, apresentamos os *boxplots* das estimativas fornecidas para $V(\gamma_i)$ sob os Modelos M1 a M4 nos Cenários G1 a G8. Observamos que a distribuição das estimativas da $V(\gamma_i)$ obtidas em todos os modelos, para os Cenários G1 a G8, estão centradas no valor real $V(\gamma_i)$. No entanto, sob os Modelos M2 e M3, as estimativas apresentam maior variabilidade do que os demais modelos, o que nos mostra que estimar o parâmetro ν aumenta a variabilidade das estimativas de $V(\gamma_i)$. Nos Cenários G1 a G8, mais uma vez observamos que sob o Modelo Logístico-T, muitas estimativas são atípicas, sendo maiores do que o valor real. Este comportamento foi observado em todos os cenários analisados. Além disso, percebemos que a distribuições das estimativas para $V(\gamma_i)$ obtidas sob os Modelos M1 e M4 foram muito similares.

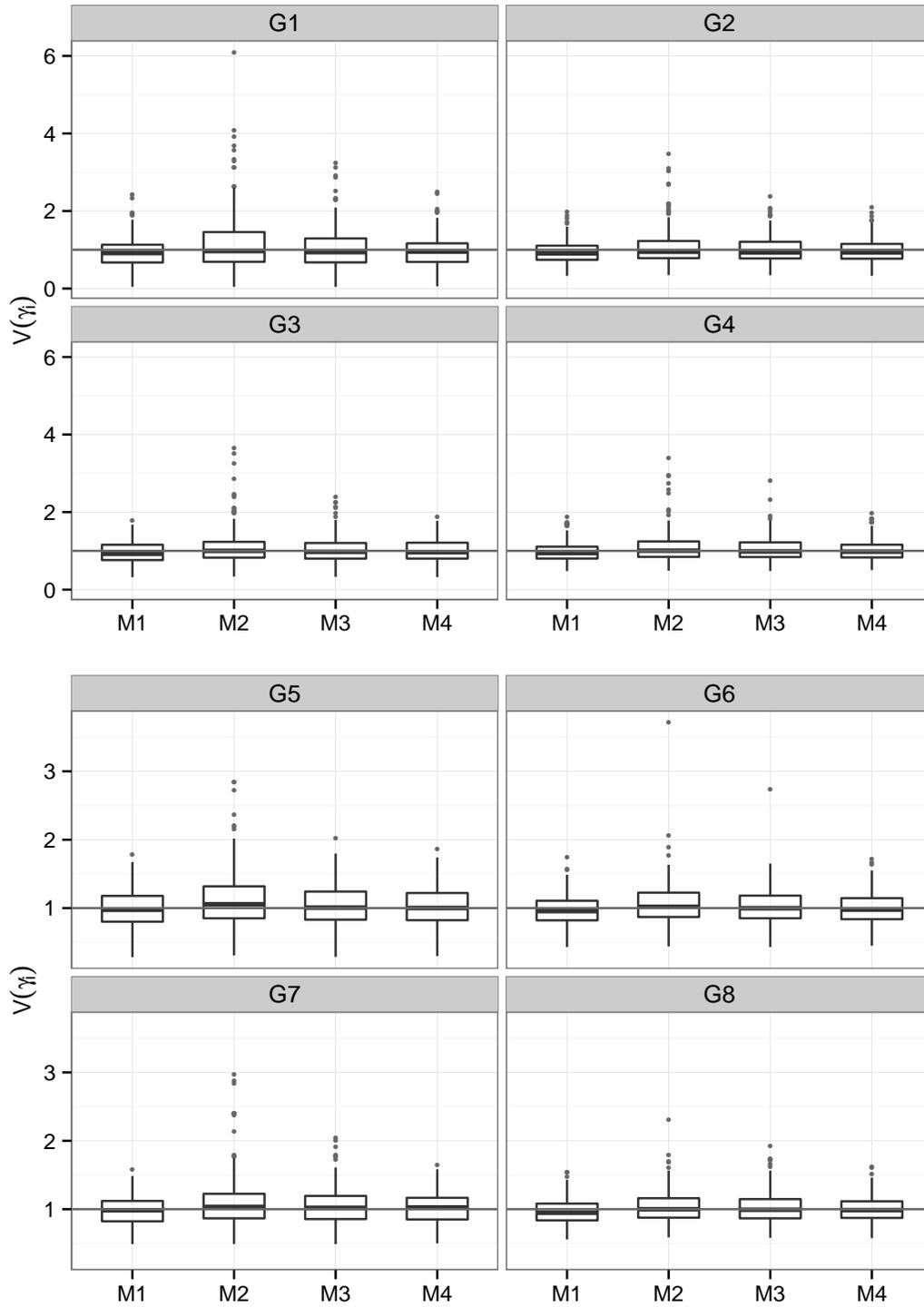


Figura 65: *Boxplots* das estimativas fornecidas para a variância dos efeitos aleatórios e valor real (linha horizontal), sob todos os modelos, nos Cenários G1 a G8.

3.7.3.2 A qualidade das estimativas da razão de chances mediana

Na Tabela 3.17, apresentamos o viés e o EQM para as estimativas da $MOR_{|12|}$ quando os Modelos M1 a M4 foram ajustados nos cenários abordados na nesta subseção. O valor real para $MOR_{|12|}$ é de aproximadamente 2,197 nos Cenários F1 a F8, e 2,51, nos Cenários G1 a G8. Notamos que, nos Cenários F1 a F8, sob o Modelo M2, melhores resultados são obtidos no que diz respeito ao viés e ao EQM das estimativas da $MOR_{|12|}$.

Em geral, para os mesmos cenários, o Modelo M4 é segundo melhor no que diz respeito ao viés, enquanto o Modelo M3 é o segundo melhor quanto ao EQM. Nos Cenários G1 a G8, o Modelo M2 continua fazendo bom papel ao avaliarmos o viés das estimativas, mas perde para os demais modelo no EQM. O Modelo M4, geralmente, é o segundo melhor modelo em termos de viés e o melhor em termos de EQM. Também nos chama a atenção que nestes cenários, as estimativas para $MOR_{|12|}$ fornecidas por todos os modelos têm valores muito próximo de EQM.

Tabela 3.17: Viés e EQM para $MOR_{|12|}$, em todos os modelos, nos Cenários F1 a F8 e G1 a G8

Cenário	Modelo							
	M1		M2		M3		M4	
	Viés	EQM	Viés	EQM	Viés	EQM	Viés	EQM
F1	0,141	0,229	0,031	0,172	0,118	0,210	0,095	0,211
F2	0,062	0,094	-0,048	0,077	0,014	0,082	0,001	0,085
F3	0,077	0,136	-0,037	0,092	0,027	0,110	0,014	0,112
F4	0,132	0,124	-0,006	0,085	0,054	0,094	0,046	0,095
F5	0,101	0,120	-0,032	0,095	0,068	0,108	0,047	0,108
F6	0,109	0,082	-0,009	0,058	0,057	0,065	0,044	0,066
F7	0,107	0,070	-0,019	0,052	0,054	0,058	0,045	0,058
F8	0,114	0,063	-0,007	0,038	0,047	0,042	0,040	0,043
G1	0,010	0,246	-0,085	0,228	-0,011	0,241	-0,021	0,239
G2	0,005	0,136	-0,065	0,141	-0,034	0,136	-0,022	0,135
G3	0,033	0,123	-0,036	0,122	0,005	0,122	0,009	0,119
G4	0,053	0,112	-0,002	0,102	0,023	0,105	0,030	0,106
G5	0,072	0,129	-0,001	0,129	0,048	0,127	0,042	0,124
G6	0,041	0,085	-0,020	0,083	0,008	0,082	0,013	0,081
G7	0,073	0,077	0,014	0,073	0,044	0,075	0,047	0,072
G8	0,041	0,053	-0,009	0,049	0,010	0,050	0,015	0,050

Comparando os cenários que são diferentes apenas quanto ao número k de *clusters*, vemos que, em cenários com maior valor de k , as estimativas para $MOR_{|12|}$ têm menor EQM em todos os modelos. Se a comparação ocorre entre cenários que diferem apenas quanto ao valor de n_i , percebemos que em cenários com maior valor de n_i as estimativas têm menor EQM apenas nos cenários em que $k = 100$. Conclusão similar é obtida ao compararmos os cenários diferentes apenas quanto ao n_{ij} .

Se comparamos os cenários em que o mesmo total de experimentos Bernoulli foram realizados, temos que os resultados são um pouco contraditórios, mas é possível observarmos algumas tendências para todos os modelos. Por exemplo, se fixamos o valor de k , os

cenários com maior valor de n_i e menor de n_{ij} tendem a fornecer estimativas com menor EQM para $MOR_{|12|}$. Algo parecido é observado para comparações fixando o valor de n_i . Neste caso, os cenários com maior valor de k e menor de n_{ij} também tendem a fornecer menores valores de EQM para as estimativas de $MOR_{|12|}$.

Nas Figuras 66 e 67 apresentamos os *boxplots* das estimativas da $MOR_{|12|}$ nos Cenários F1 a F8 e G1 a G8, respectivamente.

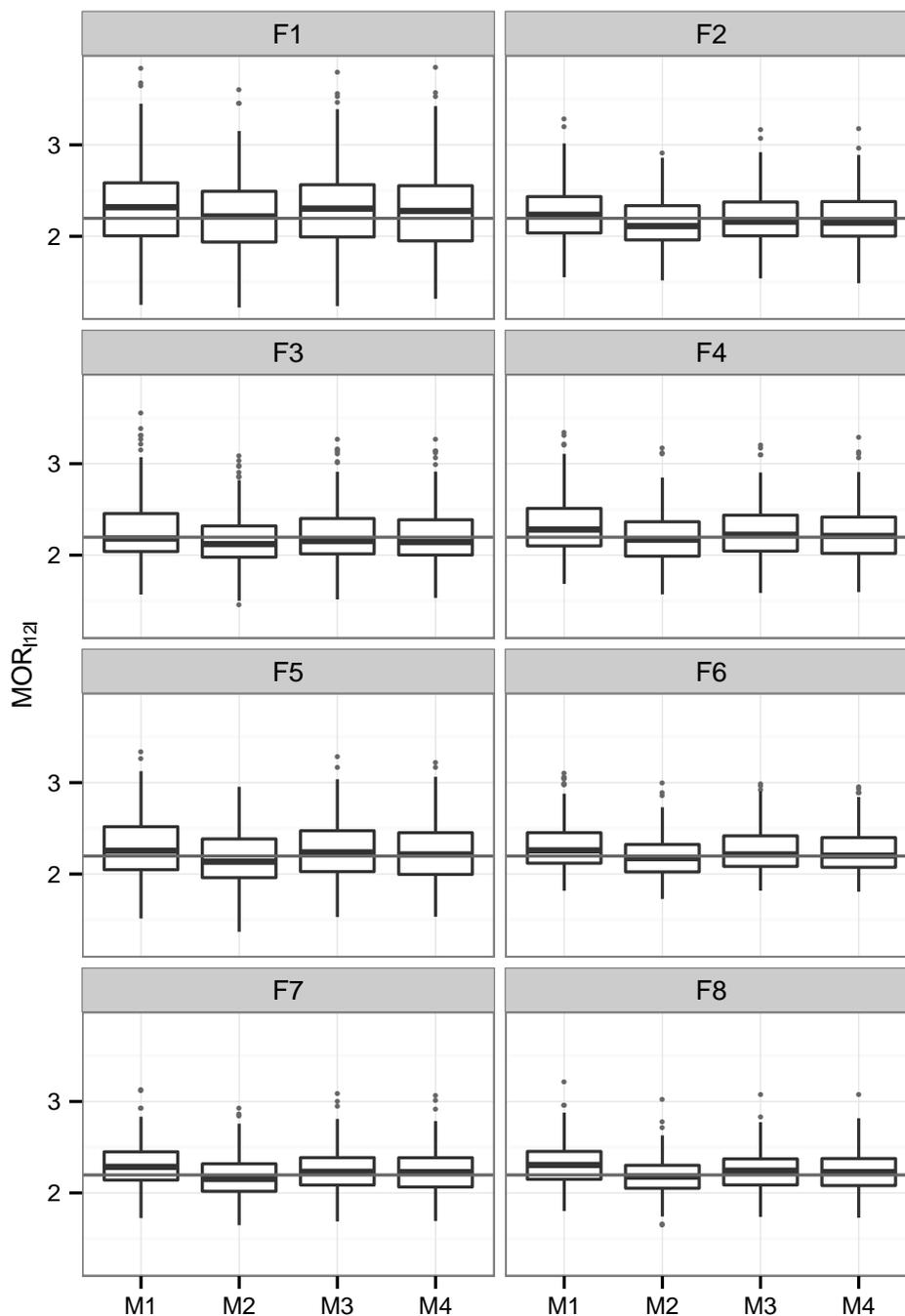


Figura 66: *Boxplots* das estimativas de $MOR_{|12|}$ e valor real (linha horizontal), sob todos os modelos, nos Cenários F1 a F8.

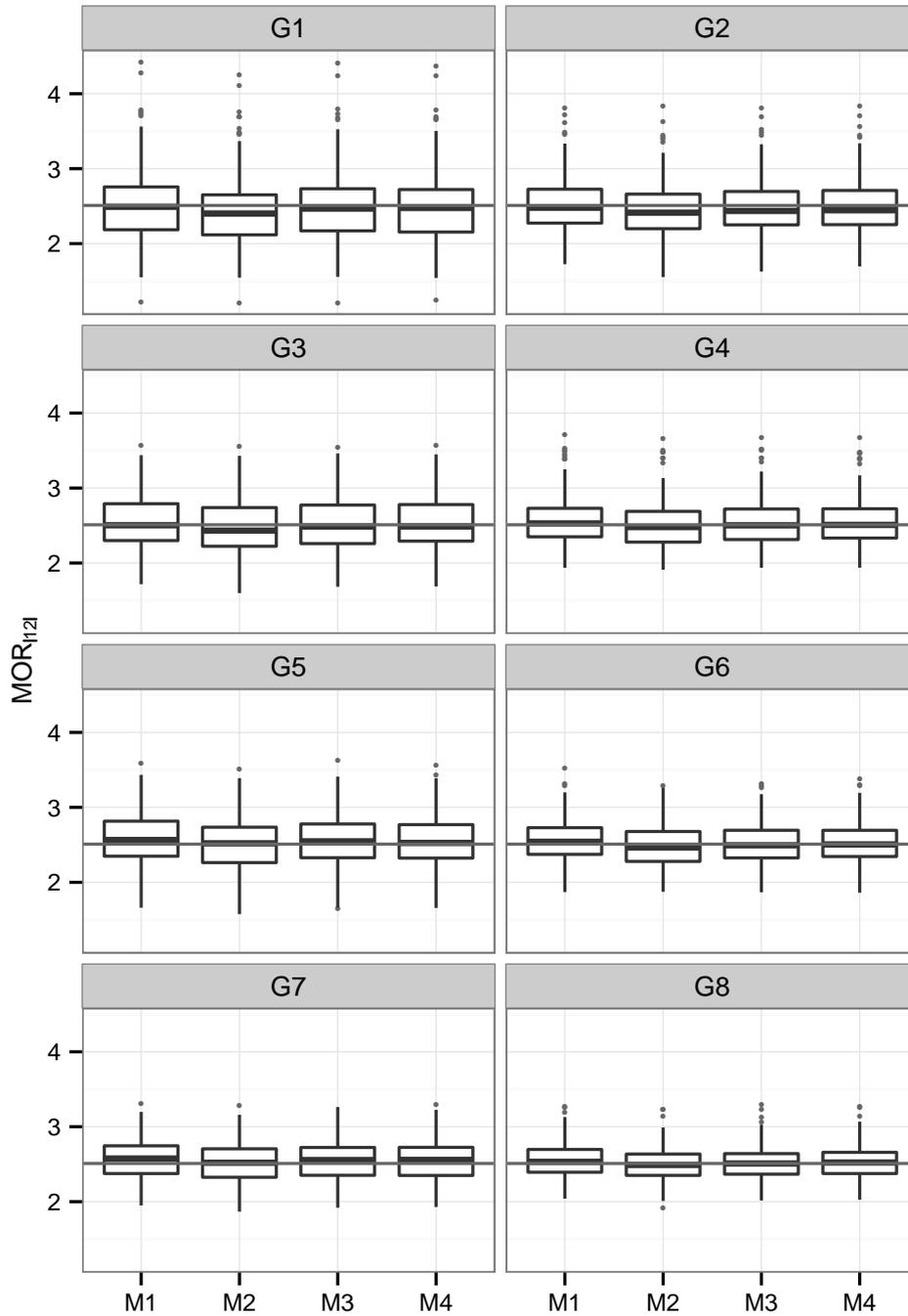


Figura 67: *Boxplots* das estimativas de $MOR_{12|}$ e valor real (linha horizontal), sob todos os modelos, nos Cenários G1 a G8.

As Figuras 66 e 67, nos permitem fazer uma melhor análise das distribuições das estimativas para $MOR_{12|}$ obtidas quando os Modelos M1 a M4 foram ajustados, nos cenários em que os dados são gerados do modelo Logístico-SL. Observamos que, nos cenários em que o valor real de ν é 1,5, sob o Modelo M1, as distribuições das estimativas

estão centradas um pouco acima do valor real. Em geral, as distribuições das estimativas da $MOR_{|12|}$ obtidas sob os Modelos M2 a M4 estão centradas muito próximas do valor real, mas sob os Modelos M3 e M4 percebemos que as medianas das estimativas estão mais próximas do valor real que aquelas obtidas sob o Modelo M2.

3.7.3.3 Comparação de modelos

Na Tabela 3.18, apresentamos a proporção de vezes em que cada modelo é escolhido como sendo o que melhor se ajustou as dados, segundo o critério o AIC, nos Cenários F1 a F8 e G1 a G8. Conduzimos a seleção dos modelos inicialmente comparando todos os modelos (seleção 1) e depois comparando apenas os modelos em que não é passada nenhuma informação sobre o parâmetro ν , ou seja, em que ν não é fixado no valor real.

Tabela 3.18: Proporção de escolha de cada modelo, usando o critério AIC, nos cenários gerados sob o modelo Logístico-SL

Cenário	Seleção 1				Seleção 2		
	M1	M2	M3	M4	M1	M2	M3
F1	0,235	0,140	0,155	0,470	0,385	0,345	0,270
F2	0,140	0,230	0,195	0,435	0,265	0,415	0,320
F3	0,100	0,245	0,165	0,490	0,265	0,470	0,265
F4	0,085	0,315	0,180	0,420	0,145	0,550	0,305
F5	0,060	0,385	0,115	0,440	0,135	0,670	0,195
F6	0,020	0,480	0,085	0,415	0,055	0,760	0,185
F7	0,020	0,485	0,095	0,400	0,050	0,815	0,135
F8	0,015	0,625	0,035	0,325	0,015	0,895	0,090
G1	0,200	0,250	0,180	0,370	0,340	0,380	0,280
G2	0,120	0,380	0,190	0,310	0,210	0,515	0,275
G3	0,135	0,330	0,215	0,320	0,215	0,505	0,280
G4	0,060	0,360	0,225	0,355	0,125	0,505	0,370
G5	0,055	0,455	0,125	0,365	0,115	0,695	0,190
G6	0,050	0,630	0,110	0,210	0,055	0,820	0,125
G7	0,055	0,560	0,125	0,260	0,075	0,770	0,155
G8	0,030	0,630	0,105	0,235	0,055	0,825	0,120

Da Tabela 3.18 concluímos que, ao compararmos todos os modelos ajustados, o Modelo M4 foi escolhido como o melhor modelo mais frequentemente nos Cenários F1 a F5 e G1. Nos demais cenários, principalmente para os cenários F6 a F8, o Modelo M4 ainda é selecionado como o melhor em uma grande proporção de vezes, mas nestes cenários, ele não supera o Modelo M2. Ressaltamos que no cenário com maior informação (F8) M2 é o melhor na maioria das vezes (62,5%). Os modelos M1 e M3 são, em geral, escolhidos com frequências pequenas em todos os cenários. Isto é esperado para o Modelo M1 que considera normalidade para os efeitos aleatórios, mas não é para o Modelo M3, visto que este assume corretamente a distribuição Slash para modelar o comportamento dos efeitos aleatórios. Quando retiramos o Modelo M4 da comparação, notamos a proporção de escolha do Modelo M2 é sempre muito alta, mas é ainda maior nos cenários em que $k = 100$.

3.7.4 Dados simulados do modelo Logístico-NC

Aqui consideramos cenários em que os dados são gerados do modelo Logístico-NC. Na geração dos bancos de dados, assumimos $k = 100$, $n_i = 10$, $n_{ij} = 4$ e três especificações para os parâmetros da distribuição Normal Contaminada, a saber,

Cenário H: $\gamma_i \stackrel{iid}{\sim} NC(0; 0,714; 0,1; 0,2)$;

Cenário I: $\gamma_i \stackrel{iid}{\sim} NC(0; 0,555; 0,2; 0,2)$;

Cenário J: $\gamma_i \stackrel{iid}{\sim} NC(0; 0,385; 0,4; 0,2)$.

A comparação entre as f.d.p das distribuições assumidas para os efeitos aleatórios pode ser vista na Figura 43 da Subseção 3.6.4. A variância teórica da distribuição dos efeitos aleatórios $V(\gamma_i)$ é igual a 1 nas três especificações consideradas. Para a razão de chances mediana $MOR_{|12|}$, o valor real é aproximadamente 2,421 no Cenário H, 2,352 no Cenário I e 2,353 no Cenário J.

Em todos os cenários, ajustamos os modelos Logístico-N (Modelo M1), Logístico-T (Modelo M2), Logístico-SL (Modelo M3), Logístico-NC com $\nu_2 = 0,1$ (Modelo M4), Logístico-NC com $\nu_2 = 0,2$ (Modelo M5) e Logístico-NC com $\nu_2 = 0,3$ (Modelo M6). Nos modelos Logístico-T e Logístico-SL, o valor do parâmetro ν é estimado uma vez que não conhecemos o valor correto. Já no modelo Logístico-NC, consideramos três valores diferentes para ν_2 para avaliar o efeito da especificação deste parâmetro. No Modelo M5 especificamos corretamente $\nu_2 = 0,2$, enquanto nos Modelos M4 e M5 especificamos um valor um pouco acima ou abaixo do correto. Note que a distribuição Normal Contaminada é a formada pela mistura de duas distribuições Normais com mesma locação. A escolha do valor $\nu_2 = 0,2$ implica que uma das distribuições componentes da mistura tem variância igual a 5 vezes variância da outra distribuição. Para as especificações $\nu_2 = 0,1$ e $\nu_2 = 0,3$, este fator multiplicativo é alterado, respectivamente, para 10 e 3,333 vezes.

3.7.4.1 A qualidade das estimativas dos parâmetros e da razão de chances mediana

Na Tabela 3.19 apresentamos os resultados de viés e EQM das estimativas obtidas para β_0 , β_1 , σ^2 , ν_1 , $V(\gamma_i)$ e $MOR_{|12|}$, sob os Modelos M1 a M6, nos Cenários H, I e J. Observando o viés das estimativas dos efeitos fixos, notamos que não existe padrão de comportamento entre os modelos. Comparando entre os cenários, os cenários com maior valor real de ν_1 têm maior viés para β_0 e menor para β_1 . Comparando apenas as estimativas obtidas sob os modelos M4 a M6, vemos que sob o modelo correto (Modelo M5) as estimativas são menos viesadas para σ^2 nos cenários em que os efeitos aleatórios foram gerados de distribuições com caudas mais pesadas, isto é, nos Cenários I e J. Para o parâmetro ν_1 , sob o Modelo M6 obtivemos o menor viés em todos os cenários. Além disso, vemos que da mesma forma que ocorreu em todos os cenários anteriores, quando os efeitos aleatórios são gerados de distribuições com caudas pesadas, sob o Modelo Logístico-N temos estimativas com viés negativo para $V(\gamma_i)$ e viés positivo para $MOR_{|12|}$. Sob o Modelo M4 as estimativas para $V(\gamma_i)$ têm menor viés nos Cenários H e I, enquanto isto ocorreu sob o Modelo M5 apenas no Cenário J. Para $MOR_{|12|}$, observamos que sob o Modelo M2, o viés das estimativas sempre foi o menor dentre os modelos comparados.

Ao analisarmos o EQM das estimativas para σ^2 e ν_1 nos Modelos M4 a M6, temos que a especificação correta do parâmetro ν_2 não trouxe ganho na estimação destes parâmetros. Analisando o EQM das estimativas de $V(\gamma_i)$ vemos que, nos Cenários H, I e J, as melhores estimativas são obtidas sob os Modelos M1, M5 e M6. Já para $MOR_{|12|}$, sob o modelo M2 sempre temos as estimativas com menor EQM.

Tabela 3.19: Viés e EQM dos parâmetros de todos os modelos nos Cenários H, I e J

		Viés					
Cenário		M1	M2	M3	M4	M5	M6
H	β_0	0,000	0,004	0,007	0,003	0,003	0,002
	β_1	-0,013	-0,011	-0,010	-0,012	-0,012	-0,012
	σ^2	-	-	-	0,144	0,147	0,145
	ν_1	-	-	-	-0,072	-0,055	-0,037
	$V(\gamma_i)$	-0,072	0,080	0,058	-0,030	-0,059	-0,064
	$MOR_{ 12 }$	0,084	-0,007	0,022	0,051	0,063	0,069
I	β_0	-0,010	-0,003	-0,001	-0,005	-0,004	-0,004
	β_1	0,008	0,009	0,010	0,009	0,009	0,009
	σ^2	-	-	-	0,190	0,158	0,169
	ν_1	-	-	-	-0,138	-0,088	-0,045
	$V(\gamma_i)$	-0,105	0,104	0,080	-0,036	-0,082	-0,097
	$MOR_{ 12 }$	0,112	-0,008	0,028	0,037	0,045	0,065
J	β_0	-0,011	-0,005	-0,004	-0,009	-0,005	-0,003
	β_1	-0,005	-0,003	-0,002	-0,003	-0,003	-0,003
	σ^2	-	-	-	0,293	0,230	0,249
	ν_1	-	-	-	-0,271	-0,171	-0,122
	$V(\gamma_i)$	-0,063	0,202	0,152	0,030	-0,028	-0,051
	$MOR_{ 12 }$	0,164	0,027	0,077	0,053	0,063	0,089
		EQM					
H	β_0	0,015	0,013	0,014	0,014	0,015	0,014
	β_1	0,021	0,021	0,021	0,021	0,020	0,021
	σ^2	-	-	-	0,080	0,087	0,081
	ν_1	-	-	-	0,011	0,016	0,020
	$V(\gamma_i)$	0,045	0,142	0,096	0,056	0,046	0,045
	$MOR_{ 12 }$	0,070	0,053	0,055	0,067	0,070	0,068
I	β_0	0,017	0,016	0,016	0,016	0,016	0,016
	β_1	0,018	0,018	0,018	0,018	0,018	0,018
	σ^2	-	-	-	0,102	0,082	0,088
	ν_1	-	-	-	0,032	0,032	0,039
	$V(\gamma_i)$	0,044	0,195	0,134	0,056	0,045	0,044
	$MOR_{ 12 }$	0,064	0,041	0,045	0,053	0,047	0,054
J	β_0	0,014	0,013	0,013	0,014	0,013	0,013
	β_1	0,018	0,018	0,018	0,018	0,018	0,018
	σ^2	-	-	-	0,192	0,137	0,122
	ν_1	-	-	-	0,107	0,078	0,064
	$V(\gamma_i)$	0,049	0,285	0,173	0,068	0,051	0,048
	$MOR_{ 12 }$	0,098	0,062	0,068	0,072	0,067	0,070

A Figura 68 apresenta os *boxplots* das estimativas de ν sob os Modelos M2 e M3. Vemos que em ambos os modelos as estimativas se concentram em torno de valores pequenos para os cenários onde as caudas da distribuição dos efeitos aleatórios são mais pesadas. No entanto, em todos os casos, muitas estimativas altas para ν são obtidas e, em alguns casos, há muitas estimativas atípicas que podem aumentar o viés.

