

Análise Bayesiana objetiva para as distribuições normal generalizada e lognormal generalizada.

SANDRA RÊGO DE JESUS

UFSCar - São Carlos/SP

Dezembro/2014

Universidade Federal de São Carlos
Centro de Ciências Exatas e de Tecnologia
Departamento de Estatística

Análise Bayesiana objetiva para as distribuições normal generalizada e lognormal generalizada.

SANDRA RÊGO DE JESUS

ORIENTADORA: PROF^A. DR^A. VERA LUCIA D. TOMAZELLA

CO-ORIENTADOR: PROF. DR. FRANCISCO LOUZADA NETO

Tese apresentada ao Programa de Pós-Graduação em Estatística do Departamento de Estatística da Universidade Federal de São Carlos - PPGes/UFSCar, como parte dos requisitos para obtenção do título de Doutor em Estatística.

UFSCar - São Carlos/SP

Dezembro/2014

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária/UFSCar**

J58ab Jesus, Sandra Rêgo de.
Análise bayesiana objetiva para as distribuições normal generalizada e lognormal generalizada / Sandra Rêgo de Jesus. -- São Carlos : UFSCar, 2014.
108 f.

Tese (Doutorado) -- Universidade Federal de São Carlos, 2014.

1. Estatística. 2. Análise de referência bayesiana. 3. Distribuição normal generalizada. 4. Distribuição lognormal generalizada. 5. Kullback-Leibler. 6. Mistura de escala uniforme. I. Título.

CDD: 519.5 (20^a)



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Estatística

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a defesa de tese de Doutorado em Estatística da candidata Sandra Rêgo de Jesus, realizada em 21/11/2014:

Profa. Dra. Vera Lucia Damasceno Tomazella - UFSCar

Prof. Dr. Francisco Louzada Neto - USP

Prof. Dr. Helio dos Santos Migon - UFRJ

Profa. Dra. Márcia D'Elia Branco - USP

Prof. Dr. Ricardo Sandes Ehlers - USP

Profa. Dra. Rosângela Helena Loschi - UFMG

Aos meus pais,

***Antônio e Disdete**, pessoas fundamentais
nesta conquista. Por todo o amor e o incentivo
nos momentos difíceis. Vocês são a razão da
minha força e inspiração para seguir sempre
em frente.*

Agradecimentos

A Deus, por ter me dado a permissão de chegar até aqui, por iluminar meu caminho e me dar forças para superar os obstáculos.

Aos meus irmãos, Antônio e Osvaldino, pelo amor, carinho e amizade.

Ao meus amores, Ariel Victor e David, que são uma preciosidade na minha vida.

À minha orientadora Vera Tomazella e ao meu coorientador Francisco Louzada, pela orientação e condução do trabalho.

Aos Professores Hélio Migon e Ricardo Ehlers, membros da banca do exame de qualificação, pelas sugestões feitas.

À minha amiga Matildes pela atenção, carinho, incentivo, preocupação e que, mesmo de longe, contribuiu de forma tão especial para a finalização deste trabalho.

A Regina, Rodrigo e Marcos pelo carinho e muito apoio, nos momentos delicados.

A Vanessa e à Camila pela amizade e por todos os momentos de alegria.

A todos que, de forma direta ou indireta, contribuíram para a realização deste trabalho.

Resumo

As distribuições normal generalizada (NG) e lognormal generalizada (logNG) são flexíveis por acomodarem características presentes nos dados que não são capturadas por distribuições tradicionais, como a normal e a lognormal, respectivamente. Essas distribuições são consideradas ferramentas para reduzir as observações aberrantes e obter estimativas robustas. Entretanto o maior obstáculo para a utilização eficiente dessas distribuições tem sido os problemas computacionais. Este trabalho propõe a metodologia da análise de referência Bayesiana para estimar os parâmetros dos modelos NG e logNG. A função *a priori* de referência para uma possível ordem dos parâmetros do modelo é obtida. Mostra-se que a função *a priori* de referência conduz a uma distribuição *a posteriori* própria, em todos os modelos propostos. Para fins de inferência, é considerado o desenvolvimento de métodos Monte Carlo em Cadeias de Markov (MCMC). Para detectar possíveis observações influentes nos modelos considerados, é utilizado o método Bayesiano de análise de influência caso a caso, baseado na divergência de Kullback-Leibler. Além disso, uma representação de mistura de escala uniforme para as distribuições NG e logNG é utilizada, como um método alternativo, para permitir o desenvolvimento de algoritmos de amostrador de Gibbs. Estudos de simulação foram desenvolvidos para analisar as propriedades frequentistas dos processos de estimação. Aplicações a conjuntos de dados reais mostraram a aplicabilidade dos modelos propostos.

Palavras-chave: Análise de referência Bayesiana, distribuição normal generalizada, distribuição lognormal generalizada, divergência de Kullback-Leibler, mistura de escala uniforme.

Abstract

The Generalized Normal (GN) and Generalized lognormal (logGN) distributions are flexible for accommodating features present in the data that are not captured by traditional distribution, such as the normal and the lognormal ones, respectively. These distributions are considered to be tools for the reduction of outliers and for the obtention of robust estimates. However, computational problems have always been the major obstacle to obtain the effective use of these distributions. This paper proposes the Bayesian reference analysis methodology to estimate the GN and logGN. The reference prior for a possible order of the model parameters is obtained. It is shown that the reference prior leads to a proper posterior distribution for all the proposed model. The development of Monte Carlo Markov Chain (MCMC) is considered for inference purposes. To detect possible influential observations in the models considered, the Bayesian method of influence analysis on a case based on the Kullback-Leibler divergence is used. In addition, a scale mixture of uniform representation of the GN and logGN distributions are exploited, as an alternative method in order, to allow the development of efficient Gibbs sampling algorithms. Simulation studies were performed to analyze the frequentist properties of the estimation procedures. Real data applications demonstrate the use of the proposed models.

Keywords: Reference Bayesian analysis, generalized normal distribution, generalized lognormal distribution, Kullback-Leibler divergence, scale mixtures of uniform.

Sumário

1	Introdução	1
1.1	Análise Bayesiana objetiva	4
1.1.1	Função <i>a priori</i> de referência para o caso multiparamétrico	5
1.1.2	Função <i>a priori</i> de Jeffreys para o caso multiparamétrico	8
1.2	Critérios de seleção de modelos	8
1.3	Diagnóstico de influência Bayesiano	10
1.4	Organização dos capítulos	12
2	Distribuição normal generalizada	13
2.1	Função de verossimilhança	15
2.2	Distribuições <i>a priori</i> e <i>a posteriori</i>	17
2.3	Estudo de simulação	21
2.4	Aplicação	24
2.4.1	Influência de observações aberrantes	26
2.5	Representação de mistura de escala uniforme	30
2.5.1	Distribuição <i>a posteriori</i>	32
2.5.2	Aplicação	35
2.6	Considerações finais	36

3	Distribuição lognormal generalizada com aplicação a dados de sobrevivência	38
3.1	Função de verossimilhança	40
3.2	Distribuições <i>a priori</i> e <i>a posteriori</i>	41
3.3	Aplicações	43
3.3.1	Dados artificiais	44
3.3.2	Dados reais	48
3.4	Representação de mistura de escala uniforme	51
3.4.1	Distribuição <i>a posteriori</i>	52
3.4.2	Aplicação	53
3.5	Considerações finais	54
4	Modelo de regressão normal generalizado	56
4.1	Função de verossimilhança	57
4.2	Distribuições <i>a priori</i> e <i>a posteriori</i>	58
4.3	Aplicações	63
4.3.1	Dados artificiais	63
4.3.2	Dados reais	66
4.4	Considerações finais	71
5	Considerações finais e propostas futuras	72
A	Demonstrações	80
A.1	Matriz de informação de Fisher	80
A.2	Distribuições <i>a posteriori</i>	86
A.3	Função de sobrevivência	92

B	Metropolis-Hasting dentro do amostrador de Gibbs	93
C	Demais resultados	95
C.1	Distribuição normal generalizada	95
C.2	Distribuição lognormal generalizada	100
C.3	Modelo de regressão normal generalizado	105

Lista de Figuras

2.1	Função densidade da distribuição dada em (2.2) considerando os parâmetros $\mu = 0$ e $\sigma = 1$	15
2.2	Variância e Erro Quadrático Médio dos estimadores Bayesianos para diferentes amostras do modelo NG.	23
2.3	Distribuição <i>a posteriori</i> predita para os dados de altura de <i>Eucalyptus sp.</i> . Linha sólida: distribuição normal generalizada, e linha pontilhada: distribuição normal.	25
2.4	Índices de $K(\pi, \pi_{(-i)})$ para os dados simulados da NG(35; 5; 3, 5).	28
2.5	Índices de $K(\pi, \pi_{(-i)})$ para os dados de altura de <i>Eucalyptus sp.</i>	29
3.1	Função densidade (a), função de sobrevivência (b) e função de risco (c) para $\mu = 0, \sigma = 1$ e diferentes valores de s	40
3.2	Índices de $K(\pi, \pi_{(-i)})$ para os dados simulados da logNG(4, 7; 1, 7; 1, 9).	48
3.3	Curva Kaplan-Meier com funções de sobrevivência estimadas pelas distribuições logNG e lognormal para os dados de câncer de pulmão inoperável.	50
3.4	Índices de $K(\pi, \pi_{(-i)})$ para os dados de câncer de pulmão inoperável.	51
4.1	Índices de $K(\pi, \pi_{(-i)})$ para os dados artificiais.	66

4.2	Dados reais. Modelo de regressão normal generalizado com $s = 2,865$. (a) Gráfico de dispersão dos dados e modelo de regressão normal generalizado ajustado. (b) Quantis residuais versus os valores ajustados. (c) Quantis residuais versus o número da observação. (d) Gráfico dos quantis da distribuição NG para os resíduos do modelo.	68
4.3	Dados reais. Modelo de regressão normal. (a) Gráfico de dispersão dos dados e modelo de regressão ajustado. (b) Quantis residuais versus os valores ajustados. (c) Quantis residuais versus o número da observação. (d) Gráfico dos quantis da distribuição normal para os resíduos do modelo.	69
4.4	Índices de $K(\pi, \pi_{(-i)})$ para os dados de altura de <i>Eucalyptus sp.</i>	70
C.1	Dados reais. Históricos das cadeias, densidades a <i>posteriori</i> marginais dos parâmetros do modelo NG com HPD a 95%, autocorrelações amostrais e médias ergódicas.	97
C.2	Dados reais. Históricos das cadeias, densidades a <i>posteriori</i> marginais dos parâmetros do modelo normal com HPD a 95%, autocorrelações amostrais e médias ergódicas.	98
C.3	Dados reais. Históricos das cadeias, densidades a <i>posteriori</i> marginais dos parâmetros do modelo NG com mistura de escala uniforme, considerando-se HPD a 95%, autocorrelações amostrais e médias ergódicas.	99
C.4	Dados artificiais. Históricos das cadeias, densidades a <i>posteriori</i> marginais dos parâmetros do modelo logNG com HPD a 95%, autocorrelações amostrais e médias ergódicas.	101
C.5	Dados reais. Históricos das cadeias, densidades a <i>posteriori</i> marginais dos parâmetros do modelo logNG com HPD a 95%, autocorrelações amostrais e médias ergódicas.	102
C.6	Dados reais. Históricos das cadeias, densidades a <i>posteriori</i> marginais dos parâmetros do modelo lognormal com HPD a 95%, autocorrelações amostrais e médias ergódicas.	103

C.7	Dados reais. Históricos das cadeias, densidades a <i>posteriori</i> marginais dos parâmetros do modelo logNG com mistura de escala uniforme, considerando-se HPD a 95%, autocorrelações amostrais e médias ergódicas.	104
C.8	Dados artificiais. Históricos das cadeias, densidades a <i>posteriori</i> marginais dos parâmetros do modelo de regressão NG com HPD a 95%, autocorrelações amostrais e médias ergódicas.	106
C.9	Dados reais. Históricos das cadeias, densidades a <i>posteriori</i> marginais dos parâmetros do modelo de regressão NG com HPD a 95%, autocorrelações amostrais e médias ergódicas.	107
C.10	Dados reais. Históricos das cadeias, densidades a <i>posteriori</i> marginais dos parâmetros do modelo de regressão normal com HPD a 95%, autocorrelações amostrais e médias ergódicas.	108

Lista de Tabelas

2.1	Probabilidade de cobertura do intervalo de credibilidade de máxima densidade de probabilidade (HPD) a 95% para cada tamanho amostral e para cada parâmetro do modelo NG.	22
2.2	Resumos <i>a posteriori</i> para os parâmetros dos modelos NG e normal, critérios de comparação Bayesiana e taxas de aceitação (TA) no algoritmo Metropolis-Hasting.	25
2.3	Dados simulados. Média <i>a posteriori</i> , $VR(\%)$ em relação aos dados originais simulados, HPD a 95% para os parâmetros da $NG(35; 5; 3, 5)$ e taxas de aceitação (TA) no algoritmo Metropolis-Hastings.	27
2.4	Medidas de diagnóstico para os dados simulados da $NG(35; 5; 3, 5)$	28
2.5	Medidas de diagnóstico para os dados de altura de <i>Eucalyptus sp.</i>	29
2.6	Estimativas <i>a posteriori</i> , $VR(\%)$ em relação aos dados originais para os dados de altura de <i>Eucalyptus sp.</i> e taxas de aceitação (TA) no algoritmo Metropolis-Hastings.	30
2.7	Resumos <i>a posteriori</i> para os parâmetros do modelo NG.	36
3.1	Dados artificiais. Resumos <i>a posteriori</i> para os parâmetros do modelo log-normal generalizado e as taxas de aceitação (TA) no algoritmo Metropolis-Hastings.	44
3.2	Dados simulados. Média <i>a posteriori</i> , $VR(\%)$ em relação aos dados originais simulados, HPD a 95% para os parâmetros da distribuição $\log NG(4; 1, 7; 1, 9)$ e taxas de aceitação no algoritmo Metropolis-Hastings.	46

3.3	Medidas de diagnóstico para os dados simulados da $\text{logNG}(4; 1, 7; 1, 9)$. . .	47
3.4	Resumos <i>a posteriori</i> para os parâmetros do modelo logNG , critérios de comparação Bayesiana e taxas de aceitação (TA) no algoritmo Metropolis-Hastings.	49
3.5	Medidas de diagnóstico para os dados de câncer de pulmão inoperável. .	51
3.6	Resumos <i>a posteriori</i> para os parâmetros do modelo logNG	54
4.1	Dados artificiais. Resumos <i>a posteriori</i> para os parâmetros do modelo de regressão NG e as taxas de aceitação (TA) no algoritmo Metropolis-Hastings.	63
4.2	Dados artificiais. Média <i>a posteriori</i> , $VR(\%)$ em relação aos dados originais simulados, HPD a 95% para os parâmetros do modelo de regressão NG e taxas de aceitação (TA) no algoritmo Metropolis-Hastings.	65
4.3	Medidas de diagnóstico para os dados artificiais.	65
4.4	Dados reais. Resumos <i>a posteriori</i> para os parâmetros do modelo de regressão NG, critérios de comparação Bayesiana e taxas de aceitação (TA) no algoritmo Metropolis-Hastings.	67
4.5	Medidas de diagnóstico para os dados de altura de <i>Eucalyptus sp.</i>	70
4.6	Estimativas <i>a posteriori</i> , $VR(\%)$ em relação aos dados originais para os dados de altura de <i>Eucalyptus sp.</i> e taxas de aceitação (TA) no algoritmo Metropolis-Hastings.	71
C.1	Dados simulados. Médias das taxas de aceitação (%) no algoritmo Metropolis-Hastings para cada tamanho amostral e para cada parâmetro do modelo NG.	95
C.2	Dados reais. Resultados do critério de Geweke para os modelos NG, normal e NG com mistura de escala uniforme.	96
C.3	Dados reais. Autocorrelações das cadeias para o modelo NG com mistura de escala uniforme.	96
C.4	Dados artificiais. Resultados do critério de Geweke para o modelo logNG	100

C.5	Dados reais. Resultados do critério de Geweke para os modelos logNG, lognormal e logNG com mistura de escala uniforme.	100
C.6	Dados reais. Autocorrelações das cadeias para o modelo logNG com mistura de escala uniforme.	100
C.7	Dados artificiais. Resultados do critério de Geweke para o modelo de regressão NG.	105
C.8	Dados reais. Resultados do critério de Geweke para os modelos de regressão NG e normal.	105

Capítulo 1

Introdução

As distribuições normal e lognormal são aplicadas em diversas áreas de estudo para modelar dados simétricos e assimétricos, respectivamente. No entanto, algumas vezes, ambas as distribuições não descrevem suficientemente bem uma situação real. Quando os dados provêm de uma distribuição com caudas mais leves ou mais pesadas do que a normal, a suposição de normalidade não é adequada. De forma semelhante, quando os dados provêm de uma distribuição com caudas com um grau significativo de assimetria positiva, a suposição de lognormalidade não é adequada. Nestas circunstâncias, é necessário utilizar-se modelos mais flexíveis, como a distribuição normal generalizada (NG) e a distribuição lognormal generalizada (logNG).

A distribuição NG, que abrange uma variedade de distribuições, como a de Laplace, a normal e a uniforme, é considerada uma ferramenta para reduzir as observações aberrantes e obter estimativas robustas (Box & Tiao, 1973; Walker & Gutiérrez-Peña, 1999; Liang *et al.*, 2007). Essa distribuição tem sido utilizada em contextos diferentes, com diferentes parametrizações, mas o maior obstáculo para a sua utilização eficiente tem sido os problemas computacionais, pois não existem formas explícitas para os estimadores do parâmetro de forma. A estimação do parâmetro de forma deve ser feita por métodos numéricos. Entretanto, no contexto clássico, existem problemas de convergências, de acordo com estudos de Varanasi & Aazhang (1989) e Agrò (1995).

No contexto Bayesiano, os métodos apresentados na literatura para estimar os parâmetros da distribuição NG são restritos e aplicados a casos particulares. Box

& Tiao (1973) utilizaram a distribuição NG sob a perspectiva Bayesiana subjetiva no contexto de modelos de regressão robustos, como uma alternativa aos pressupostos de normalidade dos erros em modelos de regressão. West (1987) provou que a distribuição NG pode ser representada por uma mistura de escala da distribuição normal. Choy & Smith (1997a) utilizaram a distribuição *a priori* normal generalizada para o parâmetro de locação no modelo Gaussiano. Obtiveram os sumários da distribuição *a posteriori*, estimado pelo método de Laplace, e examinaram as suas propriedades de robustez. Choy & Smith (1997b) utilizaram a representação da distribuição NG como uma mistura de escala da distribuição normal em modelos de efeitos aleatórios e consideraram o método Monte Carlo em Cadeia de Markov (MCMC) para a obtenção dos sumários *a posteriori* do parâmetro de interesse. Walker & Gutiérrez-Peña (1999) propuseram que a distribuição NG fosse representada por uma mistura de escala da distribuição uniforme e estimaram os parâmetros de locação e de escala pelo método Monte Carlo em Cadeia de Markov (MCMC). Recentemente, Salazar *et al.* (2012) consideraram a análise Bayesiana objetiva para o modelo de regressão normal generalizado. Eles utilizaram três diferentes funções *a priori* não-informativas de Jeffreys para os parâmetros do modelo e mostraram que apenas uma dessas funções *a priori* conduz a uma distribuição *a posteriori* própria.

A distribuição logNG, que inclui a distribuição lognormal como um caso particular, também é considerada uma ferramenta para obter estimativas robustas (Martín & Pérez, 2009). Essa distribuição tem sido utilizada com diferentes parametrizações, em contextos diferentes, mas os procedimentos inferenciais para os parâmetros do modelo apresentam problemas. Em particular, no contexto clássico, a parametrização que considera o parâmetro limiar (localização) causa um efeito não usual na função de verossimilhança, isto é, a função de verossimilhança de alguma amostra ordenada dessa distribuição torna-se ilimitada. Consequentemente, a maximização global da verossimilhança conduz a estimativa inconsistente. Hill (1963) mostrou que o parâmetro limiar tende para o menor valor observado da variável resposta, isto é, o mínimo, e as estimativas de máxima verossimilhança dos demais parâmetros são inconsistentes. No entanto Hill (1963) também manteve a ideia de que estimativas consistentes poderiam ser obtidas através das equações de verossimilhança. Neste contexto, diferentes autores estudaram como obter estimativas consistentes para essa distribuição, como pode ser visto em Harter & Moore (1966), Calitz (1973), Cohen & Whitten (1980, 1988). Entre-

tanto algumas vezes a função de verossimilhança pode não ter máximo local claramente definido para pequenas amostras (Harter & Moore, 1966; Cohen & Whitten, 1988).

No contexto Bayesiano, os métodos apresentados na literatura para estimar os parâmetros da distribuição logNG também são restritos e aplicados a casos particulares. Hill (1963) considerou funções *a priori* subjetivas. Upadhyay & Peshwani (2001) utilizaram função *a priori* não-informativa de Box & Tiao (1973). Recentemente, Martín & Pérez (2009) utilizaram a parametrização da distribuição logNG, obtida da transformação exponencial de uma variável aleatória, que segue a distribuição normal generalizada estudada por Nadarajah (2005). Representaram a distribuição logNG como uma mistura de escala da distribuição uniforme, utilizaram funções *a priori* informativas e não-informativas de Jeffreys, e estimaram os parâmetros de interesse pelo método amostrador de Gibbs.

O objetivo deste trabalho é utilizar a metodologia da análise de referência, como uma forma objetiva, para obter função *a priori* não informativa (de referência) para os parâmetros dos modelos NG e logNG a fim de fazer uma análise Bayesiana dos modelos considerados. Modelos de regressão para a distribuição NG também são considerados. Para fins de inferência, é considerado o método Monte Carlo em Cadeia de Markov (MCMC). Técnicas de diagnósticos baseadas na divergência de Kullback-Leibler são utilizadas como medida de influência local. Além disso, uma representação de mistura de escala uniforme para as distribuições NG e logNG é utilizada, como um método alternativo, para permitir o desenvolvimento de algoritmos de amostrador de Gibbs.

Uma breve revisão sobre análise Bayesiana objetiva e dos métodos de obtenção de funções *a priori* não-informativas, como as de Jeffreys (1946) e de Bernardo (1979) são apresentados na Seção 1.1. A Seção 1.2 apresenta alguns critérios de seleção de modelos. A Seção 1.3 apresenta um método Bayesiano de análise de influência caso a caso, baseado na divergência de Kullback-Leibler. A organização dos capítulos está apresentada na Seção 1.4.

1.1 Análise Bayesiana objetiva

Na abordagem Bayesiana, sabe-se que a inferência acerca de qualquer parâmetro de interesse é obtida a partir da distribuição *a posteriori*, que contém toda a informação probabilística a respeito do parâmetro de interesse. A distribuição *a posteriori* é obtida a partir da combinação do conhecimento prévio do pesquisador sobre o parâmetro (distribuição *a priori*) com a informação proveniente dos dados (função de verossimilhança). A distribuição *a priori* funciona como um peso para a informação fornecida pela função de verossimilhança. Entretanto, nas situações em que não existem informações *a priori* substanciais sobre o problema e/ou existe complexidade em elicitar as distribuições *a priori* subjetivas para os parâmetros de interesse, é necessário adotar outro procedimento que consiga captar essa noção de que a informação *a priori* tenha um efeito mínimo sobre os resultados. O uso de funções *a priori* não-informativas, como as de [Jeffreys \(1946\)](#) e de [Bernardo \(1979\)](#), é adequado, nessas circunstâncias.

A função *a priori* objetiva (ou não-informativa) mais comumente utilizada é a de [Jeffreys \(1946\)](#), que é invariante sob transformações um a um e bem sucedida no caso unidimensional. No entanto, no caso multidimensional, Jeffreys encontrou dificuldades no método e sugeriu modificações *ad hoc* para a *priori*.

O método de análise de referência foi proposto por [Bernardo \(1979\)](#) para corrigir as deficiências da aplicação da regra de Jeffreys no caso multiparamétrico. A abordagem da *priori* de referência de Bernardo foi aprimorada por [Berger & Bernardo \(1992\)](#). Essa abordagem descreve uma forma de encontrar funções *a priori* não informativas (*priori* de referência) que maximizam a falta de informações. Nesse caso, as afirmações produzidas pelo processo de estimação dependem somente do modelo assumido e dos dados observados. A função *a priori* de referência produz uma distribuição *a posteriori* de referência, a partir da aplicação formal do teorema de Bayes. A análise de referência fornece distribuições *a posteriori* com propriedades importantes, como generalidade, invariância, marginalização consistente e propriedades de amostragem consistente ([Bernardo, 2005](#)).

O método de obtenção da função *a priori* de referência, para o modelo multiparamétrico, depende da matriz de informação de Fisher e é baseado na normalidade

assintótica da distribuição *a posteriori*. [Berger & Bernardo \(1992\)](#) sugeriram organizar os parâmetros de acordo com a sua ordem de importância inferencial, e forneceram uma descrição detalhada de métodos matemáticos para obter uma função *a priori* de referência.

A seguir, apresentam-se a definição da matriz de informação esperada de Fisher no caso multiparamétrico e os métodos de Bernardo (referência) e de Jeffreys para obtenção de funções *a priori* não-informativas.

Definição 1.1. Seja Y uma variável aleatória com função densidade $f(y|\boldsymbol{\theta})$, em que $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$. Denota-se $I(\boldsymbol{\theta})$ a matriz de informação esperada de Fisher de $\boldsymbol{\theta}$, cujos elementos são definidos por

$$I_{ij}(\boldsymbol{\theta}) = -E \left(\frac{\partial^2 \log f(y|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right) = E \left(\frac{\partial \log f(y|\boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \log f(y|\boldsymbol{\theta})}{\partial \theta_j} \right) \quad i, j = 1, \dots, m$$

em que $I_{ij}(\boldsymbol{\theta}) = I_{ji}(\boldsymbol{\theta})$, θ_j o j -ésimo elemento de $\boldsymbol{\theta}$.

A medida da informação de Fisher está associada à curvatura média da função de verossimilhança. Quanto maior for a curvatura, maior será a informação contida na função de verossimilhança, portanto maior será $I(\boldsymbol{\theta})$.

1.1.1 Função *a priori* de referência para o caso multiparamétrico

Esta subseção apresenta o método de análise de referência Bayesiana para o caso de dois ou mais parâmetros perturbadores, supondo normalidade assintótica da distribuição *a posteriori*, segundo [Bernardo & Ramón \(1998\)](#).

A função *a priori* de referência maximiza a falta de informação sobre a quantidade de interesse. Uma característica importante do método de análise de referência é o tratamento diferenciado dado aos parâmetros de interesse e aos parâmetros perturbadores.

Seja $f(y|\phi, \boldsymbol{\lambda})$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)$, $\phi \in \Phi$, $\boldsymbol{\lambda} \in \Lambda = \prod_{j=1}^m \Lambda_j$, um modelo probabilístico com $m + 1$ parâmetros, a quantidade de interesse é ϕ , e $I(\phi, \lambda_1, \dots, \lambda_m)$ é a matriz de informação de Fisher $(m + 1) \times (m + 1)$. Supõe-se que a distribuição *a posteriori* conjunta de $(\phi, \lambda_1, \dots, \lambda_m)$ seja assintoticamente normal com média $\hat{\phi}, \hat{\lambda}_1, \dots, \hat{\lambda}_m$,

os correspondentes estimadores de máxima verossimilhança, e com matriz de covariância $B(\hat{\phi}, \hat{\lambda}_1, \dots, \hat{\lambda}_m)$, em que $B(\hat{\phi}, \hat{\lambda}_1, \dots, \hat{\lambda}_m) = I^{-1}(\hat{\phi}, \hat{\lambda}_1, \dots, \hat{\lambda}_m)$. Considerando-se que B_j é a submatriz superior $j \times j$ à esquerda de B , $j = 1, \dots, m+1$, $H_j = B_j^{-1}$, e $h_{jj}(\phi, \lambda_1, \dots, \lambda_m)$ é o elemento inferior à direita (j, j) de H_j , tal que $H_{m+1} = I$ e $h_{m+1, m+1} = I_{m+1, m+1}$, definem-se

$$B(\phi, \lambda_1, \dots, \lambda_m) = \begin{bmatrix} B_{11}(\phi, \lambda_1, \dots, \lambda_m) & \cdots & B_{1(m+1)}(\phi, \lambda_1, \dots, \lambda_m) \\ \vdots & \ddots & \vdots \\ B_{(m+1)1}(\phi, \lambda_1, \dots, \lambda_m) & \cdots & B_{(m+1)(m+1)}(\phi, \lambda_1, \dots, \lambda_m) \end{bmatrix}$$

e

$$H(\phi, \lambda_1, \dots, \lambda_m) = B^{-1}(\phi, \lambda_1, \dots, \lambda_m),$$

então, pelas propriedades padrão da distribuição normal multivariada tem-se que:

1. A distribuição *a posteriori* marginal de ϕ é assintoticamente normal com desvio padrão $h_{11}^{-\frac{1}{2}}(\hat{\phi}, \hat{\lambda}_1, \dots, \hat{\lambda}_m)$ e
2. A distribuição *a posteriori* condicional de λ_i dado $(\phi, \lambda_1, \dots, \lambda_{i-1})$, é assintoticamente normal com desvio padrão $h_{i+1, i+1}^{-\frac{1}{2}}(\phi, \lambda_1, \dots, \lambda_{i-1}, \hat{\lambda}_i, \dots, \hat{\lambda}_m)$.