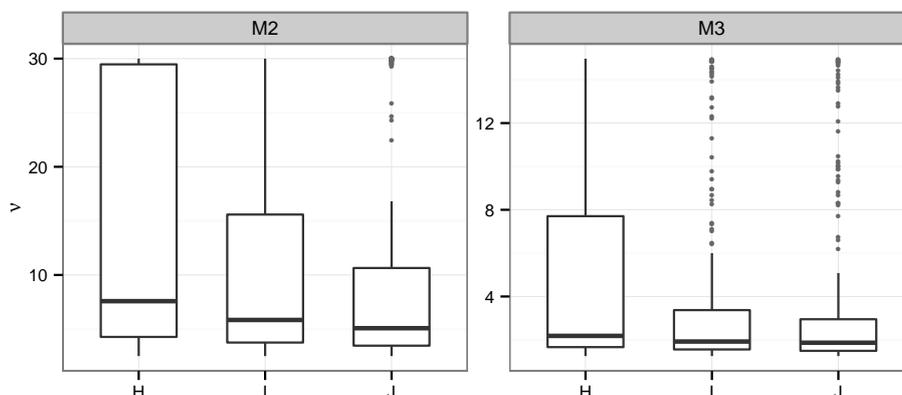


Figura 68: *Boxplots* das estimativas fornecidas para ν e valor real (linha horizontal), sob os Modelos M2 e M3, nos Cenários H, I e J.

Da Figura 69, observamos que ν_1 é mal estimado, principalmente, sob o modelo especificado corretamente (Modelo M5). Notamos que, sob os Modelos M4, M5 e M6, estimativas maiores para ν_1 são obtidas nos cenários com maior valor real para ν_1 , mas as estimativas permanecem muito distantes do valor real.

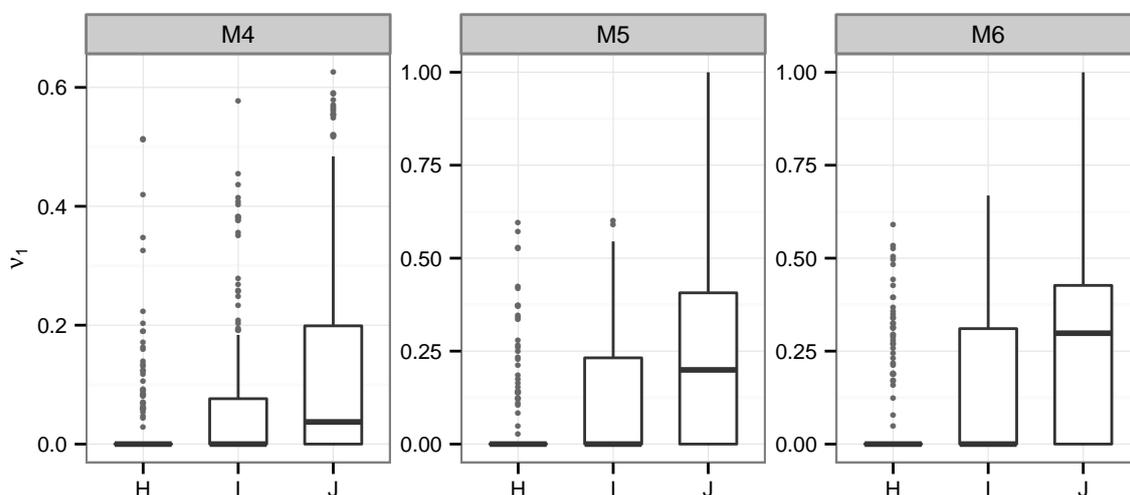


Figura 69: *Boxplots* das estimativas fornecidas para ν_1 e valor real (linha horizontal), sob os Modelos M4, M5 e M6 nos Cenários H, I e J.

Os histogramas das estimativas de ν_1 (Figura 70) fornecidas pelo Modelo M5 nos Cenários H, I e J mostram que nos três cenários avaliados, na maior parte das vezes o parâmetro ν_1 é estimado como sendo zero, indicando que a distribuição Normal é a distribuição apropriada para descrever o comportamento dos efeitos aleatórios. Poucas

estimativas se concentram em torno do valor real de ν , isto é, em 0,1 no Cenário H, 0,2 no Cenário I e 0,3 no Cenário J. Os histogramas de ν ou ν_1 com os demais modelos, nos Cenários H, I e J, são apresentados no Apêndice F.

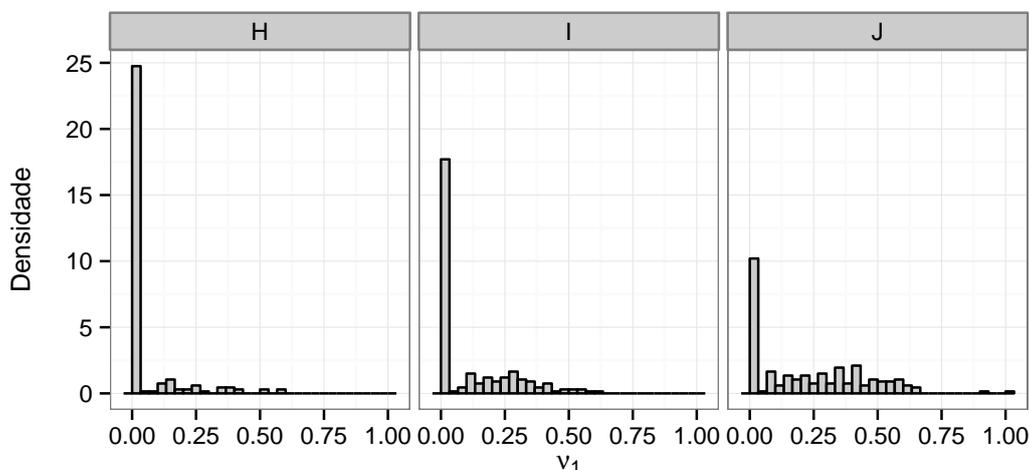


Figura 70: Histogramas das estimativas fornecidas para ν_1 , sob o Modelo M5, nos Cenários H, I e J.

Passamos então para a análise mais detalhada das estimativas para as medidas de heterogeneidade entre os *clusters*. Na Figura 71, apresentamos os *boxplots* das estimativas fornecidas para $V(\gamma_i)$ pelos Modelos M1 a M6 nos Cenários H, I e J.

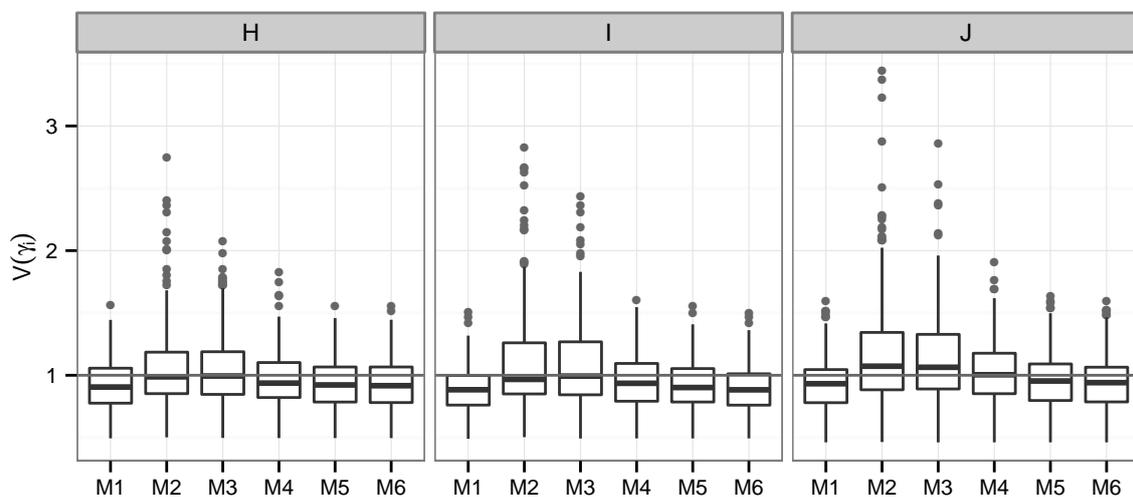


Figura 71: *Boxplots* das estimativas da variância dos efeitos aleatórios e valor real (linha horizontal), sob todos os modelos, nos Cenários H, I e J.

Da Figura 71, vemos que ao utilizarmos os Modelos M2 e M3, a distribuição das estimativas da $V(\gamma_i)$ está centrada no valor real nos cenários H e I e próximo do valor real no Cenário J. No entanto, com estes modelos, as estimativas possuem grande variabilidade e, frequentemente, estimativas consideradas *outliers* são obtidas. Sob os demais modelos,

as estimativas possuem variabilidade muito menor, sendo que no Cenários J, quando os Modelos M4 a M6 são ajustados, obtemos melhores estimativas para $V(\gamma_i)$.

Na Figura 72, apresentamos os *boxplots* das estimativas fornecidas para $MOR_{|12|}$, sob os Modelos M1 a M6, nos Cenários H, I e J.

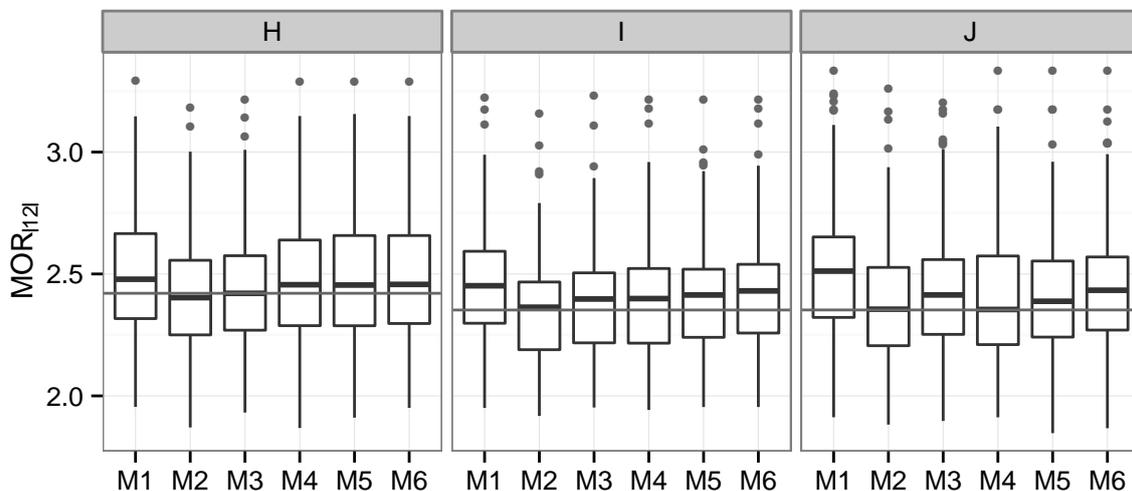


Figura 72: *Boxplots* das estimativas de $MOR_{|12|}$ e valor real (linha horizontal), sob todos os modelos, nos Cenários H, I e J.

Observando os *boxplots* das estimativas da $MOR_{|12|}$ na Figura 72, percebemos que nos cenários H, I e J, as melhores estimativas para $MOR_{|12|}$ foram obtidas quando o modelo Logístico-T foi ajustado, para o qual a distribuição das estimativas está centrada no valor real. As piores estimativas foram obtidas sob o modelo Logístico-N. Sob os modelos M4 a M6, que assumem distribuição Normal Contaminada para os efeitos aleatórios, observamos o melhor desempenho nas estimativas da $MOR_{|12|}$ nos Cenários I e J.

3.7.4.2 Comparação de modelos

Observamos da Tabela 3.20 que o modelo Logístico-T é mais frequentemente selecionado como melhor modelo, segundo o AIC (80% das vezes no Cenário H, 66,5% no Cenário I e 61% no Cenário J). Observando os resultados para os modelos que assumem a distribuição Normal Contaminada para os efeitos aleatórios, notamos que a proporção de escolha é maior nos cenários em que a distribuição dos efeitos aleatórios tem caudas mais pesadas. Nos resultados do Modelo M5, percebemos, mais uma vez, que a especificação correta do valor de ν_2 não conduziu a um melhor ajuste do modelo.

Tabela 3.20: Proporção de escolha de cada modelo, usando o critério AIC, nos cenários gerados sob o modelo Logístico-NC

	1	2	3	4	5	6
H	0,020	0,800	0,100	0,035	0,025	0,020
I	0,020	0,665	0,100	0,065	0,070	0,080
J	0,025	0,610	0,095	0,130	0,105	0,035

3.7.5 Discussão sobre o estudo Monte Carlo

Nesta subseção apresentamos as principais conclusões obtidas com o estudo de simulação Monte Carlo realizado na Seção 3.7. Inicialmente, na Subseção 3.7.1, vimos que nossa implementação do modelo Logístico-N com o algoritmo EMMC-M forneceu estimativas com a mesma qualidade daquelas obtidas via o modelo implementado no *software* R considerando métodos de integração numérica. Também sob o Modelo Logístico-N, percebemos que as estimativas de todos os parâmetros apresentam maior variabilidade quando aumentamos o valor assumido para a variância dos efeitos aleatórios $V(\gamma_i)$ na geração dos dados.

Apesar de, na Subseção 3.5.2, apresentarmos as expressões para a matriz de Informação de Fisher observada nos modelos implementados, construímos intervalos de confiança apenas nos cenários em que os bancos de dados são gerados sob o modelo Logístico-N. A razão disto é que, em nossa implementação, o cálculo da matriz de Informação de Fisher observada frequentemente retorna valores negativos, ou leva a algum erro numérico, para o erro padrão das estimativas dos parâmetros dos modelos Logístico-T, Logístico-SL e Logístico-NC.

Nas demais subseções, analisamos, entre outras coisas, o efeito da má-especificação da distribuição dos efeitos aleatórios na estimação dos parâmetros de interesse. Observamos um resultado já mencionado na literatura para a estimação dos efeitos fixos. Em todos os cenários, notamos que a má-especificação da distribuição dos efeitos aleatórios não tem grande influência nas estimativas dos efeitos fixos, isto é, para estes parâmetros não existe problema de utilizarmos o modelo Logístico-N quando os dados são gerados sob um modelo com caudas pesadas para a distribuição dos efeitos aleatórios. No entanto, se nosso interesse está em medir a heterogeneidade entre os *clusters*, temos um grande problema ao utilizar indevidamente o modelo Logístico-N. Quando os bancos de dados são gerados assumindo uma distribuição com caudas pesadas para os efeitos aleatórios, vimos que o Modelo Logístico-N subestima a $V(\gamma_i)$ e superestima a $MOR_{|12|}$.

Ao longo deste trabalho, mostramos que para medir a heterogeneidade entre os *clusters*, podemos utilizar tanto a $V(\gamma_i)$ quanto a $MOR_{|12|}$. Comparando essas duas medidas, vemos que a $MOR_{|12|}$ apresenta algumas vantagens. Uma delas é que a $MOR_{|12|}$ tem uma interpretação mais intuitiva, como foi visto na Seção 1.1. Outra vantagem é que a $MOR_{|12|}$ sempre existe, ao contrário do que ocorre para a $V(\gamma_i)$ quando assumimos as distribuições *t*-Student ou Slash para os efeitos aleatório. Se utilizamos o modelo Logístico-T, então a $V(\gamma_i)$ existe apenas quando $\nu > 2$, enquanto no modelo Logístico-SL a $V(\gamma_i)$ existe se $\nu > 1$. Além disso, observamos no estudo de simulação Monte Carlo que, em muitos bancos de dados, estimativas discrepantes para $V(\gamma_i)$ são obtidas ao ajustarmos os modelos Logístico-T e Logístico-SL quando não conhecemos o valor real de ν . A $MOR_{|12|}$ apresenta como desvantagem o fato de que em alguns casos não foi possível obter forma analítica o seu cálculo e, por isso, é necessário utilizar simulações para obter suas estimativas.

Tanto no modelo Logístico-T, quanto no modelo Logístico-SL, melhores resultados na estimação de σ^2 são obtidos ao assumirmos conhecido o valor de ν . Nos dois modelos, resultados ruins são obtidos para as estimativas do número de graus de liberdade ν . No modelo Logístico-NC a estimação do parâmetro ν_1 também é ruim. Além disso, a especificação correta do valor de ν_2 não melhorou as estimativas para σ^2 e ν_1 . Apesar destas

dificuldades na estimação, quando os efeitos aleatórios são gerados de distribuições com caudas pesadas, os modelos Logístico-T, Logístico-SL e Logístico-NC fornecem estimativas para $V(\gamma_i)$ e $MOR_{|12|}$ muito melhores do que as obtidas se o modelo Logístico-N é ajustado. Se as caudas da distribuição dos efeitos aleatórios não são muito pesadas, todos os modelos apresentam resultados similares.

Na Subseção 3.7.2, em que geramos dados sob o modelo Logístico-T, as melhores estimativas de $V(\gamma_i)$ e da $MOR_{|12|}$ foram obtidas quando ajustamos os modelos corretos. Nestes casos, o melhor modelo foi o modelo Logístico-T em que assumimos que ν é conhecido e fixado no valor real de ν , seguido do modelo Logístico-T em que ν é estimado. Quando geramos os bancos de dados sob o modelo Logístico-SL (subseção 3.7.3), os melhores resultados para as medidas de variabilidade dos efeitos aleatórios foram obtidos, respectivamente, sob os modelos Logístico-SL com ν conhecido e Logístico-T com ν estimado. Por fim, na Subseção 3.7.4, vemos que quando os bancos de dados foram gerados sob o modelo Logístico-NC, o modelo Logístico-T fornece bons resultados na estimação da $V(\gamma_i)$ e da $MOR_{|12|}$. De forma geral, quando comparamos os modelos implementados segundo o AIC, os modelos apontados como sendo os melhores são aqueles que tiveram melhor desempenho na estimação de $V(\gamma_i)$ e de $MOR_{|12|}$.

Em geral, aumentar apenas um dos valores, k , n_i ou n_{ij} , mantendo fixos os demais, quase sempre leva a estimativas com menores valores de EQM para todos os parâmetros. Exceções a isto ocorrem na estimação de ν e $V(\gamma_i)$ quando utilizamos os modelos Logístico-T e Logístico-SL, mas estas exceções só ocorrem se ν não é fixado. Comparamos os cenários em que os bancos de dados possuem a mesma quantidade de observações Bernoulli. Ao avaliarmos todas as subseções, algumas vezes os resultados foram um pouco contraditórios, mas algumas conclusões foram tiradas. Quando a comparação envolve cenários com valor igual de k , os cenários com maior valor de n_i e menor de n_{ij} tendem a fornecer estimativas com menor EQM para β_0 , β_1 e $V(\gamma_i)$ em todos os modelos. Por outro lado, se o valor n_{ij} é que está fixado, em todos os modelos, geralmente os cenários com maior k e menor n_i fornecem estimativas com menor EQM para β_0 , σ^2 e $V(\gamma_i)$. Se n_i está fixado na comparação, os cenários com maior k e menor n_{ij} tem as estimativas com menores valores de EQM para β_0 e $V(\gamma_i)$.

3.8 Estudo de caso: *Ascaris suum*

Nesta seção, vamos reanalisar o banco de dados estudado na Seção 2.7, agora ajustando um modelo logístico misto elíptico. Estes dados, originalmente apresentados em Larsen *et al.* (2000), são oriundos de um estudo com suínos na Dinamarca, cujo o objetivo é investigar a ocorrência de *Ascaris suum* em suínos criados em dois tipos diferentes de chiqueiros, a saber, chiqueiros convencional e SPF (*Specific Pathogen Free*). O banco de dados, que consiste de 1016 suínos divididos em 72 chiqueiros convencionais e 36 chiqueiros do tipo SPF, é apresentado na Tabela 2.12 com as informação do número total de suínos e do número de suínos contaminados por *Ascaris suum* em cada chiqueiro.

Na Seção 2.7, sob o paradigma bayesiano, ajustamos o modelo logístico considerando as distribuições Normal e *Skew*-Normal para representar o comportamento dos efeitos aleatórios. Naquela seção, apesar dos critérios de seleção de modelos não apresentarem resultados totalmente consistentes, existem indícios de que a distribuição dos efeitos

aleatórios seja assimétrica. Nesta seção, a análise será feita sob o paradigma clássico e serão ajustados os modelos Logístico-N, Logístico-T e Logístico-SL ao banco de dados. Duas abordagens serão adotadas para fazermos inferência no modelo Logístico-T. Inicialmente ajustamos o modelo assumindo que o número de graus de liberdade ν é um parâmetro a ser estimado. Em um segundo momento, utilizamos uma estratégia que é muito usada na análise de modelos *t*-Student, assumimos que ν é fixo e consideramos uma grade de valores de ν (aqui indo de 3 a 25 para contemplar desde f.d.p com caudas muito pesadas até f.d.p bem similares a da distribuição Normal).

Utilizamos o algoritmo EMMC-M em todos os modelos ajustados, para os quais assumimos 2400 iterações e especificamos o número de réplicas Monte Carlo L tal que $L = 50$ para as iterações 1-400, $L = 200$ para as iterações 401-1000, $L = 500$ para as iterações 1001-1500, $L = 1000$ para as iterações 1501-1800, $L = 2000$ para as iterações 1801-2100 e $L = 5000$ para as iterações 2101-2400. Os gráficos das estimativas dos parâmetros de interesse *versus* iterações, sob os modelos Logístico-N, Logístico-T e Logístico-SL, podem ser encontrados no Apêndice G. No caso do modelo Logístico-T omitimos os gráficos nos casos em que o parâmetro ν é fixado.

Na Tabela 3.21, apresentamos as estimativas dos parâmetros dos modelos e o erros padrão das estimativas sob todos os modelos, exceto sob o modelo Logístico-SL, para o qual houve algum erro numérico no cálculo da aproximação para a matriz de informação de Fisher observada. Apresentamos, também, as estimativas para a variância dos efeitos aleatórios e para as medidas de razão de chances MOR_{12} e $MOR_{|12|}$. Nesta aplicação, a única covariável incluída no modelo é o tipo de chiqueiro, assumindo o valor 0 para chiqueiro convencional e o valor 1 para chiqueiro do tipo SPF. Por isso, calculamos a razão de chances $MOR_{12} = \exp\{-\beta_1\}$, a qual representa o quanto a chance relativa de contaminação por *Ascaris suum* é maior nos chiqueiros do tipo convencional do que no tipo SPF. A razão de chances $MOR_{|12|}$, calculada como na Proposição 3.5, nos mostra o quanto a chance relativa de contaminação por *Ascaris suum* pode ser diferente em suínos criados em chiqueiros diferentes e do mesmo tipo, isto é, quantifica o impacto do chiqueiro na chance relativa de contaminação, independentemente do tipo convencional ou SPF. Além disto, na Tabela 3.21, T_ν e SL_ν representam os modelos Logístico-T e Logístico-SL com o parâmetro ν sendo estimado, enquanto T_3 , por exemplo, representa o modelo Logístico-T com ν fixado no valor 3. Para efeito de comparação, também apresentamos as estimativas obtidas para o modelo Logístico-N implementado no pacote estatístico “lme4” do *software* gratuito R, qual representamos por N_R .

Da Tabela 3.21, inicialmente avaliando apenas o modelo Logístico-N, observamos que as estimativas obtidas com o Algoritmo EMMC-M são próximas daquelas obtidas com a implementação do modelo, no *software* R, através de métodos numéricos de integração. Ao ajustarmos o modelo Logístico-T, vemos que o parâmetro ν foi estimado em 2,87, indicando que os efeitos aleatórios possuem uma distribuição com caudas muito mais pesadas do que a distribuição Normal. O mesmo ocorre sob o modelo Logístico-SL, sob o qual o ν é estimado em 1,629. Na análise realizada na Seção 2.7, existem indícios de que a distribuição dos efeitos aleatórios seja assimétrica e, por isso, acreditamos que distribuições com caudas pesadas se ajustem melhor aos efeitos aleatórios do que a distribuição Normal. Analisando as estimativas obtidas para os efeitos fixos, vemos que apesar das estimativas para β_0 e β_1 estarem próximas sob os modelos ajustados, elas não são tão similares quanto aquelas obtidas no estudo de simulação Monte Carlo da Seção 3.7.

Quando analisamos as estimativas para a variância $V(\gamma_i)$ dos efeitos aleatórios, notamos que as maiores estimativas são obtidas sob o modelo Logístico-T quando ν é estimado ou fixado no valor $\nu = 3$. O contrário ocorre para $MOR_{|12|}$, quando as estimativas obtidas sob o modelo Logístico-T com ν pequeno são as menores entre os modelos ajustados. As estimativas obtidas para MOR_{12} são similares sob todos os modelos, o que é esperado uma vez que dependem apenas das estimativas de β_1 . Ao compararmos os resultados obtidos nesta seção com aqueles obtidos na Seção 2.7, percebemos que as estimativas de máxima verossimilhança obtidas para β_0 , β_1 e $V(\gamma_i)$ sob o modelo Logístico-N se assemelham às obtidas sob o enfoque bayesiano quando assumimos normalidade para os efeitos aleatórios e distribuições *a priori* vagas para os parâmetros. Também observamos que as estimativas de máxima verossimilhança obtidas para β_0 , β_1 e $V(\gamma_i)$ sob o modelo Logístico-T com $\nu = 5$ são bem próximas das médias *a posteriori* obtidas sob o Modelo 4, o qual considera a distribuição *Skew*-Normal para os efeitos aleatórios com distribuições *a priori* pouco informativas para os parâmetros de escala e assimetria.

Tabela 3.21: Estimativas dos parâmetros (erro padrão das estimativas), da variância dos efeitos aleatórios e da razão de chances sob todos os modelos

	Modelo						
	N_R	N	T_ν	T_3	T_5	T_7	T_9
β_0	-2,529 (0,24)	-2,525 (0,27)	-2,395 (0,25)	-2,401 (0,23)	-2,449 (0,24)	-2,472 (0,25)	-2,485 (0,26)
β_1	-1,062 (0,47)	-1,063 (0,46)	-1,085 (0,43)	-1,084 (0,44)	-1,080 (0,46)	-1,072 (0,45)	-1,072 (0,45)
σ^2	2,197	2,182 (0,70)	0,868 (0,79)	0,904 (0,36)	1,274 (0,50)	1,486 (0,56)	1,625 (0,61)
ν	—	—	2,870 (2,46)	—	—	—	—
$V(\gamma_i)$	2,197	2,182	2,863	2,712	2,124	2,080	2,089
MOR_{12}	2,892	2,959	2,956	2,945	2,921	2,921	2,912
$MOR_{ 12 }$	4,112	4,092	3,145	3,171	3,444	3,600	3,732
	Modelo						
	T_{11}	T_{13}	T_{15}	T_{17}	T_{20}	T_{25}	SL_ν
β_0	-2,493 (0,26)	-2,500 (0,25)	-2,503 (0,26)	-2,505 (0,27)	-2,510 (0,26)	-2,513 (0,26)	-2,485
β_1	-1,069 (0,46)	-1,067 (0,46)	-1,070 (0,46)	-1,069 (0,46)	-1,067 (0,46)	-1,066 (0,46)	-1,078
σ^2	1,715 (0,60)	1,783 (0,59)	1,836 (0,66)	1,873 (0,66)	1,921 (0,66)	1,973 (0,65)	0,962
ν	—	—	—	—	—	—	1,629
$V(\gamma_i)$	2,096	2,108	2,119	2,122	2,134	2,144	2,492
MOR_{12}	2,907	2,915	2,912	2,907	2,904	2,895	2,939
$MOR_{ 12 }$	3,759	3,832	3,845	3,925	3,959	4,007	3,678

Observação: Apesar das estimativas para o modelo Logístico-N obtidas pelo método que implementamos coincidirem com aquelas fornecidas pelo método implementados no *software* R, se comparados àquelas obtidas em Larsen *et al.* (2000), vemos que existe muita diferença, principalmente para a variância dos efeitos aleatórios. Não há uma justificativa para tal diferença. Suspeitamos que possa haver algum erro de transcrição dos dados reais ou dos resultados obtidos em Larsen *et al.* (2000).

Apresentamos, na Figura 73, os gráficos das estimativas dos efeitos fixos β_0 e β_1 , da variância dos efeitos aleatórios $V(\gamma_i)$ e das medidas razão de chances mediana MOR_{12} e $MOR_{|12|}$. Nestes gráficos, para facilitar a análise do comportamento das estimativas, interligamos por linhas as estimativas obtidas sob o modelo Logístico-T quando assumimos ν como conhecido.

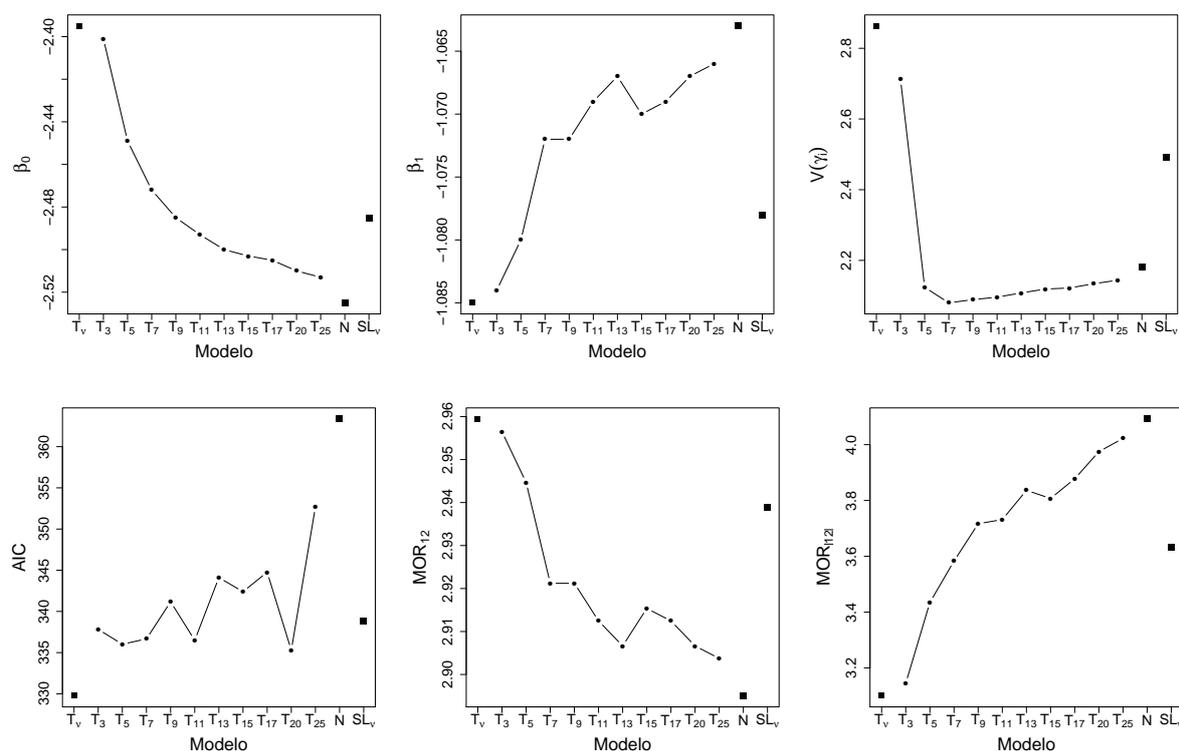


Figura 73: Estimativas, sob todos os modelos ajustados, dos efeitos fixos β_0 e β_1 e da variância $V(\gamma_i)$ dos efeitos aleatórios, das medidas de razão de chances mediana MOR_{12} e $MOR_{|12|}$ e do AIC, para dados de *Ascaris suum*.

Da Figura 73, observamos que as estimativas para β_0 obtidas sob o modelo Logístico-T com ν fixado possuem um comportamento monótono decrescente com respeito ao valor assumido para ν . Quando assumimos que $\nu = 3$, a estimativa obtida está próxima daquela obtida sob o modelo Logístico-T em que ν é estimado ($\nu = 2,870$). A medida em que fixamos ν em valores maiores, as estimativas para β_0 decrescem e se aproximam do valor estimado sob o modelo Logístico-N. Isto é esperado, uma vez que o modelo Logístico-N pode ser visto como um caso limite do modelo Logístico-T. Sob modelo Logístico-SL, a estimativa para β_0 assume um valor intermediário entre os observados nos demais modelos e é similar à obtida sob o modelo Logístico-T com $\nu = 9$. Para β_1 , observamos

que as estimativas possuem uma tendência de comportamento monótono crescente com respeito ao valor de ν . Acreditamos que esse comportamento só não é perfeito devido ao erro Monte Carlo contido nas estimativas. Mais uma vez, sob modelo o Logístico-SL a estimativa obtida foi um valor intermediário entre os obtidos sob os modelos Logístico-N e Logístico-T com ν pequeno. Comportamento similar ao observado para as estimativas dos efeitos fixos ocorre para as estimativas de MOR_{12} e $MOR_{|12|}$, principalmente porque MOR_{12} é função apenas de β_1 . Do gráfico das estimativas de $V(\gamma_i)$, observamos que sob o modelo Logístico-T, quando o parâmetro ν é estimado ou fixado em $\nu = 3$, as estimativas obtidas para $V(\gamma_i)$ são altas e bem maiores que aquela obtida sob o modelo Logístico-N. No entanto, vemos que quando fixamos ν em um valor maior ou igual a 5, as estimativas sob o Logístico-T são menores do que sob o Logístico-N. Da Figura 73 também notamos que, segundo o AIC, o modelo T_ν é o melhor modelo, seguido pelos modelos T_{20} e T_{11} , nesta ordem. Este resultado parece ser um pouco contraditório, uma vez que as estimativas obtidas ajustando-se o modelo T_3 são mais próximas daquelas obtidas sob T_ν e os modelos T_3 , T_{11} e T_{20} têm o mesmo grau de complexidade. O modelo Logístico-N é o pior modelo seguido pelo modelo T_{25} , o que é esperado uma vez que as estimativas fornecidas por ambos os modelos são próximas e estes modelos têm grau de complexidade similar. O comportamento do AIC para os demais modelos não é o esperado e, por isso, tal critério não parece ser muito confiável.

Assumindo que o modelo Logístico-T com ν estimado foi o que melhor ajustou-se aos dados, com as medidas de razão de chances mediana podemos obter as seguintes interpretações:

- Se estamos comparando suínos criados em chiqueiros de tipos diferentes, com probabilidade de 50%, a chance relativa de contaminação por *Ascaris suum* de suínos criados em chiqueiros do tipo convencional é maior ou igual a 2,956 vezes a chance relativa de contaminação de suínos criados em chiqueiros do tipo SPF;
- Se compararmos dois suínos criados em chiqueiros diferentes do mesmo tipo, seja convencional ou SPF, com probabilidade de 50%, a chance relativa de contaminação por *Ascaris suum* de um dos suínos é pelo menos 3,145 vezes a chance relativa de contaminação do outro suíno com menor risco de contaminação. Isto indica grande heterogeneidade em chiqueiros do mesmo tipo.

3.9 Conclusões e trabalhos futuros

No Capítulo 3, consideramos o modelo logístico misto assumindo distribuições elípticas para os efeitos aleatórios. Estendemos resultados obtidos em trabalhos anteriores, fornecendo resultados úteis para a interpretação dos efeitos fixos e quantificação da heterogeneidade entre *clusters* a partir da razão de chances mediana. Para tal, em várias circunstâncias, encontramos a distribuição de probabilidade da razão de chances e a razão de chances mediana. Dentre os resultados obtidos, alguns deles podem ser utilizados sob o enfoque bayesiano para determinarmos distribuições *a priori* mais adequadas para os efeitos fixos e aleatórios. Utilizando a abordagem clássica de Estatística, implementamos o modelo logístico misto com intercepto aleatório assumindo distribuições na

família Normal Independente (Subclasse da classe de distribuições elípticas) para modelarmos o comportamento dos efeitos aleatórios. Implementamos os modelos propostos utilizando o algoritmo EMMC para obtermos os estimadores de máxima verossimilhança dos parâmetros de interesse. Discutimos algumas questões relacionadas a implementação do algoritmo EMMC como, por exemplo, a geração de amostras das distribuições condicionais completas das variáveis latentes do modelo, necessárias no passo E do algoritmo. Neste sentido, implementamos três propostas de algoritmos Metropolis-Hastings para gerarmos da distribuição condicional completa do vetor de efeitos aleatórios γ , o que implica em três versões distintas do algoritmo EMMC. Além disto, avaliamos aspectos como a escolha, em cada iteração, do tamanho das amostras Monte Carlo utilizadas no passo E do algoritmo, a maximização da função de log-verossimilhança com respeito aos parâmetros de interesse e possíveis critérios de parada para algoritmos do tipo EM. Obtemos uma aproximação para a matriz de informação de Fisher observada nos modelos propostos. Por fim, realizamos um estudo com dados simulados para avaliarmos os algoritmos EMMC implementados e visando definir a configuração do algoritmo e dos dados utilizados no estudo Monte Carlo. Neste estudo Monte Carlo avaliamos a qualidade da inferência feita nos modelos ajustados. Em seguida, aplicamos os modelos propostos a dados reais referentes a contaminação de suínos por *Ascaris suum* na Dinamarca.

Em resumo, concluímos que o três algoritmos Metropolis-Hastings para gerarmos da distribuição de γ (EMMC-M, EMMC-PA e EMMC-Q) conduzem a estimativas pontuais iguais para os parâmetros de interesse nos modelos Logístico-N e Logístico-T. No entanto, os algoritmos apresentam diferença de comportamento na velocidade de convergência das estimativas, em que os algoritmos EMMC-M e EMMC-Q têm melhor desempenho. De forma geral, o algoritmo EMMC-M se mostrou a melhor opção entre as avaliadas, pois os resultados obtidos com o algoritmo EMMC-Q mostraram grande sensibilidade a especificação *ad hoc* da constante c . Quando comparamos os modelos implementados, vemos que o modelo Logístico-N é o que apresenta melhor comportamento na convergência das estimativas. Nos modelos Logístico-T e Logístico-SL, encontramos maior dificuldade na estimação do parâmetro ν , para o qual as estimativas demoram a convergir e apresentam comportamento instável quando valores altos para ν são estimados. Este problema é resolvido quando assumimos que ν é conhecido. No modelo Logístico-NC, observamos bom comportamento de convergência das estimativas dos parâmetros, exceto em alguns casos, para o parâmetro ν_1 . Em todos os modelos implementados, observamos bom comportamento, no que tange a convergência do algoritmo, das estimativas dos efeitos fixos e da variância dos efeitos aleatórios. Notamos que ao aumentarmos o tamanho dos bancos de dados, aumentando n_i e n_{ij} , melhoramos a convergência das estimativa, mas o mesmo não é observado se aumentamos o número k de clusters.

No estudo de simulação Monte Carlo, vimos que a má-especificação da distribuição dos efeitos aleatórios não influenciou na estimação dos efeitos fixos, no entanto influenciou nas estimativas das medidas de heterogeneidade entre os *clusters*. Observamos que, se os efeitos aleatórios são gerados a partir de distribuições com caudas pesadas, o modelo Logístico-N subestima a $V(\gamma_i)$ e superestima a $MOR_{|12|}$. Melhores resultados são observados quando ajustamos os modelos Logístico-SL, Logístico-NC e, principalmente, Logístico-T. Nos modelos Logístico-T e Logístico-SL, os resultados são ainda melhores quando assumimos que ν é fixo e igual ao valor real, o que não é muito realista. Quando comparamos as duas medidas de heterogeneidade $V(\gamma_i)$ e $MOR_{|12|}$, apesar da $MOR_{|12|}$

ser obtida numericamente em alguns casos, ela apresenta duas vantagens. A primeira diz respeito a simplicidade da interpretação do valor obtido. A segunda se refere ao fato de que a $MOR_{[12]}$ sempre existe nos modelos Logístico-T e Logístico-SL, ao contrário da $V(\gamma_i)$ que existe apenas se $\nu > 2$, no modelo Logístico-T, e $\nu > 1$, no modelo Logístico-SL. Quando comparamos os modelos com o AIC, vimos que os resultados relativos a escolha dos melhores modelos são similares aos obtidos para as melhores estimativas das medidas de heterogeneidade. Isto é, o modelo Logístico-T geralmente é o melhor modelo, principalmente quando ν é conhecido. O modelo Logístico-SL com ν conhecido também tem desempenho razoável. Quando comparamos a estimação em bancos de mesmo tamanho, avaliando o efeito dos valores de k , n_i e n_{ij} , a conclusão obtida é que, principalmente para β_0 e $V(\gamma_i)$, o número k de *clusters* parece ter maior impacto na qualidade das estimativas, seguido pelo número n_i de indivíduos dentro de cada *cluster*.

Na análise dos dados reais referentes a contaminação de suínos por *Ascaris suum*, observamos que as estimativas dos efeitos fixos não são tão similares sob os modelos propostos quanto o que foi observado no estudo de simulação Monte Carlo. As estimativas obtidas para β_0 são bem diferentes sob os modelos Logístico-N e Logístico-T com ν estimado ou fixado em um valor pequeno. Em geral, as estimativas dos parâmetros de interesse e da razão de chances mediana apresentam comportamento monótono com respeito ao valor fixado para ν no modelo Logístico-T.

Nosso projeto para trabalhos futuros é, inicialmente, ajustarmos os modelos implementados neste capítulo a outros conjuntos de dados reais e explorar a interpretação obtida com a razão de chances mediana para os efeitos fixos e para a variabilidade entre os *clusters*. Nos estudos de simulação, em geral, assumimos que $V(\gamma_i) = 1$ na geração dos dados. Um ponto importante é avaliarmos como os modelos Logísticos-NI se comportam quando esta condição não é utilizada. Também em estudos de simulação, pretendemos avaliar o comportamento dos modelos propostos quando os dados são gerados assumindo uma distribuição de probabilidade assimétrica para os efeitos aleatórios. Outro caminho a ser seguido é a otimização de aspectos computacionais do algoritmo EMMC implementado, para obtermos melhor desempenho quanto ao tempo de simulação. Neste caso, a determinação automática do tamanho das amostras Monte Carlo nos passos de esperança do algoritmo EMMC e a utilização de critérios de parada eficientes podem diminuir drasticamente o tempo de execução do algoritmo implementado. Na comparação dos modelos, buscamos melhorar os critérios utilizados para medir a qualidade de ajuste dos modelos. No que diz respeito a inferência realizada, pretendemos obter melhores resultados para o processo de estimação intervalar, principalmente ao obtermos a matriz de informação de Fisher observada. Outra alternativa é considerarmos métodos *bootstrap* para a obtenção dos intervalos de confiança para os parâmetros de interesse. Uma extensão natural dos modelos implementados neste trabalho é considerarmos, para os efeitos aleatórios, distribuições que pertencem a classe de distribuições *Skew-Normal Independente*. Esta classe tem a vantagem de incluir distribuições que possuem assimetria e caudas pesadas simultaneamente, como ocorre, por exemplo, com as distribuições *Skew-t-Student* e *Skew-Slash*. Por último, podemos estender os modelos Logísticos-NI implementados ao considerarmos efeitos aleatórios também associados a covariáveis, como ocorre frequentemente com o tempo em estudos longitudinais.

Apêndice A

Proposições Auxiliares

Neste apêndice introduzimos alguns resultados e suas respectivas demonstrações, os quais foram utilizados ao longo do texto para obtermos os resultados de interesse neste trabalho.

LEMA A.1. *Sejam X e Y variáveis aleatórias reais, absolutamente contínuas, e tais que $X \stackrel{d}{=} Y$ e $f_{X,Y}(x,y) = f_{X,Y}(y,x)$, $\forall x \in \mathbb{R}$ e $\forall y \in \mathbb{R}$. Então,*

$$E(X^2Y) = E(XY^2).$$

DEMONSTRAÇÃO.

Como X e Y são identicamente distribuídos temos que

$$\begin{aligned} E(X^2Y) &= \int \int x^2 y f_{X,Y}(x,y) dy dx \\ &= \int x^2 \left[\int y f_{Y|X=x}(y) dy \right] f_X(x) dx \\ &= \int x^2 E(Y | X = x) f_X(x) dx. \end{aligned}$$

Analogamente,

$$E(XY^2) = \int y^2 E(X | Y = y) f_Y(y) dy.$$

Agora,

$$\begin{aligned} E(X | Y = y) &= \int x \frac{f_{X,Y}(x,y)}{f_Y(y)} dx \\ &= \int x \frac{f_{X,Y}(x,y)}{f_X(y)} dx \\ &= \int x f_{Y|X=y}(x) dx = E(Y | X = y), \end{aligned}$$

o que conclui a prova. □

É importante salientar que se X e Y são variáveis aleatórias i.i.d o resultado apresentado no Lema A.1 segue para qualquer tipo de variáveis aleatórias.

PROPOSIÇÃO A.1. *Seja $W_{lj} = \gamma_l - \gamma_j$, com $l, j \in \{1, \dots, k\}$. Se γ_j e γ_l são variáveis aleatórias independentes e identicamente distribuídas então a f.d.p de W_{lj} é simétrica em torno de zero.*

DEMONSTRAÇÃO.

Sem perda de generalidade, assuma que γ_i e γ_j são absolutamente contínuos. Inicialmente, considerando que γ_l e γ_j são i.i.d, segue que a média da variável aleatória W_{lj} é dada por

$$E[W_{lj}] = E[\gamma_l - \gamma_j] = E[\gamma_l] - E[\gamma_j] = 0.$$

Além disto, através do terceiro momento padronizado, calculamos o grau de assimetria da distribuição da variável aleatória W_{lj} . A medida de assimetria de tal densidade é dada por $E \left[\left(\frac{W_{lj} - \mu}{\sigma} \right)^3 \right]$, em que μ e σ são a média e o desvio padrão de W_{lj} , respectivamente. Como $\mu = E[W_{lj}] = 0$, utilizando o Lema A.1, segue que

$$\begin{aligned} E \left[\left(\frac{W_{lj} - \mu}{\sigma} \right)^3 \right] &= E \left[\left(\frac{W_{lj}}{\sigma} \right)^3 \right] \\ &= \frac{E[(\gamma_l - \gamma_j)^3]}{\sigma^3} \\ &= \frac{E[\gamma_l^3] - 3E[\gamma_l^2\gamma_j] + 3E[\gamma_l\gamma_j^2] - E[\gamma_j^3]}{\sigma^3} \\ &= 0. \end{aligned}$$

Desta forma, como a medida de assimetria da distribuição de W_{lj} e $E[W_{lj}]$ são iguais a zero, concluímos que a f.d.p de W_{lj} é simétrica em torno de zero. \square

O resultado da Proposição A.1 também segue para as variáveis aleatórias que satisfazem as condições do Lema A.1 e que tem esperança igual a zero.

PROPOSIÇÃO A.2. *Se W é uma variável aleatória com f.d.p absolutamente contínua então*

$$\text{med} \{ \exp \{ W \} \} = \exp \{ \text{med} \{ W \} \}.$$

DEMONSTRAÇÃO. Temos, por definição, que $\text{med}\{W\} = F_W^{-1}(0,5)$. Em consequência disto, segue que $\exp \{ \text{med}\{W\} \} = \exp \{ F_W^{-1}(0,5) \}$.

Além disto, se $md = \text{med} \{ \exp \{ W \} \}$, então

$$P(\exp \{ W \} \leq md) = 0,5$$

$$F_W(\ln md) = 0,5$$

$$md = \exp \{ F_W^{-1}(0,5) \}.$$

Portanto, temos que $\text{med} \{ \exp \{ W \} \} = \exp \{ \text{med}\{W\} \} = \exp \{ F_W^{-1}(0,5) \}$, concluindo a demonstração. \square

Apêndice B

Distribuições de Probabilidade

Neste apêndice apresentamos algumas distribuições de probabilidades com algumas de suas características as quais serão utilizadas ao longo do texto. Para cada uma das distribuições apresentaremos a f.d.p, expressões para média e variância, e em alguns casos, quando for necessário, a expressão da f.g.m. e gráficos que ilustrem o comportamento da f.d.p. B

Distribuição Beta

Dizemos que uma variável aleatória Y tem distribuição de probabilidade Beta com parâmetros de forma $a > 0$ e $b > 0$, com notação $Y \sim Beta(a, b)$, se sua f.d.p é

$$f(y) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1} (1-y)^{b-1} \quad y > 0,$$

em que $\Gamma(\cdot)$ é a função Gama.

A esperança e variância de Y são dadas, respectivamente, por

$$E(Y) = \frac{a}{a+b} \quad \text{e} \quad V(Y) = \frac{ab}{(a+b)^2(a+b+1)}.$$

Distribuição Gama

Dizemos que uma variável aleatória Y tem distribuição de probabilidade Gama com parâmetros de forma $a > 0$ e de taxa $b > 0$, com notação $Y \sim Gama(a, b)$, se sua f.d.p é

$$f(y) = \frac{b^a}{\Gamma(a)} y^{a-1} \exp\{-by\}, \quad y > 0.$$

Neste caso, esperança e variância de Y são dadas, respectivamente, por

$$E(Y) = \frac{a}{b} \quad \text{e} \quad V(Y) = \frac{a}{b^2}.$$

Distribuição Gama Inversa

Dizemos que uma variável aleatória Y tem distribuição de probabilidade Gama Inversa com parâmetros de forma $a > 0$ e de escala $b > 0$, com notação $Y \sim GI(a, b)$, se sua f.d.p é

$$f(y) = \frac{b^a}{\Gamma(a)} y^{-(a+1)} \exp\left\{-\frac{b}{y}\right\} \quad y > 0.$$

Neste caso, esperança e variância de Y são dadas, respectivamente, por

$$E(Y) = \frac{b}{a-1}, \text{ para } a > 1,$$
$$V(Y) = \frac{b^2}{(a-1)^2(a-2)}, \text{ para } a > 2.$$

Distribuição *Skew*-Normal univariada

A distribuição *Skew*-normal no caso univariado foi introduzida por Azzalini (1985). Dizemos que uma variável aleatória Y tem distribuição *Skew*-normal com parâmetros de locação $\xi \in \mathbb{R}$, de escala $\sigma^2 \in \mathbb{R}_+$ e de assimetria $\lambda \in \mathbb{R}$, com notação $Y \sim SN(\xi, \sigma^2, \lambda)$, se sua f.d.p é escrita como

$$f(y) = \frac{2}{\sigma} \phi\left(\frac{y-\xi}{\sigma}\right) \Phi\left(\lambda\left(\frac{y-\xi}{\sigma}\right)\right), \quad -\infty < y < \infty.$$

Se os parâmetros de locação ξ e escala σ^2 são iguais a zero e a um, respectivamente, dizemos que Y tem distribuição *Skew*-Normal padrão denotada por $Y \sim SN(\lambda)$.

De forma geral, a média e a variância de Y são, respectivamente, dadas por

$$E(Y) = \xi + \sigma \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \delta \quad \text{e} \quad V(Y) = \sigma^2 \left(1 - \frac{2}{\pi} \delta^2\right),$$

em que $\delta = \lambda(1 + \lambda^2)^{-1/2}$.

Henze (1986) introduziu uma representação estocástica para a distribuição *Skew*-Normal que é um resultado importante tanto do ponto de vista teórico quanto prático. Apresentamos este resultado no lema a seguir.

LEMA B.1. (*Representação Estocástica de Henze*) Se Y_0 e Y_1 são variáveis aleatórias independentes e identicamente distribuídas com distribuição $N(0,1)$ e $\delta \in (-1,1)$, então

$$Z = \delta|Y_0| + (1 - \delta^2)^{\frac{1}{2}}Y_1$$

tem distribuição Normal assimétrica padrão com parâmetro de assimetria $\lambda = \frac{\delta}{\sqrt{1-\delta^2}}$.

A Figura 74 mostra as densidades para algumas distribuições *Skew*-normais padrão com diferentes valores de λ .