Esse resultado pode ser utilizado para calcular sequencialmente as funções *a priori* de referência condicionais $\pi^R(\lambda_m | \phi, \lambda_1, \dots, \lambda_{m-1})$, $\pi^R(\lambda_{m-1} | \phi, \lambda_1, \dots, \lambda_{m-2})$, \dots , $\pi^R(\lambda_1 | \phi)$, e a função *a priori* de referência marginal $\pi^R(\phi)$, onde $\pi^R(\cdot)$ é a função *a priori* de referência. Assim, a função *a priori* de referência conjunta, relativa à parametrização ordenada $(\phi, \lambda_1, \dots, \lambda_m)$, é obtida como

$$\pi^R(\phi, \lambda_1, \dots, \lambda_m) = \pi^R(\lambda_m | \phi, \lambda_1, \dots, \lambda_{m-1}) \times \dots \times \pi^R(\lambda_1 | \phi) \pi^R(\phi).$$

Para obter a função *a priori* de referência, no caso da presença de parâmetros perturbadores, escolhe-se o parâmetro de interesse e segue-se o procedimento abaixo (Bernardo & Ramón, 1998).

Proposição 1.1. (Priori de referência sob normalidade assintótica) Seja $f(y | \phi, \boldsymbol{\lambda})$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)$, $\phi \in \Phi$, $\boldsymbol{\lambda} \in \Lambda = \prod_{j=1}^m \Lambda_j$, um modelo probabilístico com $m+1$ parâmetros, e a quantidade de interesse é ϕ . Supõe-se que a distribuição *a posteriori* conjunta de $(\phi, \lambda_1, \dots, \lambda_m)$ seja assintoticamente normal com matriz de covariância $B(\hat{\phi}, \hat{\lambda}_1, \dots, \hat{\lambda}_m)$.

Se B_j é a submatriz superior $j \times j$ à esquerda de B , $H_j = B_j^{-1}$, e $h_{jj}(\phi, \lambda_1, \dots, \lambda_m)$ é o elemento inferior à direita (j, j) de H_j , tem-se que

1. As funções *a priori* de referência condicionais são

$$\pi^R(\lambda_m | \phi, \lambda_1, \dots, \lambda_{m-1}) \propto h_{m+1, m+1}^{\frac{1}{2}}(\phi, \lambda_1, \dots, \lambda_m),$$

e

$$\pi^R(\lambda_i | \phi, \lambda_1, \dots, \lambda_{i-1}) \propto \exp \left\{ \int_{\Lambda_{i+1}} \dots \int_{\Lambda_m} \log h_{i+1, i+1}^{\frac{1}{2}}(\phi, \lambda_1, \dots, \lambda_m) \left\{ \prod_{j=i+1}^m \pi^R(\lambda_j | \phi, \lambda_1, \dots, \lambda_{j-1}) \right\} d\lambda_{i+1} \right\},$$

em que $d\lambda_j = d\lambda_j \times \dots \times d\lambda_m$, e $\pi^R(\lambda_i | \phi, \lambda_1, \dots, \lambda_{i-1})$, $i = 1, \dots, m$ são próprias. Se qualquer dessas funções *a priori* de referência condicionais não for própria, uma aproximação compacta será necessária para as integrais correspondentes.

2. A função *a priori* de referência marginal de ϕ é

$$\pi^R(\phi) \propto \exp \left\{ \int_{\Lambda_1} \dots \int_{\Lambda_m} \log B_{11}^{-\frac{1}{2}}(\phi, \lambda_1, \dots, \lambda_m) \left\{ \prod_{j=1}^m \pi^R(\lambda_j | \phi, \lambda_1, \dots, \lambda_{j-1}) \right\} d\lambda_1 \right\},$$

onde $B_{11}^{-\frac{1}{2}}(\phi, \lambda_1, \dots, \lambda_m) = h_{11}^{\frac{1}{2}}(\phi, \lambda_1, \dots, \lambda_m)$.

Corolário 1.1. Se os espaços paramétricos dos parâmetros perturbadores Λ_i não dependem de $\{\phi, \lambda_1, \dots, \lambda_{i-1}\}$, e as funções $h_{11}, h_{22}, \dots, h_{mm}, h_{m+1, m+1}$ se fatoram da forma

$$h_{11}^{\frac{1}{2}}(\phi, \lambda_1, \dots, \lambda_m) = f_0(\phi)g_0(\lambda_1, \dots, \lambda_m) \text{ e}$$

$$h_{i+1, i+1}^{\frac{1}{2}}(\phi, \lambda_1, \dots, \lambda_m) = f_i(\lambda_i)g_i(\phi, \lambda_i, \dots, \lambda_{i-1}, \lambda_{i+1}, \dots, \lambda_m) \text{ com } i = 1, \dots, m$$

tem-se que

$$\pi^R(\phi) \propto f_0(\phi)$$

$$\pi^R(\lambda_i | \phi, \lambda_1, \dots, \lambda_{i-1}) \propto f_i(\lambda_i), \quad i = 1, \dots, m,$$

e não há necessidade de aproximações compactas, mesmo que $\pi^R(\lambda_i | \phi, \lambda_1, \dots, \lambda_{i-1})$ não seja própria (Bernardo & Ramón, 1998).

Sob condições de regularidade (ver Teorema 5.14 e Seção 5.4.5 de (Bernardo & Smith, 1994)) a distribuição *a posteriori* de $(\phi, \lambda_1, \dots, \lambda_m)$ é assintoticamente normal com média $(\hat{\phi}, \hat{\lambda}_1, \dots, \hat{\lambda}_m)$, os correspondentes estimadores de máxima verossimilhança, e com matriz de covariância $B(\hat{\phi}, \hat{\lambda}_1, \dots, \hat{\lambda}_m)$, onde $B(\hat{\phi}, \hat{\lambda}_1, \dots, \hat{\lambda}_m) = I^{-1}(\hat{\phi}, \hat{\lambda}_1, \dots, \hat{\lambda}_m)$ e $I(\phi, \lambda_1, \dots, \lambda_m)$ é a matriz de informação de Fisher $(m + 1) \times (m + 1)$, neste caso, $H(\phi, \lambda_1, \dots, \lambda_m) = I(\phi, \lambda_1, \dots, \lambda_m)$, e a *priori* de referência pode ser obtida dos elementos da matriz de informação de Fisher $I(\phi, \lambda_1, \dots, \lambda_m)$.

1.1.2 Função *a priori* de Jeffreys para o caso multiparamétrico

A regra de Jeffreys é um método para determinar *priori* não-informativa, que é invariante sob transformações injetivas (Migon, H. S. and Gamerman, D., 1999). Apesar da propriedade de invariância, a *priori* Jeffreys geralmente é imprópria e conduz a uma distribuição *a posteriori* imprópria.

A função *a priori* de Jeffreys é proporcional à raiz quadrada do determinante da matriz de informação de Fisher, isto é,

$$\pi^J(\boldsymbol{\theta}) \propto |\det I(\boldsymbol{\theta})|^{\frac{1}{2}},$$

em que $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$ e $I(\boldsymbol{\theta})$ é a matriz de informação de Fisher.

A utilização da regra de Jeffreys no caso multiparamétrico costuma ter um desempenho insatisfatório, pois a propriedade de invariância é menos rigorosa e convincente sobre a verossimilhança, especialmente, quando os parâmetros de localização e escala ocorrem simultaneamente. A suposição de independência *a priori* entre os parâmetros de natureza diferente e a utilização da regra de Jeffreys separada para a especificação das distribuições marginais é recomendada por Jeffreys.

1.2 Critérios de seleção de modelos

No contexto Bayesiano há uma variedade de critérios que podem ser adotados para selecionar o melhor ajuste entre uma coleção de modelos. Neste trabalho são considerados os seguintes critérios: o Critério de Informação de Desvio (DIC) proposto

por Spiegelhalter *et al.* (2002), o Critério de Informação de Akaike Esperado (EAIC) proposto por Brooks (2002) e o Critério de Informação Bayesiano Esperado (EBIC) proposto por Carlin & Louis (2001a). Esses critérios são baseados na média *a posteriori* do desvio, $E(D(\theta))$, que pode ser estimado a partir dos resultados das amostras MCMC, $\bar{D} = \frac{1}{Q} \sum_{q=1}^Q D(\theta^{(q)})$, onde o índice q indica a q -ésima realização de um total de Q realizações e $D(\theta) = -2 \sum_{i=1}^n \log(f(y_i|\theta))$, onde $f(\cdot)$ é a função densidade de probabilidade. Os critérios DIC, EAIC e EBIC são dados por

$$DIC = \bar{D} + p_D = 2\bar{D} - \hat{D}, \quad EAIC = \bar{D} + 2b \quad e \quad EBIC = \bar{D} + b \log(n),$$

em que b é o número de parâmetros do modelo e p_D é o número efetivo de parâmetros no modelo, que é dado por $E[D(\theta)] - D[E(\theta)]$, em que $D[E(\theta)]$ é o desvio avaliado na média *a posteriori* que pode ser estimado por $\hat{D} = D\left(\frac{1}{Q} \sum_{q=1}^Q \theta^{(q)}\right)$. Valores menores de DIC, EAIC e EBIC indicam o modelo preferido.

Outro critério bastante aceito para seleção de modelos é obtido da estatística preditiva condicional ordenada (CPO). Para uma discussão detalhada da estatística CPO e suas aplicações para seleção de modelos, consulte Gelfand *et al.* (1992), Gelfand & Dey (1994) e Ibrahim *et al.* (2001). Sejam D os dados completos e $D_{(-i)}$ os dados com a i -ésima observação excluída. Define-se a densidade *a posteriori* de θ dado $D_{(-i)}$ por $\pi(\theta|D_{(-i)})$, para $i = 1, \dots, n$. Desse modo, a i -ésima observação da CPO $_i$ pode ser escrita como

$$CPO_i = \int_{\Theta} f(y_i|\theta)\pi(\theta|D_{(-i)})d\theta = \left\{ \int_{\Theta} \frac{\pi(\theta|D)}{f(y_i|\theta)} d\theta \right\}^{-1}, \quad i = 1, \dots, n,$$

em que $f(y_i|\theta)$ é a função densidade de y_i . Valores altos de CPO $_i$ indicam um bom ajuste do modelo. Em geral, não existem expressões de formas fechadas para a CPO $_i$. Entretanto uma estimativa para a CPO $_i$ pode ser obtida através da amostra MCMC da distribuição *a posteriori* $\pi(\theta|D)$. Portanto uma estimativa Monte Carlo para para a CPO $_i$ é dada por

$$\widehat{CPO}_i = \left\{ \frac{1}{Q} \sum_{q=1}^Q \frac{1}{f(y_i|\theta_q)} \right\}^{-1},$$

em que Q denota o tamanho da amostra MCMC, como pode ser visto em Chen *et al.* (2000). O sumário estatístico da CPO $_i$ é dado por $B = \sum_{i=1}^n \log(CPO_i)$, em que um valor alto de B indica um melhor ajuste do modelo considerado.

1.3 Diagnóstico de influência Bayesiano

Ao se ajustar um modelo a um conjunto de dados, é importante estudar a robustez dos resultados obtidos no processo de estimação dos parâmetros. A análise de diagnóstico via deleção de casos pode ser utilizada para esta finalidade.

Para avaliar a sensibilidade na estimação dos parâmetros, considera-se um diagnóstico Bayesiano de influência via deleção de casos (Cook & Weisberg, 1982) baseado na divergência de Kullback-Leibler (K-L) (Cho *et al.*, 2009), o qual é calculado utilizando a densidade preditiva condicional ordinária.

Considera-se $K(\pi, \pi_{(-i)})$ a divergência de K-L entre π e $\pi_{(-i)}$, em que π denota a distribuição *a posteriori* de θ para os dados completos e $\pi_{(-i)}$ é a distribuição *a posteriori* de θ sem o i -ésimo caso. Especificamente,

$$K(\pi, \pi_{(-i)}) = \int \pi(\theta|D) \log \left\{ \frac{\pi(\theta|D)}{\pi(\theta|D_{(-i)})} \right\} d\theta, \quad (1.1)$$

em que $K(\pi, \pi_{(-i)})$ mede o efeito de remover o i -ésimo caso dos dados completos na distribuição *a posteriori* de θ (Cancho *et al.*, 2011).

A equação (1.1) pode ser simplificada em termos de uma média *a posteriori*.

Considere a seguinte relação:

$$\begin{aligned}
\frac{\pi(\theta|D)}{\pi(\theta|D_{-i})} &= \frac{\prod_{j \in D} f(y_j|\theta)\pi(\theta) \int_{\theta \in \Theta} \prod_{j \in D_{-i}} f(y_j|\theta)\pi(\theta)d\theta}{\prod_{j \in D_{-i}} f(y_j|\theta)\pi(\theta) \int_{\theta \in \Theta} \prod_{j \in D} f(y_j|\theta)\pi(\theta)d\theta} \quad (1.2) \\
&= f(y_i|\theta) \frac{\int_{\theta \in \Theta} \prod_{j \in D_{-i}} f(y_j|\theta)\pi(\theta)d\theta}{\int_{\theta \in \Theta} \prod_{j \in D} f(y_j|\theta)\pi(\theta)d\theta} \\
&= f(y_i|\theta) \frac{\int_{\theta \in \Theta} \prod_{j \in D_{-i}} f(y_j|\theta)\pi(\theta)d\theta}{\int_{\theta \in \Theta} f(y_i|\theta) \prod_{j \in D_{-i}} f(y_j|\theta)\pi(\theta)d\theta} \\
&= f(y_i|\theta) \frac{\int_{\theta \in \Theta} \prod_{j \in D_{-i}} f(y_j|\theta)\pi(\theta)d\theta}{\int_{\theta \in \Theta} f(y_i|\theta) \left(\pi(\theta|D_{-i}) \int_{\theta \in \Theta} \prod_{j \in D_{-i}} f(y_j|\theta)\pi(\theta)d\theta \right) d\theta} \\
&= f(y_i|\theta) \frac{\int_{\theta \in \Theta} \prod_{j \in D_{-i}} f(y_j|\theta)\pi(\theta)d\theta}{\int_{\theta \in \Theta} \prod_{j \in D_{-i}} f(y_j|\theta)\pi(\theta)d\theta \int_{\theta \in \Theta} f(y_i|\theta)\pi(\theta|D_{-i})d\theta} \\
&= f(y_i|\theta) \frac{1}{\int_{\theta \in \Theta} f(y_i|\theta)\pi(\theta|D_{-i})d\theta} = \frac{f(y_i|\theta)}{CPO_i}.
\end{aligned}$$

Aplicando o logaritmo natural em ambos os lados da equação (1.2), obtem-se

$$\log \left(\frac{\pi(\theta|D)}{\pi(\theta|D_{-i})} \right) = \log(f(y_i|\theta)) - \log(CPO_i). \quad (1.3)$$

Substituindo (1.3) em (1.1), a divergência de K-L pode ser escrita como uma média *a posteriori*

$$K(\pi, \pi_{(-i)}) = -\log(CPO_i) + E_{\theta|D} \{ \log[f(y_i|\theta)] \}. \quad (1.4)$$

Uma estimativa para a divergência de K-L pode ser obtida através da amostra MCMC da distribuição *a posteriori* $\pi(\theta|D)$.

Portanto uma estimativa de Monte Carlo para a equação (1.4) é dada por

$$\hat{K}(\pi, \pi_{(-i)}) = -\log(\widehat{CPO}_i) + \frac{1}{Q} \sum_{q=1}^Q \log[f(y_i|\theta_q)],$$

em que Q denota o tamanho da amostra MCMC. Valores altos da divergência de K-L indicam a presença de pontos influentes no conjunto de dados. Para verificar se a i -ésima

observação é um ponto influente ou não, é determinada uma medida de calibração da divergência de K-L dada por

$$p_i = 0,5 \left[1 + \sqrt{1 - \exp \left\{ -2\hat{K}(\pi, \pi_{(-i)}) \right\}} \right],$$

como mostrado em [McCulloch \(1989\)](#) e [Cho et al. \(2009\)](#). Se o valor de p_i é muito maior do que 0,5, o i -ésimo caso é considerado influente.

1.4 Organização dos capítulos

A abordagem Bayesiana para a distribuição normal generalizada, os estudos de simulação e uma aplicação a conjunto de dados reais são apresentados no Capítulo 2. O Capítulo 3 apresenta a abordagem Bayesiana para a distribuição lognormal generalizada, estudos de simulação e uma aplicação a conjunto de dados reais. O Capítulo 4 apresenta a abordagem Bayesiana para o modelo de regressão normal generalizado, um estudo com dados artificiais e uma aplicação a conjunto de dados reais. Por fim, o Capítulo 5 apresenta as considerações finais e propostas futuras do trabalho.

Capítulo 2

Distribuição normal generalizada

Na literatura estatística, muitas pesquisas têm sido desenvolvidas sob a suposição de que os dados seguem uma distribuição normal. Entretanto essa suposição pode não ser realista quando os dados apresentam um comportamento naturalmente simétrico, com caudas mais leves ou mais pesadas que a distribuição normal. Ao contrário da distribuição t de Student, que possui caudas mais pesadas que a distribuição normal, a distribuição NG pode fornecer caudas mais leves ou mais pesadas do que a distribuição normal.

A primeira formulação da distribuição NG foi proposta por [Subbotin \(1923\)](#) como uma generalização da distribuição normal, caracterizada pelos parâmetros locação, escala e forma. Essa distribuição é também conhecida como distribuição potência exponencial ou distribuição de erro exponencial.

A parametrização da distribuição NG considerada neste trabalho foi estudada por [Nadarajah \(2005\)](#).

Definição 2.1. Considera-se que a variável aleatória Y tem distribuição Normal Generalizada, denotada por $Y \sim \text{NG}(\mu, \tau, s)$, se sua função densidade de probabilidade (f.d.p.) for dada por

$$f_Y(y|\mu, \tau, s) = \frac{s}{2\tau\Gamma\left(\frac{1}{s}\right)} \exp\left\{-\left|\frac{y-\mu}{\tau}\right|^s\right\}, \quad y, \mu \in \mathfrak{R}, \quad \tau, s > 0. \quad (2.1)$$

O parâmetro μ é a média, τ é o fator de escala e s é o parâmetro de forma. Em

particular, a média, a variância e o coeficiente de curtose de Y são dados por

$$E(Y) = \mu, \text{Var}(Y) = \frac{\tau^2 \Gamma\left(\frac{3}{s}\right)}{\Gamma\left(\frac{1}{s}\right)} e^{-\gamma} = \frac{\Gamma\left(\frac{1}{s}\right) \Gamma\left(\frac{5}{s}\right)}{\Gamma^2\left(\frac{3}{s}\right)},$$

respectivamente.

A distribuição NG descreve distribuições leptocúrticas se $0 < s < 2$ e distribuições platicúrticas se $s > 2$. Alguns casos particulares da distribuição NG são a distribuição Laplace quando $s = 1$, a distribuição normal, quando considera-se $s = 2$ e substitui-se τ por $\sqrt{2}\tau$ e quando $s \rightarrow \infty$ a f.d.p. em (2.1) converge pontualmente para uma distribuição uniforme em $(\mu - \tau, \mu + \tau)$.

A distribuição NG é simétrica em relação à média e unimodal. Essa distribuição permite um ajuste mais flexível da curtose do que a distribuição normal. Além disso, a capacidade da distribuição NG em fornecer um ajuste preciso para os dados depende da sua forma.

A parametrização em (2.1) fornece uma versão muito complicada da matriz de informação Fisher, em que a expressão obtida para o parâmetro de forma é complexa. Devido à essa complexidade, adota-se uma reparametrização similar a [Zhu & Zinde-Walsh \(2009\)](#) para simplificar a matriz de informação de Fisher. Neste caso, vale ressaltar que [Zhu & Zinde-Walsh \(2009\)](#) propuseram uma reparametrização para a distribuição potência exponencial assimétrica.

Considerando uma reparametrização similar, $\sigma = \tau \Gamma\left(1 + \frac{1}{s}\right)$, tem-se que a distribuição NG em (2.1) é dada por

$$f_Y(y|\mu, \sigma, s) = 2^{-1} \sigma^{-1} \exp \left\{ - \left(\frac{\Gamma\left(1 + \frac{1}{s}\right) |y - \mu|}{\sigma} \right)^s \right\}, \quad y, \mu \in \mathfrak{R}, \sigma > 0, s > 0. \quad (2.2)$$

Esta função densidade é representada na Figura 2.1. Nota-se que a distribuição apresenta caudas mais pesadas do que a normal, quando $s < 2$, ou mais leves do que a normal, quando $s > 2$.

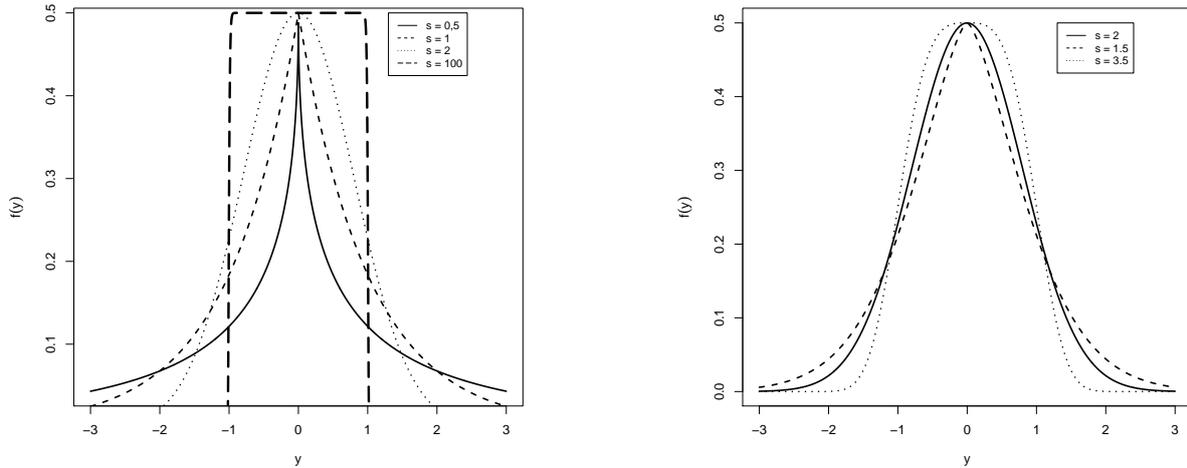


Figura 2.1: Função densidade da distribuição dada em (2.2) considerando os parâmetros $\mu = 0$ e $\sigma = 1$.

Para inferir sobre os parâmetros do modelo em (2.2), adota-se uma abordagem Bayesiana objetiva. A função de verossimilhança e as funções *a priori* e *a posteriori* para os parâmetros dos modelos são apresentadas a seguir.

2.1 Função de verossimilhança

Seja $\mathbf{y} = \{y_1, y_2, \dots, y_n\}$ uma amostra aleatória da distribuição em (2.2). A função de verossimilhança é dada por

$$L(\mu, \sigma, s|\mathbf{y}) = 2^{-n} \sigma^{-n} \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^s \right\}. \quad (2.3)$$

A função de log-verossimilhança em (2.3) é dada por

$$\log L(\mu, \sigma, s|\mathbf{y}) = -n \log 2 - n \log \sigma - \sum_{i=1}^n \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^s. \quad (2.4)$$

A derivada de primeira ordem da função de log-verossimilhança em (2.4) é dada

por

$$\frac{\partial \log L}{\partial \mu} = \frac{s\Gamma(1 + \frac{1}{s})}{\sigma} \sum_{i=1}^n \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^{s-1} \text{sign}(y_i - \mu) \quad (2.5)$$

$$\frac{\partial \log L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{s}{\sigma} \sum_{i=1}^n \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^s \quad (2.6)$$

$$\frac{\partial \log L}{\partial s} = -\sum_{i=1}^n \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^s \left[\log \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right) - \frac{\Psi(1 + \frac{1}{s})}{s} \right], \quad (2.7)$$

em que $s > 1$ e $\Psi(s) = \frac{\Gamma'(s)}{\Gamma(s)}$ é a função digamma. A restrição $s > 1$ garante que a distribuição NG seja diferenciável e unimodal, condições necessárias para estimação do parâmetro de localização. Além disso, a restrição $s > 1$ deixa de fora o caso da distribuição de Laplace quando $s = 1$.

Considerando a Definição 1.1, tem-se que os elementos da matriz de informação de Fisher para $\boldsymbol{\theta} = (\mu, \sigma, s)$ no modelo em (2.2) são dados por

$$\begin{aligned} I_{11}(\boldsymbol{\theta}) &= E \left[\left(\frac{\partial \log L}{\partial \mu} \right)^2 \right] = \frac{n\Gamma(\frac{1}{s})\Gamma(2 - \frac{1}{s})}{\sigma^2}, \\ I_{12}(\boldsymbol{\theta}) &= E \left[\left(\frac{\partial \log L}{\partial \mu} \right) \left(\frac{\partial \log L}{\partial \sigma} \right) \right] = 0, \\ I_{13}(\boldsymbol{\theta}) &= E \left[\left(\frac{\partial \log L}{\partial \mu} \right) \left(\frac{\partial \log L}{\partial s} \right) \right] = 0, \\ I_{22}(\boldsymbol{\theta}) &= E \left[\left(\frac{\partial \log L}{\partial \sigma} \right)^2 \right] = \frac{ns}{\sigma^2}, \\ I_{23}(\boldsymbol{\theta}) &= E \left[\left(\frac{\partial \log L}{\partial \sigma} \right) \left(\frac{\partial \log L}{\partial s} \right) \right] = -\frac{n}{\sigma s}, \\ I_{33}(\boldsymbol{\theta}) &= E \left[\left(\frac{\partial \log L}{\partial s} \right)^2 \right] = \frac{n}{s^3} \left\{ \left(1 + \frac{1}{s} \right) \Psi' \left(1 + \frac{1}{s} \right) \right\}, \end{aligned}$$

onde $\Psi'(s) = \frac{\partial \Psi(s)}{\partial s}$ é a função trigamma. A prova destes resultados é dada no Apêndice A.1.

A matriz de informação de Fisher e a sua inversa são dadas por

$$I(\boldsymbol{\theta}) = n \begin{bmatrix} \frac{\Gamma(\frac{1}{s})\Gamma(2 - \frac{1}{s})}{\sigma^2} & 0 & 0 \\ 0 & \frac{s}{\sigma^2} & -\frac{1}{\sigma s} \\ 0 & -\frac{1}{\sigma s} & \frac{1}{s^3} \left\{ \left(1 + \frac{1}{s} \right) \Psi' \left(1 + \frac{1}{s} \right) \right\} \end{bmatrix} \quad (2.8)$$

e

$$B(\boldsymbol{\theta}) = \frac{1}{n} \begin{bmatrix} \frac{\sigma^2}{\Gamma(\frac{1}{s})\Gamma(2-\frac{1}{s})} & 0 & 0 \\ 0 & \frac{\sigma^2}{s \left[1 - \frac{s}{(1+s)\Psi'(1+\frac{1}{s})} \right]} & \frac{\sigma s^2}{-s+(1+s)\Psi'(1+\frac{1}{s})} \\ 0 & \frac{\sigma s^2}{-s+(1+s)\Psi'(1+\frac{1}{s})} & \frac{s^4}{-s+(1+s)\Psi'(1+\frac{1}{s})} \end{bmatrix}. \quad (2.9)$$

2.2 Distribuições *a priori* e *a posteriori*

Nesta seção, aplica-se o método de análise de referência e a regra de Jeffreys, descritos nas Subseções 1.1.1 e 1.1.2, respectivamente, para obter funções *a priori* não-informativas para os parâmetros do modelo NG em (2.2).

Função *a priori* de referência

A construção da função *a priori* de referência é feita com o uso de uma parametrização ordenada, segundo sua importância inferencial (Berger & Bernardo, 1992). Neste trabalho, a função *a priori* de referência é invariante em relação à ordenação dos parâmetros de perturbação (Bernardo, 1997).

O vetor de parâmetros (μ, σ, s) é dividido e ordenado em 3 grupos distintos. Considera-se o caso em que μ é o parâmetro de interesse e os demais (σ e s) são parâmetros perturbadores. Para se obter a função *a priori* de referência conjunta dos parâmetros μ, σ e s , foi adotada a seguinte parametrização ordenada:

$$\pi^R(\mu, \sigma, s) = \pi^R(s|\mu, \sigma)\pi^R(\sigma|\mu)\pi^R(\mu).$$

Considere-se a metodologia descrita na Subseção 1.1.1. Seja $H_3(\boldsymbol{\theta})$ igual à matriz de informação de Fisher dada em (2.8), e seja, $h_{33}(\boldsymbol{\theta}) = \frac{1}{s^3} \left\{ \left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) \right\}$ o elemento (3, 3) da matriz $H_3(\boldsymbol{\theta})$. Pelo Corolário 1.1, tem-se que $h_{33}(\boldsymbol{\theta})$ pode ser fatorado da forma $h_{33}^{\frac{1}{2}}(\boldsymbol{\theta}) = \sqrt{\frac{1}{s^3} \left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right)} = f_2(s)g_2(\mu, \sigma)$.