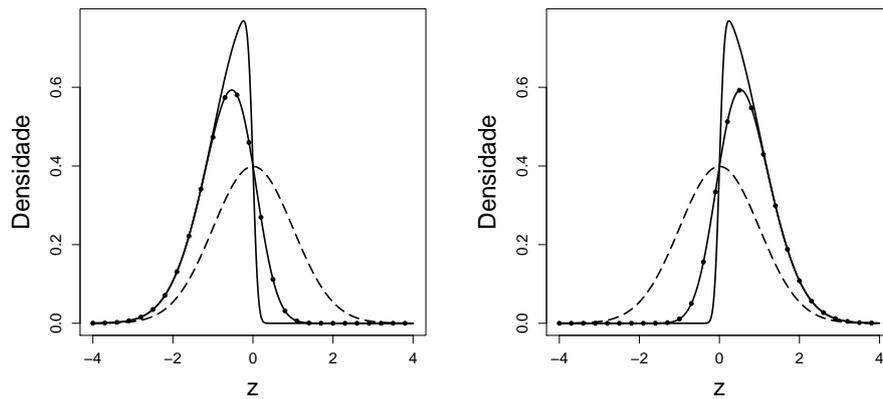


Figura 74: Densidade da distribuição *Skew-Normal* padrão com diferentes valores para o parâmetro λ . A esquerda a densidade de $SN(\lambda)$ com λ igual a -10 (linha sólida), -2 (●) e 0 (linha tracejada) e a direita com λ igual a 10 (linha sólida), 2 (●) e 0 (linha tracejada).

A Figura 75 ilustra o comportamento da f.d.p *Skew-Normal* para vários valores dos parâmetros de locação e escala.

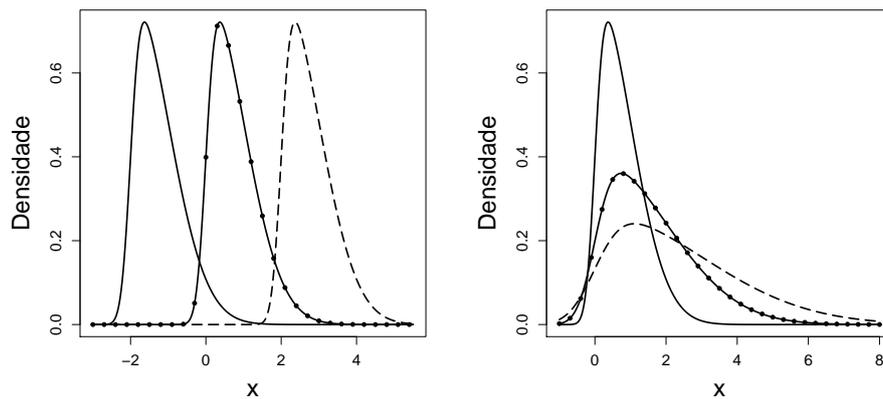


Figura 75: Densidade da distribuição *Skew-Normal* para diferentes valores de ξ e σ^2 . A esquerda a densidade de $SN(\xi, 1, 5)$ com ξ igual a -2 (linha sólida), 0 (●) e 2 (linha tracejada). A direita a densidade de $SN(0, \sigma^2, 5)$ com σ^2 igual a 1 (linha sólida), 4 (●) e 9 (linha tracejada).

Distribuição *Skew-Normal* multivariada

A distribuição Normal assimétrica multivariada é uma extensão para o caso multivariado da distribuição Normal assimétrica apresentada anteriormente. Proposta por Azzalini e Dalla-Valle (1996), esta classe de distribuições tem a propriedade de que cada uma de suas distribuições marginais univariadas tem distribuição Normal assimétrica

univariada, da mesma forma que ocorre com a distribuição Normal. Dizemos que um vetor aleatório Y tem distribuição *Skew-Normal* n -variada com parâmetros de locação $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)^t \in \mathbb{R}^k$, matriz de escala $\mathbf{S}\boldsymbol{\Omega}\mathbf{S}$ positiva definida e vetor de assimetria $\boldsymbol{\alpha} \in \mathbb{R}^n$, com notação $SN_n(\boldsymbol{\xi}, \mathbf{S}\boldsymbol{\Omega}\mathbf{S}, \boldsymbol{\alpha})$, se sua f.d.p é dada por

$$f_{Y|\boldsymbol{\xi}, \boldsymbol{\Omega}, \mathbf{S}, \boldsymbol{\alpha}}(\mathbf{y}) = 2\phi_n(\mathbf{y}|\boldsymbol{\xi}, \mathbf{S}\boldsymbol{\Omega}\mathbf{S}) \Phi(\boldsymbol{\alpha}^t \mathbf{S}^{-1}(\mathbf{y} - \boldsymbol{\xi})), \quad (\text{B.1})$$

em que $\mathbf{S} = \text{diag}\{\sigma_1, \dots, \sigma_n\}$ é uma matriz $n \times n$ e $\boldsymbol{\Omega}$ é uma matriz com características de matriz de correlação.

O vetor de médias e a matriz de variância-covariância da distribuição com densidade em (B.1) são dados, respectivamente, por

$$E[\mathbf{Y}] = \boldsymbol{\xi} + \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \boldsymbol{\Delta} \quad \text{e} \quad V[\mathbf{Y}] = \mathbf{S}\boldsymbol{\Omega}\mathbf{S} - \frac{2}{\pi} \mathbf{S}\boldsymbol{\Delta}\boldsymbol{\Delta}^t \mathbf{S},$$

em que $\boldsymbol{\Delta} = \frac{1}{(1 + \boldsymbol{\alpha}^t \boldsymbol{\Omega} \boldsymbol{\alpha})^{1/2}} \boldsymbol{\Omega} \boldsymbol{\alpha}$.

Se $\mathbf{Z} \sim SN_n(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$, então a correlação entre dois componentes genéricos do vetor aleatório \mathbf{Z} é dada por

$$\rho_{ij} = \text{Corr}(Z_i, Z_j) = \frac{\omega_{ij} - \frac{2}{\pi} \Delta_i \Delta_j}{\left[\left(1 - \frac{2}{\pi} \Delta_i^2\right)\left(1 - \frac{2}{\pi} \Delta_j^2\right)\right]^{1/2}}, \quad i, j = 1, \dots, n,$$

em que Δ_i é o i -ésimo componente do vetor $\boldsymbol{\Delta}$ e ω_{ij} é o componente da i -ésima linha e j -ésima coluna da matriz $\boldsymbol{\Omega}$.

Azzalini e Capitanio (1999) apresentam o resultado sobre as distribuições marginais da distribuição *Skew-Normal* multivariada. Este resultado diz que se $\mathbf{Y} \sim SN_n(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$, em que os parâmetros são particionados como segue

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}, \quad \boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix} \quad \text{e} \quad \boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{pmatrix},$$

em que $\boldsymbol{\Omega}$ é uma matriz de correlação e \mathbf{Y}_1 tem é um vetor $n_1 \times 1$, então

$$\mathbf{Y}_1 \sim SN_{n_1}(\mathbf{0}, \boldsymbol{\Omega}_{11}, \boldsymbol{\alpha}_1^*),$$

em que $\boldsymbol{\alpha}_1^* = \frac{\boldsymbol{\alpha}_1 + \boldsymbol{\Omega}_{11}^{-1} \boldsymbol{\Omega}_{12} \boldsymbol{\alpha}_2}{(1 + \boldsymbol{\alpha}_2^t \boldsymbol{\Omega}_{22 \cdot 1} \boldsymbol{\alpha}_2)^{1/2}}$ e $\boldsymbol{\Omega}_{22 \cdot 1} = \boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21} \boldsymbol{\Omega}_{11}^{-1} \boldsymbol{\Omega}_{12}$.

O Lema B.2 a seguir é uma forma útil de representarmos a distribuição Normal assimétrica multivariada. Tal representação também pode ser utilizada para gerar amostras da distribuição em (B.1).

LEMA B.2. (*Representação Estocástica da distribuição Skew-Normal multivariada*)
Se a variável aleatória Y_0 e o vetor aleatório \mathbf{Y}_1 são independentes de forma que $Y_0 \sim N(0, 1)$ e $\mathbf{Y}_1 \sim N_n(\mathbf{0}_k, \boldsymbol{\Psi})$, em que $\boldsymbol{\Psi} = \boldsymbol{\Omega} - c^2 \boldsymbol{\Omega} \boldsymbol{\alpha} \boldsymbol{\alpha}^t \boldsymbol{\Omega}$, $c = (1 + \boldsymbol{\alpha}^t \boldsymbol{\Omega} \boldsymbol{\alpha})^{-\frac{1}{2}}$ e $\boldsymbol{\Delta} = c \boldsymbol{\Omega} \boldsymbol{\alpha}$, então

$$\mathbf{Y} = \boldsymbol{\xi} + \mathbf{S}(\boldsymbol{\Delta}|Y_0| + \mathbf{Y}_1)$$

é um vetor aleatório com distribuição $SN_n(\boldsymbol{\xi}, \mathbf{S}\boldsymbol{\Omega}\mathbf{S}, \boldsymbol{\alpha})$.

A Figura 76 mostra curvas de níveis das f.d.p de duas distribuições *Skew-normais* bivariadas diferentes.

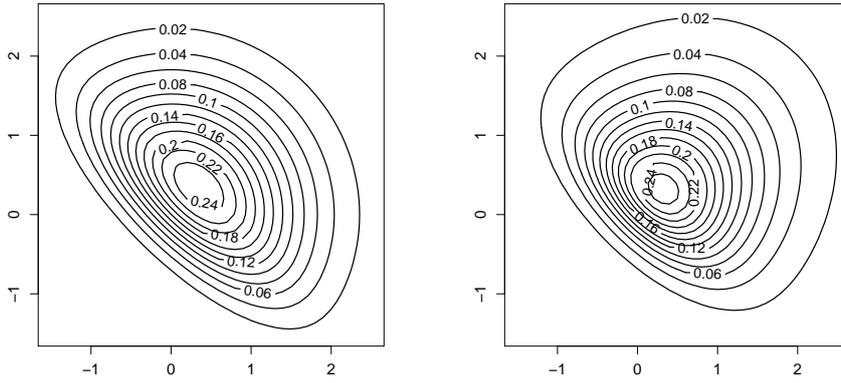


Figura 76: Curvas de nível da f.d.p da distribuição *Skew-Normal* bvariada $SN_2(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$ com vetor de locação $\boldsymbol{x} = (0, 0)^t$, vetor de assimetria $\boldsymbol{\alpha} = (2, 2)^t$ e matriz de escala $\boldsymbol{\Sigma}$, em que $\boldsymbol{\Sigma} = \mathbf{I}_2$ (à esquerda) ou $\boldsymbol{\Sigma}$ é uma matriz com diagonal principal com valor 1 e diagonal secundária com valor 0,3 (à direita).

Skew-Normal Unificada (SUN)

A família de distribuições *Skew-Normal Unificada (SUN)* foi introduzida por Arellano-Valle e Azzalini (2006). Ao assumirmos que $\mathbf{Y} \in \mathbb{R}^d$ tem comportamento probabilístico descrito por uma distribuição pertencente a esta família SUN, isto é, ao considerarmos que $\mathbf{Y} \sim SUN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \boldsymbol{\Omega}^*)$, em que $\bar{\boldsymbol{\omega}} = \boldsymbol{\omega} \mathbf{1}_d$, $\boldsymbol{\Omega} = \boldsymbol{\omega} \bar{\boldsymbol{\Omega}} \boldsymbol{\omega}$ e $\boldsymbol{\Omega}^* = \begin{pmatrix} \boldsymbol{\Delta} & \boldsymbol{\Gamma}^t \\ \boldsymbol{\Gamma} & \bar{\boldsymbol{\Omega}} \end{pmatrix}$, sua f.d.p e função geradora de momentos (f.g.m.) são, respectivamente, dadas por

$$f(\mathbf{y}) = \phi_d(\mathbf{y} - \boldsymbol{\xi}; \boldsymbol{\Omega}) \frac{\Phi_m(\boldsymbol{\gamma} + \boldsymbol{\Gamma}^t \bar{\boldsymbol{\Omega}}^{-1} \boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi}); \boldsymbol{\Delta} - \boldsymbol{\Gamma}^t \bar{\boldsymbol{\Omega}} \boldsymbol{\Gamma})}{\Phi_m(\boldsymbol{\gamma} \mid \boldsymbol{\Delta})}, \quad \mathbf{y} \in \mathbb{R}^d,$$

$$M_{\mathbf{Y}}(\mathbf{r}) = \exp \left\{ \boldsymbol{\xi}^t \mathbf{r} + \frac{1}{2} \mathbf{r}^t \boldsymbol{\Omega} \mathbf{r} \right\} \Phi_m(\boldsymbol{\gamma} + \boldsymbol{\Gamma}^t \boldsymbol{\omega} \mathbf{r} \mid \boldsymbol{\Delta}) [\Phi_m(\boldsymbol{\gamma} \mid \boldsymbol{\Delta})]^{-1}, \quad \mathbf{r} \in \mathbb{R}^d. \quad (\text{B.2})$$

A média e a variância da distribuição SUN podem ser obtidas utilizando a expressão da f.g.m. em (B.2). Para mais detalhes ver Arellano-Valle e Azzalini (2006), em que as expressões para média e variância são apresentadas para casos particulares da distribuição SUN.

Apêndice C

Distribuições condicionais completas no modelo logístico misto com efeitos aleatórios com distribuição *Skew-Normal* multivariada

Neste apêndice apresentamos as distribuições condicionais completas (d.c.c.) para o modelo logístico com intercepto aleatório quando consideramos efeitos aleatórios correlacionados tal que, dados σ^2 e α , $\boldsymbol{\gamma} \sim SN(-\Delta\sigma\sqrt{2/\pi}\mathbf{1}_k, \sigma^2\mathbf{I}_k, \alpha\mathbf{1}_k)$ em que $\Delta = \alpha[1 + k\alpha^2]^{-1/2}$, o qual foi abordado no Capítulo 2 deste trabalho. As distribuições *a priori* assumidas são $\boldsymbol{\beta} \sim N_{p+1}(\mathbf{m}, b^2\mathbf{I}_{p+1})$, $\sigma^2 \sim GI(a, d)$ e $\alpha \sim N(h, \tau^2)$, em que $\mathbf{m} \in \mathbb{R}^{p+1}$, $h, \theta \in \mathbb{R}$, e τ^2, b^2, a e d são números reais positivos.

Para $\vartheta = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \alpha)$, a distribuição *a posteriori* é dada por

$$f(\vartheta|\mathbf{D}) \propto \left[\prod_{i=1}^k \prod_{j=1}^{n_i} (\pi_{ij})^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}} \right] \left(\frac{1}{\sigma^2} \right)^{\frac{k+2a+2}{2}} \Phi \left(\frac{\alpha}{\sigma} \sum_{i=1}^k \left(\gamma_i + \Delta\sigma\sqrt{\frac{2}{\pi}} \right) \right) \\ \times \exp \left\{ -\frac{\sum_{i=0}^p (\beta_i - m_i)^2}{2b^2} - \frac{\sum_{i=1}^k \left(\gamma_i + \Delta\sigma\sqrt{\frac{2}{\pi}} \right)^2}{2\sigma^2} - \frac{(\alpha - h)^2}{2\tau^2} - \frac{d}{\sigma^2} \right\}.$$

Para obtermos as d.c.c. *a posteriori*, consideramos a representação estocástica para a distribuição multivariada SN obtida em Azzalini e Dalla-Valle (1996)(ver Apêndice B), que estabelece o seguinte resultado: se $\boldsymbol{\gamma} \sim SN_k(\mathbf{0}_k, \mathbf{I}_k, \alpha\mathbf{1}_k)$ então $\boldsymbol{\gamma} \stackrel{d}{=} \Delta|U| + \mathbf{V}$, onde $U \sim N(0, 1)$ e $\mathbf{V} \sim N_k(\mathbf{0}_k, \mathbf{I}_k - (1 + k\alpha^2)^{-1}\alpha^2\mathbf{1}_k\mathbf{1}_k^t)$. Assumimos que $T = |U|$. Consequentemente, se $\boldsymbol{\gamma} \sim SN_k(-\Delta\sigma\sqrt{2/\pi}, \sigma^2\mathbf{I}_k, \alpha\mathbf{1}_k)$, então $\boldsymbol{\gamma} \stackrel{d}{=} \Delta\sigma \left(T - \sqrt{2/\pi} \right) + \sigma\mathbf{V}$. Seja $\mathbf{F} = \sigma \left(T - \sqrt{2/\pi} \right)$ e defina $\vartheta^* = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \alpha, T)$. Consequentemente, as d.c.c.

a posteriori para γ , σ^2 , α e T são dadas respectivamente por

$$f(\gamma|\mathbf{D}, \vartheta_{-\gamma}^*) \propto \left[\prod_{i=1}^k \prod_{j=1}^{n_i} \left[\frac{\exp\{\eta_{ij}\}}{1 + \exp\{\eta_{ij}\}} \right]^{y_{ij}} \left[\frac{1}{1 + \exp\{\eta_{ij}\}} \right]^{1-y_{ij}} \right] \\ \times \exp \left\{ -\frac{1}{2\sigma^2} [\gamma - \Delta \mathbf{F}]^t \left[\mathbf{I}_k + \frac{\Delta^2}{1 - k\Delta^2} \mathbf{1}_k \mathbf{1}_k^t \right] [\gamma - \Delta \mathbf{F}] \right\},$$

$$f(T|\mathbf{D}, \vartheta_{-T}^*) \propto \exp \left\{ -\frac{1}{2(1 - k\Delta^2)} \left[T^2 - 2T \left(\frac{\Delta}{\sigma} \sum_{i=1}^k \gamma_i + k\Delta^2 \sqrt{2/\pi} \right) \right] \right\} \mathbf{1}_{(0, \infty)}(T),$$

$$f(\alpha|\mathbf{D}, \vartheta_{-\alpha}^*) \propto (1 + k\alpha^2)^{\frac{1}{2}} \exp \left\{ -\frac{(\alpha - h)^2}{2\tau^2} \right\} \\ \times \exp \left\{ -\frac{1}{2\sigma^2} \left[\gamma - \frac{\alpha \mathbf{F}}{\sqrt{1 + k\alpha^2}} \mathbf{1}_k \right]^t \left[\mathbf{I}_k + \alpha^2 \mathbf{1}_k \mathbf{1}_k^t \right] \left[\gamma - \frac{\alpha \mathbf{F}}{\sqrt{1 + k\alpha^2}} \mathbf{1}_k \right] \right\},$$

$$f(\sigma^2|\mathbf{D}, \vartheta_{-\sigma^2}^*) \propto \left(\frac{1}{\sigma^2} \right)^{\frac{k}{2} + a + 1} \exp \left\{ -\frac{d}{\sigma^2} \right\} \\ \times \exp \left\{ -\frac{1}{2\sigma^2} [\gamma - \Delta \mathbf{F}]^t \left[\mathbf{I}_k + \frac{\Delta^2}{1 - k\Delta^2} \mathbf{1}_k \mathbf{1}_k^t \right] [\gamma - \Delta \mathbf{F}] \right\}.$$

A d.c.c a posteriori para β é a mesma obtida no caso i.i.d e pode ser encontrada no Capítulo 2, Seção 2.3.

Apêndice D

Gráficos de viés e erro quadrático médio: Seção 2.5

No Apêndice D apresentamos gráficos referentes a Seção 2.5. São apresentados os gráficos de MSE e vício das estimativas dos parâmetros de efeitos fixos e variância dos efeitos aleatórios ao utilizarmos mediana e moda *a posteriori* como estimadores.

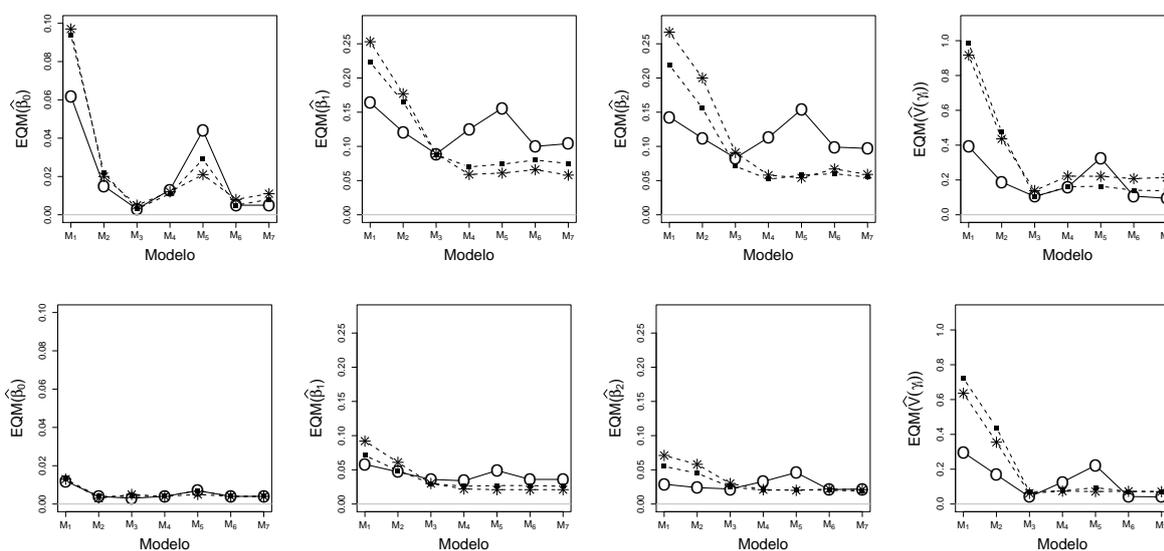


Figura 77: Erro quadrático médio (MSE) para as medianas *a posteriori* para os efeitos fixos e variância dos efeitos aleatórios, para os cenários com $k = 25$ (acima) e $k = 100$ (abaixo) e verdadeiros valores $\lambda = 0$ (\circ), $\lambda = 5$ (\blacksquare) e $\lambda = 30$ (*).

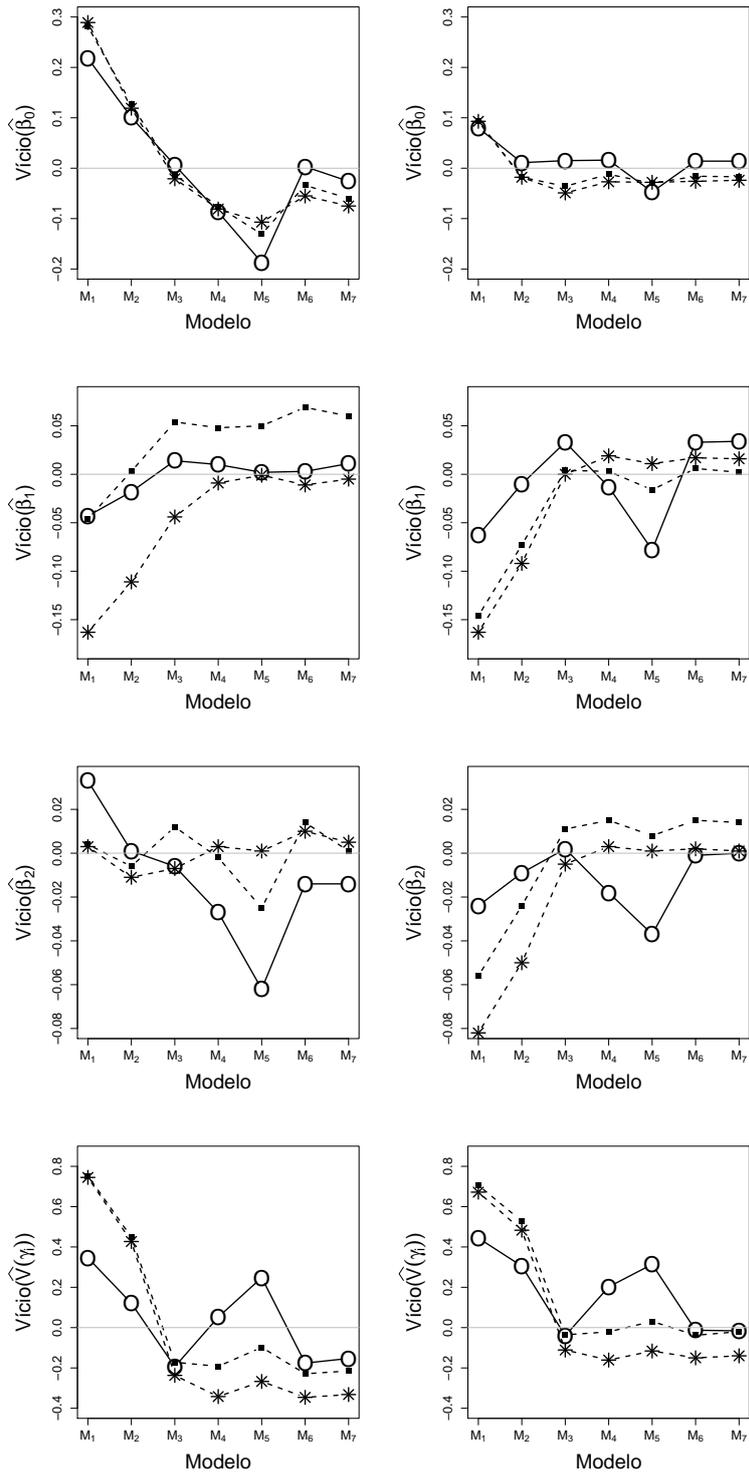


Figura 78: Vício para as medianas *a posteriori* para os efeitos fixos e variância dos efeitos aleatórios, para os cenários com $k = 25$ (esquerda) e $k = 100$ (direita) e verdadeiros valores $\lambda = 0$ (\circ), $\lambda = 5$ (\blacksquare) e $\lambda = 30$ (*).

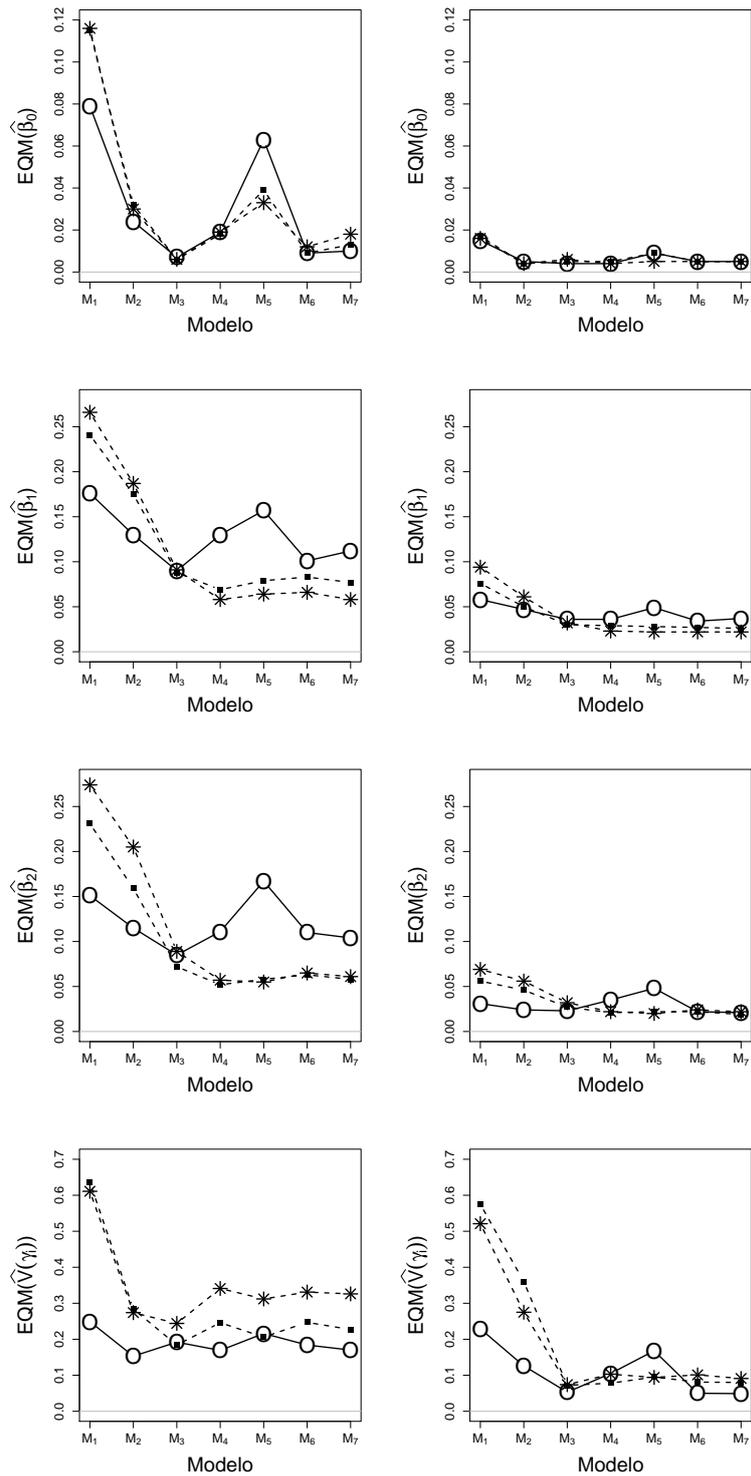


Figura 79: Erro quadrático médio (MSE) para as modas *a posteriori* para os efeitos fixos e variância dos efeitos aleatórios, para os cenários com $k = 25$ (esquerda) e $k = 100$ (direita) e verdadeiros valores $\lambda = 0$ (\circ), $\lambda = 5$ (\blacksquare) e $\lambda = 30$ ($*$).

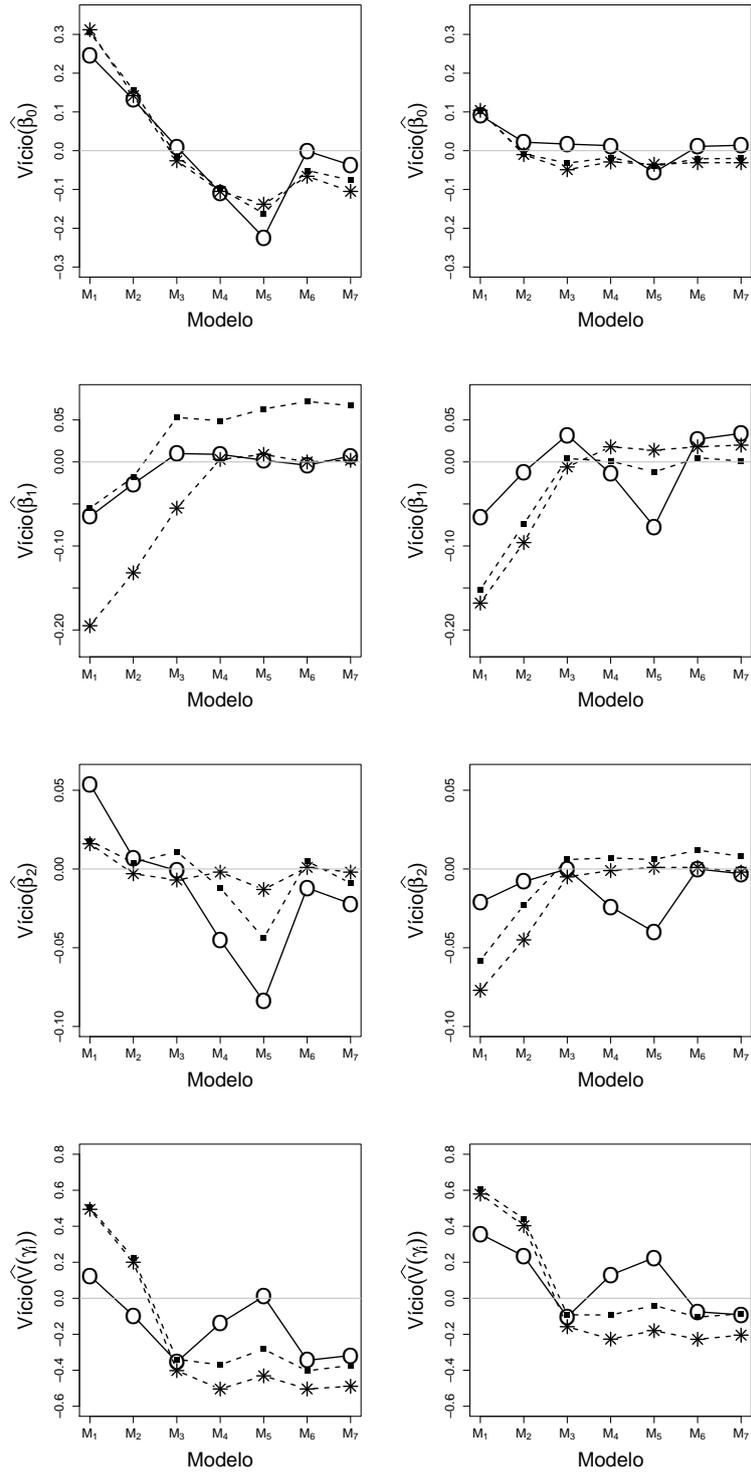


Figura 80: Vício para as modas *a posteriori* para os efeitos fixos e variância dos efeitos aleatórios, para os cenários com $k = 25$ (esquerda) e $k = 100$ (direita) e verdadeiros valores $\lambda = 0$ (\circ), $\lambda = 5$ (\blacksquare) e $\lambda = 30$ (*).

Apêndice E

Gráficos dos Estudos com dados simulados da Seção 3.6

No Apêndice E apresentamos gráficos referentes a Seção 3.6. São apresentados os histogramas dos efeitos aleatórios gerados em alguns cenários e os gráficos das estimativas dos parâmetros de interesse *versus* iteração nos bancos de dados analisados nas Subseções 3.6.1, 3.6.2 e 3.6.3.

Gráficos referentes a Subseção 3.6.1

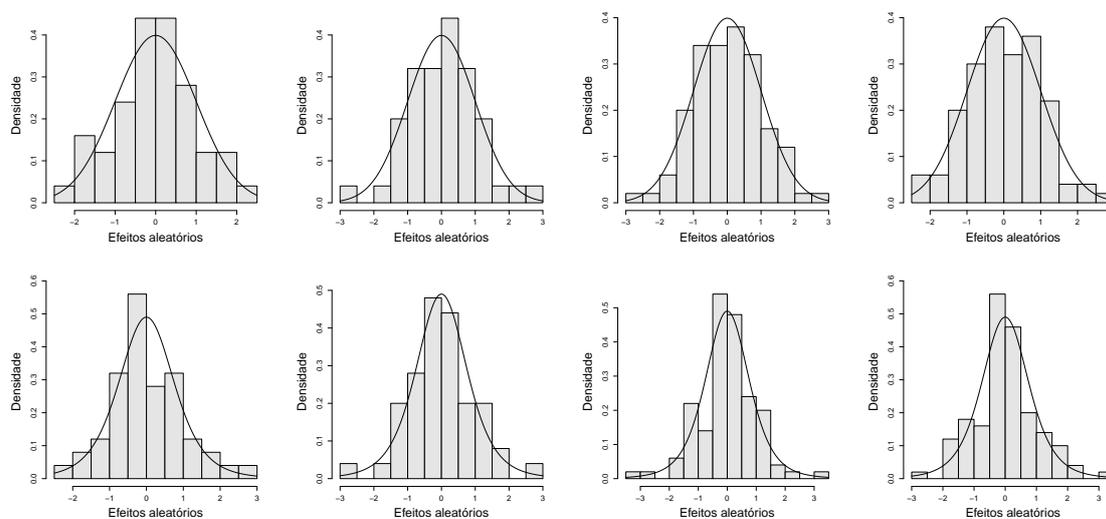


Figura 81: Histograma e f.d.p teórica dos efeitos aleatórios gerados no Cenário 1 para Dados 1 a 8 (da esquerda para direita).

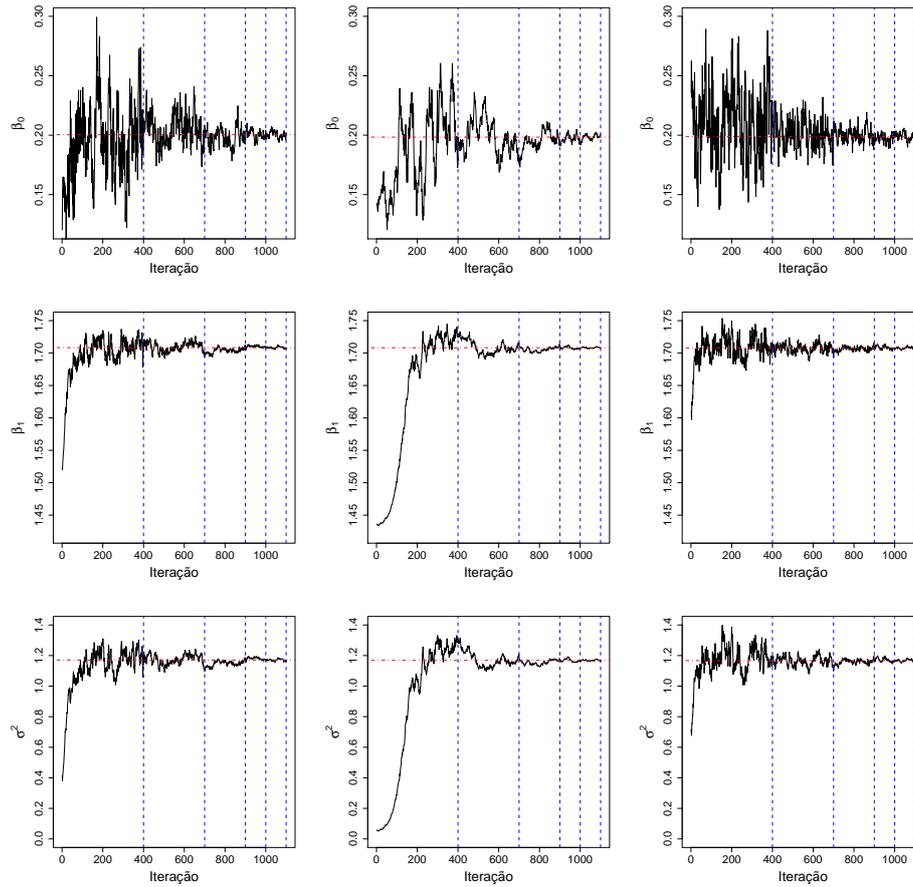


Figura 82: Gráficos das estimativas dos parâmetros *versus* iterações, e média das estimativas nas últimas 100 iterações (linha tracejada), obtidas com os algoritmos EMMC-M (esquerda), EMMC-PA (centro) e EMMC-Q (direita) para Dados 3 e Cenário 1.

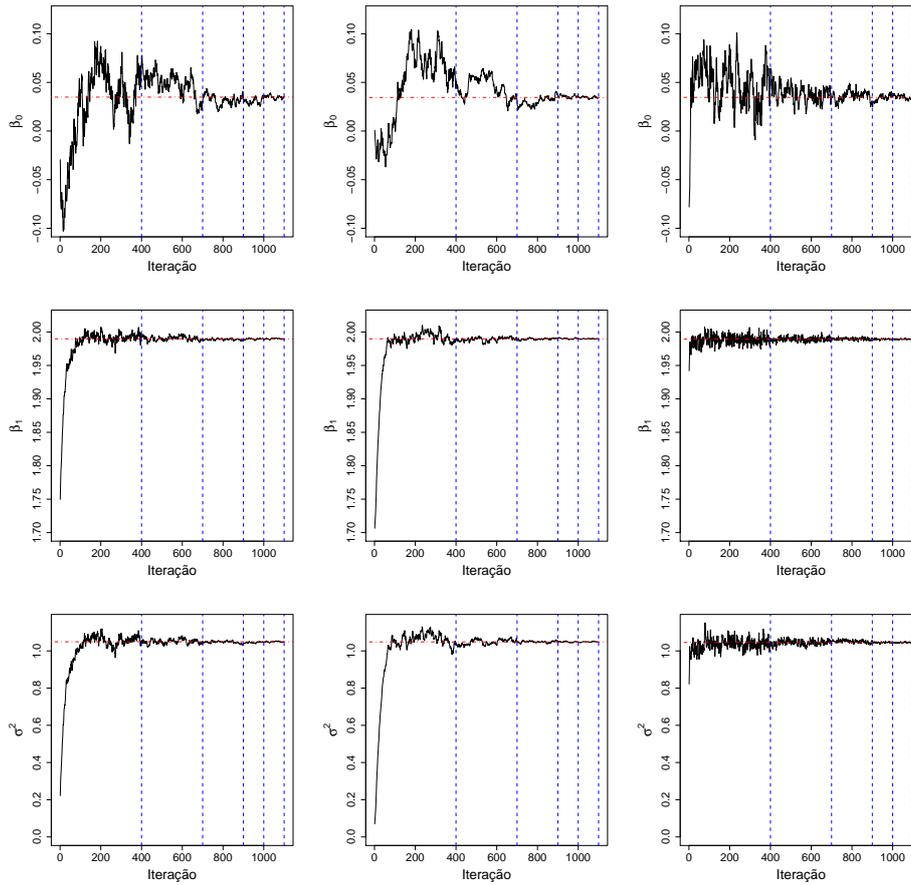


Figura 83: Gráficos das estimativas dos parâmetros *versus* iterações, e média das estimativas nas últimas 100 iterações (linha tracejada), obtidas com os algoritmos EMMC-M (esquerda), EMMC-PA (centro) e EMMC-Q (direita) para Dados 4 e Cenário 1.

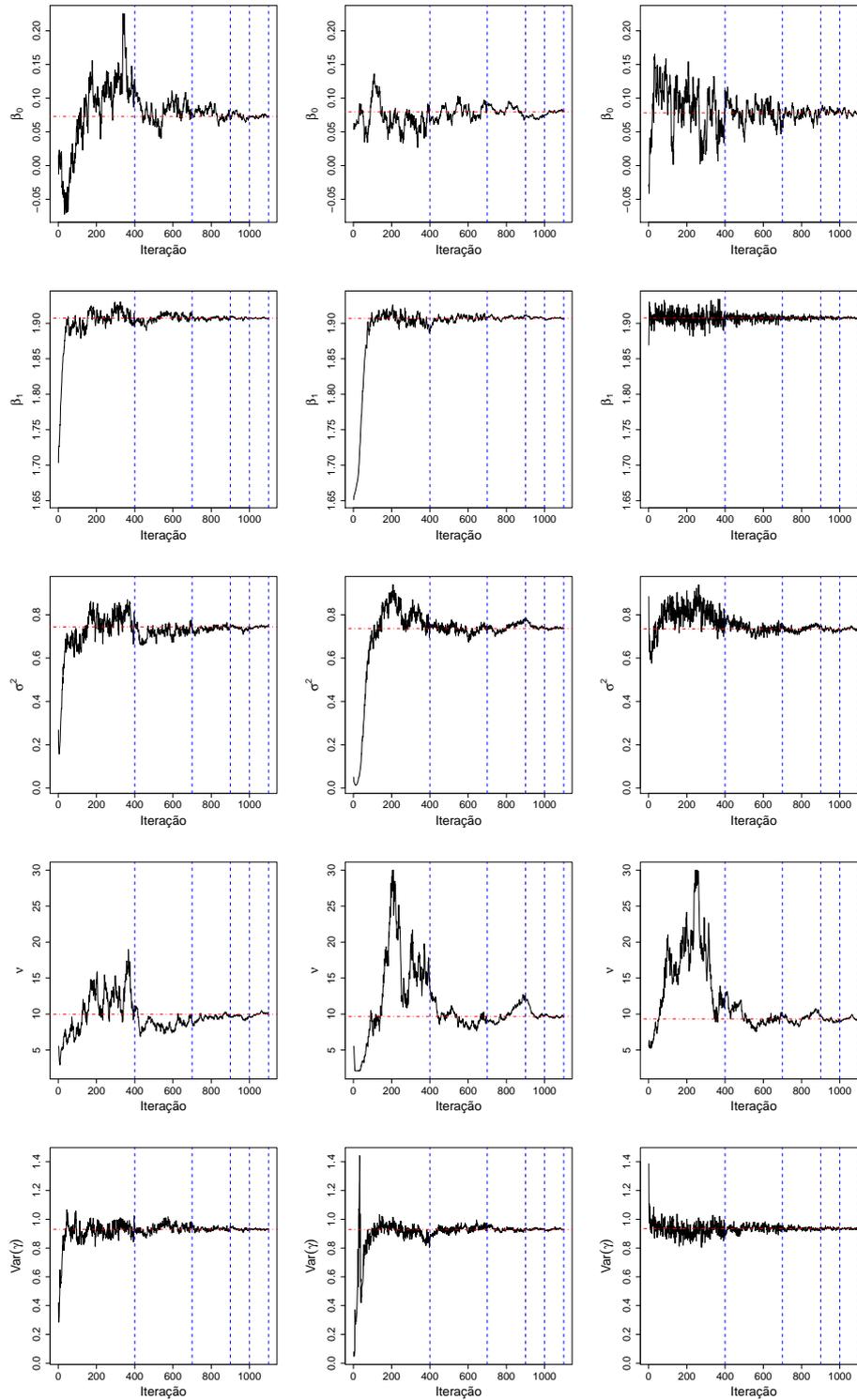


Figura 84: Gráficos das estimativas dos parâmetros *versus* iterações, e média das estimativas nas últimas 100 iterações (linha tracejada), obtidas com os algoritmos EMMC-M (esquerda), EMMC-PA (centro) e EMMC-Q (direita) para Dados 6 e Cenário 1.

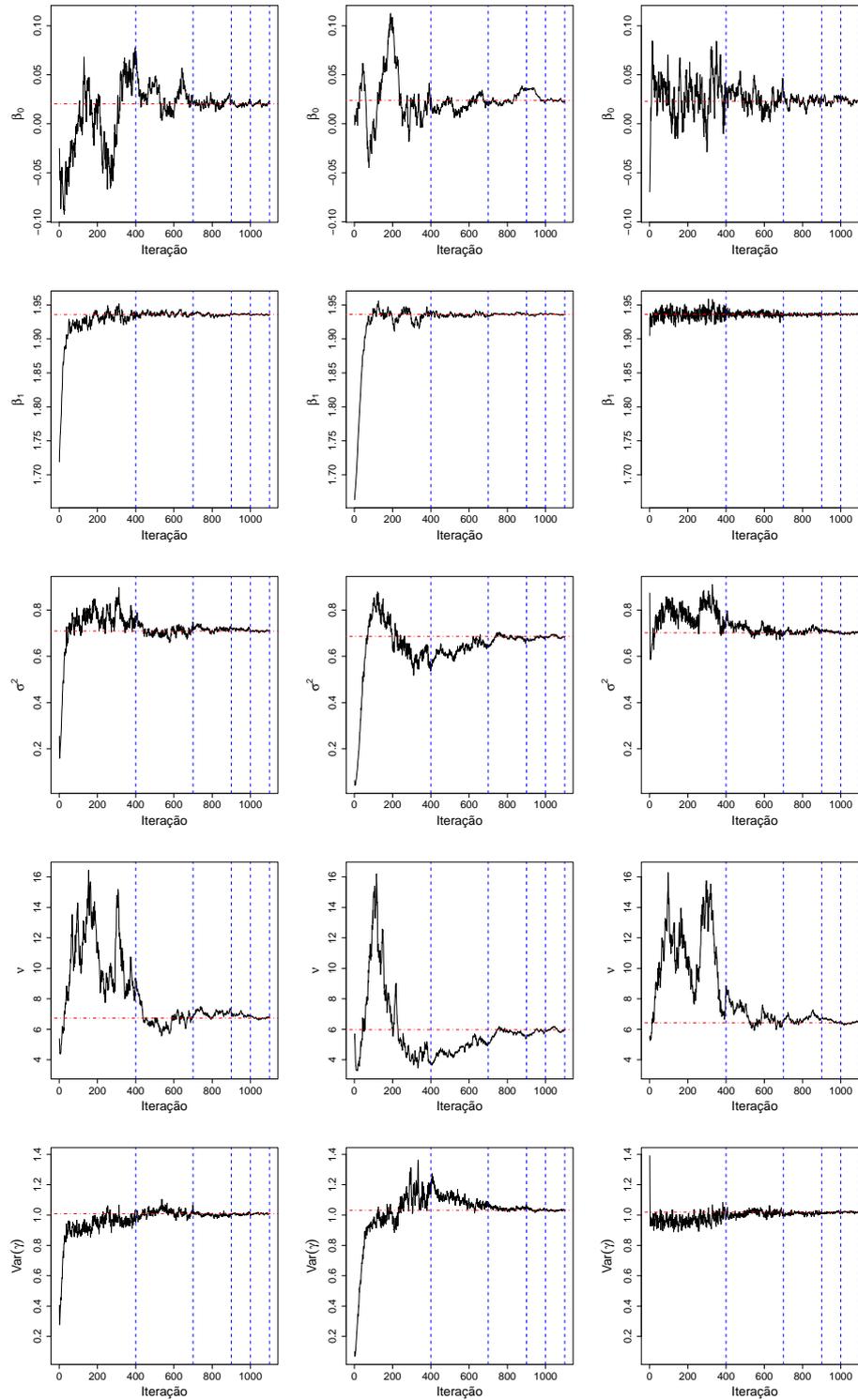


Figura 85: Gráficos das estimativas dos parâmetros *versus* iterações, e média das estimativas nas últimas 100 iterações (linha tracejada), obtidas com os algoritmos EMMC-M (esquerda), EMMC-PA (centro) e EMMC-Q (direita) para Dados 8 e Cenário 1.

Gráficos referentes a Subseção 3.6.2

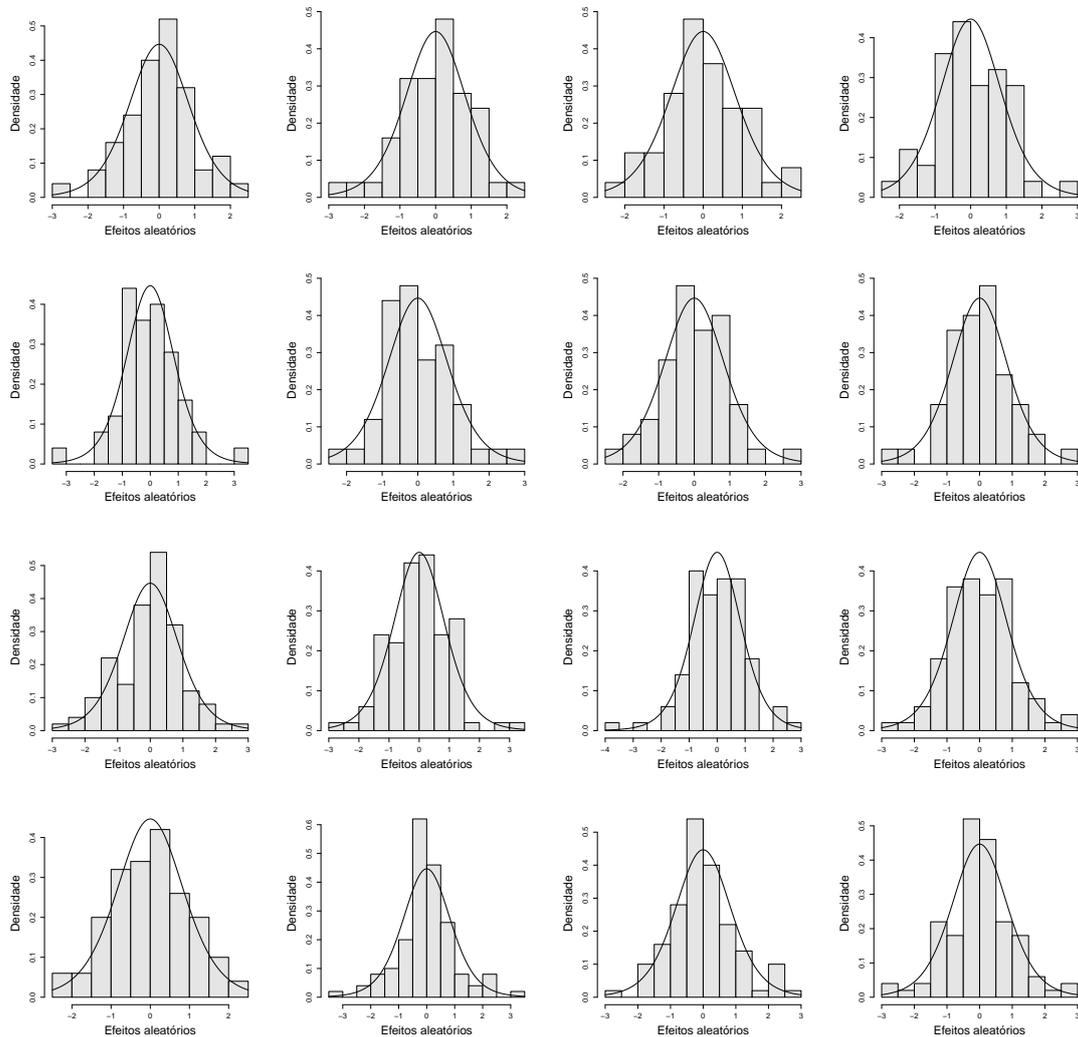


Figura 86: Histograma e f.d.p teórica dos efeitos aleatórios gerados no Cenário 2 para Dados 13 a 28 (da esquerda para direita).