Portanto

$$\pi^R(s|\mu, \sigma) \propto s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) \right]^{\frac{1}{2}}.$$

Seja $B_2(\boldsymbol{\theta})$ a submatriz superior 2×2 de $B(\boldsymbol{\theta})$ e $H_2(\boldsymbol{\theta}) = B_2^{-1}(\boldsymbol{\theta})$. O objetivo é encontrar $h_{22}(\boldsymbol{\theta})$, o elemento inferior à direita (2, 2) da matriz $H_2(\boldsymbol{\theta})$, e mostrar que esse se fatora como $h_{22}^{\frac{1}{2}}(\boldsymbol{\theta}) = f_1(\sigma)g_1(\mu, s)$.

Tem-se que

$$B_2(\boldsymbol{\theta}) = \frac{1}{n} \begin{bmatrix} \frac{\sigma^2}{\Gamma(\frac{1}{s})\Gamma(2-\frac{1}{s})} & 0 \\ 0 & \frac{\sigma^2}{s \left[1 - \frac{s}{(1+s)\Psi'(1+\frac{1}{s})} \right]} \end{bmatrix}$$

e

$$H_2(\boldsymbol{\theta}) = n \begin{bmatrix} \frac{\Gamma(\frac{1}{s})\Gamma(2-\frac{1}{s})}{\sigma^2} & 0 \\ 0 & \frac{s \left[1 - \frac{s}{(1+s)\Psi'(1+\frac{1}{s})} \right]}{\sigma^2} \end{bmatrix}.$$

Portanto $h_{22}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} s \left[1 - \frac{s}{(1+s)\Psi'(1+\frac{1}{s})} \right]$ pode ser fatorado da forma

$$h_{22}^{\frac{1}{2}}(\boldsymbol{\theta}) = \frac{1}{\sigma} \sqrt{s \left[1 - \frac{s}{(1+s)\Psi'(1+\frac{1}{s})} \right]} = f_1(\sigma)g_1(\mu, s).$$

Então,

$$\pi^R(\sigma|\mu) \propto \frac{1}{\sigma}.$$

Finalmente, seja $B_{11}(\boldsymbol{\theta})$ o elemento (1, 1) da matriz $B(\boldsymbol{\theta})$ e $h_{11}(\boldsymbol{\theta}) = B_{11}^{-1}(\boldsymbol{\theta})$. O objetivo é mostrar que $h_{11}(\boldsymbol{\theta})$ se fatora como $h_{11}^{\frac{1}{2}}(\boldsymbol{\theta}) = f_0(\mu)g_0(\sigma, s)$.

Invertendo $B_{11}(\boldsymbol{\theta})$, tem-se que $h_{11}(\boldsymbol{\theta}) = \frac{\Gamma(\frac{1}{s})\Gamma(2-\frac{1}{s})}{\sigma^2}$ pode ser fatorado da forma $h_{11}^{\frac{1}{2}}(\boldsymbol{\theta}) = \sqrt{\frac{\Gamma(\frac{1}{s})\Gamma(2-\frac{1}{s})}{\sigma^2}} = f_0(\mu)g_0(\sigma, s)$.

Portanto

$$\pi^R(\mu) \propto 1.$$

A função *a priori* de referência conjunta relativa à parametrização ordenada é dada por

$$\pi^R(\mu, \sigma, s) \propto \frac{1}{\sigma} \times s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s} \right) \Psi' \left(1 + \frac{1}{s} \right) \right]^{\frac{1}{2}}, \quad (2.10)$$

em que $\mu \in \mathfrak{R}$, $\sigma \in \mathfrak{R}^+$ e $s > 1$. Neste caso, vale ressaltar que a função *a priori* de referência coincide com a função *a priori* de Jeffreys independente para o grupo $\{(\mu), (\sigma), (s)\}$, como mostra o Teorema a seguir.

Teorema 2.1. A função a priori em (2.10) coincide com a função priori de Jeffreys independente para o grupo $\{(\mu), (\sigma), (s)\}$.

Prova: Para a função a priori de Jeffreys independente, π^{JI} , considera-se cada um dos parâmetros como independente, isto é, $\pi^{\text{JI}}(\mu, \sigma, s) \propto \pi^{\text{JI}}(\mu) \pi^{\text{JI}}(\sigma) \pi^{\text{JI}}(s)$. Para cada grupo de parâmetros calcula-se uma priori aplicando a regra de Jeffreys como se os outros parâmetros fossem fixados. Sendo assim, a partir da matriz de informação de Fisher dada em (2.8), tem-se que $\pi^{\text{JI}}(\mu) \propto 1$, $\pi^{\text{JI}}(\sigma) \propto \frac{1}{\sigma}$ e $\pi^{\text{JI}}(s) \propto s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) \right]^{\frac{1}{2}}$. Assim,

$$\pi^{\text{JI}}(\mu, \sigma, s) \propto \frac{1}{\sigma} \times s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) \right]^{\frac{1}{2}}.$$

Portanto a função a priori em (2.10) coincide com a função a priori de Jeffreys independente para o grupo $\{(\mu), (\sigma), (s)\}$, completando a prova. \square

Função a priori de Jeffreys

A função a priori não-informativa de Jeffreys, obtida da raiz quadrada do determinante da matriz de informação de Fisher em (2.8), para (μ, σ, s) , é dada por

$$\pi^{\text{J}}(\mu, \sigma, s) \propto \sigma^{-2} s^{-1} \left[\Gamma \left(\frac{1}{s}\right) \Gamma \left(2 - \frac{1}{s}\right) \left\{ \left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) - 1 \right\} \right]^{\frac{1}{2}}. \quad (2.11)$$

No entanto, a função a priori em (2.11) conduz a uma distribuição a posteriori imprópria.

A seguir, apresentam-se definições e Corolários, de maneira similar a [Salazar et al. \(2012\)](#), para provar que as funções a priori em (2.10) e em (2.11) produzem distribuições a posteriori própria e imprópria, respectivamente.

As funções priori $\pi^{\text{R}}(\mu, \sigma, s)$ e $\pi^{\text{J}}(\mu, \sigma, s)$ pertencem à classe de distribuições a priori impróprias dada por

$$\pi(\mu, \sigma, s) \propto \frac{\pi(s)}{\sigma^a}, \quad a \in \mathfrak{R}, \quad (2.12)$$

onde a é um hiperparâmetro e $\pi(s)$ é a função a priori marginal do parâmetro de forma, para várias escolhas de $\pi(s)$ e a .

A função a priori de referência e a função a priori de Jeffreys são da forma (2.12)

com, respectivamente,

$$a = 1 \text{ e } \pi^R(s) \propto s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) \right]^{\frac{1}{2}}, \quad (2.13)$$

$$a = 2 \text{ e } \pi^J(s) \propto s^{-1} \left[\Gamma \left(\frac{1}{s}\right) \Gamma \left(2 - \frac{1}{s}\right) \left\{ \left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) - 1 \right\} \right]^{\frac{1}{2}}. \quad (2.14)$$

A distribuição *a posteriori* associada com a função *a priori* (2.12) é própria se

$$\int_1^{\infty} L(s|\mathbf{y})\pi(s)ds < \infty, \quad (2.15)$$

onde $L(s|\mathbf{y})$ é a função de verossimilhança integrada de s , dada por

$$\int_{\mathbb{R}} \int_0^{\infty} L(\mu, \sigma, s|\mathbf{y}) \sigma^{-a} d\sigma d\mu. \quad (2.16)$$

Proposição 2.1. *A função a priori marginal para s dada nas equações (2.13) e (2.14) é uma função contínua em $[1, \infty)$ e quando $s \rightarrow \infty$ tem-se que $\pi^R(s) = O(s^{-3/2})$ e $\pi^J(s) = O(s^{-1})$.*

Prova: *A inspeção direta das funções (2.13) e (2.14) mostra que estas funções são contínuas em $[1, \infty)$. Quando $s \rightarrow \infty$ tem-se que $\Gamma \left(\frac{1}{s}\right) = O(s)$ e $\Psi' \left(1 + \frac{1}{s}\right) \rightarrow 1$, 6449.*

Proposição 2.2. *A função de verossimilhança integrada de s para $n > 2 - a$, sob a classe de priori (2.12) é uma função contínua em $[1, \infty)$ e $L(s|\mathbf{y}) = O(1)$ quando $s \rightarrow \infty$.*

Prova: *Ver Apêndice A.2.*

Proposição 2.3. *A função a priori de referência dada em (2.10) conduz a uma distribuição a posteriori própria e a função a priori de Jeffreys dada em (2.11) conduz a uma distribuição a posteriori imprópria.*

Prova: *Ver Apêndice A.2.*

Distribuição *a posteriori*

Neste trabalho, a função *a priori* de referência definida em (2.10), será adotada para obter as inferências *a posteriori* do modelo proposto em (2.2). Portanto, utilizando-se a função de verossimilhança em (2.3) e a função *a priori* de referência conjunta em (2.10), obtém-se a distribuição *a posteriori* conjunta para μ, σ e s ,

$$\pi^R(\mu, \sigma, s|\mathbf{y}) \propto \sigma^{-(n+1)} s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) \right]^{\frac{1}{2}} \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma \left(1 + \frac{1}{s}\right) |y_i - \mu|}{\sigma} \right)^s \right\}. \quad (2.17)$$

As distribuições condicionais completas, associadas com a distribuição *a posteriori* conjunta em (2.17), são dadas por

$$\pi^R(\mu|\sigma, s, \mathbf{y}) \propto \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^s \right\}, \quad (2.18)$$

$$\pi^R(\sigma|\mu, s, \mathbf{y}) \propto \sigma^{-(n+1)} \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^s \right\}, \quad (2.19)$$

$$\pi^R(s|\mu, \sigma, \mathbf{y}) \propto s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) \right]^{\frac{1}{2}} \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^s \right\}. \quad (2.20)$$

As distribuições condicionais completas em (2.18) e em (2.20) não têm um núcleo de uma distribuição conhecida, e a distribuição em (2.19) pode ser reduzida facilmente a uma distribuição gama inversa pela transformação $\lambda = \sigma^s$. As estimativas dos parâmetros de interesse são obtidas por métodos de Monte Carlo em Cadeia de Markov (MCMC), como a utilização do algoritmo Metropolis-Hasting dentro do amostrador de Gibbs (Chib & Greenberg, 1995). Detalhes do algoritmo são dados no Apêndice B.

2.3 Estudo de simulação

Esta seção apresenta algumas propriedades frequentistas dos estimadores Bayesianos, considerando a *priori* de referência dada em (2.10). Foram gerados 1.000 conjuntos de dados com tamanhos de amostras $n = 30, 100, 300$, e 500 de acordo com (2.2), com parâmetros $\mu = 0, \sigma = 1$ e $s = \{1, 05; 2; 3; 5\}$. As amostras *a posteriori* foram geradas pelo algoritmo Metropolis-Hasting dentro do amostrador de Gibbs, através do método MCMC implementado no software R. Uma única cadeia de dimensão 15.000 foi considerada para cada parâmetro. Adotou-se um aquecimento de 5.000 com o objetivo de eliminar o efeito dos valores iniciais, obtendo uma amostra de tamanho 10.000 da distribuição *a posteriori*. Além disso, o valor da constante de escala (k) que controla a taxa de aceitação (TA) do algoritmo foi definido igual a 1 para cada cenário.

Para analisar as propriedades frequentistas dos estimadores Bayesianos, considerou-se as 1.000 amostras MCMC das distribuições *a posteriori* obtidas para cada cenário,

como descrito no início da Seção 2.3. Em cada amostra MCMC obtida, calculou-se os intervalos de credibilidade de máxima densidade de probabilidade (HPD) para uma credibilidade fixada em 95%. De posse destes intervalos, avaliou-se as probabilidades de cobertura de cada parâmetro. Além disso, para cada amostra MCMC, obteve-se as taxas de aceitação (TA) no algoritmo Metropolis-Hasting para cada parâmetro. As médias destas taxas são apresentadas na Tabela C.1 (ver Apêndice C).

A Tabela 2.1 mostra a probabilidade de cobertura do intervalo de credibilidade de máxima densidade de probabilidade (HPD) a 95% para cada tamanho amostral e para cada parâmetro do modelo NG. As convergências empíricas são similares e próximas do nível nominal quando $n > 100$. Entretanto quando $s = 1,05$ as probabilidades de cobertura são maiores do que o nível nominal. Uma explicação para este fato pode ser dada com base no comportamento suave da função *a posteriori* em torno da sua moda, resultando em uma matriz de covariância mal condicionada, que produz uma estimativa grande da variância do estimador.

Tabela 2.1: Probabilidade de cobertura do intervalo de credibilidade de máxima densidade de probabilidade (HPD) a 95% para cada tamanho amostral e para cada parâmetro do modelo NG.

Parâmetros	n			
	30	100	300	500
$\mu = 0$	93,9	95,1	95,2	95,1
$\sigma = 1$	90,3	93,4	94,6	94,9
$s = 1,05$	99,6	97,8	98,3	98,1
$\mu = 0$	93,2	95,1	95,6	94,7
$\sigma = 1$	94,3	95,5	96,2	94,6
$s = 2$	97,9	95,4	95,7	94,8
$\mu = 0$	92,8	94,1	94,5	93,2
$\sigma = 1$	94,3	93,7	95,9	94,3
$s = 3$	96,5	93,7	95,1	95,4
$\mu = 0$	94,5	94,2	95,8	94,4
$\sigma = 1$	95,6	94,1	94,3	96,0
$s = 5$	97,9	95,4	96,2	96,7

A Figura 2.2 mostra a variância e o erro quadrático médio dos estimadores de σ e s para diferentes amostras. Para determinar estes valores, considerou-se as 1.000 amostras MCMC das distribuições *a posteriori* dos parâmetros σ e s . Em cada amostra MCMC, calculou-se a variância (Var) e o erro quadrático médio (EQM). De posse destas estimativas, calculou-se as médias das variâncias e dos EQMs. Estas medidas foram obtidas, como por exemplo, para o parâmetro σ , com as seguintes equações:

$$\text{Var}(\hat{\sigma}) = \frac{\sum_{i=1}^N (\hat{\sigma}_i - \bar{\sigma})^2}{N - 1} \quad \text{e} \quad \text{EQM}(\hat{\sigma}) = \frac{\sum_{i=1}^N (\hat{\sigma}_i - \sigma)^2}{N},$$

em que $\bar{\sigma} = \frac{\sum_{i=1}^N \hat{\sigma}_i}{N}$ e $N = 1.000$. Como esperado, à medida que o tamanho da amostra cresce, a variância e o erro quadrático médio dos estimadores decrescem. Entretanto a redução é mais acentuada para o estimador de σ , em todas as situações consideradas, quando comparada com a redução do estimador de s . Além disso, observa-se que quanto maior for o valor do parâmetro de forma, s , maior será a variância e o erro quadrático médio da estimativa.

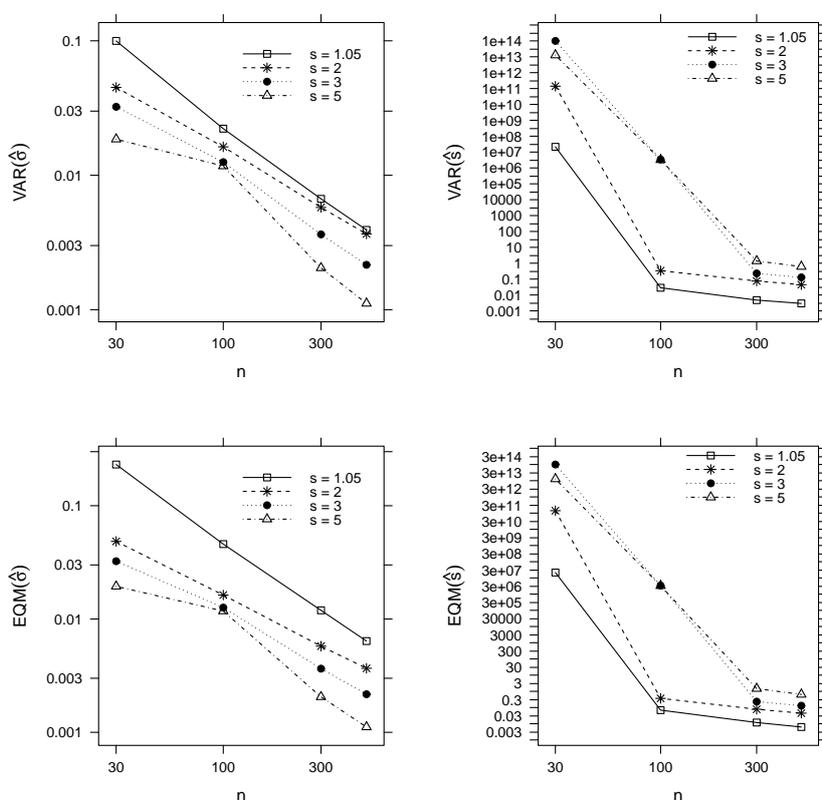


Figura 2.2: Variância e Erro Quadrático Médio dos estimadores Bayesianos para diferentes amostras do modelo NG.

2.4 Aplicação

Com a finalidade de testar a metodologia proposta, considera-se um conjunto de dados com 1.309 informações sobre altura de clones de *Eucalyptus sp.*, plantados em florestas no Sul do Estado de São Paulo. O objetivo é ajustar o modelo normal generalizado e o modelo normal ao conjunto de dados e comparar a eficiência dos respectivos modelos.

Os modelos normal generalizado e normal foram ajustados aos dados pelo procedimento de referência Bayesiana. As amostras da distribuição *a posteriori* foram geradas pelo algoritmo Metropolis-Hasting dentro do amostrador de Gibbs, de maneira similar ao estudo de simulação. Uma única cadeia de dimensão 120.000 foi considerada para cada parâmetro, descartando as primeiras 60.000 iterações (*burn in*) para eliminar o efeito dos primeiros valores. Para evitar a correlação nas séries, considerou-se um espaçamento de tamanho 10, obtendo-se uma amostra final de tamanho 6.000 da distribuição *a posteriori*. Para controlar a taxa de aceitação (TA) do algoritmo, considerou-se as constantes de escala $k = 0,5$ e $k = 1$, para os modelos NG e normal, respectivamente. A convergência das cadeias foi verificada pelo critério de Geweke (1992) (ver Tabela C.2 no Apêndice C). As cadeias resultantes são apresentadas nas Figuras C.1 e C.2 (ver Apêndice C). O tempo de computação requerido por esta aplicação foi de 2,66 minutos, em um computador Core i7 – 2.90GHz com 6.0GB de memória RAM.

A Tabela 2.2 mostra os resumos *a posteriori* para os parâmetros de ambas as distribuições, os critérios de seleção de modelos e as taxas de aceitação (TA) no algoritmo Metropolis-Hasting. O modelo normal generalizado é o mais indicado para representar os dados, pois apresenta um desempenho melhor do que o modelo normal para todos os critérios utilizados. Em particular, a análise do parâmetro de forma ($s > 2$) fornece fortes evidências de uma distribuição platicúrtica para os dados de altura de clones de *Eucalyptus sp.*, o que favorece o modelo NG. Isso é confirmado no gráfico da distribuição *a posteriori* predita apresentado na Figura 2.3.

Tabela 2.2: Resumos *a posteriori* para os parâmetros dos modelos NG e normal, critérios de comparação Bayesiana e taxas de aceitação (TA) no algoritmo Metropolis-Hasting.

Modelo	Parâmetros	Média	Desvio Padrão	HPD (95%)	DIC	EAIC	EBIC	B	TA (%)
NG	μ	21,640	0,093	(21,456; 21,821)	7100,025	7099,552	7107,906	-3555,254	67,79
	σ	5,646	0,132	(5,389; 5,903)					66,50
	s	3,549	0,234	(3,088; 4,003)					66,25
Normal	μ	21,820	0,105	(21,612; 22,024)	7576,504	7575,816	7586,170	-3789,546	58,14
	σ	3,313	0,067	(3,183; 3,444)					58,30

A Figura 2.3 apresenta as distribuições *a posteriori* preditas para ambos os modelos. Percebe-se que o ajuste do modelo é mais pobre quando se considera a distribuição normal. Portanto, há evidência de que a distribuição normal generalizada supera a distribuição normal no ajuste dos dados.

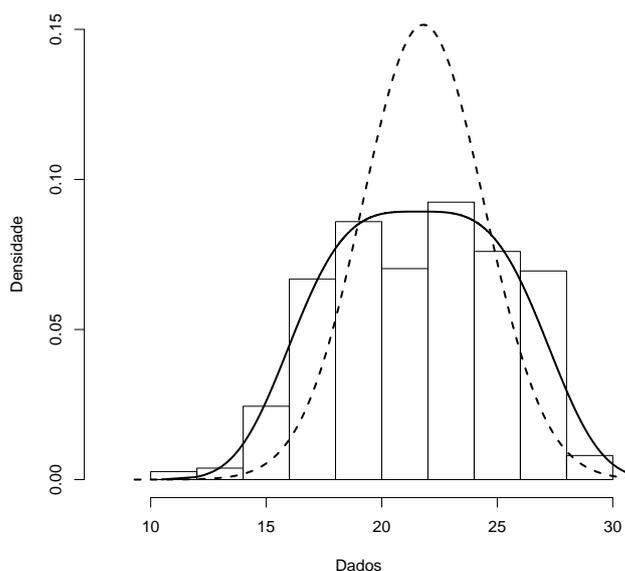


Figura 2.3: Distribuição *a posteriori* predita para os dados de altura de *Eucalyptus sp.*. Linha sólida: distribuição normal generalizada, e linha pontilhada: distribuição normal.

2.4.1 Influência de observações aberrantes

Para investigar a medida de divergência de K-L na detecção de observações influentes no ajuste de um modelo escolhido, considerou-se um conjunto de dados simulados, em que a variável y_i , altura da árvore, foi gerada de uma distribuição NG com parâmetros $\mu = 35, \sigma = 5$ e $s = 3, 5$; para $i = 1, \dots, 500$. Foram selecionados os casos 205 e 465 para perturbação. Escolheu-se um ou dois desses casos selecionados e perturbou-se a variável resposta da seguinte maneira: $\tilde{y}_i = y_i + 5S_y$, em que S_y são os desvios padrão das alturas y_i . As estimativas MCMC foram obtidas de uma maneira similar à seção anterior.

Para mostrar o impacto das observações influentes detectadas no modelo, [Lee et al. \(2006\)](#), em contexto clássico, definem uma medida de mudança relativa total. Neste trabalho, o interesse é mostrar o impacto das observações influentes nas estimativas de cada parâmetro do modelo. Para tanto, calculou-se a medida de variação relativa (VR), a qual é obtida por

$$VR = \left| \frac{\hat{\theta} - \hat{\theta}_0}{\hat{\theta}} \right| \times 100\%, \quad (2.21)$$

em que $\hat{\theta}$ e $\hat{\theta}_0$ são as médias *a posteriori* dos parâmetros do modelo, considerando os dados originais e os dados perturbados, respectivamente.

A Tabela 2.3 mostra as estimativas *a posteriori* para os dados simulados, as variações relativas (VR) das estimativas dos parâmetros, em relação aos dados originais simulados e as taxas de aceitação (TA) no algoritmo Metropolis-Hastings. O conjunto de dados (a) denota os dados originais simulados sem perturbação e os conjuntos de dados de (b) a (d) denotam os dados com casos perturbados. Notam-se maiores valores de VR para as estimativas dos parâmetros σ e s , evidenciando a presença de pontos influentes no conjunto de dados, em que s passa a ser não significativo no modelo com casos perturbados (b) e (d).

Tabela 2.3: Dados simulados. Média *a posteriori*, VR (%) em relação aos dados originais simulados, HPD a 95% para os parâmetros da $NG(35; 5; 3, 5)$ e taxas de aceitação (TA) no algoritmo Metropolis-Hastings.

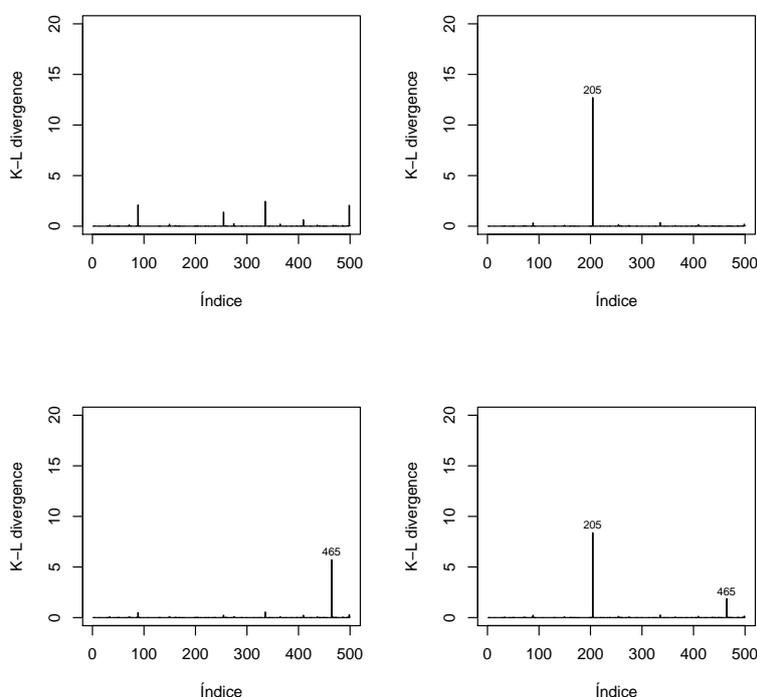
Dados	Casos	Parâmetros	Média	VR (%)	HPD (95%)	TA (%)
a	Nenhum	μ	35,150	–	(34,886; 35,404)	66,14
		σ	4,876	–	(4,439; 5,311)	64,17
		s	3,344	–	(2,521; 4,157)	63,80
b	205	μ	35,200	0,142	(34,911; 35,477)	73,94
		σ	4,634	4,963	(4,213; 5,052)	72,31
		s	2,694	19,438	(2,170; 3,215)	72,97
c	465	μ	35,190	0,114	(34,912; 35,471)	70,68
		σ	4,740	2,789	(4,305; 5,177)	68,92
		s	2,918	12,739	(2,285; 3,533)	69,00
d	{205; 465}	μ	35,220	0,199	(34,928; 35,506)	74,25
		σ	4,589	5,886	(4,156; 5,019)	72,63
		s	2,539	33,044	(2,074; 3,020)	73,41

Considerando as amostras geradas da distribuição *a posteriori* dos parâmetros do modelo normal generalizado, estimaram-se as medidas de divergência de K-L e suas respectivas calibrações para cada um dos casos considerados (a-d), como descrito na seção 1.3. Os resultados da Tabela 2.4 mostram que para os dados sem perturbação (a), os casos selecionados não são influentes, pois apresentam valores pequenos para $K(\pi, \pi_{(-i)})$ e calibração próxima de 0,64. Entretanto quando os dados são perturbados (b - d), os valores de $K(\pi, \pi_{(-i)})$ são maiores e suas respectivas calibrações são próximas ou iguais a 1, indicando que estes dados são influentes.

A Figura 2.4 mostra as medidas de divergência de K-L para o modelo proposto. Nota-se que a medida de K-L tem um bom desempenho para identificar os casos influentes e apresenta valores maiores de $K(\pi, \pi_{(-i)})$ para os casos perturbados quando comparado com os outros casos.

Tabela 2.4: Medidas de diagnóstico para os dados simulados da NG(35; 5; 3, 5).

Dados	Casos	$K(\pi, \pi_{(-i)})$	Calibração
a	205	0,0380	0,6353
	465	0,0023	0,5339
b	205	12,7042	1,0000
c	465	5,7077	1,0000
d	205	8,3793	1,0000
	465	1,8643	0,9940

Figura 2.4: Índices de $K(\pi, \pi_{(-i)})$ para os dados simulados da NG(35; 5; 3, 5).

Com o intuito de investigar a influência dos dados de altura de *Eucalyptus sp.* no ajuste do modelo normal generalizado escolhido, calcularam-se as medidas de divergência de K-L e suas respectivas calibrações. A Figura 2.5 apresenta as medidas de K-L para cada observação. Nota-se que as observações 706, 719, 732, 882, 897 e 912 apresentam os maiores valores de K-L quando comparado com as outras observações. As divergências de K-L e as calibrações referentes às 6 observações que apresentaram os maiores valores de calibração são apresentadas na Tabela 2.5. Percebe-se que as

observações 719 e 882 são possíveis casos influentes. Para avaliar se estas observações alteram as estimativas dos parâmetros do modelo normal generalizado, realizou-se uma análise de sensibilidade.