Gráficos referentes a Subseção 3.6.3

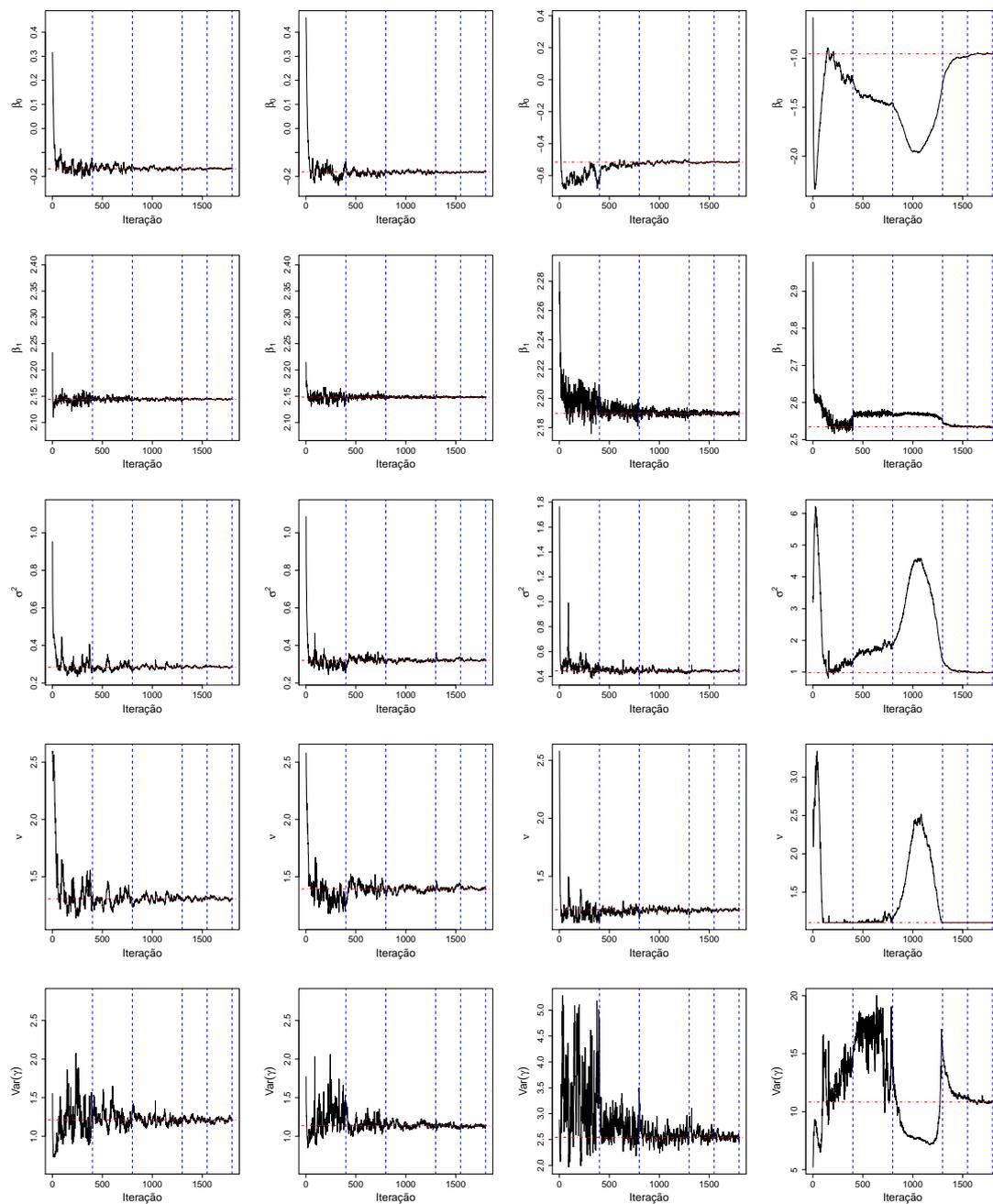


Figura 87: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-Q com constante $c = 0,1$ (primeira coluna), $0,3$ (segunda coluna), $0,5$ (terceira coluna) e 1 (quarta coluna) e média das estimativas nas últimas 100 iterações (linha tracejada) com Dados 1 no Cenário 3.

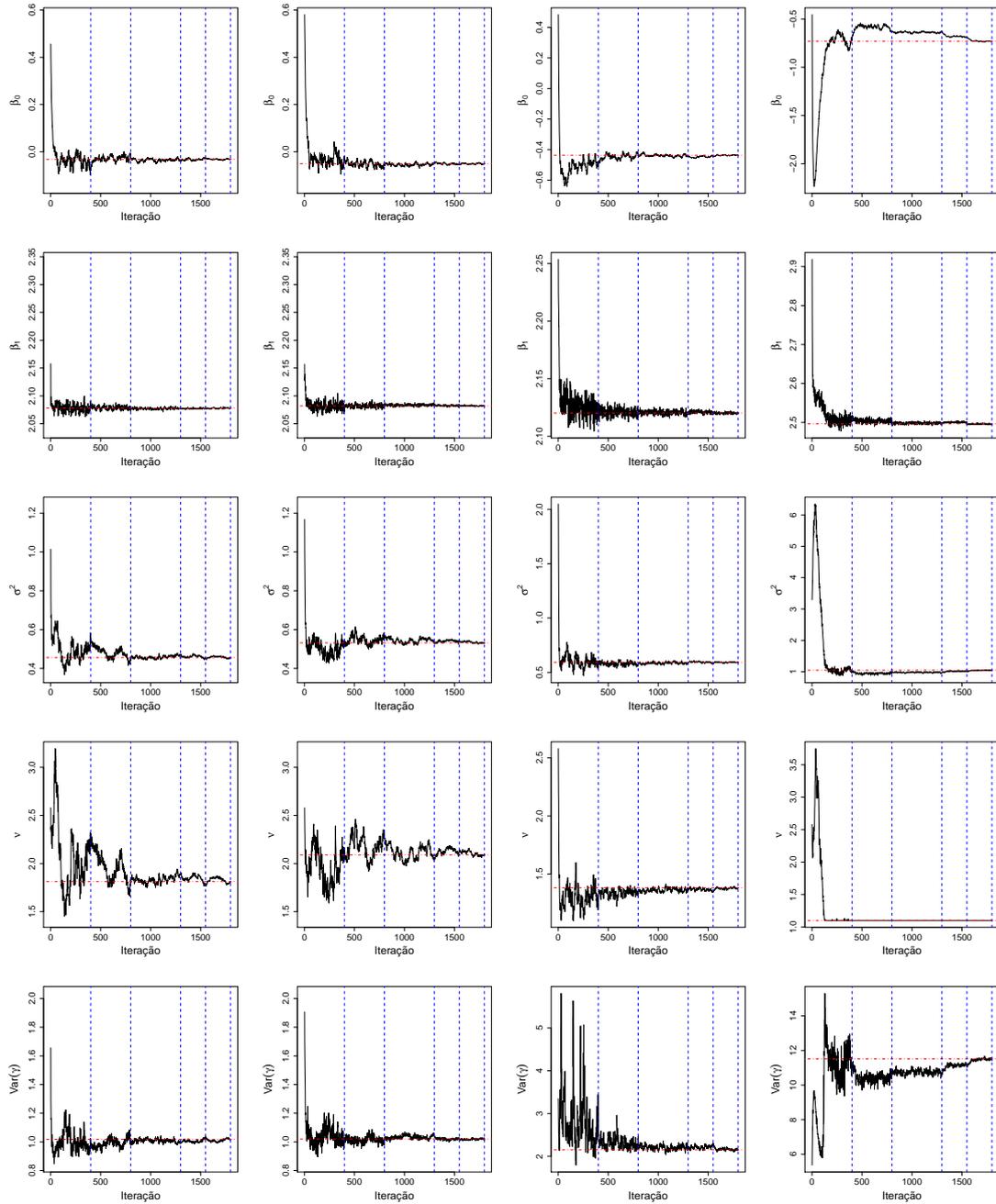


Figura 88: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMM-Q com constante $c = 0,1$ (primeira coluna), $0,3$ (segunda coluna), $0,5$ (terceira coluna) e 1 (quarta coluna) e média das estimativas nas últimas 100 iterações (linha tracejada) com Dados 2 no Cenário 3.

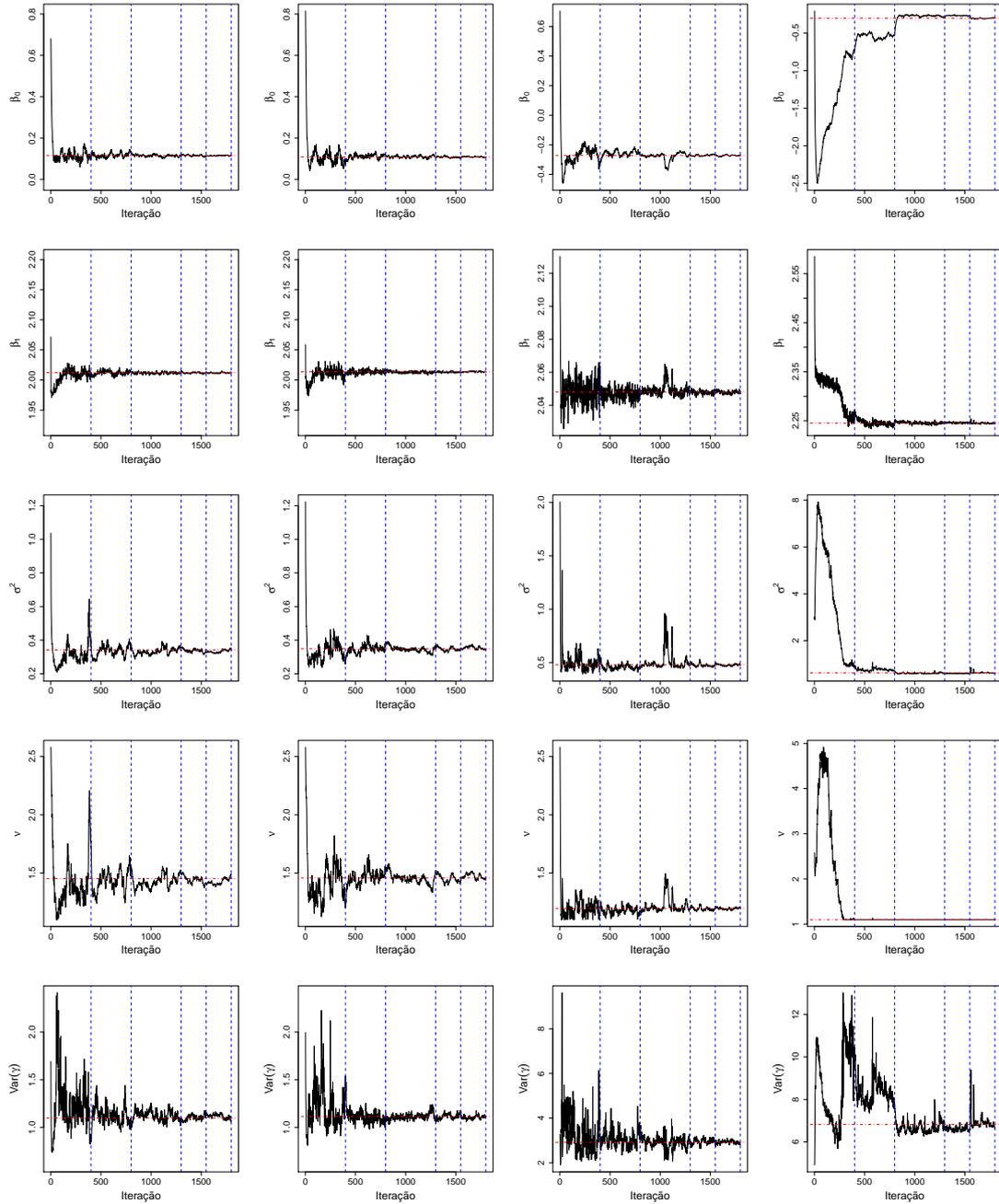


Figura 89: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-Q com constante $c = 0,1$ (primeira coluna), $0,3$ (segunda coluna), $0,5$ (terceira coluna) e 1 (quarta coluna) e média das estimativas nas últimas 100 iterações (linha tracejada) com Dados 3 no Cenário 3.

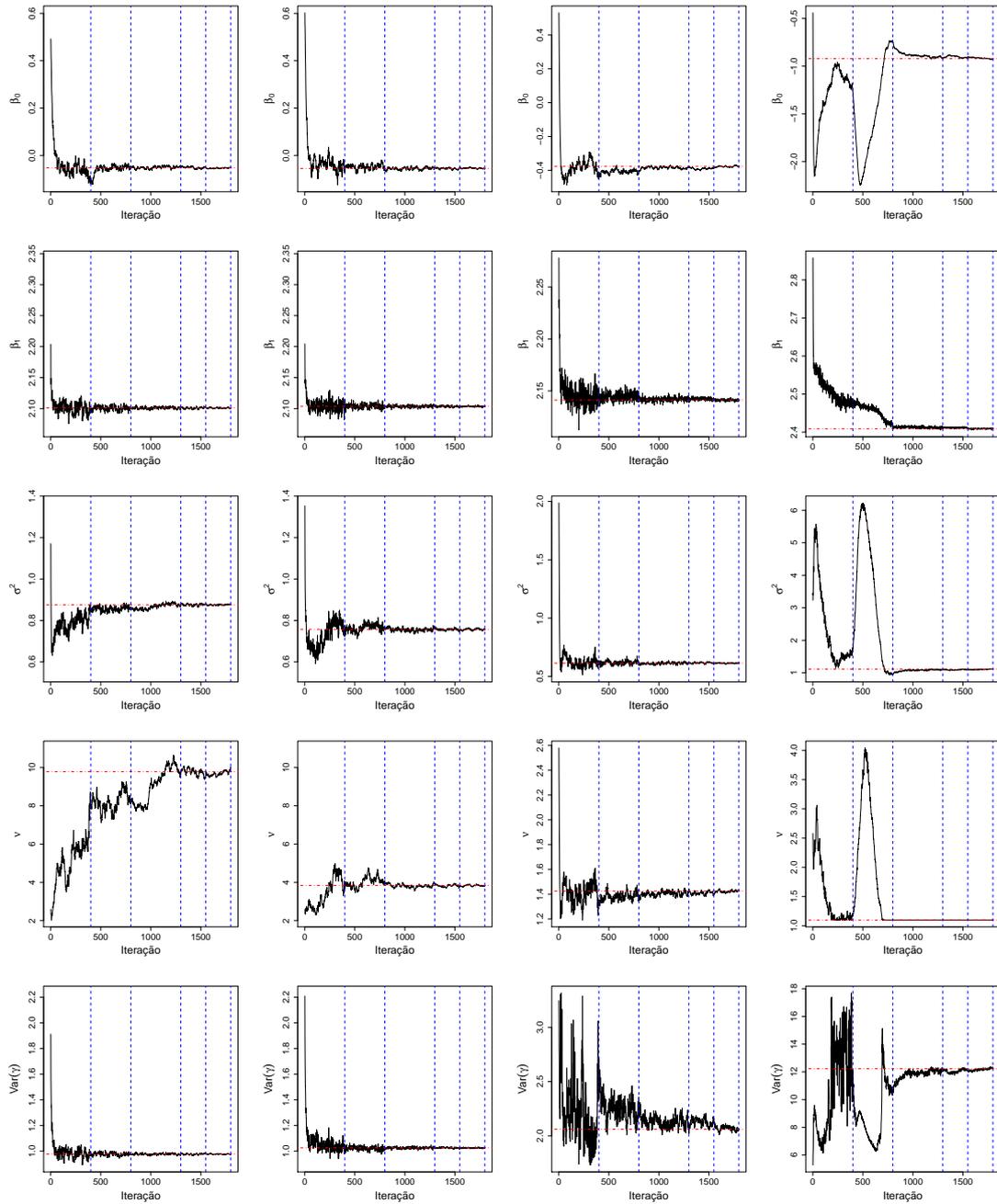


Figura 90: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-Q com constante $c = 0,1$ (primeira coluna), $0,3$ (segunda coluna), $0,5$ (terceira coluna) e 1 (quarta coluna) e média das estimativas nas últimas 100 iterações (linha tracejada) com Dados 6 no Cenário 3.

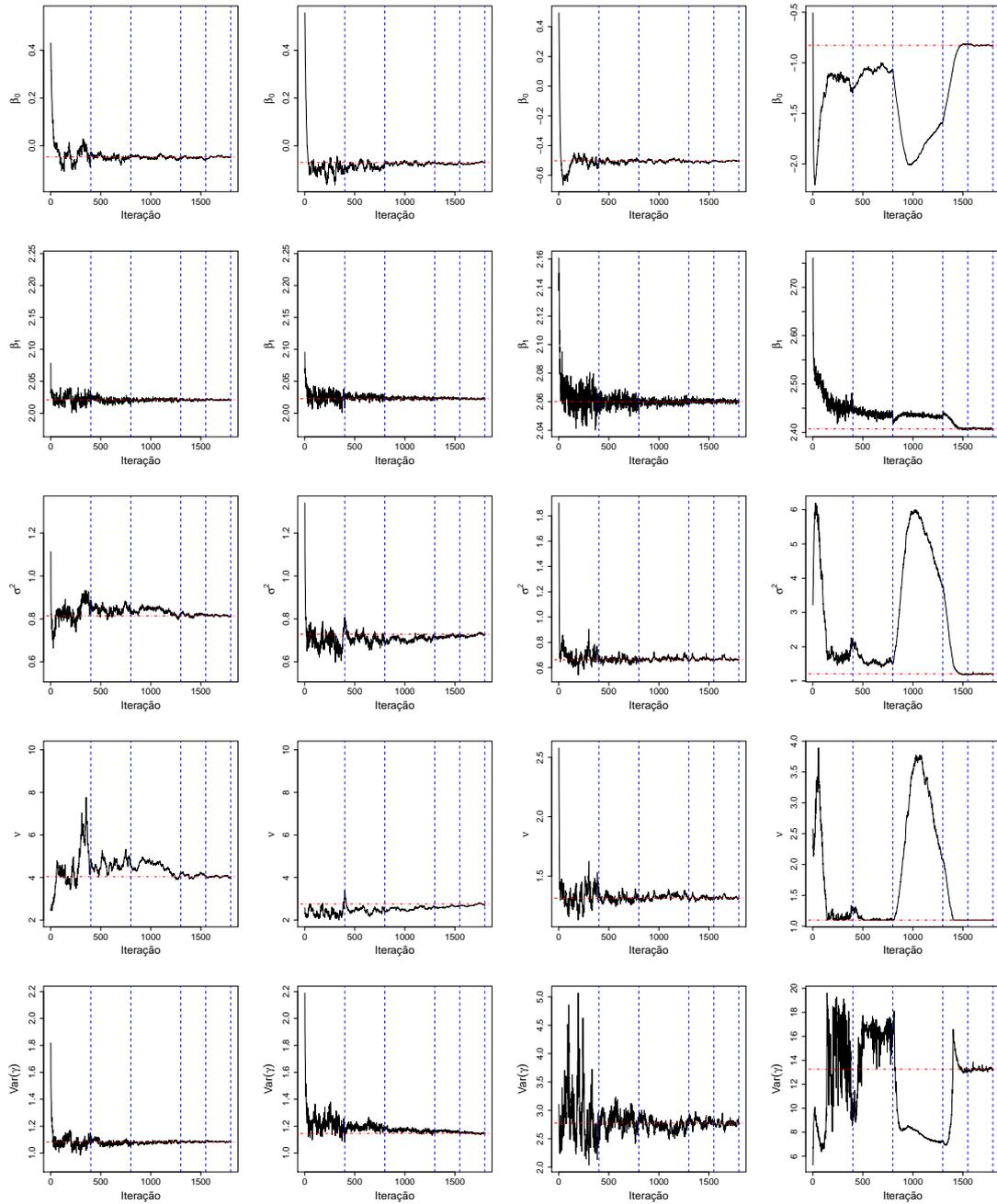


Figura 91: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-Q com constante $c = 0,1$ (primeira coluna), $0,3$ (segunda coluna), $0,5$ (terceira coluna) e 1 (quarta coluna) e média das estimativas nas últimas 100 iterações (linha tracejada) com Dados 7 no Cenário 3.

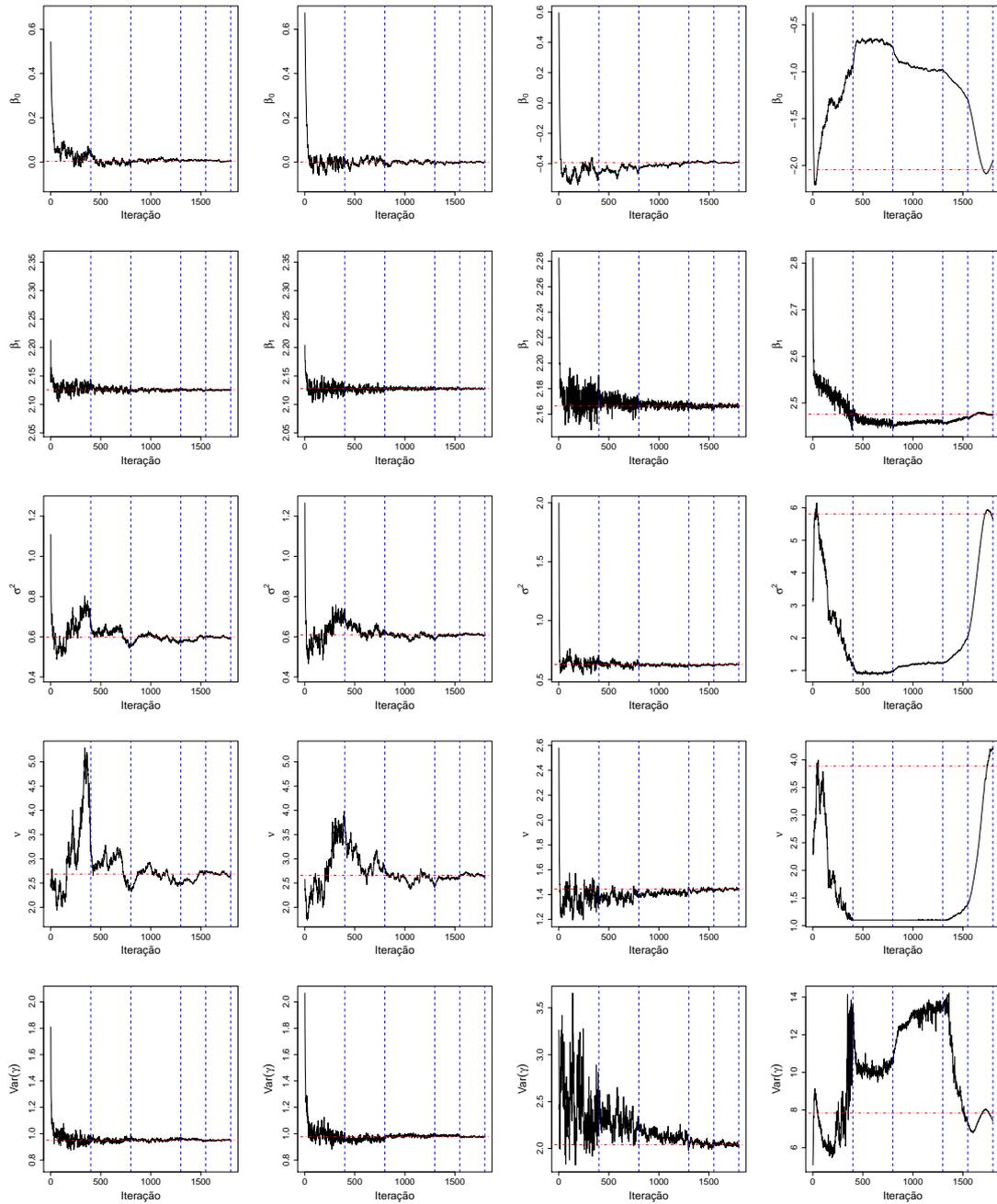


Figura 92: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-Q com constante $c = 0,1$ (primeira coluna), $0,3$ (segunda coluna), $0,5$ (terceira coluna) e 1 (quarta coluna) e média das estimativas nas últimas 100 iterações (linha tracejada) com Dados 8 no Cenário 3.

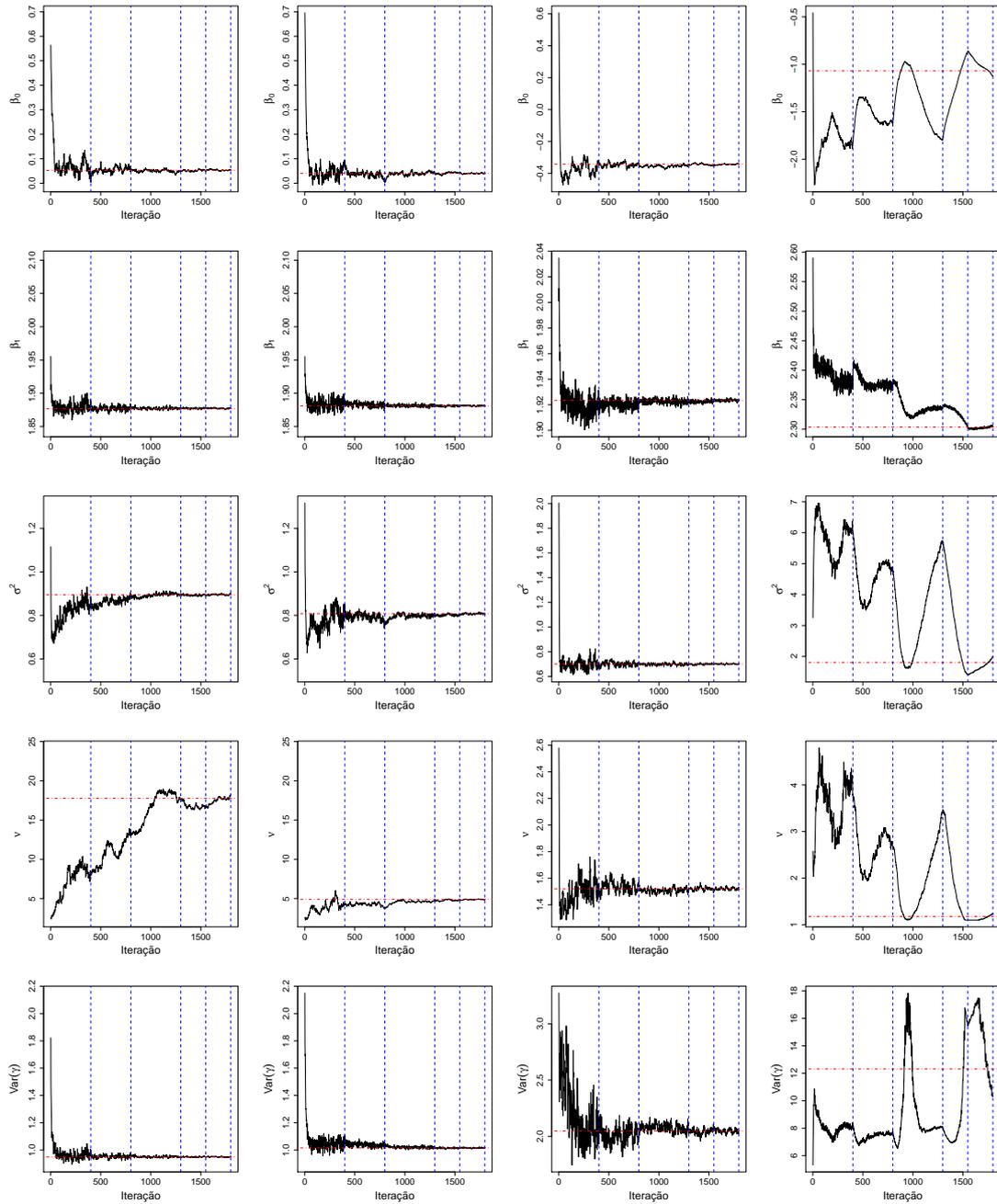


Figura 93: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-Q com constante $c = 0,1$ (primeira coluna), $0,3$ (segunda coluna), $0,5$ (terceira coluna) e 1 (quarta coluna) e média das estimativas nas últimas 100 iterações (linha tracejada) com Dados 10 no Cenário 3.

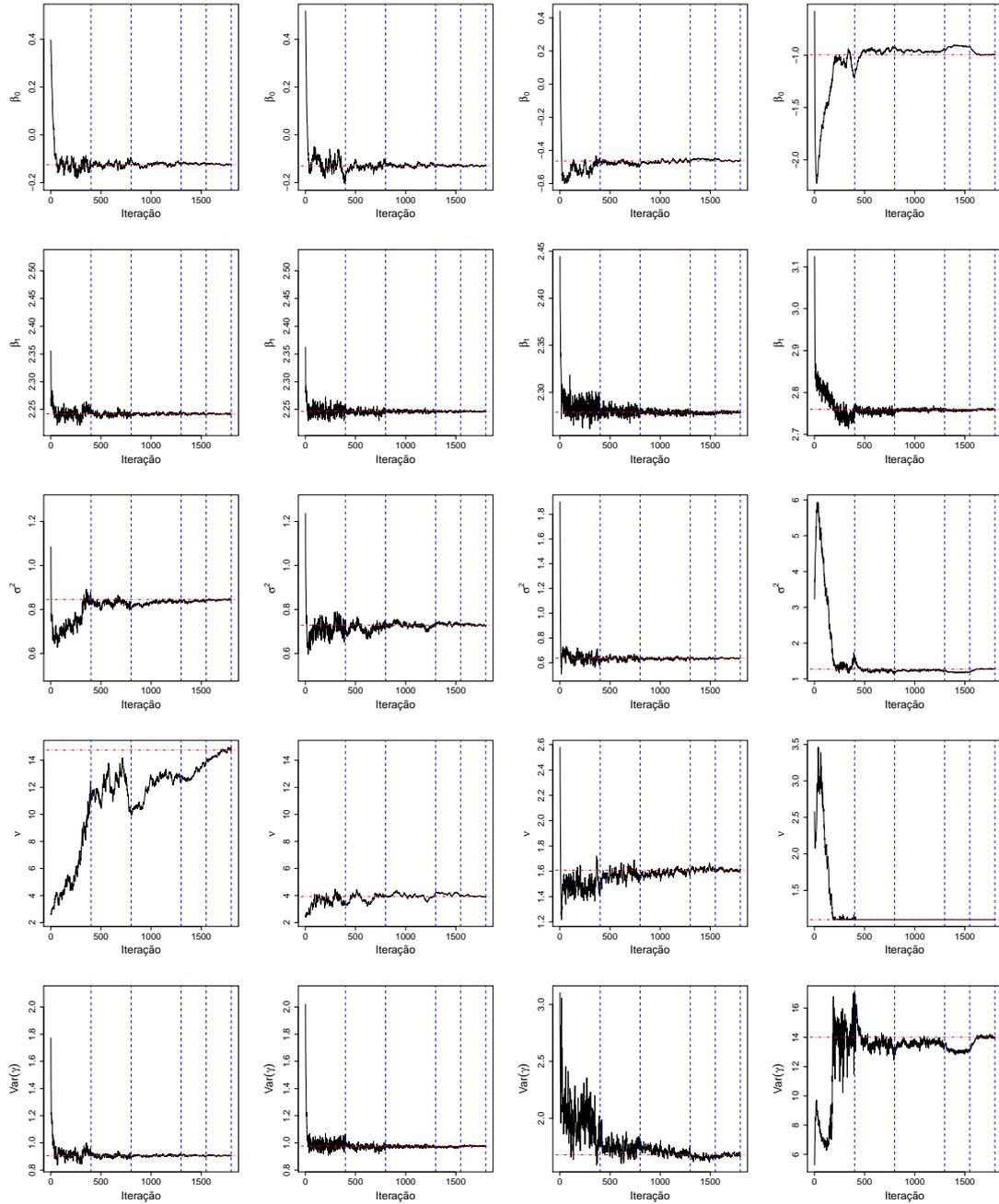


Figura 94: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-Q com constante $c = 0,1$ (primeira coluna), $0,3$ (segunda coluna), $0,5$ (terceira coluna) e 1 (quarta coluna) e média das estimativas nas últimas 100 iterações (linha tracejada) com Dados 11 no Cenário 3.

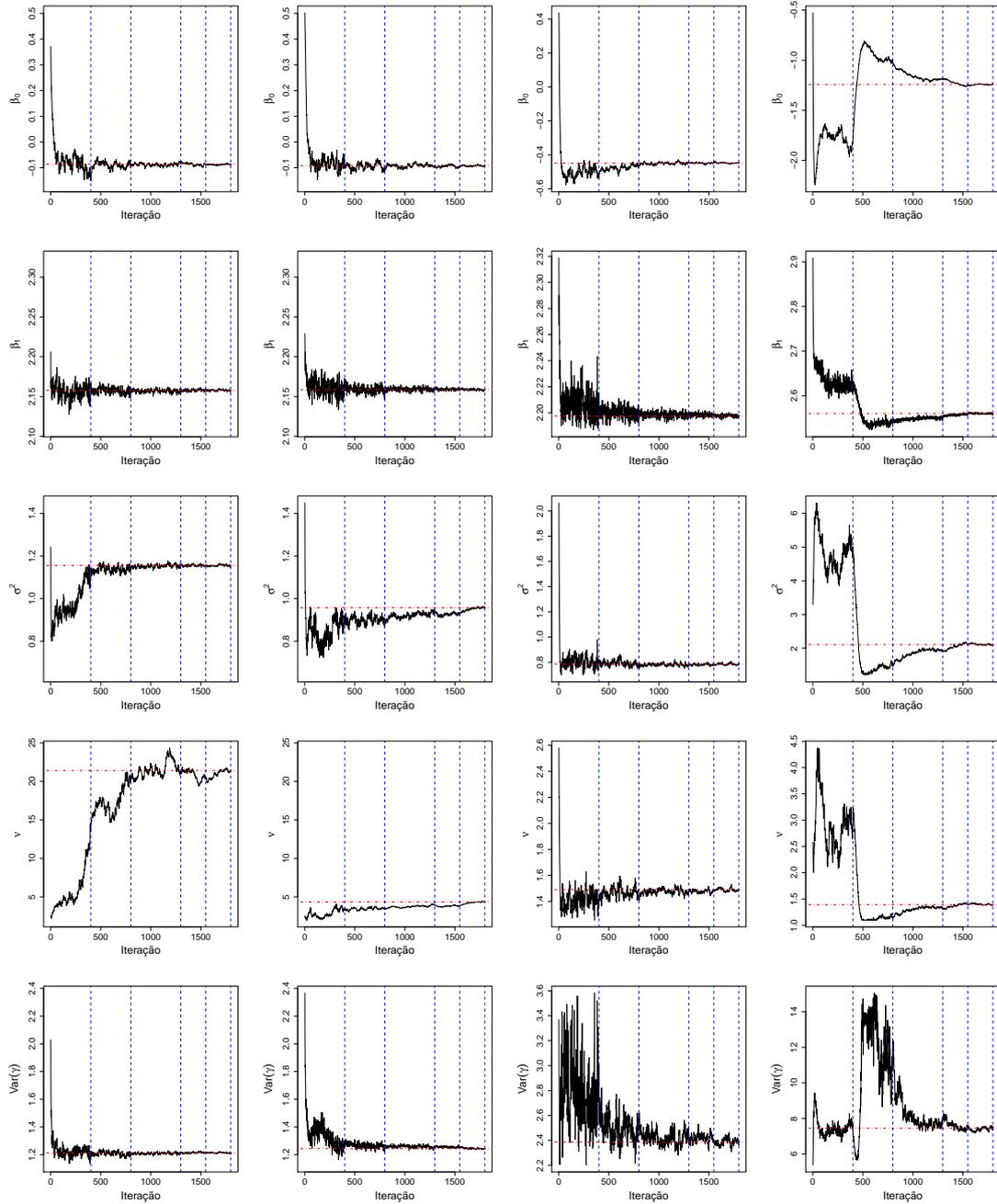


Figura 95: Gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-Q com constante $c = 0,1$ (primeira coluna), $0,3$ (segunda coluna), $0,5$ (terceira coluna) e 1 (quarta coluna) e média das estimativas nas últimas 100 iterações (linha tracejada) com Dados 12 no Cenário 3.

Apêndice F

Gráficos do Estudo Monte Carlo da Seção 3.7

No Apêndice F apresentamos gráficos referentes a Seção 3.7. São apresentados alguns *boxplots* e histogramas para as estimativas de ν , obtidas sob os modelos Logístico-T e Logístico-SL, nos cenários em que os bancos de dados são gerados sob os modelos Logístico-T (Subseção 3.7.2), Logístico-SL (Subseção 3.7.3) e Logístico-NC (Subseção 3.7.4). São apresentados também os histogramas para as estimativas de ν_1 obtidas sob o modelo Logístico-NC nos cenários em que os dados são gerados sob este modelo (Subseção 3.7.4).

Gráficos referentes a Subseção 3.7.2

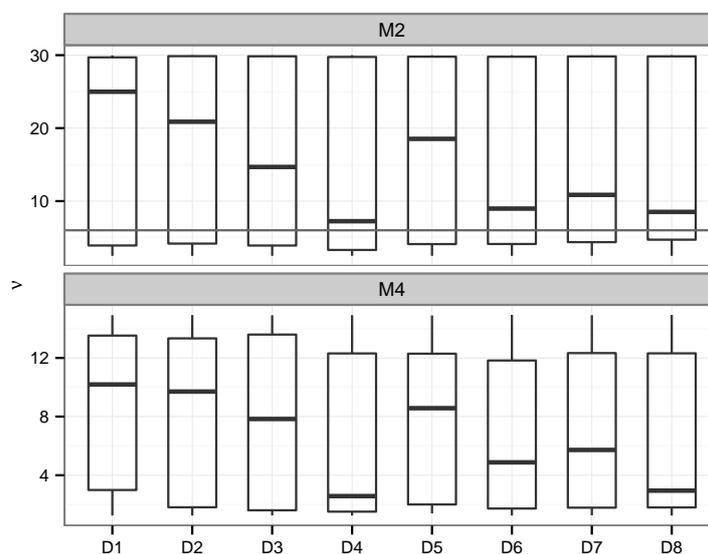


Figura 96: *Boxplots* das estimativas de ν e valor real (linha horizontal), sob os Modelos M2 e M4, nos Cenários D1 a D8.

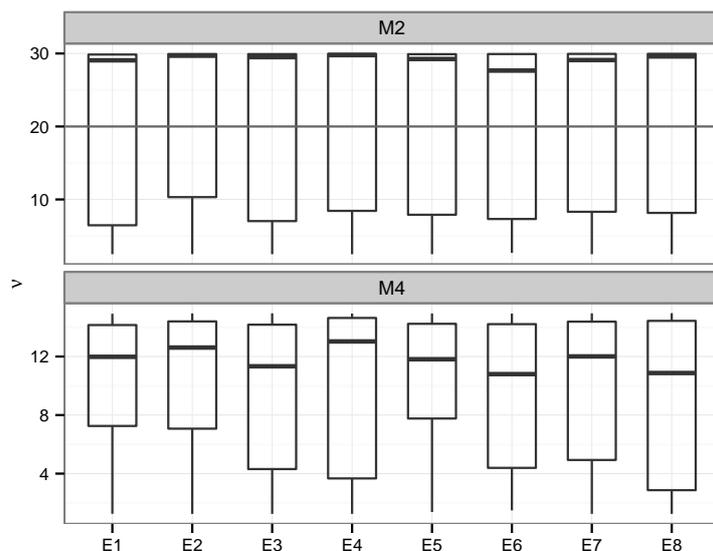


Figura 97: *Boxplots* das estimativas de ν e valor real (linha horizontal), sob os Modelos M2 e M4, nos Cenários E1 a E8.

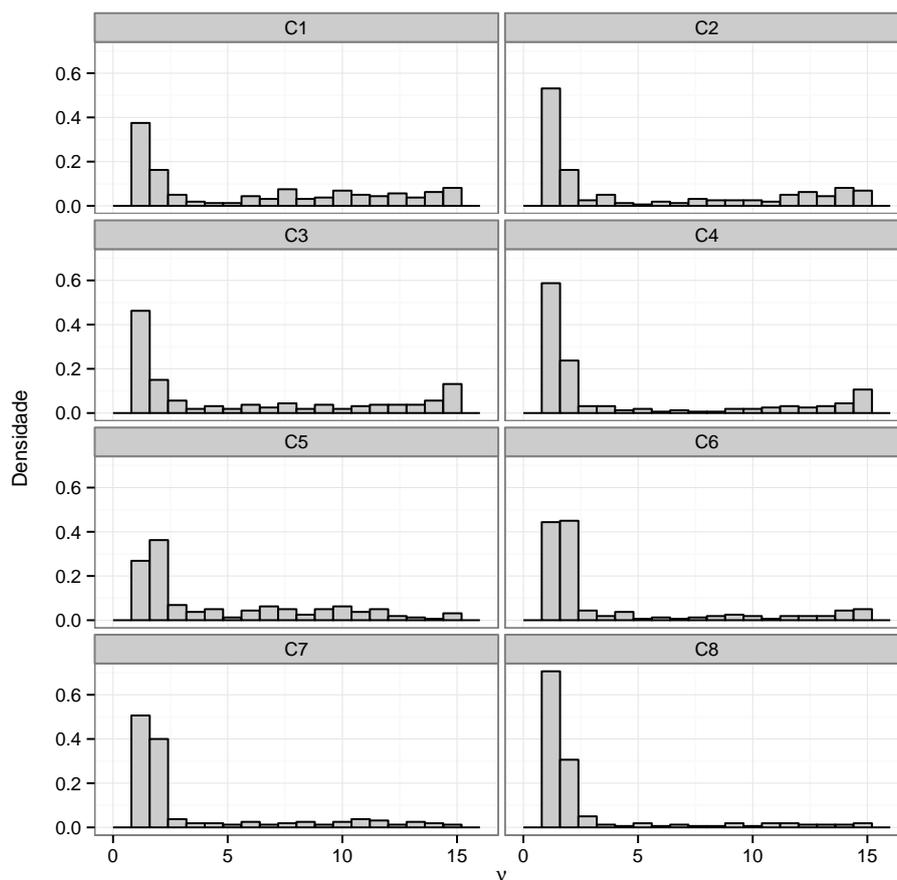


Figura 98: Histogramas das estimativas fornecidas para ν , sob o Modelo M4, nos Cenários C1 a C8.

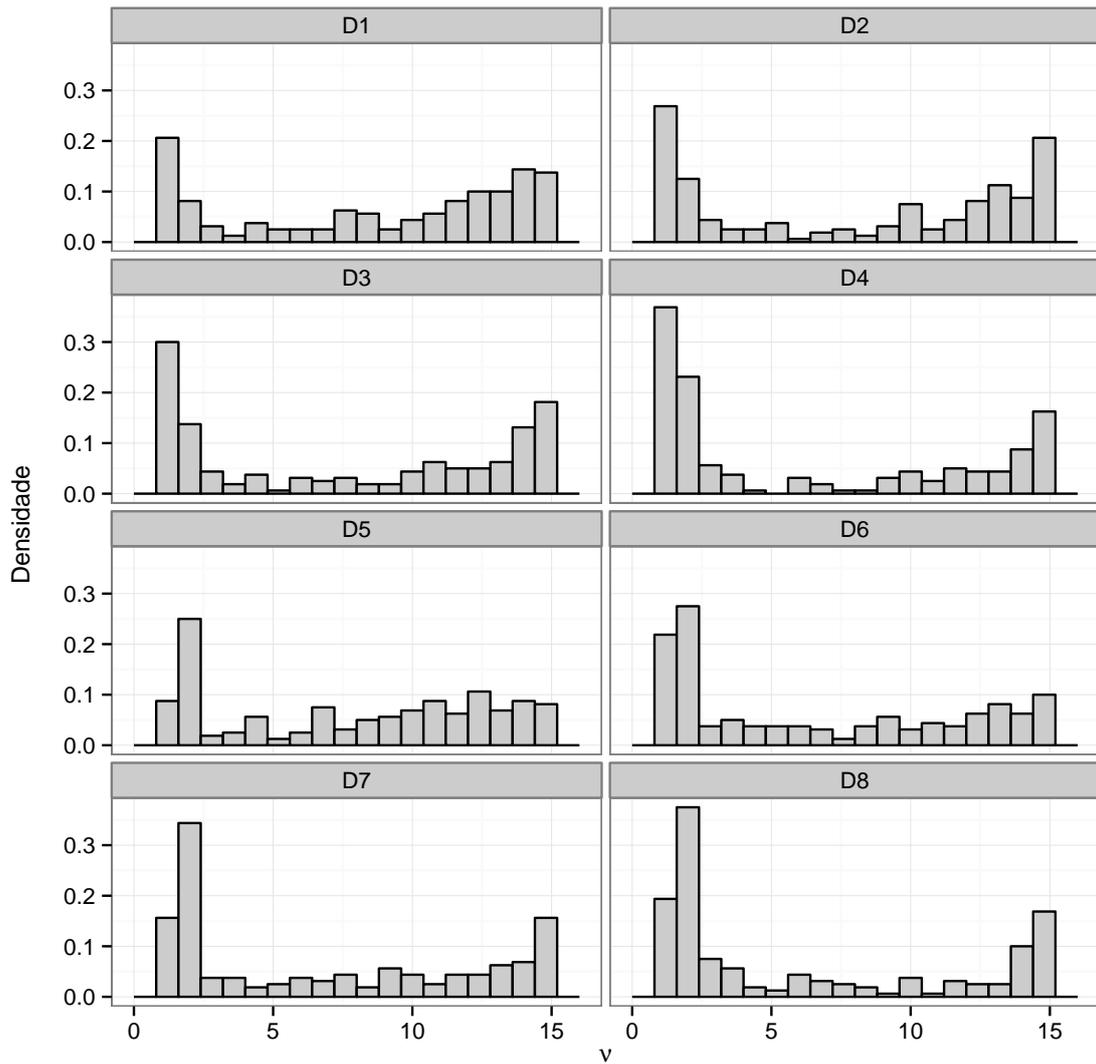


Figura 99: Histogramas das estimativas fornecidas para ν , sob o Modelo M4, nos Cenários D1 a D8.

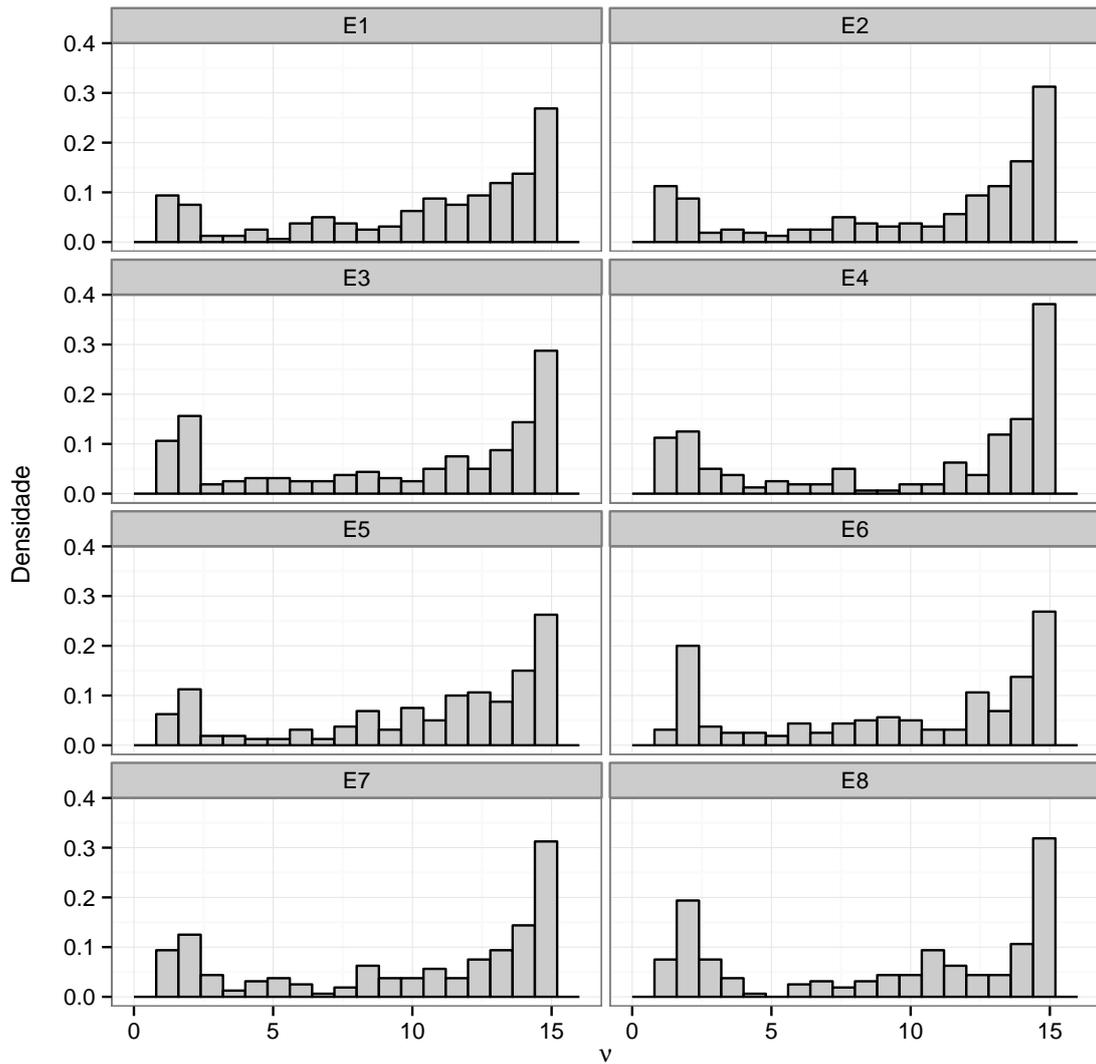


Figura 100: Histogramas das estimativas fornecidas para ν , sob o Modelo M4, nos Cenários E1 a E8.

Gráficos referentes a Subseção 3.7.3

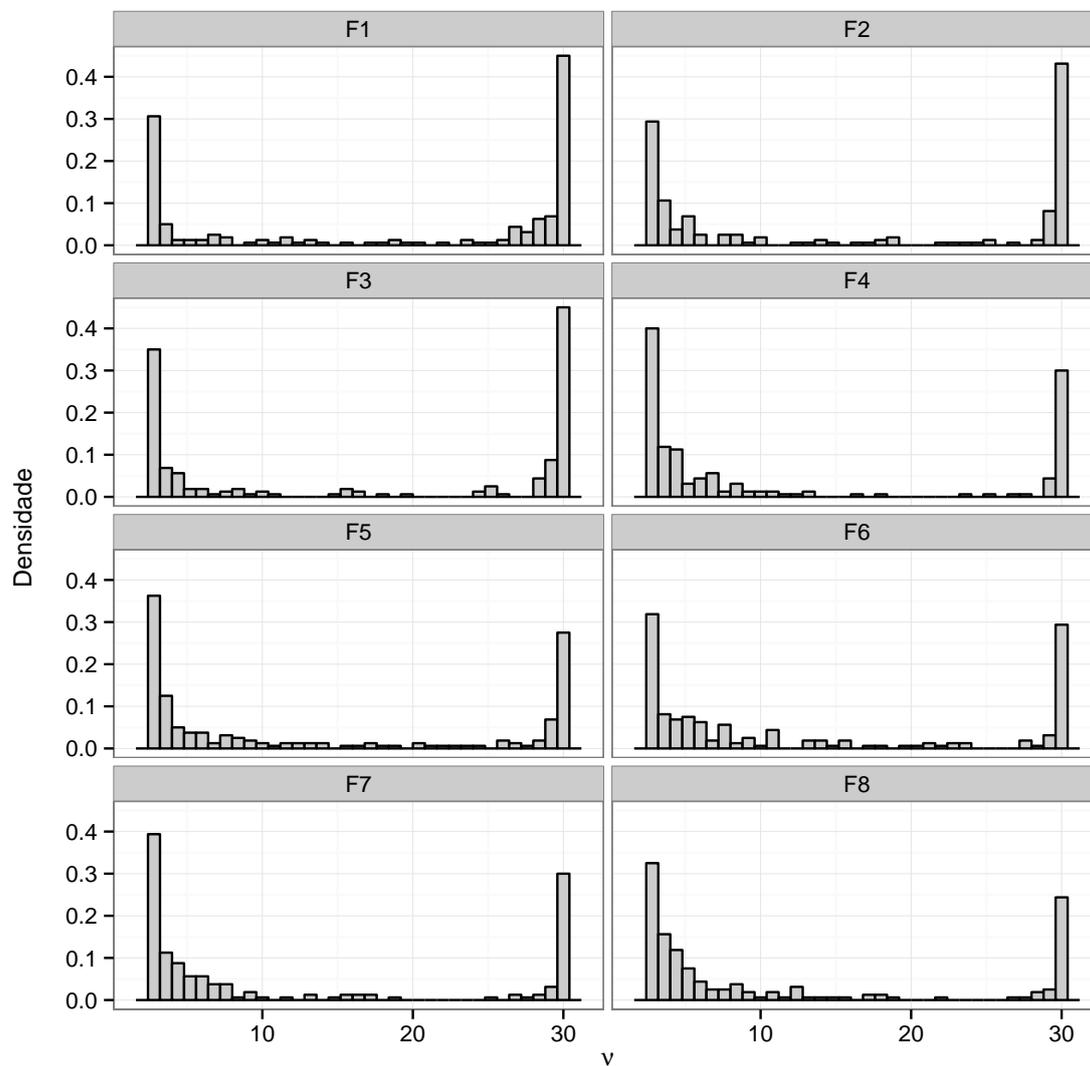


Figura 101: Histogramas das estimativas fornecidas para ν , sob o Modelo M2, nos Cenários F1 a F8.