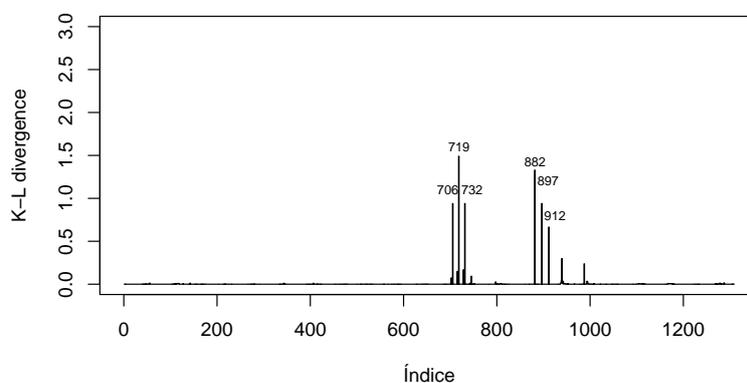


Figura 2.5: Índices de $K(\pi, \pi_{(-i)})$ para os dados de altura de *Eucalyptus sp.*

Tabela 2.5: Medidas de diagnóstico para os dados de altura de *Eucalyptus sp.*

Casos	$K(\pi, \pi_{(-i)})$	Calibração
706	0,9395	0,9602
719	1,4904	0,9871
732	0,9395	0,9602
882	1,3286	0,9821
897	0,9395	0,9602
912	0,6649	0,9288

A Tabela 2.6 mostra as novas estimativas dos parâmetros do modelo após a exclusão dos casos influentes, as variações relativas destas estimativas, em relação aos dados originais para os dados de altura de *Eucalyptus sp.* e as taxas de aceitação (TA) no algoritmo Metropolis-Hastings. Nesse caso, a variação relativa é obtida por (2.21), em que $\hat{\theta}$ e $\hat{\theta}_0$ são as médias a posteriori dos parâmetros do modelo, obtidas dos dados originais e dos dados sem a observação influente, respectivamente. Nota-se uma leve mudança na VR do parâmetro s quando excluimos qualquer observação influente, no entanto a referida mudança não é significativa, pois a exclusão de qualquer observação

praticamente não altera as estimativas dos parâmetros, indicando que modelo normal generalizado é adequado para acomodar observações influentes.

Tabela 2.6: Estimativas *a posteriori*, VR (%) em relação aos dados originais para os dados de altura de *Eucalyptus sp.* e taxas de aceitação (TA) no algoritmo Metropolis-Hastings.

Casos retirados	Parâmetros	Média	VR (%)	HPD (95%)	TA (%)
719	μ	21,640	–	(21,464; 21,824)	67,08
	σ	5,666	0,354	(5,411; 5,921)	65,80
	s	3,672	3,466	(3,181; 4,156)	65,30
882	μ	21,640	–	(21,467; 21,820)	67,19
	σ	5,664	0,319	(5,408; 5,921)	65,78
	s	3,666	3,297	(3,183; 4,143)	65,42
{719; 882}	μ	21,650	0,046	(21,471; 21,824)	66,79
	σ	5,695	0,868	(5,443; 5,947)	65,48
	s	3,819	7,608	(3,298; 4,331)	65,10

2.5 Representação de mistura de escala uniforme

A família de distribuições de mistura de escala uniforme (Qin *et al.*, 1998) tem sido utilizada para representar distribuições que são consideradas alternativas robustas ao modelo normal. Essa família de distribuições também fornece uma abordagem flexível para uma variedade de modelos estatísticos, como por exemplo, modelos robustos, modelos de regressão heterocedástico, modelos econométricos e modelos não-lineares (Qin, 2000). A estrutura hierárquica da representação de mistura de escala uniforme resulta em estratégias computacionais eficientes, principalmente em análise Bayesiana, onde métodos MCMC, como o amostrador de Gibbs, são utilizados. Além disso, o uso de distribuições de mistura de escala uniforme não só simplifica as distribuições condicionais completas exigidas para o amostrador de Gibbs, mas também fornece um meio para diagnóstico de *outliers* (Choy & Chan, 2008).

Embora o algoritmo Metropolis-Hastings tenha sido utilizado para modelar a distribuição NG, o uso de variáveis latentes resulta em um amostrador de Gibbs de fácil implementação. Com o objetivo de implementar análises *a posteriori* pelo método amostrador de Gibbs e avaliar esta estratégia computacional, utiliza-se a distribuição NG como uma mistura de escala uniforme. Neste trabalho, o papel da variável latente não foi explorada como indicador de possíveis *outliers*.

A representação de mistura de escala uniforme para a distribuição NG dada em (2.2), associada com a função *a priori* de referência conjunta dada em (2.10), não fornece uma distribuição condicional completa para o parâmetro s que seja fácil de amostrar, pois o espaço paramétrico do respectivo parâmetro depende do próprio parâmetro. Devido à essa dificuldade, considera-se uma representação de mistura de escala uniforme para a distribuição NG dada em (2.1).

A seguir, considera-se a representação de mistura proposta por [Walker & Gutiérrez-Peña \(1999\)](#) para representar a distribuição NG em (2.1).

Teorema 2.2. *Se $Y|[U = u] \sim U(\mu - \tau u^{\frac{1}{s}}, \mu + \tau u^{\frac{1}{s}})$ e $U \sim Ga(1 + \frac{1}{s}, 1)$, então $Y \sim NG(\mu, \tau, s)$.*

Prova: *Escrevendo a densidade conjunta de Y e U como*

$$f(y|u)f(u) \propto u^{-\frac{1}{s}} \Upsilon\left(u > \left|\frac{y-\mu}{\tau}\right|^s\right) u^{\frac{1}{s}} \exp\{-u\},$$

onde $\Upsilon(\cdot)$ é a função indicadora, tem-se que a densidade marginal de Y é dada por

$$\begin{aligned} f(y) &\propto \int_0^{\infty} \exp\{-u\} \Upsilon\left(u > \left|\frac{y-\mu}{\tau}\right|^s\right) du \\ &\propto \int_{\left|\frac{y-\mu}{\tau}\right|^s}^{\infty} \exp\{-u\} du \\ &\propto \exp\left\{-\left|\frac{y-\mu}{\tau}\right|^s\right\}. \end{aligned}$$

Portanto pode-se concluir que $Y \sim NG(\mu, \tau, s)$, como definido em (2.1).

2.5.1 Distribuição *a posteriori*

Nesta subsecção, utiliza-se a representação de mistura dada no Teorema 2.2 para implementar análises *a posteriori* pelo método amostrador de Gibbs.

Considerando-se a representação de mistura dada no Teorema 2.2, tem-se que para uma amostra $\mathbf{y} = \{y_1, y_2, \dots, y_n\}$ e para o vetor de variáveis latentes $\mathbf{u} = \{u_1, u_2, \dots, u_n\}$, a função de verossimilhança é dada por

$$\begin{aligned} L(\mu, \tau, s | \mathbf{u}, \mathbf{y}) &= \prod_{i=1}^n f(y_i | \mu, \tau, s, u_i) f(u_i | s) \\ &= \prod_{i=1}^n \frac{1}{2\tau u_i^{\frac{1}{s}}} \Upsilon(\mu - \tau u_i^{\frac{1}{s}} < y_i < \mu + \tau u_i^{\frac{1}{s}}) \times \frac{1}{\Gamma(1 + \frac{1}{s})} u_i^{(1 + \frac{1}{s}) - 1} e^{-u_i}. \end{aligned} \quad (2.22)$$

A distribuição *a posteriori* dos parâmetros de interesse e das variáveis latentes é dada por

$$\pi(\mu, \tau, s, \mathbf{u} | \mathbf{y}) \propto \pi(\mu) \pi(\tau) \pi(s) L(\mu, \tau, s | \mathbf{u}, \mathbf{y}).$$

As funções *a priori* de referência e de Jeffreys podem ser obtidas pelos métodos descritos nas Subsecções 1.1.1 e 1.1.2, respectivamente, utilizando-se a matriz de informação de Fisher dada por [Martín & Pérez \(2009\)](#). Considerando a relação estocástica $Y = \log(T)$, onde $T \sim \log NG(\mu, \tau, s)$, tem-se que a matriz de informação de Fisher do modelo NG coincide com a matriz de informação de Fisher do modelo logNG encontrada por [Martín & Pérez \(2009\)](#), e é dada por

$$I(\mu, \tau, s) = n \begin{bmatrix} \frac{(s-1)s\Gamma(1-\frac{1}{s})}{\tau^2\Gamma(\frac{1}{s})} & 0 & 0 \\ 0 & \frac{s}{\tau^2} & -\frac{1+\Psi(1+\frac{1}{s})}{\tau s} \\ 0 & -\frac{1+\Psi(1+\frac{1}{s})}{\tau s} & \frac{1}{s^3} \left\{ (1 + \frac{1}{s}) \Psi' (1 + \frac{1}{s}) + (1 + \Psi (1 + \frac{1}{s}))^2 - 1 \right\} \end{bmatrix}.$$

A sua inversa é dada por

$$B(\mu, \tau, s) = \frac{1}{n} \begin{bmatrix} \frac{\tau^2\Gamma(\frac{1}{s})}{(s-1)s\Gamma(1-\frac{1}{s})} & 0 & 0 \\ 0 & \frac{\tau^2 \{ s\Psi(1+\frac{1}{s})(2+\Psi(1+\frac{1}{s})) + (1+s)\Psi'(1+\frac{1}{s}) \}}{s \{ (s+1)\Psi'(1+\frac{1}{s}) - s \}} & \frac{\tau s^2 (1+\Psi(1+\frac{1}{s}))}{-s+(1+s)\Psi'(1+\frac{1}{s})} \\ 0 & \frac{\tau s^2 (1+\Psi(1+\frac{1}{s}))}{-s+(1+s)\Psi'(1+\frac{1}{s})} & \frac{s^4}{-s+(1+s)\Psi'(1+\frac{1}{s})} \end{bmatrix}.$$

Priori de referência

Para se obter a função *a priori* de referência conjunta dos parâmetros μ, τ e s , foi adotada a seguinte parametrização ordenada:

$$\pi^R(\mu, \tau, s) = \pi^R(s|\mu, \tau)\pi^R(\tau|\mu)\pi^R(\mu).$$

Aplicando a metodologia descrita na Subseção 1.1.1, as funções *a priori* para os parâmetros são dadas por

$$\pi^R(s|\mu, \tau) \propto s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) + \left(1 + \Psi \left(1 + \frac{1}{s}\right)\right)^2 - 1 \right]^{\frac{1}{2}},$$

$$\pi^R(\tau|\mu) \propto \frac{1}{\tau}, \quad \pi^R(\mu) \propto 1.$$

Portanto a função *a priori* de referência conjunta relativa à parametrização ordenada é dada por

$$\pi^R(\mu, \tau, s) \propto \frac{1}{\tau} \times s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) + \left(1 + \Psi \left(1 + \frac{1}{s}\right)\right)^2 - 1 \right]^{\frac{1}{2}}, \quad (2.23)$$

em que $\mu \in \mathfrak{R}, \tau \in \mathfrak{R}^+$ e $s > 1$. Neste caso, vale ressaltar que a função *a priori* de referência para cada parâmetro do modelo proposto coincide com a função *a priori* de Jeffreys independente, encontrada por [Martín & Pérez \(2009\)](#). No entanto, o referido autor considera a função *a priori* de Jeffreys para s muito complicada e utiliza uma função *a priori* diferente da que foi encontrada, ele considera $\pi(s|\mu, \tau) \propto s^{-1}$.

Neste trabalho, diferentemente de [Martín & Pérez \(2009\)](#), considera-se a função *a priori* conjunta dada (2.23). Além disso, essa função *a priori* conduz a uma distribuição *a posteriori* própria.

Função *a priori* de Jeffreys

A função *a priori* não-informativa de Jeffreys, obtida da raiz quadrada do determinante da matriz de informação de Fisher, para (μ, τ, s) , é dada por

$$\pi^J(\mu, \tau, s) \propto \tau^{-2} s^{-1} \left[\frac{(s-1)\Gamma(1-\frac{1}{s})}{\Gamma(1+\frac{1}{s})} \left\{ \left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) - 1 \right\} \right]^{\frac{1}{2}}. \quad (2.24)$$

No entanto, a função *a priori* em (2.24) conduz a uma distribuição *a posteriori* imprópria.

De maneira similar a Seção 2.2, quando $s \rightarrow \infty$ tem-se que $L(s|\mathbf{y}) = O(1)$, $\pi^R(s) = O(s^{-3/2})$ e $\pi^J(s) = O(s^{-1})$. Consequentemente, as funções *a priori* em (2.23) e (2.24) produzem distribuições *a posteriori* própria e imprópria, respectivamente.

Neste trabalho, a função *a priori* de referência definida em (2.23), será adotada para obter as inferências *a posteriori* do modelo proposto em (2.1).

Utilizando-se a função de verossimilhança em (2.22) e a função *a priori* de referência conjunta em (2.23), a distribuição *a posteriori* dos parâmetros de interesse e das variáveis latentes é dada por

$$\begin{aligned} \pi^R(\mu, \tau, s, \mathbf{u}|\mathbf{y}) &\propto \frac{\left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) + \left(1 + \Psi \left(1 + \frac{1}{s}\right)\right)^2 - 1 \right]^{\frac{1}{2}}}{\tau^{(n+1)} \Gamma^n \left(1 + \frac{1}{s}\right) s^{\frac{3}{2}}} \\ &\times \prod_{i=1}^n e^{-u_i} \Upsilon \left(\mu - \tau u_i^{\frac{1}{s}} < y_i < \mu + \tau u_i^{\frac{1}{s}} \right). \end{aligned} \quad (2.25)$$

As distribuições condicionais completas, associadas com a distribuição *a posteriori* conjunta em (2.25), são dadas por

$$\pi^R(\mu|\tau, s, \mathbf{u}, \mathbf{y}) \propto 1, \quad \Upsilon(\mu \in (\underline{\mu}, \bar{\mu})), \quad (2.26)$$

onde

$$\underline{\mu} = \max_i \left\{ y_i - \tau u_i^{\frac{1}{s}} \right\}, \quad \bar{\mu} = \min_i \left\{ y_i + \tau u_i^{\frac{1}{s}} \right\}.$$

$$\pi^R(\tau|\mu, s, \mathbf{u}, \mathbf{y}) \propto \tau^{-(n+1)}, \quad \Upsilon \left(\tau > \max_i \left\{ \frac{|y_i - \mu|}{u_i^{\frac{1}{s}}} \right\} \right), \quad (2.27)$$

$$\pi^R(s|\mu, \tau, \mathbf{u}, \mathbf{y}) \propto \frac{\left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) + \left(1 + \Psi \left(1 + \frac{1}{s}\right)\right)^2 - 1 \right]^{\frac{1}{2}}}{\Gamma^n \left(1 + \frac{1}{s}\right) s^{\frac{3}{2}}}, \quad \Upsilon(s \in (\underline{s}, \bar{s})), \quad (2.28)$$

onde

$$\underline{s} = \max_{i \in s^-} \left\{ 1, \frac{\log(u_i)}{\log \left(\frac{|y_i - \mu|}{\tau} \right)} \right\}, \quad s^- = \left\{ i : \log \left(\frac{|y_i - \mu|}{\tau} \right) < 0 \right\},$$

$$\bar{s} = \min_{i \in s^+} \left\{ \frac{\log(u_i)}{\log \left(\frac{|y_i - \mu|}{\tau} \right)} \right\}, \quad s^+ = \left\{ i : \log \left(\frac{|y_i - \mu|}{\tau} \right) > 0 \right\}.$$

$$\pi^R(u_i|\mu, \tau, s, \mathbf{y}) \propto e^{-u_i}, \quad \Upsilon \left(u_i > \left(\frac{|y_i - \mu|}{\tau} \right)^s \right), \quad i = 1, \dots, n. \quad (2.29)$$

As distribuições condicionais completas em (2.26), (2.27) e (2.29) são Uniforme,

Pareto tipo I e Exponencial, respectivamente. A distribuição condicional completa em (2.28) não tem um núcleo de uma distribuição conhecida, mas pode ser facilmente gerada através do método de aceitação-rejeição.

O algoritmo consiste na escolha de valores iniciais de $\mu^{(0)}, \tau^{(0)}$ e $s^{(0)}$, e da amostragem iterativa de $u^{(j)}, \mu^{(j)}, \tau^{(j)}$ e $s^{(j)}$ das distribuições condicionais completas de (2.29), (2.26), (2.27) e (2.28), respectivamente.

2.5.2 Aplicação

Com a finalidade de testar a metodologia proposta, considera-se um conjunto de dados com 1.309 informações sobre altura de clones de *Eucalyptus sp.*, descrito na Seção 2.4.

O modelo normal generalizado foi ajustado aos dados pelo procedimento de referência Bayesiana. As amostras da distribuição *a posteriori* foram geradas pelo método amostrador de Gibbs. Uma única cadeia de dimensão 180.000 foi considerada para cada parâmetro, descartando as primeiras 100.000 iterações (*burn in*) para eliminar o efeito dos primeiros valores. A convergência das cadeias foi verificada pelo critério de Geweke (1992) (ver Tabela C.2 no Apêndice C). No entanto, as cadeias resultantes são altamente correlacionadas (Ver Tabela C.3 e Figura C.3 no Apêndice C). Estes resultados já eram esperados, tendo em vista o uso de variáveis latentes no modelo, isto é, dados aumentados (Chen *et al.*, 2000). Diante destes resultados, é interessante lembrar também que uma das principais vantagens do uso de espaçamento (salto) em uma cadeia de sequência longa é a facilidade de se calcular as variâncias dos estimadores obtidos a partir da amostra da distribuição *a posteriori* sob a hipótese de uma amostra aproximadamente independente (Carlin & Louis, 2001a). Entretanto o procedimento de espaçamento de uma cadeia sempre aumenta a variância dos estimadores obtidos por média amostral (MacEachern & Berliner, 1994; Carlin & Louis, 2001a). Além disso, este procedimento demanda elevado tempo computacional (Kacperczyk *et al.*, 2013). Portanto todas as iterações após o *burn in* têm sido utilizadas. Neste caso, vale ressaltar que uma decisão semelhante foi tomada por Naranjo *et al.* (2014) ao utilizar uma representação de mistura escala da distribuição exponencial assimétrica. O tempo de computação requerido por esta aplicação foi de 1,28 dias, em um computador Core i7 – 2.90GHz com 6.0GB de

m moria RAM.

A Tabela 2.7 mostra os resumos *a posteriori* para os par metros da distribui o NG. Como esperado, considerando-se a transforma o $\sigma = \tau\Gamma\left(1 + \frac{1}{s}\right)$, tem-se que os valores das m dias na Tabela 2.7 est o bem pr ximas das estimativas obtidas com o algoritmo Metropolis-Hastings dentro do amostrador de Gibbs, apresentadas na Tabela 2.2. Atrav s desta tabela pode-se observar tamb m que o par metro de forma ($s > 2$) fornece fortes evid ncias de uma distribui o platic rtica para os dados de altura de clones de *Eucalyptus sp.*, o que favorece o modelo NG.

Tabela 2.7: Resumos *a posteriori* para os par metros do modelo NG.

Modelo	Par�metros	M�dia	Desvio Padr�o	HPD (95%)
	μ	21,640	0,094	(21,454; 21,823)
NG	τ	6,256	0,133	(5,982; 6,509)
	s	3,582	0,242	(3,130; 4,051)

2.6 Considera es finais

Neste cap tulo, apresentou-se a distribui o normal generalizada do ponto de vista da an lise Bayesiana objetiva. Estudos de simula es foram feitos para verificar a adequa o do m todo inferencial proposto, considerando diferentes tamanhos amostrais. Os resultados simulados mostraram melhores propriedades frequentistas para tamanhos de amostra acima de 100. Al m disso, verificou-se, por meio do estudo de simula o e dos dados reais, que a diverg ncia de Kullback-Leibler detecta de maneira eficaz os pontos influentes no ajuste do modelo normal generalizado. A retirada dos pontos influentes do conjunto de dados reais mostra que o modelo normal generalizado   adequado para acomodar observa es influentes. A distribui o normal generalizada tamb m foi representada como uma mistura de escala uniforme. Estudos com dados reais mostraram, como esperado, que as estimativas do modelo proposto, obtidas pelo m todo amostrador de Gibbs, ficaram bem pr ximas das estimativas do modelo normal generalizado com reparametriza o, obtidas pelo algoritmo Metropolis-Hastings dentro do amostrador de Gibbs. No entanto, no estudo com dados reais, n o foi observado um ganho computacional com a inclus o da vari vel latente. A inclus o desta vari vel

aumentou as correlações nas amostras estimadas, de modo que foi necessário aumentar o número de iterações da cadeia de MCMC para atingir convergência. Consequentemente, o tempo de processamento desse algoritmo foi consideravelmente maior do que a do algoritmo Metropolis-Hastings. No capítulo 3, apresenta-se a distribuição lognormal generalizada do ponto de vista da análise Bayesiana objetiva.

Capítulo 3

Distribuição lognormal generalizada com aplicação a dados de sobrevivência

A distribuição lognormal generalizada, ao contrário das distribuições Weibull e lognormal, conduz a inferência que é robusta a observações aberrantes. Essa distribuição foi introduzida por [Vianelli \(1982, 1983\)](#) como sendo uma família de distribuições de erros multiplicativos de ordem r . Outras parametrizações podem ser encontradas na literatura. Ver, por exemplo, [Hill \(1963\)](#) e [Chen \(1995\)](#).

Neste trabalho, considera-se que a distribuição logNG é obtida através da transformação exponencial de uma variável aleatória que segue a distribuição normal generalizada estudada por [Nadarajah \(2005\)](#), de maneira similar a [Martín & Pérez \(2009\)](#).

Definição 3.1. Uma variável aleatória T tem distribuição logNG, denotada por $T \sim \text{logNG}(\mu, \tau, s)$, se sua função densidade de probabilidade (f.d.p.) for dada por

$$f_T(t|\mu, \tau, s) = \frac{s}{2t\tau\Gamma\left(\frac{1}{s}\right)} \exp\left\{-\left|\frac{\log(t) - \mu}{\tau}\right|^s\right\}, \quad (3.1)$$

com $t > 0$, $\mu \in \mathfrak{R}$, $\tau, s > 0$.

A distribuição logNG é a distribuição de $T = g(Y) = \exp(Y)$, onde $Y \sim \text{NG}(\mu, \tau, s)$, como foi definido em (2.1). Assim, a função densidade de T pode ser

escrita, através de (2.1), como

$$\begin{aligned} f_T(t|\mu, \tau, s) &= f_Y(g^{-1}(t)) \left| \frac{d}{dt} g^{-1}(t) \right| \\ &= f_Y(\log(t)) \frac{1}{t} \\ &= \frac{s}{2t\tau\Gamma\left(\frac{1}{s}\right)} \exp\left\{-\left|\frac{\log(t) - \mu}{\tau}\right|^s\right\}. \end{aligned}$$

Através dessa relação é possível obter uma série de propriedades para a distribuição logNG a partir dos resultados conhecidos da distribuição NG. Na equação (3.1), e^μ é um parâmetro de escala, τ e s são parâmetros de forma.

A distribuição logNG, tem como caso particular, a distribuição lognormal quando considera-se $s = 2$ e substitui-se τ por $\sqrt{2}\tau$. Para o caso em que $s = 1$, obtém-se a distribuição log-Laplace. Segundo [Martín & Pérez \(2009\)](#), a distribuição logNG apresenta um ajuste melhor aos dados em relação à distribuição lognormal quando $s \in (1, 2) \cup (2, 3)$. A distribuição logNG permite um ajuste mais flexível da curtose do que a distribuição lognormal. Além disso, a capacidade da distribuição logNG fornecer um ajuste preciso para os dados depende da sua forma.

Considerando uma reparametrização similar ao caso da NG, $\sigma = \tau\Gamma\left(1 + \frac{1}{s}\right)$, tem-se que a distribuição logNG em (3.1) é dada por

$$f_T(t|\mu, \sigma, s) = 2^{-1} t^{-1} \sigma^{-1} \exp\left\{-\left(\frac{\Gamma\left(1 + \frac{1}{s}\right) |\log(t) - \mu|}{\sigma}\right)^s\right\}, \quad t > 0, \mu \in \mathfrak{R}, \sigma, s > 0. \quad (3.2)$$

As funções de sobrevivência e de risco associadas à densidade em (3.2) são dadas, respectivamente, por:

$$S(t|\mu, \sigma, s) = \begin{cases} 1 - \frac{\Gamma\left(\frac{1}{s}, \left(\frac{\Gamma\left(1 + \frac{1}{s}\right)(\mu - \log(t))}{\sigma}\right)^s\right)}{2s\Gamma\left(1 + \frac{1}{s}\right)} & \text{se } 0 < t \leq \exp(\mu) \\ \frac{\Gamma\left(\frac{1}{s}, \left(\frac{\Gamma\left(1 + \frac{1}{s}\right)(\log(t) - \mu)}{\sigma}\right)^s\right)}{2s\Gamma\left(1 + \frac{1}{s}\right)} & \text{se } t > \exp(\mu), \end{cases} \quad (3.3)$$

e

$$h(t|\mu, \sigma, s) = \begin{cases} \frac{s \Gamma(1+\frac{1}{s}) \exp\left\{-\left(\frac{\Gamma(1+\frac{1}{s})|\log(t)-\mu|}{\sigma}\right)^s\right\}}{t \sigma \left\{2s\Gamma(1+\frac{1}{s})-\Gamma\left(\frac{1}{s}, \left(\frac{\Gamma(1+\frac{1}{s})(\mu-\log(t))}{\sigma}\right)^s\right)\right\}} & \text{se } 0 < t \leq \exp(\mu) \\ \frac{s \Gamma(1+\frac{1}{s}) \exp\left\{-\left(\frac{\Gamma(1+\frac{1}{s})|\log(t)-\mu|}{\sigma}\right)^s\right\}}{t \sigma \Gamma\left(\frac{1}{s}, \left(\frac{\Gamma(1+\frac{1}{s})(\log(t)-\mu)}{\sigma}\right)^s\right)} & \text{se } t > \exp(\mu), \end{cases} \quad (3.4)$$

em que $\Gamma(\cdot, \cdot)$ é a função gama incompleta. A prova destes resultados é dada no Apêndice A.3.

A Figura 3.1 apresenta os gráficos da função densidade, da função de sobrevivência e da função de risco da distribuição logNG para diferentes valores de s , assumindo $\mu = 0$ e $\sigma = 1$. Propriedades das funções de sobrevivência e de risco podem ser encontradas em [Gupta & Lvin \(2005\)](#).

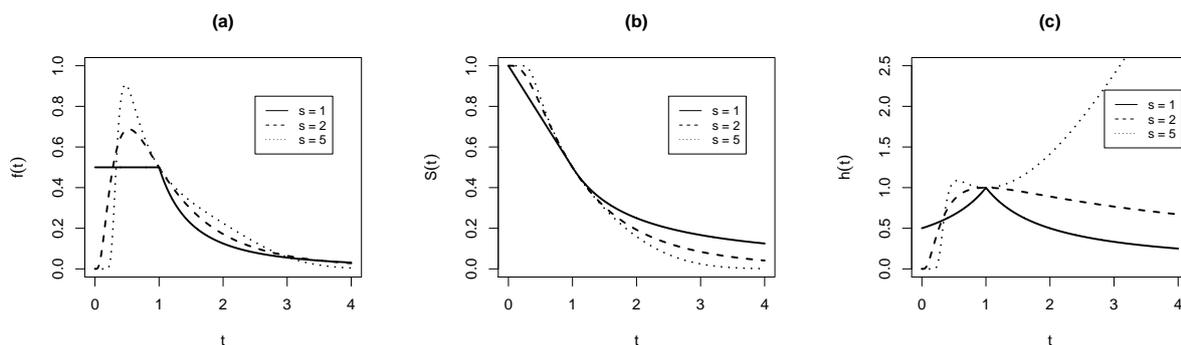


Figura 3.1: Função densidade (a), função de sobrevivência (b) e função de risco (c) para $\mu = 0, \sigma = 1$ e diferentes valores de s .

Para inferir sobre os parâmetros do modelo em (3.2), adota-se uma abordagem Bayesiana objetiva. A função de verossimilhança e as funções *a priori* e *a posteriori* para os parâmetros do modelos são apresentadas a seguir.

3.1 Função de verossimilhança

Nesta seção, apresenta-se a função de verossimilhança para o modelo logNG, considerando-se a presença de dados completos e a presença de dados censurados à

direita.