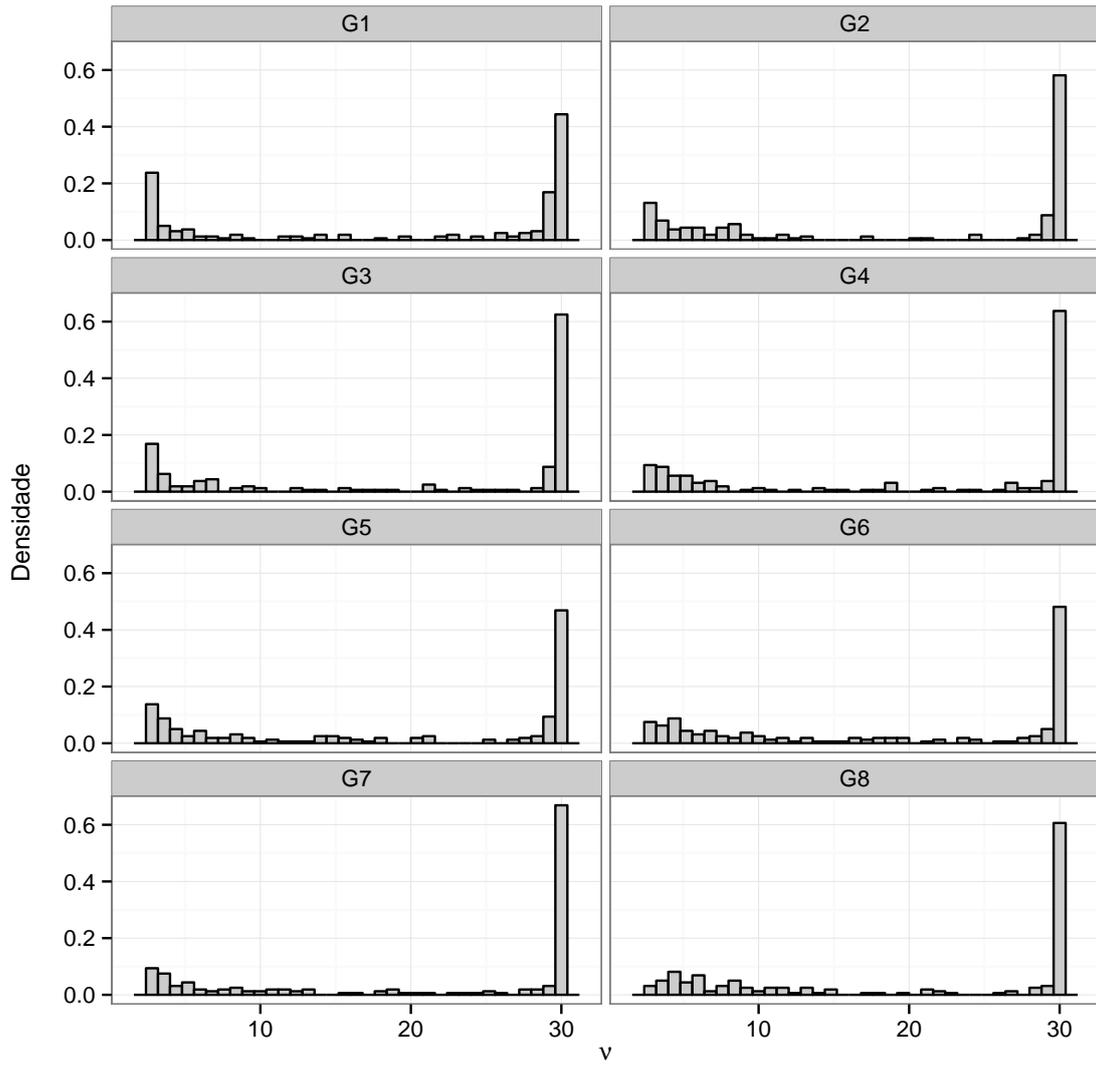


Figura 102: Histogramas das estimativas fornecidas para ν , sob o Modelo M2, nos Cenários G1 a G8.

Gráficos referentes a Subseção 3.7.4

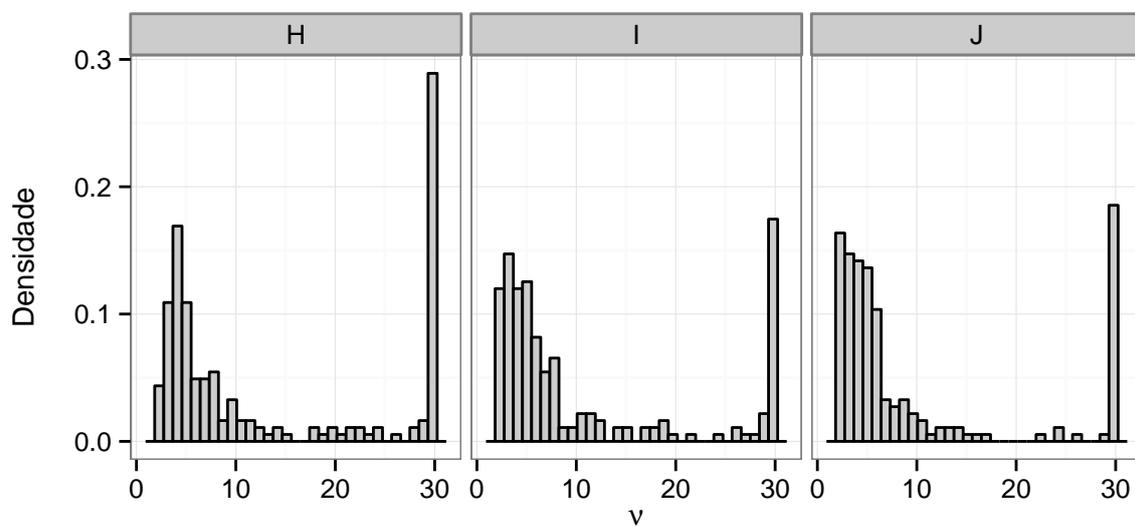


Figura 103: Histogramas das estimativas fornecidas para ν , sob o Modelo M2, nos Cenários H, I e J.

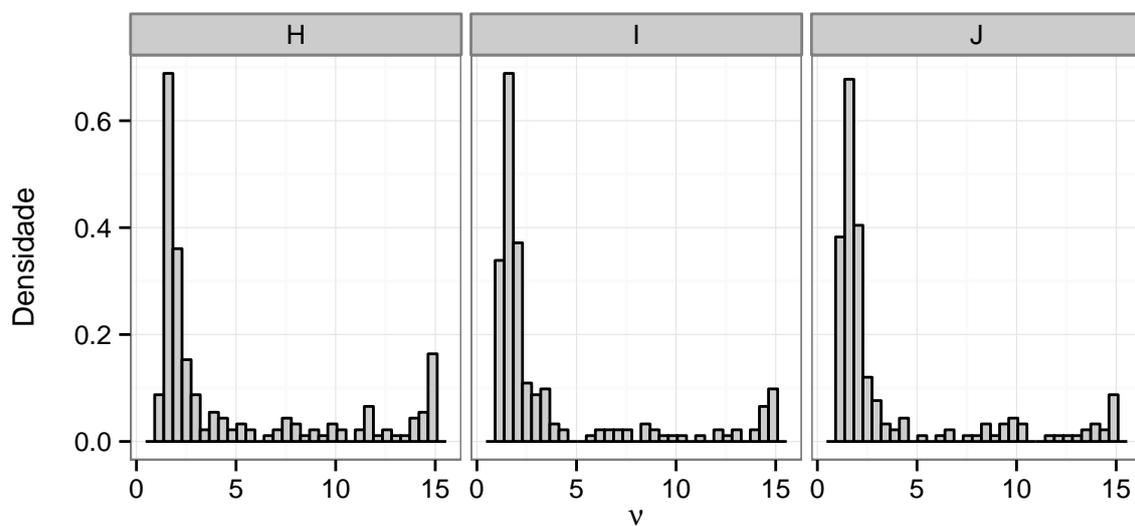


Figura 104: Histogramas das estimativas fornecidas para ν , sob o Modelo M3, nos Cenários H, I e J.

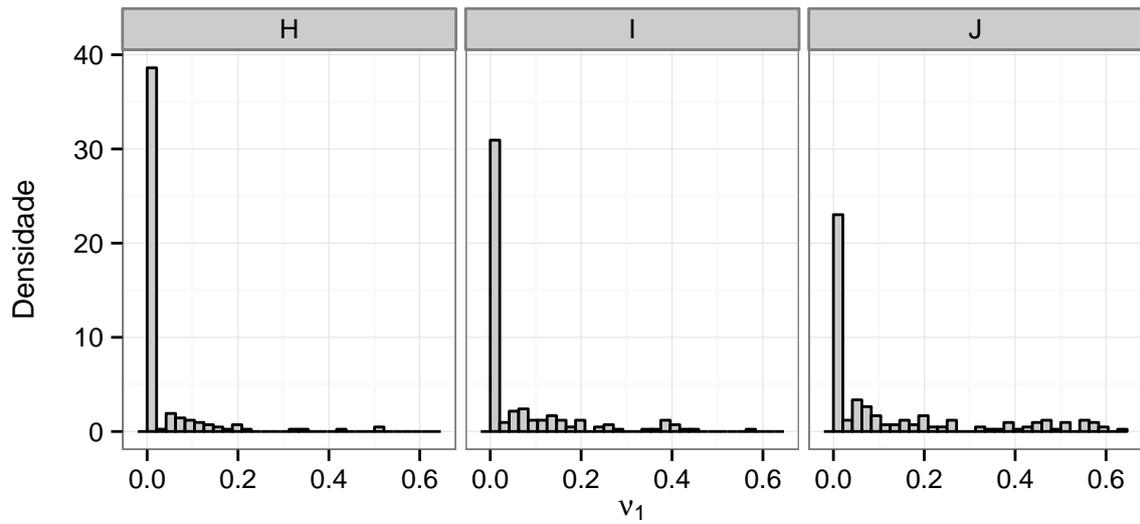


Figura 105: Histogramas das estimativas fornecidas para ν_1 , sob o Modelo M4, nos Cenários H, I e J.

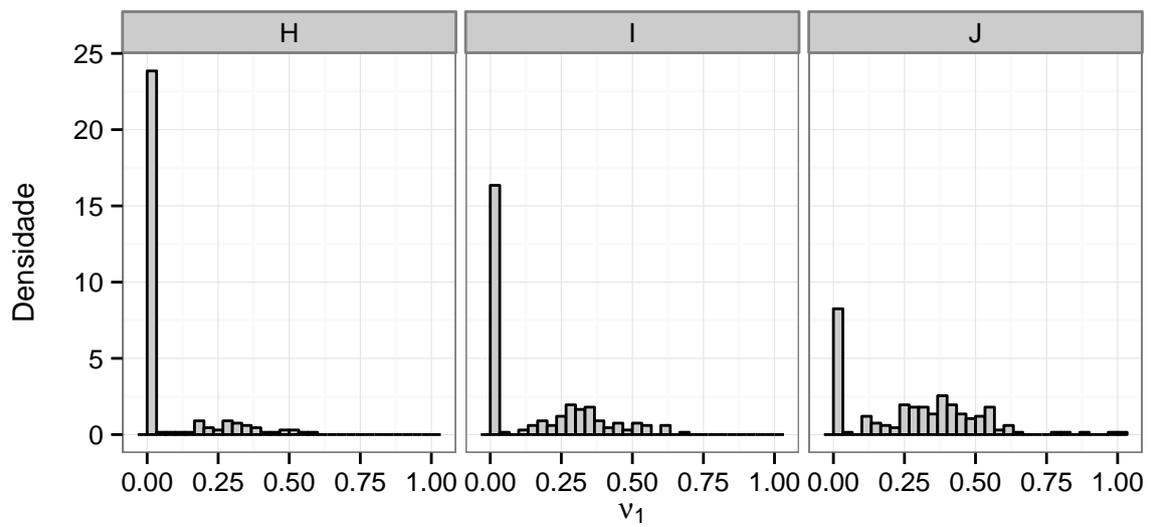


Figura 106: Histogramas das estimativas fornecidas para ν_1 , sob o Modelo M6, nos Cenários H, I e J.

Apêndice G

Gráficos da análise de dados reais da Seção 3.8

Apresentamos no Apêndice G os gráficos referentes a Seção 3.8. São apresentados os gráficos das estimativas dos parâmetros *versus* iterações do algoritmo EMMC-Q para dados de *Ascaris suum* sob os modelos Logístico-N, Logístico-T e Logístico-SL, sendo que para os dois últimos o parâmetro ν é estimado.

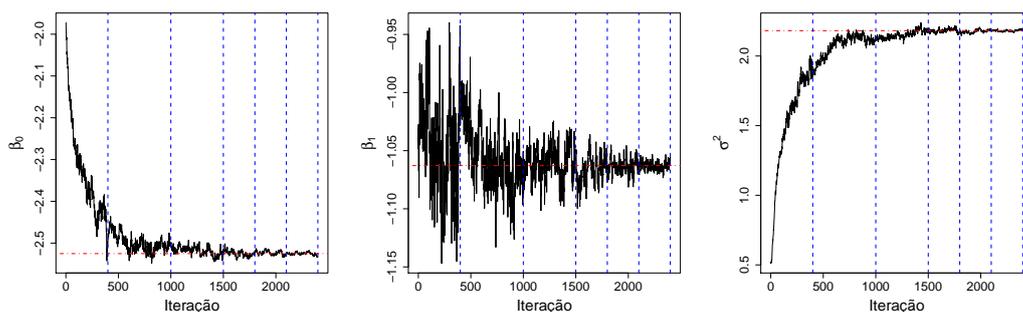


Figura 107: Gráficos das estimativas dos parâmetros sob o modelo Logístico-N *versus* iterações do algoritmo EMMC-Q e média das estimativas nas últimas 300 iterações (linha tracejada) com para dados de *Ascaris suum*.

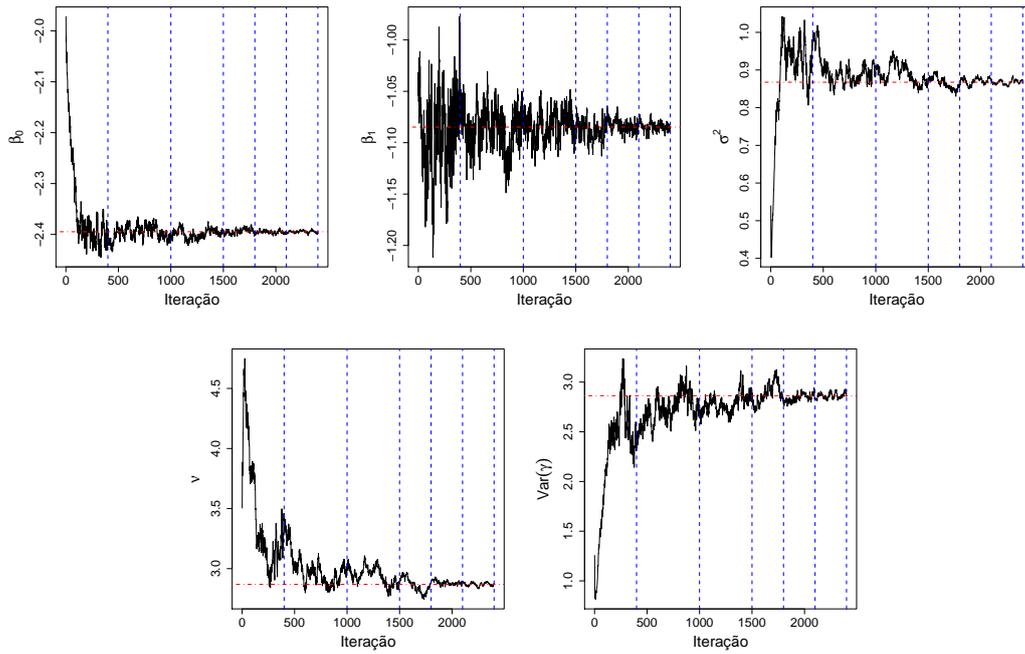


Figura 108: Gráficos das estimativas dos parâmetros sob o modelo Logístico-T *versus* iterações do algoritmo EMMC-Q e média das estimativas nas últimas 300 iterações (linha tracejada) com para dados de *Ascaris suum*.

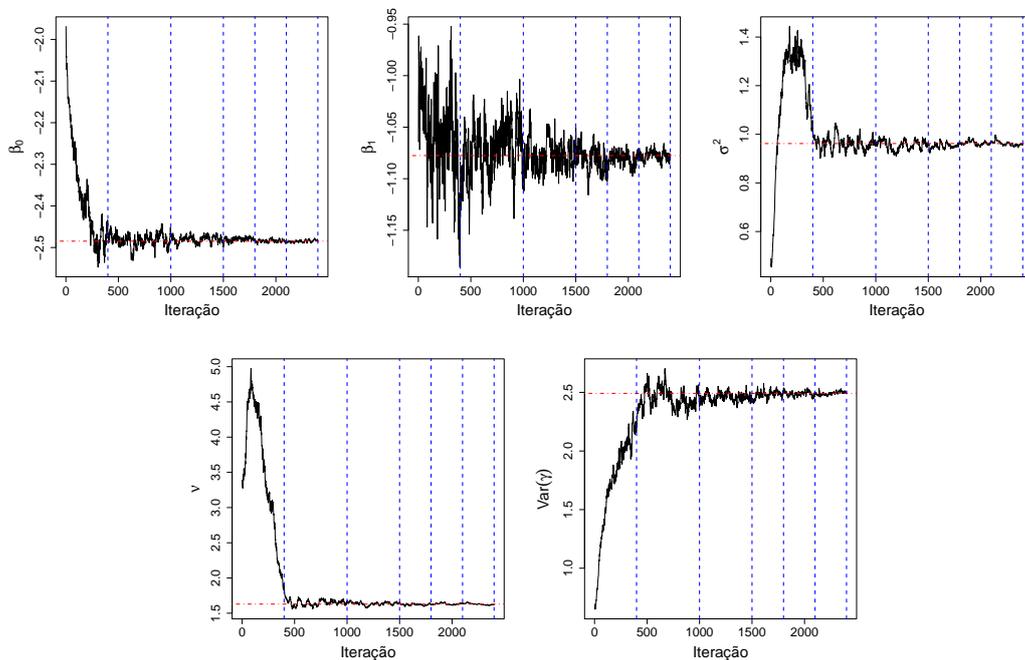


Figura 109: Gráficos das estimativas dos parâmetros sob o modelo Logístico-SL *versus* iterações do algoritmo EMMC-Q e média das estimativas nas últimas 300 iterações (linha tracejada) com para dados de *Ascaris suum*.

Referências Bibliográficas

- Andrews, D. F. e Mallows, C. L. (1974). Scale mixtures of normal distributions, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **36**(1): 99–102.
- Arellano-Valle, R. B., Castro, L. M. e Loschi, R. H. (2013). Change point detection in the skew-normal model parameters, *Communications in Statistics - Theory and Methods* **42**: 603–618.
- Arellano-Valle, R. B., Genton, M. G. e Loschi, R. H. (2009). Shape mixtures of multivariate skew-normal distributions, *Journal of Multivariate Analysis* **100**: 91–101.
- Arellano-Valle, R. e Azzalini, A. (2006). On the unification of families of skew-normal distributions, *Scandinavian Journal of Statistics* **33**: 561–574.
- Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics* **12**: 171–178.
- Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones, *Statistica* **46**: 199–208.
- Azzalini, A. (2005). The skew-normal distribution and related multivariate families, *Scandinavian Journal of Statistics* **32**: 159–188.
- Azzalini, A. e Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **61**: 579–602.
- Azzalini, A. e Dalla-Valle, A. (1996). The multivariate skew-normal distribution, *Biometrika* **83**: 715–726.
- Booth, J. G. e Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated monte carlo em algorithm, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **61**(1): 265–285.
- Caffo, B. S., Jank, W. e Jones, G. L. (2005). Ascent-based monte carlo expectation-maximization, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(2): 235–251.
- Celeux, G. e Diebolt, J. (1985). The SEM algorithm: a probabilistic teacher algorithm derived from EM algorithm for the mixture problem, *Computational Statistics Quarterly* **2**: 73–82.

- Celeux, G., Forbes, F., Robert, C. P. e Titterton, D. M. (2006). Deviance information criteria for missing data models, *Bayesian Analysis* **1**: 651–674.
- Chan, J. S. K. e Kuk, A. Y. C. (1997). Maximum likelihood estimation for probit-linear mixed models with correlated random effects, *Biometrics* **53**: 86–97.
- Chen, J., Zhang, D., e Davidian, M. (2002). A Monte Carlo EM algorithm for generalized linear mixed models with flexible random effects distributions, *Biostatistics* **3**: 347–360.
- de Queiroz, M. M., Loschi, R. H. e Silva, R. W. C. (2015). Multivariate log-skewed distributions with normal kernel and their applications, *Statistics* (ahead-of-print): 1–19.
- Dempster, A. P., Laird, N. M. e Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society: Series B (Methodological)* **39**: 1–38.
- Dey, D. K., Chen, M. H. e Chang, H. (1997). Bayesian approach for nonlinear random effects models, *Biometrics* **53**: 1239–1252.
- Doornik, J. A. (2007). *Ox 5 - An Object-oriented Matrix Programming Language*, Timberlake Consultants Ltd.
- Fang, K. T., Kotz, S. e Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*, Chapman & Hall, London/New York.
- Gamerman, D. e Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, second edn, Chapman & Hall/CRC.
- Gelfand, A. E. (1996). *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London., chapter Model determination using sampling-based methods, pp. 145–161.
- Gelfand, A. E. e Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **56**: 501–514.
- Gibbons, R. D., Hedeker, D., Charles, S. C. e Frisch, P. (1994). A random-effects probit model for predicting medical malpractice claims, *Journal of the American Statistical Association* **89**: 760–767.
- Gilks, W. R., Best, N. G. e Tan, K. K. C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling, *Applied Statistics* **44**: 455–472.
- Gilks, W. R. e Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling, *Applied Statistics* **41**: 337–348.
- Gilks, W. R., Richardson, S. e Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics*, Chapman & Hall.
- Greven, S. e Kneib, T. (2009). On the behavior of marginal and conditional akaike information criteria in linear mixed models, *Technical report*, Johns Hopkins University, Dept. of Biostatistics.

- Henze, N. (1986). A probabilistic representaton of the ‘skew-normal’ distribution, *Scandinavian Journal of Statistics* **13**: 271–275.
- Jank, W. (2005). Stochastic variants of EM: Monte Carlo, Quasi-Monte Carlo and more, *Proceedings of the American Statistical Association*, Mineapolis, Minnesota, pp. 1–6.
- Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization, *Sankhya A* **32**: 419–430.
- Kuhn, E. e Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models, *Computational Statistics & Data Analysis* **49**: 1020–1038.
- Laird, N. M. e Ware, J. H. (1982). Random-effects models for longitudinal data, *Biometrics* **38**: 963–974.
- Lange, K. e Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression, *Journal of Computational and Graphical Statistics* **2**(2): 175–198.
- Larsen, K. e Merlo, J. (2005). Appropriate assessment of neighborhood effects on individual health: Integrating random and fixed effects in multilevel logistic regression, *American Journal of Epidemiology* **161**: 81–88.
- Larsen, K., Petersen, J. H., Budtz-Jørgensen, E. e Endahl, L. (2000). Interpreting parameters in the logistic regression model with random effects, *Biometrics* **56**: 909–914.
- Levine, R. A. e Casella, G. (2001). Implementations of the Monte carlo EM algorithm, *Journal of Computational and Graphical Statistics* **10**(3): 422–439.
- Levine, R. A. e Fan, J. (2004). An automated (markov chain) monte carlo em algorithm, *Journal of Statistical Computation and Simulation* **74**(5): 349–360.
- Liu, J. e Dey, D. K. (2008). Skew random effects in multilevel binomial models: An alternative to nonparametric approach, *Statistical Modelling* **8**: 221–241.
- Liu, L. e Yu, Z. (2008). A likelihood reformulation method in non-normal random effects models, *Statistics in Medicine* **27**: 3105–3124.
- Louis, T. A. (1982). Finding the observed information matrix when using the em algorithm, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **44**: 226–233.
- Marchenko, Y. V. e Genton, M. G. (2010). Multivariate log-skew-elliptical distributions with applications to precipitation data, *Environmetrics* **21**: 318–340.
- Martins-Filho, S., Loschi, R. H. e Colosimo, E. A. (2010). Bayesian interpretation for the parameters of the mixed logistic model: A seed germination application., Unpublished manuscript.
- McCullagh, P. e Nelder, J. A. (1989). *Generalized Linear Models*, London: Chapman & Hall.

- McCulloch, C. E. (1994). Maximum likelihood variance components estimation for binary data, *Journal of the American Statistical Association* **89**: 330–335.
- McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models, *Journal of the American Statistical Association* **92**: 162–170.
- McCulloch, C. e Searle, S. R. (2001). *Generalized, Linear, and Mixed Models*, Wiley: New York.
- Neath, R. (2006). *Monte Carlo Methods for Likelihood-based Inference in Hierarchical Models*, ProQuest.
- Nelson, K. P., Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J., Parzen, M. e Strawderman, R. (2006). Use of the probability integral transformation to fit nonlinear mixed-effects models with nonnormal random effects, *Journal of Computational and Graphical Statistics* **15**: 39–57.
- Newton, M. A. e Raftery, A. E. (1994). Approximate bayesian inference with the weighted likelihood bootstrap, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **56**: 3–48.
- Paulino, C. D., Silva, G. e Achcar, J. A. (2005). Bayesian analysis of correlated misclassified binary data, *Computational Statistics & Data Analysis* **49**: 1120–1131.
- Philippe, A. (1997). Simulation of right and left truncated gamma distributions by mixtures, *Statistics and Computing* **7**(3): 173–181.
- Prates, M. O., Costa, D. R. e Lachos, V. H. (2014). Generalized linear mixed models for correlated binary data with t-link, *Statistics and Computing* **24**(6): 1111–1123.
- Quintana, F. A., Liu, J. S. e Pino, G. E. d. (1999). Monte carlo em with importance reweighting and its applications in random effects models, *Computational Statistics & Data Analysis* **29**(4): 429–444.
- R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Rogers, W. H. e Tukey, J. W. (1972). Understanding some long-tailed symmetrical distributions, *Statistica Neerlandica* **26**(3): 211–226.
- Saefken, B., Kneib, T., van Waveren, C.-S., Greven, S. *et al.* (2014). A unifying approach to the estimation of the conditional akaike information in generalized linear mixed models, *Electronic Journal of Statistics* **8**: 201–225.
- Santos, C. C. (2011). *Modelos logísticos com classes estendidas de distribuições normais para os efeitos aleatórios*, Dissertação de mestrado, Universidade Federal de Minas Gerais.
- Santos, C. C. e Loschi, R. H. (2015). *Current Trends in Bayesian Methodology with Applications*, Chapman and Hall/CRC Press, chapter Aspects of Bayesian Inference in Skewed Mixed Logistic Regression Models, pp. 569–590.

- Santos, C. C., Loschi, R. H. e Arellano-Valle, R. B. (2013). Parameters interpretation in skewed logistic regression with random intercept, *Bayesian Analysis* **8**: 381–410.
- Souza, A. D. P. e Migon, H. S. (2010). Bayesian outlier analysis in binary regression, *Journal of Applied Statistics* **37**: 1355–1368.
- Vaida, F. e Blanchard, S. (2005). Conditional akaike information for mixed-effects models, *Biometrika* **92**(2): 351–370.
- Wei, G. C. e Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms, *Journal of the American Statistical Association* **85**(411): 699–704.
- Yu, D. e Yau, K. K. (2012). Conditional akaike information criterion for generalized linear mixed models, *Computational Statistics & Data Analysis* **56**(3): 629–644.