Dados completos

Seja $\mathbf{t} = \{t_1, t_2, \dots, t_n\}$ uma amostra aleatória da distribuição em (3.2). A função de verossimilhança para dados completos é dada por

$$L(\mu, \sigma, s|\mathbf{t}) = 2^{-n} \sigma^{-n} \prod_{i=1}^n t_i^{-1} \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma(1 + \frac{1}{s}) |\log(t_i) - \mu|}{\sigma} \right)^s \right\}. \quad (3.5)$$

Dados censurados à direita

Seja $T_i, i = 1, \dots, n$, o tempo de falha do i -ésimo indivíduo e seja C_i o tempo de censura correspondente. Observam-se os pares (t_i, δ_i) , em que $t_i = \min(T_i, C_i)$ e $\delta_i = 1$ se $T_i \leq C_i$ ou $\delta_i = 0$ se $T_i > C_i$. Consideram-se os tempos de falha e de censura independentes e com distribuição logNG dada em (3.2). Sob a suposição de que o mecanismo de censura é não informativa, a função de verossimilhança é dada por

$$L(\mu, \sigma, s|\mathbf{t}) = \prod_{i=1}^n [f(t_i|\mu, \sigma, s)]^{\delta_i} [S(t_i|\mu, \sigma, s)]^{1-\delta_i},$$

ou equivalentemente por

$$L(\mu, \sigma, s|\mathbf{t}) = \prod_{i=1}^r f(t_i|\mu, \sigma, s) \prod_{j=r+1}^n S(t_j|\mu, \sigma, s),$$

onde as r primeiras observações, $\mathbf{t}_o = (t_1, \dots, t_r)^\top$, são as não censuradas (falhas), as $n - r$ observações seguintes, $\mathbf{t}_c = (t_{r+1}, \dots, t_n)^\top$, são as censuradas, em que $\mathbf{t} = \mathbf{t}_o \cup \mathbf{t}_c$, $f(t_i|\mu, \sigma, s)$ e $S(t_j|\mu, \sigma, s)$ são função densidade e função de sobrevivência, respectivamente.

Assim, a função de verossimilhança pode ser escrita como

$$L(\mu, \sigma, s|\mathbf{t}) = \prod_{i=1}^r f(t_i|\mu, \sigma, s) \prod_{j=r+1}^n \int_{t_j^*}^{\infty} f(t_j|\mu, \sigma, s) dt_j, \quad (3.6)$$

onde $f(\cdot)$ é dada em (3.2).

3.2 Distribuições *a priori* e *a posteriori*

Considerando-se a relação estocástica $T = \exp(Y)$, onde $Y \sim \text{NG}(\mu, \sigma, s)$, tem-se que as funções *a priori* de referência e de Jeffreys do modelo logNG coincidem

com as funções *a priori* encontradas para o modelo NG, dadas em (2.10) e (2.11), respectivamente. Além disso, essas funções *a priori* produzem distribuições *a posteriori* própria e imprópria, respectivamente, para o modelo logNG. A prova é feita de maneira similar ao modelo NG.

A função *a priori* de referência definida em (2.10), será adotada para obter as inferências *a posteriori* do modelo proposto em (3.2), considerando-se a presença de dados completos e a presença de dados censurados.

Para inferir sobre os parâmetros do modelo logNG (3.2), é necessário obter a distribuição *a posteriori*. Assim, considerando a presença de dados completos, e utilizando-se a função de verossimilhança em (3.5) e a função *a priori* de referência conjunta em (2.10), obtém-se a distribuição *a posteriori* conjunta para μ, σ e s ,

$$\pi^R(\mu, \sigma, s | \mathbf{t}) \propto \sigma^{-(n+1)} s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) \right]^{\frac{1}{2}} \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma \left(1 + \frac{1}{s}\right) |\log(t_i) - \mu|}{\sigma} \right)^s \right\}. \quad (3.7)$$

As distribuições condicionais completas, associadas com a distribuição *a posteriori* conjunta em (3.7), são dadas por

$$\pi^R(\mu | \sigma, s, \mathbf{t}) \propto \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma \left(1 + \frac{1}{s}\right) |\log(t_i) - \mu|}{\sigma} \right)^s \right\}, \quad (3.8)$$

$$\pi^R(\sigma | \mu, s, \mathbf{t}) \propto \sigma^{-(n+1)} \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma \left(1 + \frac{1}{s}\right) |\log(t_i) - \mu|}{\sigma} \right)^s \right\}, \quad (3.9)$$

$$\pi^R(s | \mu, \sigma, \mathbf{t}) \propto s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) \right]^{\frac{1}{2}} \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma \left(1 + \frac{1}{s}\right) |\log(t_i) - \mu|}{\sigma} \right)^s \right\}. \quad (3.10)$$

As distribuições condicionais completas em (3.8) e em (3.10) não têm um núcleo de uma distribuição conhecida e a distribuição em (3.9) pode ser reduzida facilmente a uma distribuição gama inversa pela transformação $\lambda = \sigma^s$. As estimativas dos parâmetros de interesse são obtidas pelo algoritmo Metropolis-Hastings dentro do amostrador de Gibbs, de maneira similar ao esquema apresentado no Apêndice B, considerando $\mathbf{y} = \log(\mathbf{t})$.

Quando considera-se a presença de dados censurados à direita, a distribuição *a posteriori* para μ, σ e s associada com a função *a priori* de referência em (2.10) é dada por

$$\pi^R(\mu, \sigma, s|\mathbf{t}) \propto L(\mu, \sigma, s|\mathbf{t}) \pi^R(\mu, \sigma, s), \quad (3.11)$$

onde $L(\mu, \sigma, s|\mathbf{t})$ é a função de verossimilhança para dados censurados, dada em (3.6).

Devido à complexidade da função de sobrevivência dada em (3.3), a determinação das distribuições condicionais completas a partir de (3.11) não é uma tarefa fácil. Portanto considera-se \mathbf{t}_c como um conjunto de não observáveis (Upadhyay *et al.*, 2001; Upadhyay & Peshwani, 2001). A distribuição *a posteriori* conjunta dos parâmetros e a distribuição *a posteriori* das observações com censuras são dadas, respectivamente, por

$$\pi^R(\mu, \sigma, s|\mathbf{t}_o, \mathbf{t}_c) = \pi^R(\mu, \sigma, s|\mathbf{t}) \quad (3.12)$$

e

$$\pi^R(\mathbf{t}_c|\mu, \sigma, s, \mathbf{t}_o) = f(\mathbf{t}_c|\mu, \sigma, s) = \prod_{j=r+1}^n \int_{t_j^*}^{\infty} f(t_j|\mu, \sigma, s) dt_j. \quad (3.13)$$

A expressão em (3.12) corresponde à distribuição *a posteriori* conjunta dada em (3.7) e a sua geração procede, como no caso não censurado, por MCMC. A expressão em (3.13) corresponde à distribuição conjunta de observações censuradas independentes e sua geração é feita da distribuição logNG truncada no valor censurado, ou seja, somente são possíveis valores gerados maiores que o valor observado da variável censurada no momento da análise.

3.3 Aplicações

Nesta seção, a metodologia desenvolvida é aplicada em dados artificiais e em dados reais.

3.3.1 Dados artificiais

Considerou-se um conjunto de dados artificiais, em que a variável T_i , o tempo de sobrevivência em dias, foi gerada de uma distribuição logNG com parâmetros $\mu = 4$, $\sigma = 1,7$ e $s = 1,9$; para $i = 1, \dots, 150$. O tempo de censura C_i também foi gerado da distribuição logNG com parâmetros $\mu = \alpha$, $\sigma = 1,7$ e $s = 1,9$; em que α é escolhido de forma a controlar a quantidade de dados censurados. Neste estudo, considerou-se 10% de dados censurados e observou-se o par (t_i, δ_i) , onde $t_i = \min(T_i, C_i)$ e $\delta_i = 1$ se $T_i \leq C_i$ ou $\delta_i = 0$ se $T_i > C_i$.

Novos valores para as observações censuradas foram gerados, utilizando-se a expressão em (3.13), que corresponde à distribuição logNG truncada no valor censurado. As amostras *a posteriori* foram geradas da expressão em (3.12), através do método MCMC. Uma única cadeia de dimensão 20.000 foi considerada para cada parâmetro, descartando as primeiras 10.000 iterações (*burn in*) para eliminar o efeito dos primeiros valores. Para evitar a correlação nas séries, considerou-se um espaçamento de tamanho 5, obtendo-se uma amostra final de tamanho 2.000 da distribuição *a posteriori*. Para controlar a taxa de aceitação (TA) do algoritmo, considerou-se a constante de escala $k = 1$. A convergência das cadeias foi verificada pelo critério de Geweke (1992) (ver Tabela C.4 no Apêndice C). As cadeias resultantes são apresentadas na Figura C.4 (ver Apêndice C).

A Tabela 3.1 mostra os resumos *a posteriori* para os parâmetros do modelo lognormal generalizado e as taxas de aceitação (TA) no algoritmo Metropolis-Hastings. Percebe-se que as estimativas estão próximas aos valores verdadeiros e os intervalos de credibilidade de máxima densidade de probabilidade (HPD) a 95% cobrem os verdadeiros valores dos parâmetros.

Tabela 3.1: Dados artificiais. Resumos *a posteriori* para os parâmetros do modelo lognormal generalizado e as taxas de aceitação (TA) no algoritmo Metropolis-Hastings.

Parâmetros	Média	Mediana	Desvio Padrão	HPD (95%)	TA (%)
μ	3,969	3,968	0,071	(3,821; 4,121)	73,11
σ	1,781	1,780	0,082	(1,610; 1,944)	42,49
s	1,895	1,906	0,183	(1,528; 2,226)	32,88

Para investigar a medida de divergência de K-L na detecção de observações influentes no ajuste de um modelo lognormal generalizado, considerou-se a mesma amostra previamente simulada. Foram selecionados os casos 19 e 148 para perturbação, em que estes casos representam o tempo de falha e o tempo de censura, respectivamente. Escolheu-se um ou dois desses casos selecionados e perturbou-se a variável resposta da seguinte maneira: $\tilde{t}_i = t_i + 50S_t$, em que S_t são os desvios padrão dos tempos t_i . As estimativas MCMC foram obtidas de uma maneira similar à informação acima.

Para mostrar o impacto das observações influentes nas estimativas de μ, σ e s , calculou-se a medida de variação relativa (VR), que é obtida por (2.21), em que $\hat{\theta}$ e $\hat{\theta}_0$ são as médias *a posteriori* dos parâmetros do modelo, considerando os dados originais e os dados perturbados, respectivamente.

A Tabela 3.2 mostra as estimativas *a posteriori* para os dados simulados, as variações relativas (VR) das estimativas dos parâmetros em relação aos dados originais simulados e as taxas de aceitação (TA) no algoritmo Metropolis-Hastings. O conjunto de dados (a) denota os dados originais simulados sem perturbação e os conjuntos de dados de (b) a (d) denotam os dados com casos perturbados. Notam-se maiores valores de VR para as estimativas dos parâmetros σ e s , evidenciando a presença de pontos influentes no conjunto de dados, em que s passa a ser não significativo no modelo com casos perturbados (b) a (d).

Tabela 3.2: Dados simulados. Média *a posteriori*, VR (%) em relação aos dados originais simulados, HPD a 95% para os parâmetros da distribuição $\text{logNG}(4; 1, 7; 1, 9)$ e taxas de aceitação no algoritmo Metropolis-Hastings.

Dados	Casos	Parâmetros	Média	VR (%)	HPD (95%)	TA (%)
a	Nenhum	μ	3,969	–	(3,821; 4,121)	73,11
		σ	1,781	–	(1,610; 1,944)	42,49
		s	1,895	–	(1,528; 2,226)	32,88
b	19	μ	3,958	0,277	(3,800; 4,117)	72,45
		σ	1,690	5,110	(1,534; 1,843)	39,29
		s	1,617	14,670	(1,330; 1,862)	29,34
c	148	μ	3,947	0,554	(3,792; 4,101)	73,25
		σ	1,666	6,457	(1,501; 1,827)	41,62
		s	1,594	15,884	(1,311; 1,842)	32,12
d	{19; 148}	μ	3,935	0,857	(3,784; 4,092)	72,22
		σ	1,611	9,455	(1,446; 1,775)	41,24
		s	1,446	23,694	(1,221; 1,654)	33,21

Considerando as amostras geradas da distribuição *a posteriori* dos parâmetros do modelo lognormal generalizado, estimaram-se as medidas de divergência de K-L e suas respectivas calibrações para cada um dos casos considerados (a-d), como descrito na Seção 1.3. Os resultados da Tabela 3.3 mostram que, para os dados sem perturbação (a), os casos selecionados não são influentes, pois apresentam valores pequenos para $K(\pi, \pi_{(-i)})$ e calibração próxima de 0,69. Entretanto quando os dados são perturbados (b - d), os valores de $K(\pi, \pi_{(-i)})$ são maiores e suas respectivas calibrações são próximas a 0,9, indicando que estes dados são influentes.

Tabela 3.3: Medidas de diagnóstico para os dados simulados da $\text{logNG}(4; 1, 7; 1, 9)$.

Dados	Casos	$K(\pi, \pi_{(-i)})$	Calibração
a	19	0,0113	0,5747
	148	0,0763	0,6881
b	19	0,5313	0,9045
c	148	1,4083	0,9848
d	19	0,2770	0,8261
	148	0,5139	0,9007

A Figura 3.2 mostra as medidas de divergência de K-L para o modelo proposto. Nota-se que a medida de K-L tem um bom desempenho para identificar os casos influentes, apresenta valores maiores de $K(\pi, \pi_{(-i)})$ para os casos perturbados, quando comparado com os outros casos.

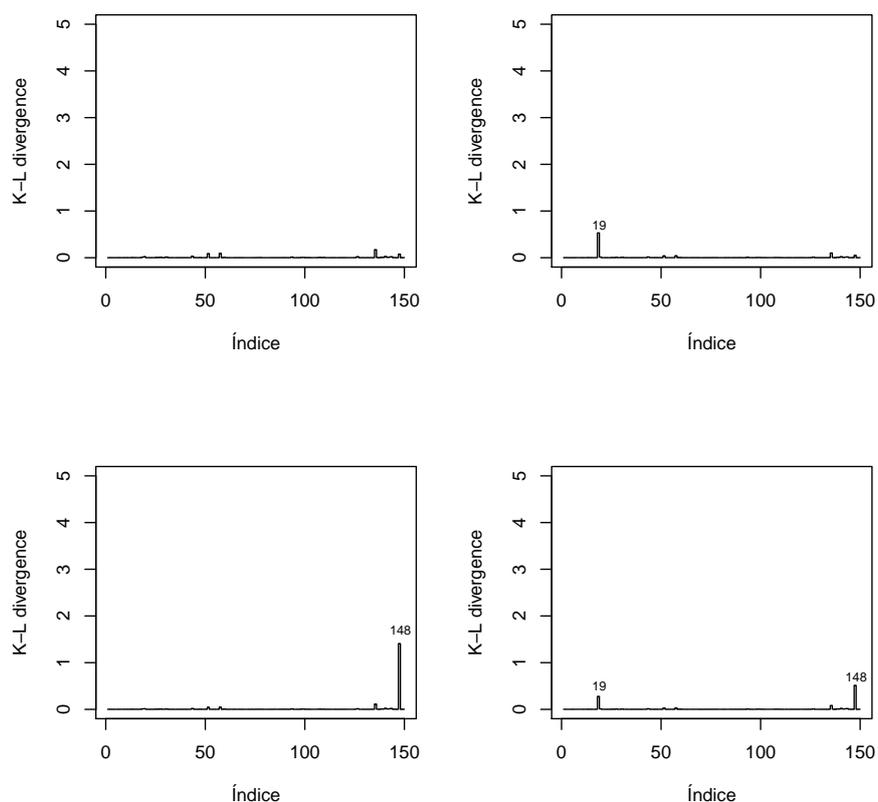


Figura 3.2: Índices de $K(\pi, \pi_{(-i)})$ para os dados simulados da $\text{logNG}(4, 7; 1, 7; 1, 9)$.

3.3.2 Dados reais

Considera-se o conjunto de dados retirado de [Kalbfleisch & Prentice \(2002\)](#). O conjunto de dados refere-se ao estudo de dois tipos de tratamento quimioterápico aplicado em 137 pacientes com câncer de pulmão inoperável. O tempo de sobrevivência dos pacientes foi medido em dias. Os dados apresentam 9 observações censuradas. O objetivo é ajustar o modelo logNG e o modelo lognormal ao conjunto de dados e comparar a eficiência dos respectivos modelos.

Os modelos logNG e lognormal foram ajustados aos dados pelo procedimento de inferência Bayesiana. As amostras das distribuições *a posteriori* foram geradas pelo algoritmo Metropolis-Hastings dentro do amostrador de Gibbs, de maneira similar ao estudo de simulação. Uma única cadeia de dimensão 60.000 foi considerada para cada parâmetro, descartando as primeiras 10.000 iterações (*burn in*) para eliminar o efeito dos primeiros valores. Para evitar a correlação nas séries, considerou-se um espaçamento de

tamanho 10, obtendo-se uma amostra final de tamanho 5.000 da distribuição *a posteriori*. Para controlar a taxa de aceitação (TA) do algoritmo, considerou-se a constante de escala $k = 1, 2$. A convergência das cadeias foi verificada pelo critério de Geweke (1992) (ver Tabela C.5 no Apêndice C). As cadeias resultantes são apresentadas nas Figuras C.5 e C.6 (ver Apêndice C). O tempo de computação requerido por esta aplicação foi de 4,61 minutos, em um computador Core i7 – 2.90GHz com 6.0GB de memória RAM.

A Tabela 3.4 mostra os resumos *a posteriori* para os parâmetros de ambas as distribuições, os critérios de seleção de modelos e as taxas de aceitação (TA) no algoritmo Metropolis-Hastings. O modelo logNG é o mais indicado para representar os conjuntos de dados, pois apresenta um desempenho melhor do que o modelo lognormal para todos os critérios utilizados. Em particular, a análise do parâmetro de forma ($s < 2$) fornece fortes evidências de uma distribuição leptocúrtica para os dados de sobrevivência dos pacientes com câncer de pulmão inoperável, o que favorece o modelo logNG. Isso é confirmado pela curva Kaplan-Meier com função de sobrevivência estimada pela distribuição logNG apresentada na Figura 3.3.

Tabela 3.4: Resumos *a posteriori* para os parâmetros do modelo logNG, critérios de comparação Bayesiana e taxas de aceitação (TA) no algoritmo Metropolis-Hastings.

Modelo	Parâmetros	Média	Desvio Padrão	HPD (95%)	DIC	EAIC	EBIC	B	TA (%)
logNG	μ	4,203	0,080	(4,030; 4,372)					64,28
	σ	1,786	0,067	(1,635; 1,937)	1502,197	1506,849	1510,689	-751,5037	33,42
	s	1,961	0,157	(1,600; 2,236)					23,81
lognormal	μ	4,148	0,073	(3,989; 4,311)	1542,583	1545,044	1550,884	-772,2264	64,87
	σ	1,213	0,042	(1,123; 1,313)					50,24

A Figura 3.3 apresenta as curvas de sobrevivência estimadas por Kaplan-Meier e por ambos os modelos para os dados de câncer de pulmão inoperável. Percebe-se que o modelo indicado apresenta um ajuste melhor do que o modelo lognormal.

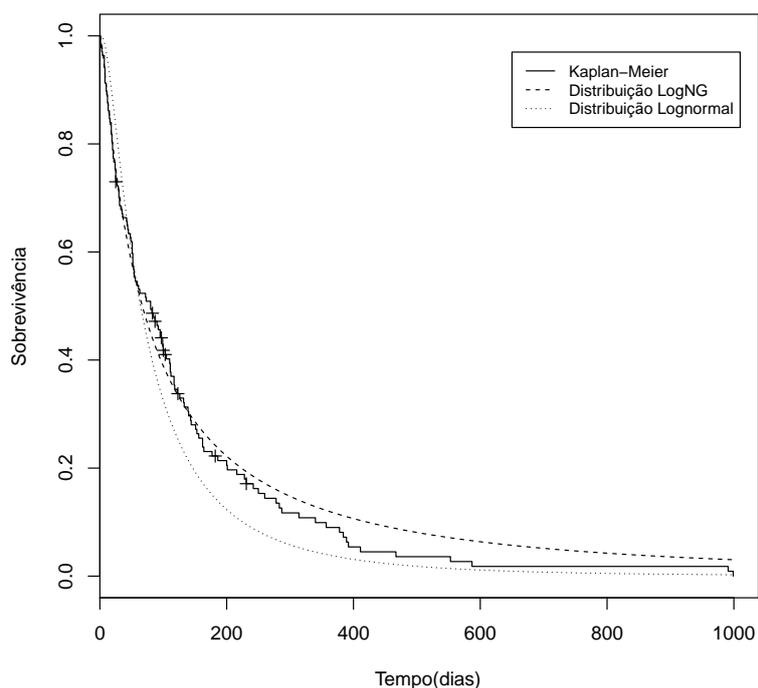


Figura 3.3: Curva Kaplan-Meier com funções de sobrevivência estimadas pelas distribuições logNG e lognormal para os dados de câncer de pulmão inoperável.

Com o intuito de investigar a influência dos dados de tempo de sobrevivência dos pacientes com câncer de pulmão inoperável no ajuste do modelo lognormal generalizado escolhido, calcularam-se as medidas de divergência de K-L e suas respectivas calibrações. A Figura 3.4 apresenta as medidas de K-L para cada observação. Percebe-se que as observações 77, 85 e 95 apresentam os maiores valores de K-L, quando comparado com as outras observações. As divergências de K-L e as calibrações referentes às 3 observações que apresentaram os maiores valores de calibração são apresentadas na Tabela 3.5. Nota-se, pelos valores da calibração, que nenhuma das observações é influente, pois seria necessário obter-se resultados similares aos apresentados na Tabela 3.3.

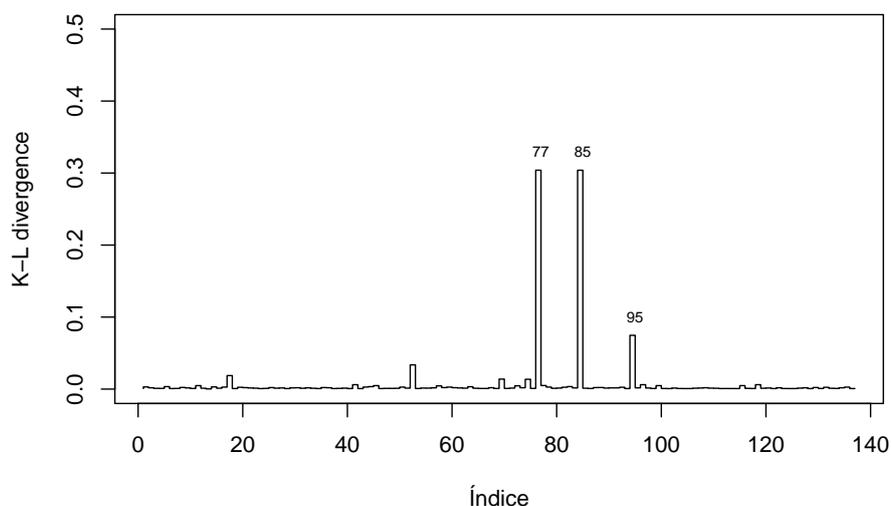


Figura 3.4: Índices de $K(\pi, \pi_{(-i)})$ para os dados de câncer de pulmão inoperável.

Tabela 3.5: Medidas de diagnóstico para os dados de câncer de pulmão inoperável.

Casos	$K(\pi, \pi_{(-i)})$	Calibração
77	0,3040	0,8374
85	0,3040	0,8374
95	0,0747	0,6862

3.4 Representação de mistura de escala uniforme

Nesta seção, a distribuição logNG é representada como uma mistura de escala uniforme para implementar análises *a posteriori* pelo método amostrador de Gibbs e avaliar esta estratégia computacional, considerando a função *a priori* de referência (2.23) do modelo NG.

Uma representação de mistura de escala uniforme para a distribuição logNG em (3.1) é mostrada no Teorema 3.1, utilizando-se a definição de Martín & Pérez (2009).

Teorema 3.1. Se T e U são duas variáveis aleatórias tais que, $f(t|u) = \frac{1}{2t\tau u^{\frac{1}{s}}} I\left(e^{\mu-\tau u^{\frac{1}{s}}} < t < e^{\mu+\tau u^{\frac{1}{s}}}\right)$ e $U \sim Ga\left(1 + \frac{1}{s}, 1\right)$, então $T \sim \text{logNG}(\mu, \tau, s)$.

Prova: Escrevendo a densidade conjunta de Y e U como

$$f(t|u)f(u) \propto u^{-\frac{1}{s}} I\left(u > \left|\frac{\log(t) - \mu}{\tau}\right|^s\right) u^{\frac{1}{s}} \exp\{-u\},$$

onde $I(\cdot)$ é a função indicadora, tem-se que a densidade marginal de T é dada por

$$\begin{aligned} f(t) &\propto \int_0^{\infty} \exp\{-u\} I\left(u > \left|\frac{\log(t) - \mu}{\tau}\right|^s\right) du \\ &\propto \int_{\left|\frac{\log(t) - \mu}{\tau}\right|^s}^{\infty} \exp\{-u\} du \\ &\propto \exp\left\{-\left|\frac{\log(t) - \mu}{\tau}\right|^s\right\}. \end{aligned}$$

Portanto pode-se concluir que $T \sim \text{logNG}(\mu, \tau, s)$, como definido em (3.1).

3.4.1 Distribuição *a posteriori*

Nesta subseção, utiliza-se a representação de mistura dada no Teorema 3.1 para implementar análises *a posteriori* pelo método amostrador de Gibbs.

Considerando-se a representação de mistura dada no Teorema 3.1, tem-se que para uma amostra $\mathbf{t} = \{t_1, t_2, \dots, t_n\}$ e para o vetor de variáveis latentes $\mathbf{u} = \{u_1, u_2, \dots, u_n\}$, a função de verossimilhança é dada por

$$\begin{aligned} L(\mu, \tau, s | \mathbf{u}, \mathbf{t}) &= \prod_{i=1}^n f(t_i | \mu, \tau, s, u_i) f(u_i | s) \\ &= \prod_{i=1}^n \frac{1}{2t_i \tau u_i^{\frac{1}{s}}} I\left(e^{\mu-\tau u_i^{\frac{1}{s}}} < t_i < e^{\mu+\tau u_i^{\frac{1}{s}}}\right) \times \frac{1}{\Gamma\left(1 + \frac{1}{s}\right)} u_i^{\left(1 + \frac{1}{s}\right)-1} e^{-u_i}. \end{aligned} \quad (3.14)$$

Utilizando-se a função de verossimilhança em (3.14) e a função *a priori* de referência conjunta em (2.23), a distribuição *a posteriori* dos parâmetros de interesse e das variáveis latentes é dada por

$$\begin{aligned} \pi^R(\mu, \tau, s, \mathbf{u} | \mathbf{t}) &\propto \frac{\left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) + \left(1 + \Psi \left(1 + \frac{1}{s}\right)\right)^2 - 1\right]^{\frac{1}{2}}}{\tau^{(n+1)} \Gamma^n \left(1 + \frac{1}{s}\right) s^{\frac{3}{2}}} \\ &\times \prod_{i=1}^n \frac{e^{-u_i}}{t_i} I\left(e^{\mu-\tau u_i^{\frac{1}{s}}} < t_i < e^{\mu+\tau u_i^{\frac{1}{s}}}\right). \end{aligned} \quad (3.15)$$

As distribuições condicionais completas, associadas com a distribuição *a posteriori* conjunta em (3.15), são dadas por

$$\pi^R(\mu|\tau, s, \mathbf{u}, \mathbf{t}) \propto 1, \Upsilon(\mu \in (\underline{\mu}, \bar{\mu})), \quad (3.16)$$

onde

$$\underline{\mu} = \max_i \left\{ \log(t_i) - \tau u_i^{\frac{1}{s}} \right\}, \bar{\mu} = \min_i \left\{ \log(t_i) + \tau u_i^{\frac{1}{s}} \right\},$$

$$\pi^R(\tau|\mu, s, \mathbf{u}, \mathbf{t}) \propto \tau^{-(n+1)}, \Upsilon \left(\tau > \max_i \left\{ \frac{|\log(t_i) - \mu|}{u_i^{\frac{1}{s}}} \right\} \right), \quad (3.17)$$

$$\pi^R(s|\mu, \tau, \mathbf{u}, \mathbf{t}) \propto \frac{\left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) + \left(1 + \Psi \left(1 + \frac{1}{s}\right)\right)^2 - 1 \right]^{\frac{1}{2}}}{\Gamma^n \left(1 + \frac{1}{s}\right) s^{\frac{3}{2}}}, \Upsilon(s \in (\underline{s}, \bar{s})), \quad (3.18)$$

onde

$$\underline{s} = \max_{i \in s^-} \left\{ 1, \frac{\log(u_i)}{\log \left(\frac{|\log(t_i) - \mu|}{\tau} \right)} \right\}, s^- = \left\{ i : \log \left(\frac{|\log(t_i) - \mu|}{\tau} \right) < 0 \right\},$$

$$\bar{s} = \min_{i \in s^+} \left\{ \frac{\log(u_i)}{\log \left(\frac{|\log(t_i) - \mu|}{\tau} \right)} \right\}, s^+ = \left\{ i : \log \left(\frac{|\log(t_i) - \mu|}{\tau} \right) > 0 \right\}.$$

$$\pi^R(u_i|\mu, \tau, s, \mathbf{t}) \propto e^{-u_i}, \Upsilon \left(u_i > \left(\frac{|\log(t_i) - \mu|}{\tau} \right)^s \right), \quad i = 1, \dots, n. \quad (3.19)$$

As distribuições condicionais completas em (3.16), (3.17) e (3.19) são Uniforme, Pareto tipo I e Exponencial, respectivamente. A distribuição condicional completa em (3.18) não tem um núcleo de uma distribuição conhecida, mas pode ser facilmente gerada através do método de aceitação-rejeição.

O algoritmo consiste na escolha de valores iniciais de $\mu^{(0)}, \tau^{(0)}$ e $s^{(0)}$, e da amostragem iterativa de $u^{(j)}, \mu^{(j)}, \tau^{(j)}$ e $s^{(j)}$ das distribuições condicionais completas de (3.19), (3.16), (3.17) e (3.18), respectivamente.

3.4.2 Aplicação

Com a finalidade de testar a metodologia proposta, considera-se o conjunto de dados retirado de [Kalbfleisch & Prentice \(2002\)](#), descritos na Seção 3.4. Os dados referem-se ao tempo de sobrevivência de 137 pacientes com câncer de pulmão inoperável e apresentam 9 observações censuradas.

O modelo lognormal generalizado foi ajustado aos dados pelo procedimento de referência Bayesiana. As amostras da distribuição *a posteriori* foram geradas pelo método amostrador de Gibbs. Uma única cadeia de dimensão 200.000 foi considerada para cada parâmetro, descartando as primeiras 100.000 iterações (*burn in*) para eliminar o efeito dos primeiros valores. A convergência das cadeias foi verificada pelo critério de Geweke (1992) (ver Tabela C.5 no Apêndice C). No entanto, as cadeias resultantes são altamente correlacionadas (Ver Tabela C.6 e Figura C.7 no Apêndice C). De maneira similar a Seção 2.5.2, todas as iterações após o *burn in* têm sido utilizadas. O tempo de computação requerido por esta aplicação foi de 1,13 horas, em um computador Core i7 – 2.90GHz com 6.0GB de memória RAM.

A Tabela 3.6 mostra os resumos *a posteriori* para os parâmetros da distribuição logNG. Como esperado, considerando a transformação $\sigma = \tau\Gamma\left(1 + \frac{1}{s}\right)$, tem-se que os valores das médias na Tabela 3.6 estão bem próximas das estimativas obtidas com o algoritmo Metropolis-Hastings dentro do amostrador de Gibbs, apresentadas na Tabela 3.4. Através desta tabela pode-se observar também que o parâmetro de forma ($s < 2$) fornece fortes evidências de uma distribuição leptocúrtica para os dados de sobrevivência dos pacientes com câncer de pulmão inoperável, o que favorece o modelo logNG.

Tabela 3.6: Resumos *a posteriori* para os parâmetros do modelo logNG.

Modelo	Parâmetros	Média	Desvio Padrão	HPD (95%)
logNG	μ	4,172	0,130	(3,915; 4,427)
	τ	1,918	0,220	(1,456; 2,360)
	s	1,986	0,348	(1,229; 2,665)

3.5 Considerações finais

Neste capítulo, apresentou-se a distribuição lognormal generalizada do ponto de vista da análise Bayesiana objetiva na presença de dados de sobrevivência. Estudos de simulação foram feitos para verificar a adequação do método inferencial proposto, considerando a presença de dados completos e de dados censurados, para diferentes tamanhos amostrais. Os resultados simulados para ambos os dados mostram melhores

propriedades frequentistas para tamanhos de amostra acima de 100. Além disso, verificou-se, por meio do estudo de simulação e dos dados reais, que a divergência de Kullback-Leibler detecta de maneira eficaz os pontos influentes no ajuste do modelo lognormal generalizado. A distribuição lognormal generalizada também foi representada como uma mistura de escala uniforme. Estudos com dados reais mostraram, como esperado, que as estimativas do modelo proposto, obtidas pelo método amostrador de Gibbs, ficaram bem próximas das estimativas do modelo lognormal generalizado com reparametrização, obtidas pelo algoritmo Metropolis-Hastings dentro do amostrador de Gibbs. No entanto, no estudo com dados reais, não foi observado um ganho computacional com a inclusão da variável latente. A inclusão desta variável aumentou as correlações nas amostras estimadas, de modo que foi necessário aumentar o número de iterações da cadeia de MCMC para atingir convergência. Consequentemente, o tempo de processamento desse algoritmo foi consideravelmente maior do que a do algoritmo Metropolis-Hastings. No capítulo 4, apresenta-se o modelo de regressão normal generalizado do ponto de vista da análise Bayesiana objetiva.

Capítulo 4

Modelo de regressão normal generalizado

Os modelos de regressão lineares são aplicados em diversas áreas de estudo para investigar e modelar a relação entre variáveis, sendo comum a suposição de que os dados provêm de uma distribuição normal. Entretanto é reconhecido que o modelo de regressão linear normal pode ser influenciado por observações aberrantes (Belsley *et al.*, 1980; Barnett & Lewis, 1994). Nestas circunstâncias, é necessária a utilização de modelos em que a distribuição do erro apresente caudas mais flexíveis em relação à distribuição normal, como a distribuição Normal Generalizada (NG).

Neste trabalho, diferentemente de Salazar *et al.* (2012), considera-se a parametrização da distribuição NG estudada por Nadarajah (2005) e utiliza-se o método de análise de referência para fazer inferências sobre os parâmetros do modelo. Além disso, técnicas de diagnósticos baseadas na divergência de Kullback-Leibler são utilizadas como medida de influência local.

O modelo de regressão linear é definido como

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \dots, n, \quad (4.1)$$

onde Y_i é a variável resposta para o i -ésimo caso, $\mathbf{x}_i^\top = (x_{i1}, \dots, x_{ip})$ são os valores de p variáveis preditoras, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ é o vetor com coeficientes de regressão e ϵ_i são variáveis aleatórias independentes, que seguem uma distribuição NG com média zero, parâmetro de escala σ e parâmetro de forma s .

Para inferir sobre os parâmetros do modelo em (4.1), adota-se uma abordagem Bayesiana objetiva. A função de verossimilhança e as funções *a priori* e *a posteriori* para os parâmetros do modelo são apresentadas a seguir.

4.1 Função de verossimilhança

A função de verossimilhança para uma amostra aleatória $\mathbf{y} = (y_1, \dots, y_n)^\top$ do modelo (4.1) é dada por

$$L(\boldsymbol{\beta}, \sigma, s | \mathbf{y}) = 2^{-n} \sigma^{-n} \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^\top \boldsymbol{\beta}|}{\sigma} \right)^s \right\}. \quad (4.2)$$

A função de log-verossimilhança em (4.2) é dada por

$$\log L(\boldsymbol{\beta}, \sigma, s | \mathbf{y}) = -n \log 2 - n \log \sigma - \sum_{i=1}^n \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^\top \boldsymbol{\beta}|}{\sigma} \right)^s. \quad (4.3)$$

A derivada de primeira ordem da função de log-verossimilhança em (4.3) é dada por

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \frac{s \Gamma(1 + \frac{1}{s})}{\sigma} \sum_{i=1}^n \mathbf{x}_i^\top \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^\top \boldsymbol{\beta}|}{\sigma} \right)^{s-1} \text{sign}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) \quad (4.4)$$

$$\frac{\partial \log L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{s}{\sigma} \sum_{i=1}^n \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^\top \boldsymbol{\beta}|}{\sigma} \right)^s \quad (4.5)$$

$$\frac{\partial \log L}{\partial s} = - \sum_{i=1}^n \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^\top \boldsymbol{\beta}|}{\sigma} \right)^s \left[\log \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^\top \boldsymbol{\beta}|}{\sigma} \right) - \frac{\Psi(1 + \frac{1}{s})}{s} \right] \quad (4.6)$$

em que $s > 1$ e $\Psi(s) = \frac{\Gamma'(s)}{\Gamma(s)}$ é a função digamma.

Considerando a Definição 1.1, tem-se que os elementos da matriz de informação de Fisher para $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma, s)$ no modelo em (4.1) são dados por

$$\begin{aligned}
I_{11}(\boldsymbol{\theta}) &= E \left[\left(\frac{\partial \log L}{\partial \boldsymbol{\beta}} \right) \left(\frac{\partial \log L}{\partial \boldsymbol{\beta}^\top} \right) \right] = \frac{\Gamma\left(\frac{1}{s}\right) \Gamma\left(2 - \frac{1}{s}\right)}{\sigma^2} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top, \\
I_{12}(\boldsymbol{\theta}) &= E \left[\left(\frac{\partial \log L}{\partial \boldsymbol{\beta}} \right) \left(\frac{\partial \log L}{\partial \sigma} \right) \right] = \mathbf{0}; \\
I_{13}(\boldsymbol{\theta}) &= E \left[\left(\frac{\partial \log L}{\partial \boldsymbol{\beta}} \right) \left(\frac{\partial \log L}{\partial s} \right) \right] = \mathbf{0}, \\
I_{22}(\boldsymbol{\theta}) &= E \left[\left(\frac{\partial \log L}{\partial \sigma} \right)^2 \right] = \frac{ns}{\sigma^2}, \\
I_{23}(\boldsymbol{\theta}) &= E \left[\left(\frac{\partial \log L}{\partial \sigma} \right) \left(\frac{\partial \log L}{\partial s} \right) \right] = -\frac{n}{\sigma s}, \\
I_{33}(\boldsymbol{\theta}) &= E \left[\left(\frac{\partial \log L}{\partial s} \right)^2 \right] = \frac{n}{s^3} \left\{ \left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) \right\},
\end{aligned}$$

onde $\Psi'(s) = \frac{\partial \Psi(s)}{\partial s}$ é a função trigamma. A prova destes resultados é dada no Apêndice A.1.

A matriz de informação de Fisher e a sua inversa são dadas por

$$I(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\Gamma\left(\frac{1}{s}\right) \Gamma\left(2 - \frac{1}{s}\right)}{\sigma^2} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{ns}{\sigma^2} & -\frac{n}{\sigma s} \\ \mathbf{0} & -\frac{n}{\sigma s} & \frac{n}{s^3} \left\{ \left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) \right\} \end{bmatrix} \quad (4.7)$$

e

$$B(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\sigma^2}{\Gamma\left(\frac{1}{s}\right) \Gamma\left(2 - \frac{1}{s}\right) \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\sigma^2}{ns \left[1 - \frac{s}{(1+s) \Psi' \left(1 + \frac{1}{s}\right)} \right]} & \frac{\sigma s^2}{n \left[-s + (1+s) \Psi' \left(1 + \frac{1}{s}\right) \right]} \\ \mathbf{0} & \frac{\sigma s^2}{n \left[-s + (1+s) \Psi' \left(1 + \frac{1}{s}\right) \right]} & \frac{s^4}{n \left[-s + (1+s) \Psi' \left(1 + \frac{1}{s}\right) \right]} \end{bmatrix}. \quad (4.8)$$

A matriz em (4.7) coincide com a matriz de informação de Fisher encontrada por [Salazar et al. \(2012\)](#).

4.2 Distribuições *a priori* e *a posteriori*

Nesta seção, aplica-se o método de análise de referência e o método de Jeffreys, descritos nas Subseções 1.1.1 e 1.1.2, respectivamente, para obter funções *a priori* não-informativas para os parâmetros do modelo de regressão NG em (4.1).

Função *a priori* de referência

A construção da função *a priori* de referência é feita com o uso de uma parametrização ordenada, segundo sua importância inferencial (Berger & Bernardo, 1992). Neste trabalho, a função *a priori* de referência é invariante em relação à ordenação dos parâmetros de perturbação (Bernardo, 1997).

O vetor de parâmetros (β, σ, s) é dividido e ordenado em 3 grupos distintos. Considera-se o caso em que β é o parâmetro de interesse e os demais (σ e s) são parâmetros perturbadores. Para se obter a função *a priori* de referência conjunta dos parâmetros β, σ e s , foi adotada a seguinte parametrização ordenada:

$$\pi^R(\beta, \sigma, s) = \pi^R(s|\beta, \sigma)\pi^R(\sigma|\beta)\pi^R(\beta).$$

Considerando-se a metodologia descrita na Subseção 1.1.1. Seja $H_3(\theta)$ igual à matriz de informação de Fisher dada em (4.7). Pelo Corolário 1.1, tem-se que $h_{33}(\theta)$ pode ser fatorado da forma $h_{33}^{\frac{1}{2}}(\theta) = \sqrt{\frac{n}{s^3} \left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right)} = f_2(s)g_2(\beta, \sigma)$.

Portanto

$$\pi^R(s|\beta, \sigma) \propto s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) \right]^{-\frac{1}{2}}.$$

Seja $B_2(\theta)$ a submatriz superior 2×2 de $B(\theta)$ e $H_2(\theta) = B_2^{-1}(\theta)$. O objetivo é encontrar $h_{22}(\theta)$, o elemento (2, 2) da matriz $H_2(\theta)$, e mostrar que esse se fatora como $h_{22}^{\frac{1}{2}}(\theta) = f_1(\sigma)g_1(\beta, s)$.

Tem-se que

$$B_2(\theta) = \begin{bmatrix} \frac{\sigma^2}{\Gamma\left(\frac{1}{s}\right)\Gamma\left(2-\frac{1}{s}\right)\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top} & \mathbf{0} \\ \mathbf{0} & \frac{\sigma^2}{ns \left[1 - \frac{s}{(1+s)\Psi'\left(1+\frac{1}{s}\right)}\right]} \end{bmatrix}.$$

e

$$H_2(\theta) = \begin{bmatrix} \frac{\Gamma\left(\frac{1}{s}\right)\Gamma\left(2-\frac{1}{s}\right)\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top}{\sigma^2} & \mathbf{0} \\ \mathbf{0} & \frac{ns \left[1 - \frac{s}{(1+s)\Psi'\left(1+\frac{1}{s}\right)}\right]}{\sigma^2} \end{bmatrix}.$$

Portanto $h_{22}(\theta) = \frac{1}{\sigma^2} ns \left[1 - \frac{s}{(1+s)\Psi'\left(1+\frac{1}{s}\right)}\right]$ pode ser fatorado da forma

$$h_{22}^{\frac{1}{2}}(\boldsymbol{\theta}) = \frac{1}{\sigma} \sqrt{ns \left[1 - \frac{s}{(1+s)\Psi'(1+\frac{1}{s})} \right]} = f_1(\sigma)g_1(\boldsymbol{\beta}, s).$$

Então,

$$\pi^R(\sigma|\boldsymbol{\beta}) \propto \frac{1}{\sigma}.$$

Finalmente, seja $B_{11}(\boldsymbol{\theta})$ o elemento (1, 1) da matriz $B(\boldsymbol{\theta})$ e $h_{11}(\boldsymbol{\theta}) = B_{11}^{-1}(\boldsymbol{\theta})$. O objetivo é mostrar que $h_{11}(\boldsymbol{\theta})$ se fatora como $h_{11}^{\frac{1}{2}}(\boldsymbol{\theta}) = f_0(\boldsymbol{\beta})g_0(\sigma, s)$.

Invertendo $B_{11}(\boldsymbol{\theta})$, tem-se que $h_{11}(\boldsymbol{\theta}) = \frac{\Gamma(\frac{1}{s})\Gamma(2-\frac{1}{s})\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top}{\sigma^2}$ pode ser fatorado da forma $h_{11}^{\frac{1}{2}}(\boldsymbol{\theta}) = \sqrt{\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top} \sqrt{\frac{\Gamma(\frac{1}{s})\Gamma(2-\frac{1}{s})}{\sigma^2}} = f_0(\boldsymbol{\beta})g_0(\sigma, s)$.

Portanto

$$\pi^R(\boldsymbol{\beta}) \propto 1.$$

A função *a priori* de referência conjunta relativa à parametrização ordenada é dada por

$$\pi^R(\boldsymbol{\beta}, \sigma, s) \propto \frac{1}{\sigma} \times s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s} \right) \Psi' \left(1 + \frac{1}{s} \right) \right]^{\frac{1}{2}} \quad (4.9)$$

em que $\boldsymbol{\beta} \in \mathbb{R}^p$, $\sigma \in \mathbb{R}^+$ e $s > 1$. Como esperado, a função *a priori* em (4.9) coincide com a função *a priori* de referência conjunta do modelo NG em (2.10).

Função *a priori* de Jeffreys

A função *a priori* não-informativa de Jeffreys, obtida da raiz quadrada do determinante da matriz de informação de Fisher em (4.7),

$$\pi^J(\boldsymbol{\beta}, \sigma, s) \propto \sqrt{\det(I(\boldsymbol{\theta}))} = \sqrt{\det(I_{11})[I_{22} I_{33} - I_{23}^2]},$$

onde $\det(I_{11}) = \left[\frac{\Gamma(\frac{1}{s})\Gamma(2-\frac{1}{s})}{\sigma^2} \right]^p \det(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top)$ e $[I_{22} I_{33} - I_{23}^2] = \frac{n^2}{\sigma^2 s^2} \left[\left(1 + \frac{1}{s} \right) \Psi' \left(1 + \frac{1}{s} \right) - 1 \right]$, é dada por

$$\pi^J(\boldsymbol{\beta}, \sigma, s) \propto \sigma^{-(p+1)} \left[\Gamma \left(\frac{1}{s} \right) \Gamma \left(2 - \frac{1}{s} \right) \right]^{\frac{p}{2}} s^{-1} \left[\left(1 + \frac{1}{s} \right) \Psi' \left(1 + \frac{1}{s} \right) - 1 \right]^{\frac{1}{2}}, \quad (4.10)$$

onde p é o número de regressores.

No entanto, a função *a priori* em (4.10) conduz a uma distribuição *a posteriori* imprópria.

A seguir apresentam-se definições e Corolários, de maneira similar à Seção 2.2, para provar que as funções *a priori* em (4.9) e em (4.10) produzem distribuições *a posteriori* própria e imprópria, respectivamente.

As funções *priori* $\pi^R(\boldsymbol{\beta}, \sigma, s)$ e $\pi^J(\boldsymbol{\beta}, \sigma, s)$ pertencem à classe de distribuições *a priori* impróprias, dada por

$$\pi(\boldsymbol{\beta}, \sigma, s) \propto \frac{\pi(s)}{\sigma^a}, \quad a \in \mathfrak{R}, \quad (4.11)$$

onde a é um hiperparâmetro e $\pi(s)$ é a função *a priori* marginal do parâmetro de forma, para várias escolhas de $\pi(s)$ e a .

A função *a priori* de referência e a função *a priori* de Jeffreys são da forma (4.11) com, respectivamente,

$$a = 1 \text{ e } \pi^R(s) \propto s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) \right]^{\frac{1}{2}}, \quad (4.12)$$

$$a = p + 1 \text{ e } \pi^J(s) \propto s^{-1} \left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) - 1 \right]^{\frac{1}{2}} \left[\Gamma \left(\frac{1}{s}\right) \Gamma \left(2 - \frac{1}{s}\right) \right]^{\frac{p}{2}}. \quad (4.13)$$

A distribuição *a posteriori* associada com a função *priori* em (4.11) é própria se

$$\int_1^\infty L(s|\mathbf{y})\pi(s)ds < \infty, \quad (4.14)$$

onde $L(s|\mathbf{y})$ é a verossimilhança integrada de s , dada por

$$\int_{\mathfrak{R}^p} \int_0^\infty L(\boldsymbol{\beta}, \sigma, s|\mathbf{y}) \sigma^{-a} d\sigma d\boldsymbol{\beta}. \quad (4.15)$$

Proposição 4.1. *A função a priori marginal para s dada nas equações (4.12) e (4.13) é uma função contínua em $[1, \infty)$ e quando $s \rightarrow \infty$ tem-se que $\pi^R(s) = O(s^{-3/2})$ e $\pi^J(s) = O(s^{\frac{p-2}{2}})$.*

Prova: A inspeção direta das funções (4.12) e (4.13) mostra que estas funções são contínuas em $[1, \infty)$. Quando $s \rightarrow \infty$ tem-se que $\Gamma \left(\frac{1}{s}\right) = O(s)$ e $\Psi' \left(1 + \frac{1}{s}\right) \rightarrow 1$, 6449.

Proposição 4.2. *A função de verossimilhança integrada de s para $n > p + 1 - a$, sob a classe de priori (4.11) é uma função contínua em $[1, \infty)$ e $L(s|\mathbf{y}) = O(1)$, quando $s \rightarrow \infty$.*

Prova: Ver Apêndice A.2.

Proposição 4.3. *A função a priori de referência dada em (4.9) conduz a uma distribuição a posteriori própria e a função a priori de Jeffreys dada em (4.10) conduz a uma distribuição a posteriori imprópria.*

Prova: Ver Apêndice A.2.

Distribuição a posteriori

Neste trabalho, a função a priori de referência definida em (4.9), será adotada para obter as inferências a posteriori do modelo proposto em (4.1).

Portanto, utilizando-se a função de verossimilhança em (4.2) e a função a priori de referência conjunta em (4.9), obtém-se a distribuição a posteriori conjunta para β , σ e s ,

$$\pi^R(\beta, \sigma, s | \mathbf{y}) \propto \sigma^{-(n+1)} s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) \right]^{\frac{1}{2}} \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma \left(1 + \frac{1}{s}\right) |y_i - \mathbf{x}_i^\top \beta|}{\sigma} \right)^s \right\} \quad (4.16)$$

As distribuições condicionais completas, associadas com a distribuição a posteriori conjunta em (4.16), são dadas por:

$$\pi^R(\beta | \sigma, s, \mathbf{y}) \propto \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma \left(1 + \frac{1}{s}\right) |y_i - \mathbf{x}_i^\top \beta|}{\sigma} \right)^s \right\}, \quad (4.17)$$

$$\pi^R(\sigma | \beta, s, \mathbf{y}) \propto \sigma^{-(n+1)} \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma \left(1 + \frac{1}{s}\right) |y_i - \mathbf{x}_i^\top \beta|}{\sigma} \right)^s \right\}, \quad (4.18)$$

$$\pi^R(s | \beta, \sigma, \mathbf{y}) \propto s^{-\frac{3}{2}} \left[\left(1 + \frac{1}{s}\right) \Psi' \left(1 + \frac{1}{s}\right) \right]^{\frac{1}{2}} \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma \left(1 + \frac{1}{s}\right) |y_i - \mathbf{x}_i^\top \beta|}{\sigma} \right)^s \right\}. \quad (4.19)$$

As distribuições condicionais completas em (4.17) e em (4.19) não têm um núcleo de uma distribuição conhecida e a distribuição em (4.18) pode ser reduzida facilmente a uma distribuição gama inversa pela transformação $\lambda = \sigma^s$. As estimativas dos parâmetros de interesse são obtidas pelo algoritmo Metropolis-Hastings dentro do amostrador de Gibbs, de maneira similar ao esquema apresentado no Apêndice B

4.3 Aplicações

Nesta seção, a metodologia desenvolvida é aplicada em dados artificiais e em dados reais.

4.3.1 Dados artificiais

Foi gerada uma amostra artificial de tamanho $n = 500$ de acordo com (4.1), com $p = 2$, $\mathbf{x}_i^\top = (1, x_{i1})$, $x_{i1} \sim N(2.5, 1)$, $\boldsymbol{\beta} = (2, -1.5)^\top$, $\sigma = 1$ e $s = 2.5$. As amostras *a posteriori* foram geradas pelo algoritmo Metropolis-Hastings dentro do amostrador de Gibbs, através do método MCMC implementado no software R. Uma única cadeia de dimensão 300.000 foi considerada para cada parâmetro, descartando-se as primeiras 150.000 iterações (*burn in*) para eliminar o efeito dos primeiros valores. Para evitar a correlação nas séries, considerou-se um espaçamento de tamanho 15, obtendo-se uma amostra final de tamanho 10.000 da distribuição *a posteriori*. Para controlar a taxa de aceitação (TA) do algoritmo, considerou-se a constante de escala $k = 0,55$. A convergência das cadeias foi verificada pelo critério de Geweke (1992) (ver Tabela C.7 no Apêndice C). As cadeias resultantes são apresentadas na Figura C.8 (ver Apêndice C).

A Tabela 4.1 mostra os resumos *a posteriori* para os parâmetros do modelo de regressão normal generalizado e as taxas de aceitação (TA) no algoritmo Metropolis-Hastings. Percebe-se que as estimativas estão próximas aos valores verdadeiros e os intervalos de credibilidade de máxima densidade de probabilidade (HPD) a 95% cobrem os verdadeiros valores dos parâmetros.

Tabela 4.1: Dados artificiais. Resumos *a posteriori* para os parâmetros do modelo de regressão NG e as taxas de aceitação (TA) no algoritmo Metropolis-Hastings.

Parâmetros	Média	Mediana	Desvio Padrão	HPD (95%)	TA (%)
β_1	1,995	1,996	0,086	(1,826; 2,164)	42,31
β_2	-1,510	-1,510	0,032	(-1,572; -1,447)	42,16
σ	1,027	1,028	0,055	(0,922; 1,135)	43,03
s	2,657	2,631	0,320	(2,042; 3,293)	43,83

Para investigar a medida de divergência de K-L na detecção de observações influentes no ajuste de um modelo de regressão normal generalizado, considerou-se a mesma amostra previamente simulada. Foram selecionados os casos 50 e 250 para perturbação. Escolheu-se um ou dois desses casos selecionados e perturbou-se a variável resposta da seguinte maneira: $\tilde{y}_i = y_i + 5S_y$, em que S_y são os desvios padrão de y_i . As estimativas MCMC foram obtidas de uma maneira similar à informação acima.

Para mostrar o impacto das observações influentes nas estimativas de β_1, β_2, σ e s , calculou-se a medidade de variação relativa (VR), que é obtida por (2.21), em que $\hat{\theta}$ e $\hat{\theta}_0$ são as médias *a posteriori* dos parâmetros do modelo, considerando-se os dados originais e os dados perturbados, respectivamente.

A Tabela 4.2 mostra as estimativas *a posteriori* para os dados simulados, as variações relativas (VR) das estimativas dos parâmetros em relação aos dados originais simulados e as taxas de aceitação (TA) no algoritmo Metropolis-Hastings. O conjunto de dados (a) denota os dados originais simulados sem perturbação e os conjuntos de dados de (b) a (d) denotam os dados com casos perturbados. Notam-se maiores valores de VR para as estimativas dos parâmetros σ e s , evidenciando a presença de pontos influentes no conjunto de dados, em que s passa a não ser significativo nos modelos com casos perturbados (c) e (d).

Considerando-se as amostras geradas da distribuição *a posteriori* dos parâmetros do modelo de regressão normal generalizado, estimaram-se as medidas de divergência de K-L e suas respectivas calibrações para cada um dos casos considerados (a-d), como descrito na Seção 1.3. Os resultados da Tabela 4.3 mostram que, para os dados sem perturbação (a), os casos selecionados não são influentes, pois apresentam valores pequenos para $K(\pi, \pi_{(-i)})$ e calibração próxima de 0,577. Entretanto quando os dados são perturbados (b - d), os valores de $K(\pi, \pi_{(-i)})$ são maiores e suas respectivas calibrações são próximas ou iguais a 1, indicando que estes dados são influentes.

A Figura 4.1 mostra as medidas de divergência de K-L para o modelo proposto. Nota-se que a medida de K-L tem um bom desempenho para identificar os casos influentes, pois apresenta valores maiores de $K(\pi, \pi_{(-i)})$ para os casos perturbados quando comparado com os outros casos.

Tabela 4.2: Dados artificiais. Média *a posteriori*, VR (%) em relação aos dados originais simulados, HPD a 95% para os parâmetros do modelo de regressão NG e taxas de aceitação (TA) no algoritmo Metropolis-Hastings.

Dados	Casos	Parâmetros	Média	VR (%)	HPD (95%)	TA (%)
a	Nenhum	β_1	1,995	–	(1,826; 2,164)	42,31
		β_2	–1,510	–	(–1,572; –1,447)	42,16
		σ	1,027	–	(0,922; 1,135)	43,03
		s	2,657	–	(2,041; 3,293)	43,83
b	50	β_1	2,026	1,554	(1,841; 2,211)	63,85
		β_2	–1,517	0,464	(–1,585; –1,449)	63,54
		σ	0,970	5,550	(0,866; 1,073)	64,88
		s	2,188	17,651	(1,763; 2,613)	66,29
c	250	β_1	1,966	1,454	(1,775; 2,154)	64,71
		β_2	–1,495	0,993	(–1,564; –1,423)	64,26
		σ	0,916	10,808	(0,816; 1,016)	65,96
		s	1,867	29,733	(1,565; 2,165)	67,51
d	{50, 250}	β_1	2,002	0,351	(1,807; 2,198)	64,18
		β_2	–1,506	0,265	(–1,578; –1,436)	63,76
		σ	0,902	12,171	(0,801; 1,003)	65,52
		s	1,773	33,271	(1,496; 2,052)	67,08

Tabela 4.3: Medidas de diagnóstico para os dados artificiais.

Dados	Casos	$K(\pi, \pi_{(-i)})$	Calibração
a	50	0,0121	0,5774
	250	0,0014	0,5262
b	50	16,1593	1,0000
c	250	19,2236	1,0000
d	50	2,8796	0,9992
	250	18,2292	1,0000

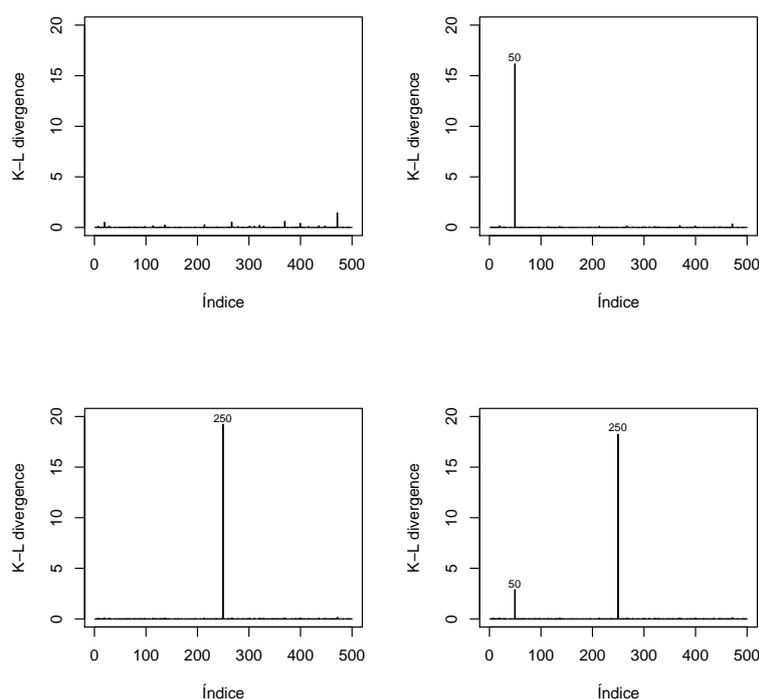


Figura 4.1: Índices de $K(\pi, \pi_{(-i)})$ para os dados artificiais.

4.3.2 Dados reais

Considera-se um conjunto de dados com 1.309 informações sobre altura e diâmetro de clones de *Eucalyptus sp.*, plantados em florestas no Sul do Estado de São Paulo. A variável resposta é altura (m), e a variável explicativa é diâmetro (cm). O objetivo é ajustar o modelo de regressão normal generalizado e o modelo de regressão normal ao conjunto de dados e comparar a eficiência dos respectivos modelos.

Os modelos de regressão normal generalizado e normal foram ajustados aos dados pelo procedimento de referência Bayesiana. As amostras *a posteriori* foram geradas pelo algoritmo Metropolis-Hastings dentro do amostrador de Gibbs, de maneira similar ao estudo de simulação. Uma única cadeia de dimensão 300.000 foi considerada para cada parâmetro, descartando as primeiras 150.000 iterações (*burn in*) para eliminar o efeito dos primeiros valores. Para evitar a correlação nas séries, considerou-se um espaçamento de tamanho 15, obtendo-se uma amostra final de tamanho 10.000 da distribuição *a posteriori*. Para controlar a taxa de aceitação (TA) do algoritmo, considerou-se a constante de escala $k = 0,2$ para ambos os modelos. A convergência das cadeias foi verificada pelo critério de

Geweke (1992) (ver Tabela C.8 no Apêndice C). As cadeias resultantes são apresentadas nas Figuras C.9 e C.10 (ver Apêndice C).

A Tabela 4.4 mostra os resumos *a posteriori* para os parâmetros, os critérios de seleção de modelos e as taxas de aceitação (TA) no algoritmo Metropolis-Hastings. O modelo de regressão normal generalizado é o mais indicado para representar os dados, pois apresenta um desempenho melhor do que o modelo de regressão normal para todos os critérios utilizados. Em particular, a análise do parâmetro de forma ($s > 2$) fornece fortes evidências de uma distribuição platicúrtica para os erros, o que favorece o modelo de regressão normal generalizado. Isso é confirmado por análises gráficas dos quantis resíduais do modelo, apresentado na Figura 4.2 (d).

Tabela 4.4: Dados reais. Resumos *a posteriori* para os parâmetros do modelo de regressão NG, critérios de comparação Bayesiana e taxas de aceitação (TA) no algoritmo Metropolis-Hastings.

Modelo	Parâmetros	Média	Desvio Padrão	HPD (95%)	DIC	EAIC	EBIC	B	TA (%)
NG	β_1	7,066	0,339	(6,398; 7,724)	5829,180	5827,252	5833,606	-2914,595	57,18
	β_2	0,948	0,021	(0,904; 0,989)					56,56
	σ	3,237	0,105	(3,034; 3,440)					60,36
	s	2,865	0,222	(2,436; 3,305)					62,99
Normal	β_1	7,002	0,337	(6,331; 7,671)	6248,652	6247,445	6255,799	-3126,624	60,13
	β_2	0,949	0,020	(0,907; 0,991)					59,53
	σ	1,993	0,037	(1,916; 2,069)					65,23

As Figuras 4.2(a) e 4.3(a) mostram o gráfico de dispersão dos dados e os modelos ajustados de regressão normal generalizado e normal. Observa-se que, em média, as alturas estimadas estão próximas das observadas, indicando que os modelos considerados apresentam um bom ajuste. O gráfico dos resíduos pelos valores ajustados e o gráfico dos resíduos pelas observações também ficaram bastante similares para os dois modelos. Nota-se a presença de heterocedasticidade (vide Figuras 4.2(b) e 4.3(b)) e a tendência quadrática (vide Figuras 4.2(c) e 4.3(c)) para a altura em relação ao diâmetro de *Eucalyptus sp.*. Os gráficos dos quantis da distribuição NG e da distribuição normal para os resíduos dos modelos são apresentados pelas Figuras 4.2(d) e 4.3(d), respectivamente. Percebe-se que, no ajuste do modelo de regressão normal, muitos pontos ficam distantes de suas caudas e fora da faixa de confiança, indicando que a especificação da distribuição do erro para o modelo parece ser inadequada. Portanto há evidência de que o modelo

escolhido supera o modelo de regressão normal no ajuste dos dados.

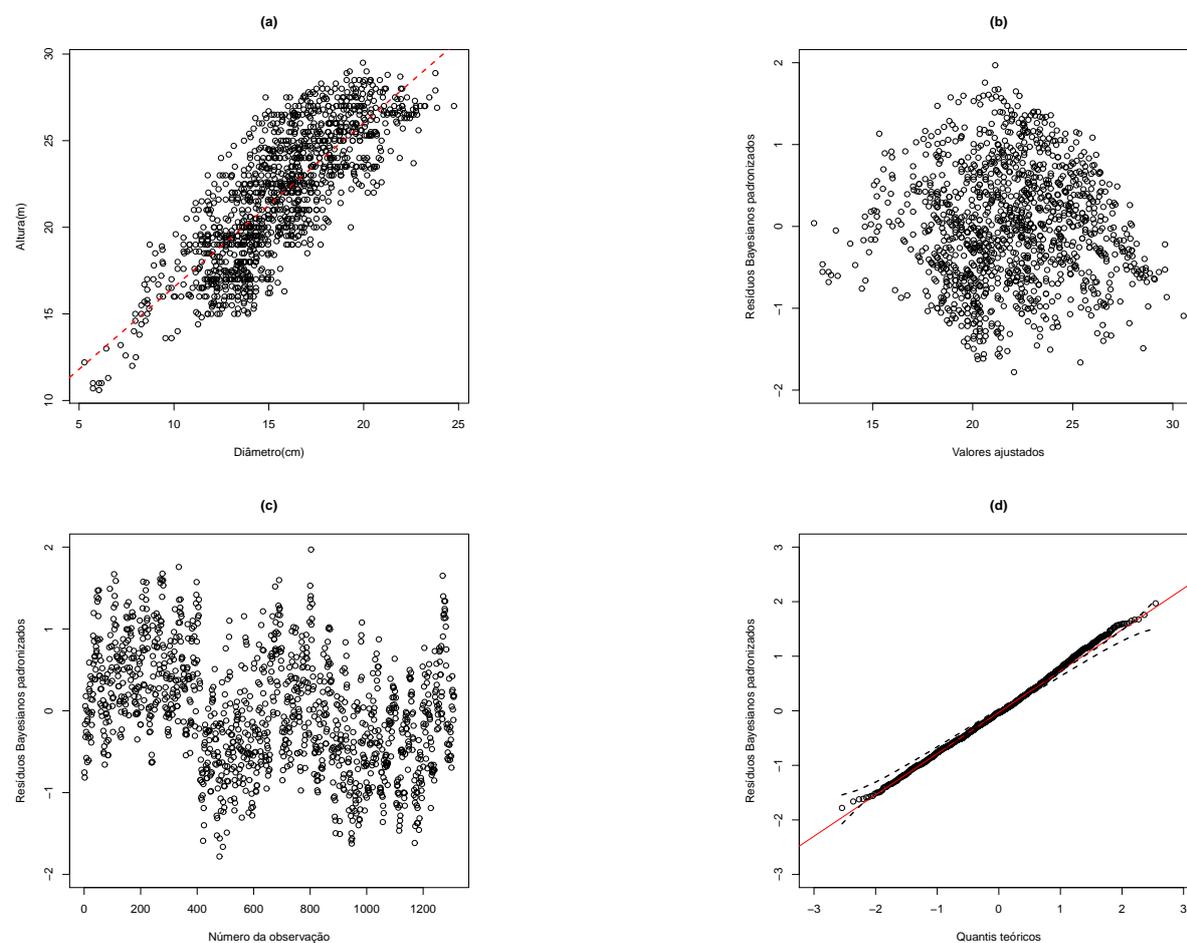


Figura 4.2: Dados reais. Modelo de regressão normal generalizado com $s = 2,865$. (a) Gráfico de dispersão dos dados e modelo de regressão normal generalizado ajustado. (b) Quantis residuais versus os valores ajustados. (c) Quantis residuais versus o número da observação. (d) Gráfico dos quantis da distribuição NG para os resíduos do modelo.

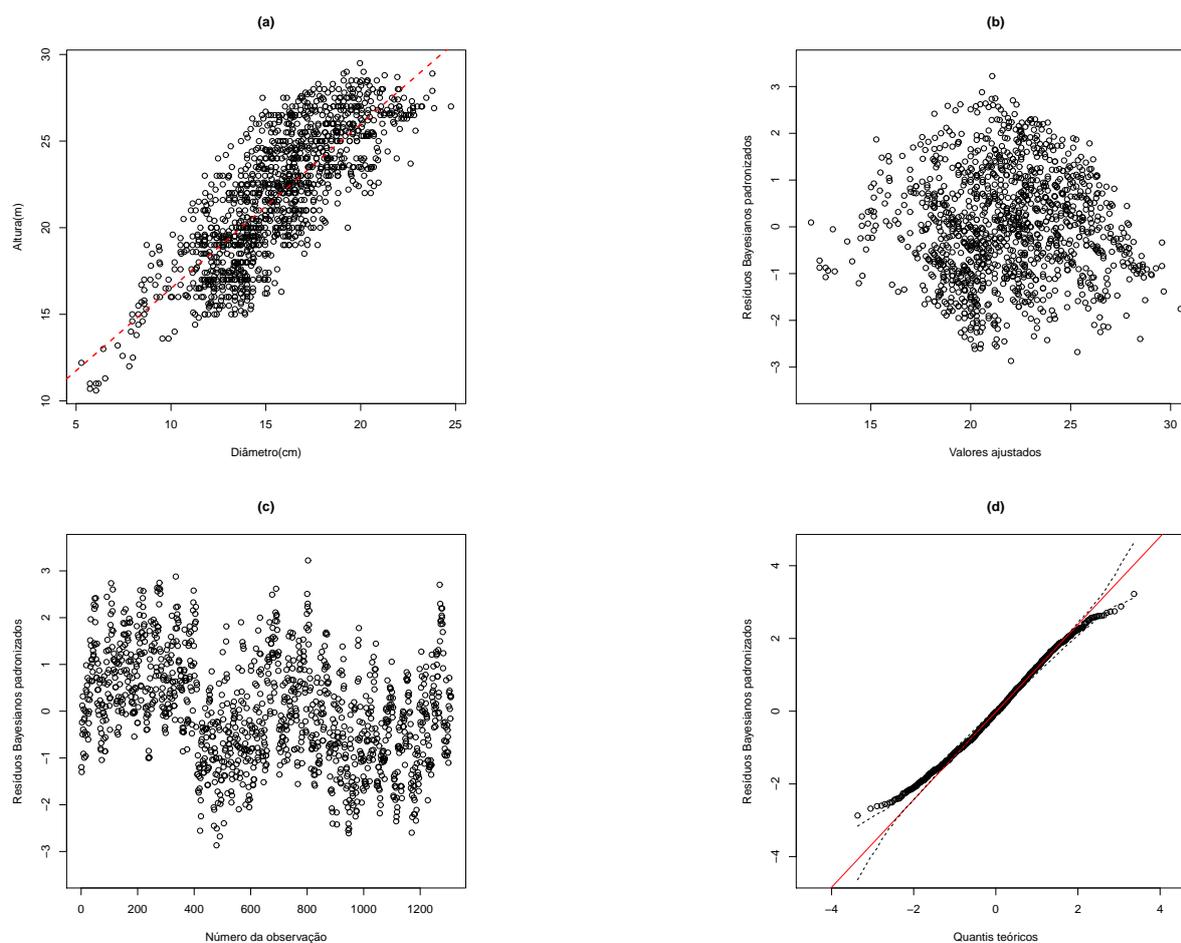


Figura 4.3: Dados reais. Modelo de regressão normal. (a) Gráfico de dispersão dos dados e modelo de regressão ajustado. (b) Quantis residuais versus os valores ajustados. (c) Quantis residuais versus o número da observação. (d) Gráfico dos quantis da distribuição normal para os resíduos do modelo.

Com o intuito de investigar a influência dos dados de altura de *Eucalyptus sp.*, calcularam-se as medidas de divergência de K-L e suas respectivas calibrações. A Figura 4.4 apresenta as medidas de K-L para cada observação. Nota-se que as observações 335, 479 e 803 apresentam os maiores valores de K-L quando comparado com as outras observações. As divergências de K-L e as calibrações referentes às 3 observações que apresentaram os maiores valores de calibração são apresentadas na Tabela 4.5. Percebe-se que a observação 803 é possível caso influente. Para avaliar se esta observação altera as estimativas dos parâmetros do modelo de regressão normal generalizado, realizou-se uma análise de sensibilidade.

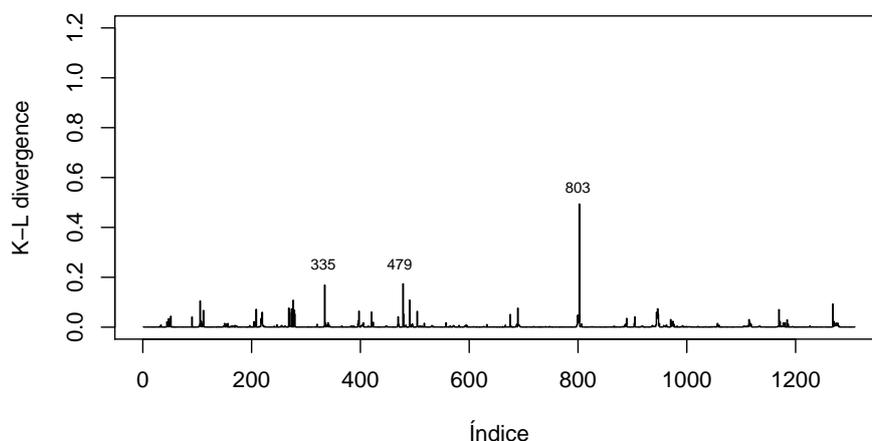


Figura 4.4: Índices de $K(\pi, \pi_{(-i)})$ para os dados de altura de *Eucalyptus sp.*

Tabela 4.5: Medidas de diagnóstico para os dados de altura de *Eucalyptus sp.*

Casos	$K(\pi, \pi_{(-i)})$	Calibração
335	0,1683	0,7634
479	0,1734	0,7656
803	0,4936	0,8960

A Tabela 4.6 mostra as novas estimativas dos parâmetros do modelo após a exclusão do caso influente, com o maior valor de calibração, as variações relativas destas estimativas em relação aos dados originais para os dados de altura de *Eucalyptus sp.* e as taxas de aceitação no algoritmo Metropolis-Hastings. Neste caso, a variação relativa é obtida por (2.21), em que $\hat{\theta}$ e $\hat{\theta}_0$ são as médias *a posteriori* dos parâmetros do modelo, obtidas dos dados originais e dos dados sem a observação influente, respectivamente. Nota-se uma leve mudança na *VR* do parâmetro s quando excluimos a observação influente, no entanto a referida mudança não é significativa, pois a exclusão da observação praticamente não altera as estimativas dos parâmetros, indicando que o modelo de regressão normal generalizado é adequado para acomodar observações influente.

Tabela 4.6: Estimativas *a posteriori*, VR (%) em relação aos dados originais para os dados de altura de *Eucalyptus sp.* e taxas de aceitação (TA) no algoritmo Metropolis-Hastings.

Observação retirada	Parâmetros	Média	VR (%)	HPD (95%)	TA (%)
803	β_1	7,067	0,014	(6,375; 7,753)	57,00
	β_2	0,948	—	(0,905; 0,991)	56,38
	σ	3,252	0,463	(3,054; 3,454)	60,11
	s	2,936	2,478	(2,486; 3,390)	62,67

4.4 Considerações finais

Neste capítulo, apresentou-se uma abordagem de análise de referência Bayesiana objetiva para o modelo de regressão normal generalizado. Técnicas de diagnósticos baseadas na divergência de Kullback-Leibler foram construídas para o modelo de regressão normal generalizado. Estudos com dados artificiais e reais foram feitos para verificar a adequação do método proposto. Além disso, verificou-se, por meio do estudo de dados artificiais e de dados reais, que a divergência de Kullback-Leibler detecta de maneira eficaz os pontos influentes no ajuste do modelo de regressão normal generalizado. A retirada do ponto influente do conjunto de dados reais mostra que o modelo de regressão normal generalizado é adequado para acomodar observações influentes.

Capítulo 5

Considerações finais e propostas futuras

No presente trabalho apresentou-se o modelo NG, o modelo logNG e o modelo de regressão NG sob a perspectiva da análise Bayesiana objetiva. As funções *a priori* de referência e a de Jeffreys para esses modelos foram encontradas, no entanto a função *a priori* de Jeffreys conduziu a uma distribuição *a posteriori* imprópria, em todos os casos estudados. Este resultado é corroborado pela dificuldade do método no caso multiparamétrico. Em contraste, a função *a priori* de referência conduziu a uma distribuição *a posteriori* própria, em todos os modelos considerados.

A análise de referência, no entanto, proporciona um método geral para encontrar funções *a priori* não informativas que maximizam a falta de informações. As funções da *priori* de referência não levam em conta as opiniões de especialistas, mas consideram o modelo assumido e os dados observados. Portanto a função *a priori* de referência pode ser convenientemente utilizada como um padrão de comunicação.

Os resultados dos estudos de simulação e a aplicação em dados reais mostraram que o modelo NG, o modelo logNG e o modelo de regressão NG são adequados para acomodar observações influentes. Estes resultados são corroborados pelo fato das distribuições NG e logNG serem consideradas ferramentas para reduzir a influência das observações aberrantes e obter estimativas robustas. Além disso, verificou-se que a divergência de Kullback-Leibler detectou de maneira eficaz os pontos influentes no

ajuste dos modelos considerados.

As distribuições NG e logNG também foram apresentadas como uma mistura de escala uniforme para permitir o desenvolvimento de algoritmos de amostrador de Gibbs, como uma alternativa ao método Metropolis-Hastings dentro do amostrador de Gibbs. Duas aplicações a dados reais mostraram a metodologia proposta. Além disso, verificou-se que podem existir altas correlações nas amostras estimadas, de modo que são necessárias grandes cadeias de MCMC para atingir convergência. Este resultado é corroborado pelo uso de uma variável latente ([Damien et al., 1999](#)).

Como perspectiva de continuidade deste trabalho, propõe-se alguns tópicos:

- O desenvolvimento da metodologia Bayesiana objetiva para o modelo de regressão lognormal generalizada e outros modelos de regressão, como, por exemplo, o modelo de regressão log-Birbaum-Saunders e o modelo de regressão potência exponencial assimétrico (AEP).
- Utilizar a distribuição lognormal generalizada para os tempos latentes em modelos de sobrevivência com fração de cura.

Referências

- Abramowitz, M. & Stegun, I. A. (1972). *Handbook of Mathematical Functions*. New York: Dover, p.256. [89](#)
- Agrò, G. (1995). Maximum likelihood estimation for the exponential power function parameters. *Communications in Statistics - Simulation and Computation*, **24**(2), 523–536. [1](#)
- Barnett, V. & Lewis, T. (1994). *Outliers in Statistical Data*. New York: John Wiley. [56](#)
- Belsley, D. A., Kuh, E. & Welsch, R. E. (1980). *Regression Diagnostic: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley. [56](#)
- Berger, J. O. & Bernardo, J. M. (1992). On the development of reference priors. In *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, 35-60 (with discussion). [4](#), [5](#), [17](#), [59](#)
- Bernardo, J. M. (1979). Reference Posterior Distributions for Bayesian-Inference. *Journal of the Royal Statistical Society Series B-Methodological*, **41**, 113–147 (with discussion). Reprinted in *Bayesian Inference* (N. G. Polson and G. C. Tiao, eds.), Brookfield, VT: Edward Elgar, (1995), 229–263. [3](#), [4](#)
- Bernardo, J. M. (1997). Noninformative Priors Do Not Exist: A Discussion. *Journal of Statistical Planning and Inference*, **65**, 159–189. [17](#), [59](#)
- Bernardo, J. M. (2005). Reference Analysis. In D. Dey & C. Rao, editors, *In Handbook of Statistics 25*, volume 25, pages 17–90. Elsevier. [4](#)
- Bernardo, J. M. & Ramón, J. M. (1998). An Introduction to Bayesian Reference Analysis: Inference on the Ratio of Multinomial Parameters. *Journal of the Royal Statistical Society. Series D*, **47**, 101–135. [5](#), [6](#), [7](#)

-
- Bernardo, J. M. & Smith, A. F. M. (1994). *Bayesian Theory*. Chichester: Wiley. 8
- Box, G. E. P. & Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley. 1, 3
- Brooks, S. P. (2002). Discussion on the paper by Spiegelhalter, Best, Carlin, and van der Linde (2002). *Journal of the Royal Statistical Society. Series B - Statistical Methodology*, 64, 616–618. 9
- Calitz, F. (1973). Maximum likelihood estimation of the parameters of the three-parameter lognormal distribution - A reconsideration. *Australian Journal of Statistics*, 15, 185–199. 2
- Cancho, V. G., Dey, D. K., Lachos, V. H. & Andrade, M. G. (2011). Bayesian nonlinear regression models with scale mixtures of skew-normal distributions: Estimation and case influence diagnostics. *Computational Statistics & Data Analysis*, 55(1), 588–602. 10
- Carlin, B. P. & Louis, T. A. (2001a). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall/CRC, Boca Raton, second edition. 9, 35
- Carlin, B. P. & Louis, T. A. (2001b). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall. 94
- Chen, G. (1995). Generalized log-normal distributions with reliability application. *Computational Statistics and Data Analysis*, 19, 309–319. 38
- Chen, M. H., Shao, Q. M. & Ibrahim, J. G. (2000). *Monte Carlo methods in Bayesian Computation*. Springer-Verlag, New York. 9, 35
- Chib, S. & Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4), p.327–335. 21
- Cho, H., Ibrahim, J. G., Sinha, D. & Zhu, H. (2009). Bayesian case influence diagnostics for survival models. *Biometrics*, 65(1), 116–124. 10, 12
- Choy, S. & Chan, J. S. (2008). Scale mixtures distributions in statistical modelling. *Australian & New Zealand Journal of Statistics*, 50(2), 135–146. 30

-
- Choy, S. T. B. & Smith (1997a). On Robust Analysis of a Normal Location Parameter. *Journal of the Royal Statistical Society, Series B*, **59**, 463–474. [2](#)
- Choy, S. T. B. & Smith, A. (1997b). Hierarchical models with scale mixtures of normal distributions. *TEST*, **6**, 205–221. [2](#)
- Cohen, A. C. & Whitten, B. J. (1980). Estimation in the three parameter lognormal distribution. *Journal of the American Statistical Association*, **75**, 399–404. [2](#)
- Cohen, A. C. & Whitten, B. J. (1988). *Parameter Estimation in Reliability and Life Span Models*. Marcel Dekker, New York. [2](#), [3](#)
- Cook, R. D. & Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, Boca Raton. [10](#)
- Damien, P., Wakefield, J. & Walker, S. (1999). Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. *Journal of the Royal Statistical Society: Series B*, **61**(2), 331–344. [73](#)
- Gelfand, A. E. & Dey, K. D. (1994). Bayesian Model Choice: Asymptotics and Exact Calculations. *Journal of the Royal Statistician Society Series B*, **56**(3), 501–514. [9](#)
- Gelfand, A. E., Dey, D. K. & Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods (with discussion). In *In Bayesian Statistics (Edited by Kotz and N. L. Johnson)*, volume 4, pages 147–167. Oxford University Press, New York. [9](#)
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics 4 (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.)*. Oxford: University Press, 625-631. [24](#), [35](#), [44](#), [49](#), [54](#), [63](#), [67](#), [94](#)
- Gupta, R. & Lvin, S. (2005). Reliability functions of generalized log-normal model. *Mathematical and Computer Modelling*, **42**, 939 – 946. [40](#)
- Harter, H. L. & Moore, A. H. (1966). Local-maximum-likelihood estimation of the parameters of the three-parameter lognormal populations from complete and censored samples. *Journal of the American Statistical Association*, **61**, 842–851. [2](#), [3](#)

-
- Hill, B. M. (1963). The three-parameter lognormal distribution and Bayesian analysis of a point-source epidemic. *Journal of the American Statistical Association*, **58**, 72–84. [2](#), [3](#), [38](#)
- Ibrahim, J. G., Chen, M.-H. & Sinha, D. (2001). *Bayesian Survival Analysis*. New York: Springer-Verlag. [9](#)
- Jeffreys, H. (1946). An Invariant Form for the Prior Probability in Estimation Problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **186**, 453–461. [3](#), [4](#)
- Kacperczyk, M., Damien, P. & Walker, S. G. (2013). A new class of Bayesian semi-parametric models with applications to option pricing. *Quantitative Finance*, **13**(6), 967–980. [35](#)
- Kalbfleisch, J. D. & Prentice, L. R. (2002). *The Statistical Analysis of Failure Time Data*. Wiley Series in Probability and Statistics. Wiley, second edition. [48](#), [53](#)
- Lee, S. Y., Lu, B. & Song, X. Y. (2006). Assessing local influence for nonlinear structural equation models with ignorable missing data. *Computational Statistics & Data Analysis*, **50**(5), 1356–1377. [26](#)
- Liang, F., Liu, C. & Wang, N. (2007). A robust sequential Bayesian method for identification of differentially expressed genes. *Statistica Sinica*, **17**, 571–595. [1](#)
- MacEachern, S. N. & Berliner, L. M. (1994). Subsampling the Gibbs Sampler. *The American Statistician*, **48**(3), 188–190. [35](#)
- Martín, J. & Pérez, C. J. (2009). Bayesian analysis of a generalized lognormal distribution. *Computational Statistics and Data Analysis*, **53**, 1377–1387. [2](#), [3](#), [32](#), [33](#), [38](#), [39](#), [51](#)
- McCulloch, R. E. (1989). Local model influence. *Journal of the American Statistical Association*, **84**, 473–478. [12](#)
- Migon, H. S. and Gamerman, D. (1999). *Statistical Inference: An Integrated Approach*. Arnold, New York, NY. [8](#)
- Muller, P. (1991). A generic approach to posterior integration and gibbs sampling. Technical report, Department of Statistics, Purdue University. [94](#)

-
- Nadarajah, S. (2005). A generalized normal distribution. *Journal of Applied Statistics*, **32**(7), 685–694. [3](#), [13](#), [38](#), [56](#)
- Naranjo, L., Pérez, C. & Martín, J. (2014). Bayesian analysis of some models that use the asymmetric exponential power distribution. *Statistics and Computing*, pages 1–18. [35](#)
- Qin, Z. (2000). *Uniform Scale Mixture Models with Applications to Bayesian Inference*. Ph.D. thesis, University of Michigan, Michigan, USA. [30](#)
- Qin, Z., Damien, P. & Walker, S. (1998). *Uniform Scale Mixture Models with Applications to Bayesian Inference*. Working paper (Research Support). University of Michigan, Business School. [30](#)
- Salazar, E., Ferreira, M. A. & Migon, H. S. (2012). Objective Bayesian analysis for exponential power regression models. *Sankhyā. Series B*, **74**(1), 107–125. [2](#), [19](#), [56](#), [58](#), [90](#)
- Spiegelhalter, D., Best, N., Carlin, B. & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, **64** (4), 583–639. [9](#)
- Subbotin, M. T. (1923). On the law of frequency of errors. *Mathematicheskii Sbornik*, **31**(2), 296–301. [13](#)
- Upadhyay, S. K. & Peshwani, M. (2001). Full posterior analysis of three parameter lognormal distribution using Gibbs sampler. *Journal of Statistical Computation and Simulation*, **71** (3), 215–230. [3](#), [43](#)
- Upadhyay, S. K., Vasishtha, N. & Smith, A. F. M. (2001). Bayes Inference in life testing and reliability via Markov chain Monte Carlo simulation. *The Indian Journal of Statistics. Series A*, **63** (1), 15–40. [43](#)
- Varanasi, M. k. & Aazhang, B. (1989). Parametric generalized Gaussian density estimation. *The Journal of the Acoustical Society of America*, **86**, 1404–1415. [1](#)
- Vianelli, S. (1982). Sulle curve lognormali di ordine r quali famiglie di distribuzioni di errori di proporzione. *Statistica*, **42**, 155–176. [38](#)

- Vianelli, S. (1983). The Family of Normal and Lognormal Distributions of Order r . *Metron*, **41**, 3–10. [38](#)
- Walker, S. G. & Gutiérrez-Peña, E. (1999). Robustifying Bayesian procedures. In *Bayesian Statistics 6* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, 685-710 (with discussion). [1](#), [2](#), [31](#)
- West, M. (1987). On scale mixtures of normal distributions. *Biometrika*, **79**, 646–648. [2](#)
- Zhu, D. & Zinde-Walsh, V. (2009). Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics*, **148**, 86–99. [14](#)

Apêndice A

Demonstrações

A.1 Matriz de informação de Fisher

Para obter os elementos da matriz de informação de Fisher, os seguintes resultados são necessários:

$$\Gamma(x+1) = x\Gamma(x), \quad \frac{\Gamma'(x+1)}{\Gamma(x)} = 1 + x\psi(x),$$
$$\frac{\Gamma''(x)}{\Gamma(x)} = \psi'(x) + \psi^2(x), \quad \frac{\Gamma''(x+1)}{\Gamma(x)} = 2\psi(x) + x\psi^2(x) + x\psi'(x),$$

onde $\Gamma(x)$ é função gamma e $\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ é a função digamma.

Além disso, as mudanças de variáveis $u = \frac{\Gamma(1 + \frac{1}{s})|y - \mu|}{\sigma}$, $u = \frac{\Gamma(1 + \frac{1}{s})|\log(t) - \mu|}{\sigma}$ e $u = \frac{\Gamma(1 + \frac{1}{s})|y - x^T\beta|}{\sigma}$ são consideradas para o modelo NG, o modelo logNG e o modelo de regressão NG, respectivamente.

A seguir, apresentam-se também 8 resultados necessários para obter os elementos da matriz de informação de Fisher. Os resultados apresentados abaixo são obtidos para a mudança de variável do modelo NG. Vale ressaltar que quando os demais modelos são considerados, com as suas respectivas mudanças de variáveis, os resultados obtidos são equivalentes.

Considera-se a mudança de variável $u = \frac{\Gamma(1 + \frac{1}{s})|y - \mu|}{\sigma}$.

$$\text{I. } E \left[\left(\frac{\Gamma(1 + \frac{1}{s})|y_i - \mu|}{\sigma} \right)^{2(s-1)} \text{sign}^2(y_i - \mu) \right] = \frac{\Gamma\left(2 - \frac{1}{s}\right)}{\Gamma\left(1 + \frac{1}{s}\right)}.$$

Prova:

$$\begin{aligned} E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^{2(s-1)} \text{sign}^2(y_i - \mu) \right] &= \frac{1}{2\Gamma\left(1 + \frac{1}{s}\right)} \int_{-\infty}^{\infty} u^{2(s-1)} e^{-u^s} \text{sign}(u) du \\ &= \frac{1}{\Gamma\left(1 + \frac{1}{s}\right)} \int_0^{\infty} u^{2(s-1)} e^{-u^s} \text{sign}(u) du = \frac{\Gamma\left(2 - \frac{1}{s}\right)}{\Gamma\left(1 + \frac{1}{s}\right) s}. \end{aligned}$$

$$\text{II. } E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^{(s-1)} \text{sign}(y_i - \mu) \right] = 0$$

$$\text{Prova: } E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^{(s-1)} \text{sign}(y_i - \mu) \right] = \frac{1}{2\Gamma\left(1 + \frac{1}{s}\right)} \int_{-\infty}^{\infty} u^{s-1} e^{-u^s} du = 0$$

$$\text{III. } E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^{2s-1} \text{sign}(y_i - \mu) \right] = 0$$

$$\text{Prova: } E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^{2s-1} \text{sign}(y_i - \mu) \right] = \frac{1}{2\Gamma\left(1 + \frac{1}{s}\right)} \int_{-\infty}^{\infty} u^{2s-1} e^{-u^s} du = 0$$

$$\text{IV. } E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^s \right] = \frac{1}{s}$$

Prova:

$$\begin{aligned} E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^s \right] &= \frac{1}{2\Gamma\left(1 + \frac{1}{s}\right)} \int_{-\infty}^{\infty} \frac{u^s}{\text{sign}(u)} e^{-u^s} du \\ &= \frac{1}{\Gamma\left(1 + \frac{1}{s}\right)} \int_0^{\infty} \frac{u^s}{\text{sign}(u)} e^{-u^s} du = \frac{1}{\Gamma\left(1 + \frac{1}{s}\right)} \frac{\Gamma\left(\frac{1}{s}\right)}{s^2} = \frac{1}{s} \end{aligned}$$

$$\text{V. } E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^{2s} \right] = \frac{(1 + \frac{1}{s})}{s}$$

Prova:

$$\begin{aligned} E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^{2s} \right] &= \frac{1}{2\Gamma\left(1 + \frac{1}{s}\right)} \int_{-\infty}^{\infty} \frac{u^{2s}}{\text{sign}(u)} e^{-u^s} du \\ &= \frac{1}{\Gamma\left(1 + \frac{1}{s}\right)} \int_0^{\infty} \frac{u^{2s}}{\text{sign}(u)} e^{-u^s} du = \frac{1}{\Gamma\left(1 + \frac{1}{s}\right)} \frac{\Gamma\left(2 + \frac{1}{s}\right)}{s} = \frac{(1 + \frac{1}{s})}{s} \end{aligned}$$

$$\text{VI. } E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^s \log \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right) \right] = \frac{\psi(1 + \frac{1}{s})}{s^2}$$

Prova:

$$\begin{aligned} E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^s \log \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right) \right] &= \frac{1}{2\Gamma(1 + \frac{1}{s})} \int_{-\infty}^{\infty} \frac{u^s}{\text{sign}(u)} \log(u) e^{-u^s} du \\ &= \frac{1}{\Gamma(1 + \frac{1}{s})} \int_0^{\infty} \frac{u^s}{\text{sign}(u)} \log(u) e^{-u^s} du = \frac{1}{\Gamma(1 + \frac{1}{s})} \frac{\Gamma\left(\frac{1}{s}\right) \psi\left(1 + \frac{1}{s}\right)}{s^3} = \frac{\psi\left(1 + \frac{1}{s}\right)}{s^2} \end{aligned}$$

$$\text{VII. } E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^{2s} \log \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right) \right] = \frac{\Gamma(2 + \frac{1}{s}) \psi(2 + \frac{1}{s})}{\Gamma(1 + \frac{1}{s}) s^2}$$

Prova:

$$\begin{aligned} E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^{2s} \log \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right) \right] &= \frac{1}{2\Gamma(1 + \frac{1}{s})} \int_{-\infty}^{\infty} \frac{u^{2s}}{\text{sign}(u)} \log(u) e^{-u^s} du \\ &= \frac{1}{\Gamma(1 + \frac{1}{s})} \int_0^{\infty} \frac{u^{2s}}{\text{sign}(u)} \log(u) e^{-u^s} du = \frac{1}{\Gamma(1 + \frac{1}{s})} \frac{\Gamma\left(2 + \frac{1}{s}\right) \psi\left(2 + \frac{1}{s}\right)}{s^2} = \frac{\Gamma\left(2 + \frac{1}{s}\right) \psi\left(2 + \frac{1}{s}\right)}{\Gamma\left(1 + \frac{1}{s}\right) s^2} \end{aligned}$$

$$\text{VIII. } E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^{2s} \log^2 \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right) \right] = \frac{(1 + \frac{1}{s}) [\psi^2(2 + \frac{1}{s}) + \psi'(2 + \frac{1}{s})]}{s^3}$$

Prova:

$$\begin{aligned} E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^{2s} \log^2 \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right) \right] &= \frac{1}{2\Gamma(1 + \frac{1}{s})} \int_{-\infty}^{\infty} \frac{u^{2s}}{\text{sign}(u)} \log^2(u) e^{-u^s} du \\ &= \frac{1}{\Gamma(1 + \frac{1}{s})} \int_0^{\infty} \frac{u^{2s}}{\text{sign}(u)} \log^2(u) e^{-u^s} du = \frac{1}{\Gamma(1 + \frac{1}{s})} \frac{\Gamma\left(2 + \frac{1}{s}\right) [\psi^2\left(2 + \frac{1}{s}\right) + \psi'\left(2 + \frac{1}{s}\right)]}{s^3} \end{aligned}$$

Após algumas manipulações algébricas, tem-se que

$$E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^{2s} \log^2 \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right) \right] = \frac{(1 + \frac{1}{s}) [\psi^2(2 + \frac{1}{s}) + \psi'(2 + \frac{1}{s})]}{s^3}.$$

Prova dos elementos da matriz da Seção 2.1:

A seguir, calculam-se os elementos da matriz de informação de Fisher em (2.8).

$$E \left[\left(\frac{\partial \log L}{\partial \mu} \right)^2 \right] = \frac{s^2 \Gamma^2(1 + \frac{1}{s})}{\sigma^2} \sum_{i=1}^n E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^{2(s-1)} \text{sign}^2(y_i - \mu) \right],$$

considerando (I), tem-se que
$$E \left[\left(\frac{\partial \log L}{\partial \mu} \right)^2 \right] = \frac{s^2 \Gamma^2(1 + \frac{1}{s})}{\sigma^2} n \frac{\Gamma\left(2 - \frac{1}{s}\right)}{\Gamma\left(1 + \frac{1}{s}\right)} = \frac{n \Gamma\left(\frac{1}{s}\right) \Gamma\left(2 - \frac{1}{s}\right)}{\sigma^2}.$$

$$\begin{aligned}
E \left[\left(\frac{\partial \log L}{\partial \mu} \right) \left(\frac{\partial \log L}{\partial \sigma} \right) \right] &= E \left[\left(\frac{\partial \log L}{\partial \sigma} \right) \left(\frac{\partial \log L}{\partial \mu} \right) \right] \\
&= -\frac{s\Gamma(1+\frac{1}{s})}{\sigma^2} \sum_{i=1}^n E \left[\left(\frac{\Gamma(1+\frac{1}{s})|y_i-\mu|}{\sigma} \right)^{(s-1)} \text{sign}(y_i-\mu) \right] \\
&\quad + \frac{s^2\Gamma(1+\frac{1}{s})}{\sigma^2} \sum_{i=1}^n E \left[\left(\frac{\Gamma(1+\frac{1}{s})|y_i-\mu|}{\sigma} \right)^{2s-1} \text{sign}(y_i-\mu) \right],
\end{aligned}$$

considerando (II) e (III), tem-se que $E \left[\left(\frac{\partial \log L}{\partial \mu} \right) \left(\frac{\partial \log L}{\partial \sigma} \right) \right] = E \left[\left(\frac{\partial \log L}{\partial \sigma} \right) \left(\frac{\partial \log L}{\partial \mu} \right) \right] = 0$.

$$\begin{aligned}
E \left[\left(\frac{\partial \log L}{\partial \sigma} \right)^2 \right] &= \frac{n}{\sigma^2} - 2\frac{s}{\sigma^2} \sum_{i=1}^n E \left[\left(\frac{\Gamma(1+\frac{1}{s})|y_i-\mu|}{\sigma} \right)^s \right] \\
&\quad + \frac{s^2}{\sigma^2} \sum_{i=1}^n E \left[\left(\frac{\Gamma(1+\frac{1}{s})|y_i-\mu|}{\sigma} \right)^{2s} \right],
\end{aligned}$$

considerando (IV) e (V), tem-se que $E \left[\left(\frac{\partial \log L}{\partial \sigma} \right)^2 \right] = \frac{n}{\sigma^2} - 2\frac{s}{\sigma^2} \frac{n}{s} + \frac{s^2}{\sigma^2} \frac{n(1+\frac{1}{s})}{s} = \frac{ns}{\sigma^2}$.

$$\begin{aligned}
E \left[\left(\frac{\partial \log L}{\partial \sigma} \right) \left(\frac{\partial \log L}{\partial s} \right) \right] &= E \left[\left(\frac{\partial \log L}{\partial s} \right) \left(\frac{\partial \log L}{\partial \sigma} \right) \right] \\
&= \frac{1}{\sigma} \sum_{i=1}^n E \left[\left(\frac{\Gamma(1+\frac{1}{s})|y_i-\mu|}{\sigma} \right)^s \log \left(\frac{\Gamma(1+\frac{1}{s})|y_i-\mu|}{\sigma} \right) \right] \\
&\quad - \frac{\psi(1+\frac{1}{s})}{\sigma s} \sum_{i=1}^n E \left[\left(\frac{\Gamma(1+\frac{1}{s})|y_i-\mu|}{\sigma} \right)^s \right] \\
&\quad - \frac{s}{\sigma} \sum_{i=1}^n E \left[\left(\frac{\Gamma(1+\frac{1}{s})|y_i-\mu|}{\sigma} \right)^{2s} \log \left(\frac{\Gamma(1+\frac{1}{s})|y_i-\mu|}{\sigma} \right) \right] \\
&\quad + \frac{\psi(1+\frac{1}{s})}{\sigma} \sum_{i=1}^n E \left[\left(\frac{\Gamma(1+\frac{1}{s})|y_i-\mu|}{\sigma} \right)^{2s} \right],
\end{aligned}$$

considerando (VI), (IV), (VII) e (V), tem-se que

$$\begin{aligned}
E \left[\left(\frac{\partial \log L}{\partial \sigma} \right) \left(\frac{\partial \log L}{\partial s} \right) \right] &= E \left[\left(\frac{\partial \log L}{\partial s} \right) \left(\frac{\partial \log L}{\partial \sigma} \right) \right] \\
&= \frac{1}{\sigma} n \frac{\psi(1+\frac{1}{s})}{s^2} - \frac{\psi(1+\frac{1}{s})}{\sigma s} n \frac{1}{s} \\
&\quad - \frac{s}{\sigma} n \frac{\Gamma(2+\frac{1}{s})\psi(2+\frac{1}{s})}{\Gamma(1+\frac{1}{s})s^2} + \frac{\psi(1+\frac{1}{s})}{\sigma} n \frac{(1+\frac{1}{s})}{s}.
\end{aligned}$$

Após algumas manipulações algébricas, tem-se que

$$E \left[\left(\frac{\partial \log L}{\partial \sigma} \right) \left(\frac{\partial \log L}{\partial s} \right) \right] = E \left[\left(\frac{\partial \log L}{\partial s} \right) \left(\frac{\partial \log L}{\partial \sigma} \right) \right] = -\frac{n}{\sigma s}$$

$$\begin{aligned}
E \left[\left(\frac{\partial \log L}{\partial s} \right)^2 \right] &= \sum_{i=1}^n E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^{2s} \log^2 \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right) \right] \\
&\quad - \frac{2\psi(1 + \frac{1}{s})}{s} \sum_{i=1}^n E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^{2s} \log \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right) \right] \\
&\quad + \frac{\psi^2(1 + \frac{1}{s})}{s^2} \sum_{i=1}^n E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^{2s} \right],
\end{aligned}$$

considerando (VIII), (VII) e (V), tem-se que

$$\begin{aligned}
E \left[\left(\frac{\partial \log L}{\partial s} \right)^2 \right] &= n \frac{(1 + \frac{1}{s}) \left[\psi^2(2 + \frac{1}{s}) + \psi'(2 + \frac{1}{s}) \right]}{s^3} \\
&\quad - \frac{2\psi(1 + \frac{1}{s})}{s} n \frac{\Gamma(2 + \frac{1}{s}) \psi(2 + \frac{1}{s})}{\Gamma(1 + \frac{1}{s}) s^2} + \frac{\psi^2(1 + \frac{1}{s})}{s^2} n \frac{(1 + \frac{1}{s})}{s}.
\end{aligned}$$

Após algumas manipulações algébricas, tem-se que

$$E \left[\left(\frac{\partial \log L}{\partial s} \right)^2 \right] = \frac{n}{s^3} \left\{ \left(1 + \frac{1}{s} \right) \Psi' \left(1 + \frac{1}{s} \right) \right\}.$$

Prova dos elementos da matriz da Seção 4:

A seguir, calculam-se os elementos da matriz de informação de Fisher em (4.7).

$$E \left[\left(\frac{\partial \log L}{\partial \boldsymbol{\beta}} \right) \left(\frac{\partial \log L}{\partial \boldsymbol{\beta}^T} \right) \right] = \frac{s^2 \Gamma^2(1 + \frac{1}{s})}{\sigma^2} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^T \boldsymbol{\beta}|}{\sigma} \right)^{2(s-1)} \text{sign}^2(y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \right],$$

considerando (I), tem-se que

$$\begin{aligned}
E \left[\left(\frac{\partial \log L}{\partial \boldsymbol{\beta}} \right) \left(\frac{\partial \log L}{\partial \boldsymbol{\beta}^T} \right) \right] &= \frac{s^2 \Gamma^2(1 + \frac{1}{s})}{\sigma^2} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T \frac{\Gamma(2 - \frac{1}{s})}{\Gamma(1 + \frac{1}{s})} \\
&= \frac{\Gamma(\frac{1}{s}) \Gamma(2 - \frac{1}{s})}{\sigma^2} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T.
\end{aligned}$$

$$\begin{aligned}
E \left[\left(\frac{\partial \log L}{\partial \boldsymbol{\beta}} \right) \left(\frac{\partial \log L}{\partial \sigma} \right) \right] &= E \left[\left(\frac{\partial \log L}{\partial \sigma} \right) \left(\frac{\partial \log L}{\partial \boldsymbol{\beta}} \right) \right] \\
&= -\frac{s\Gamma(1 + \frac{1}{s})}{\sigma^2} \sum_{i=1}^n \mathbf{x}_i^T E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^T \boldsymbol{\beta}|}{\sigma} \right)^{(s-1)} \text{sign}(y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \right] \\
&\quad + \frac{s^2 \Gamma(1 + \frac{1}{s})}{\sigma^2} \sum_{i=1}^n \mathbf{x}_i^T E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^T \boldsymbol{\beta}|}{\sigma} \right)^{2s-1} \text{sign}(y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \right],
\end{aligned}$$

considerando (II) e (III), tem-se que $E \left[\left(\frac{\partial \log L}{\partial \beta} \right) \left(\frac{\partial \log L}{\partial \sigma} \right) \right] = E \left[\left(\frac{\partial \log L}{\partial \sigma} \right) \left(\frac{\partial \log L}{\partial \beta} \right) \right] = \mathbf{0}$.

$$E \left[\left(\frac{\partial \log L}{\partial \sigma} \right)^2 \right] = \frac{n}{\sigma^2} - 2 \frac{s}{\sigma^2} \sum_{i=1}^n E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^T \beta|}{\sigma} \right)^s \right] + \frac{s^2}{\sigma^2} \sum_{i=1}^n E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^T \beta|}{\sigma} \right)^{2s} \right],$$

considerando (IV) e (V), tem-se que $E \left[\left(\frac{\partial \log L}{\partial \sigma} \right)^2 \right] = \frac{n}{\sigma^2} - 2 \frac{s}{\sigma^2} \frac{n}{s} + \frac{s^2}{\sigma^2} \frac{n(1 + \frac{1}{s})}{s} = \frac{ns}{\sigma^2}$.

$$\begin{aligned} E \left[\left(\frac{\partial \log L}{\partial \sigma} \right) \left(\frac{\partial \log L}{\partial s} \right) \right] &= E \left[\left(\frac{\partial \log L}{\partial s} \right) \left(\frac{\partial \log L}{\partial \sigma} \right) \right] \\ &= \frac{1}{\sigma} \sum_{i=1}^n E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^T \beta|}{\sigma} \right)^s \log \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^T \beta|}{\sigma} \right) \right] \\ &\quad - \frac{\psi(1 + \frac{1}{s})}{\sigma s} \sum_{i=1}^n E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^T \beta|}{\sigma} \right)^s \right] \\ &\quad - \frac{s}{\sigma} \sum_{i=1}^n E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^T \beta|}{\sigma} \right)^{2s} \log \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^T \beta|}{\sigma} \right) \right] \\ &\quad + \frac{\psi(1 + \frac{1}{s})}{\sigma} \sum_{i=1}^n E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^T \beta|}{\sigma} \right)^{2s} \right], \end{aligned}$$

considerando (VI), (IV), (VII) e (V), tem-se que

$$\begin{aligned} E \left[\left(\frac{\partial \log L}{\partial \sigma} \right) \left(\frac{\partial \log L}{\partial s} \right) \right] &= E \left[\left(\frac{\partial \log L}{\partial s} \right) \left(\frac{\partial \log L}{\partial \sigma} \right) \right] \\ &= \frac{1}{\sigma} n \frac{\psi(1 + \frac{1}{s})}{s^2} - \frac{\psi(1 + \frac{1}{s})}{\sigma s} n \frac{1}{s} \\ &\quad - \frac{s}{\sigma} n \frac{\Gamma(2 + \frac{1}{s}) \psi(2 + \frac{1}{s})}{\Gamma(1 + \frac{1}{s}) s^2} + \frac{\psi(1 + \frac{1}{s})}{\sigma} n \frac{(1 + \frac{1}{s})}{s}. \end{aligned}$$

Após algumas manipulações algébricas, tem-se que

$$E \left[\left(\frac{\partial \log L}{\partial \sigma} \right) \left(\frac{\partial \log L}{\partial s} \right) \right] = E \left[\left(\frac{\partial \log L}{\partial s} \right) \left(\frac{\partial \log L}{\partial \sigma} \right) \right] = -\frac{n}{\sigma s}$$

$$\begin{aligned} E \left[\left(\frac{\partial \log L}{\partial s} \right)^2 \right] &= \sum_{i=1}^n E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^T \beta|}{\sigma} \right)^{2s} \log^2 \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^T \beta|}{\sigma} \right) \right] \\ &\quad - \frac{2\psi(1 + \frac{1}{s})}{s} \sum_{i=1}^n E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^T \beta|}{\sigma} \right)^{2s} \log \left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^T \beta|}{\sigma} \right) \right] \\ &\quad + \frac{\psi^2(1 + \frac{1}{s})}{s^2} \sum_{i=1}^n E \left[\left(\frac{\Gamma(1 + \frac{1}{s}) |y_i - \mathbf{x}_i^T \beta|}{\sigma} \right)^{2s} \right], \end{aligned}$$

considerando (VIII), (VII) e (V), tem-se que

$$E \left[\left(\frac{\partial \log L}{\partial s} \right)^2 \right] = n \frac{(1 + \frac{1}{s}) \left[\psi^2 (2 + \frac{1}{s}) + \psi' (2 + \frac{1}{s}) \right]}{s^3} - \frac{2\psi (1 + \frac{1}{s})}{s} n \frac{\Gamma (2 + \frac{1}{s}) \psi (2 + \frac{1}{s})}{\Gamma (1 + \frac{1}{s}) s^2} + \frac{\psi^2 (1 + \frac{1}{s})}{s^2} n \frac{(1 + \frac{1}{s})}{s}.$$

Após algumas manipulações algébricas, tem-se que

$$E \left[\left(\frac{\partial \log L}{\partial s} \right)^2 \right] = \frac{n}{s^3} \left\{ \left(1 + \frac{1}{s} \right) \Psi' \left(1 + \frac{1}{s} \right) \right\}.$$

A.2 Distribuições *a posteriori*

Prova da Proposição 2.2: Integrando em σ , obtém-se a verossimilhança para (μ, s) .

$$\begin{aligned} L(\mu, s|\mathbf{y}) &= \int_0^\infty L(\mu, \sigma, s|\mathbf{y})\pi(\sigma)\mathbf{d}\sigma \\ &= \int_0^\infty 2^{-n}\sigma^{-n} \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma (1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^s \right\} \sigma^{-a} d\sigma \\ &\propto \int_0^\infty \sigma^{-(n+a)} \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma (1 + \frac{1}{s}) |y_i - \mu|}{\sigma} \right)^s \right\} d\sigma. \end{aligned}$$

Considera-se a mudança de variável $z = \frac{1}{\sigma^s} \sum_{i=1}^n (\Gamma (1 + \frac{1}{s}) |y_i - \mu|)^s$. A função de verossimilhança para (μ, s) reduz-se a

$$L(\mu, s|\mathbf{y}) \propto \int_0^\infty z^{\frac{n+a-1}{s}-1} e^{-z} dz$$

Utiliza-se a definição da função gama, dada por $\Gamma(a) = \int_0^\infty y^{a-1} e^{-y} dy$. A função de verossimilhança para (μ, s) pode ser reescrita como

$$L(\mu, s|\mathbf{y}) \propto s^{-1} \Gamma \left(\frac{n+a-1}{s} \right) \left[\sum_{i=1}^n \left(\Gamma \left(1 + \frac{1}{s} \right) |y_i - \mu| \right)^s \right]^{-\frac{(n+a-1)}{s}} \quad (\text{A.1})$$

Considerando-se a verossimilhança $L(\mu, s|\mathbf{y})$, integra-se em μ

$$\begin{aligned} L(s|\mathbf{y}) &= \int_{\mathfrak{R}} L(\mu, s|\mathbf{y})\pi(\mu)\mathbf{d}\mu \\ &= 2^{-n} s^{-1} \Gamma \left(\frac{n+a-1}{s} \right) \Gamma \left(1 + \frac{1}{s} \right)^{-(n+a-1)} \int_{\mathfrak{R}} \left[\sum_{i=1}^n |y_i - \mu|^s \right]^{-\frac{(n+a-1)}{s}} d\mu \end{aligned} \quad (\text{A.2})$$

A seguir, definem-se duas funções para provar que $\int_{\mathfrak{R}} \left[\sum_{i=1}^n |y_i - \mu|^s \right]^{-\frac{(n+a-1)}{s}} d\mu$ é limitada.

Sejam

$$A(\mu, s) = n (\max |y_i| + |\mu|)^s$$

e

$$B(\mu, s) = \begin{cases} n|\bar{y} - \mu^*|^s, & \text{se } \mu \in L1, \\ n|\bar{y} - \mu|^s, & \text{se } \mu \in L2, \end{cases}$$

onde $\mu^* = \arg \min_{\mu} \sum_{i=1}^n |y_i - \mu|^s$, $L1 = \{\mu \in \mathfrak{R} : |\mu| < |\bar{y}|\}$ e $L2 = \{\mu \in \mathfrak{R} : |\mu| \geq |\bar{y}|\}$.

Observe-se que

$$\text{I) } \sum_{i=1}^n |y_i - \mu|^s \leq \sum_{i=1}^n (|y_i| + |\mu|)^s \leq \sum_{i=1}^n (\max |y_i| + |\mu|)^s = n(\max |y_i| + |\mu|)^s = A(\mu, s)$$

II) A função $|\cdot|^s$, para $s \geq 1$, é convexa em \mathfrak{R} . Considerando a desigualdade de Jensen, tem-se que

$$\sum_{i=1}^n |y_i - \mu|^s \geq \left| \sum_{i=1}^n (y_i - \mu) \right|^s = n |\bar{y} - \mu|^s.$$

Se $\mu \in L1$, segue que $\sum_{i=1}^n |y_i - \mu|^s \geq n|\bar{y} - \mu^*|^s$. Se $\mu \in L2$, segue que $\sum_{i=1}^n |y_i - \mu|^s \geq n|\bar{y} - \mu|^s$.

$$\text{Assim, } \sum_{i=1}^n |y_i - \mu|^s \geq B(\mu, s).$$

Considerando I e II, segue que $B(\mu, s) \leq \sum_{i=1}^n |y_i - \mu|^s \leq A(\mu, s)$. Portanto

$$\int_{\mathfrak{R}} [A(\mu, s)]^{-\frac{(n+a-1)}{s}} d\mu \leq \int_{\mathfrak{R}} \left[\sum_{i=1}^n |y_i - \mu|^s \right]^{-\frac{(n+a-1)}{s}} d\mu \leq \int_{\mathfrak{R}} [B(\mu, s)]^{-\frac{(n+a-1)}{s}} d\mu.$$

Calcula-se a integral acima do lado esquerdo e direito. Para a integral do lado esquerdo, tem-se

$$\int_{\mathfrak{R}} [A(\mu, s)]^{-\frac{(n+a-1)}{s}} d\mu = \int_{\mathfrak{R}} [n(\max |y_i| + |\mu|)^s]^{-\frac{(n+a-1)}{s}} d\mu = 2 n^{-\frac{(n+a-1)}{s}} \int_0^{\infty} (\max |y_i| + \mu)^{-\frac{(n+a-1)}{s}} d\mu$$

seja $z = \max |y_i| + \mu$, então $dz = d\mu$. Assim, tem-se que

$$\begin{aligned} \int_{\mathfrak{R}} [n(\max |y_i| + |\mu|)^s]^{-\frac{(n+a-1)}{s}} d\mu &= 2 \left. \frac{n^{-\frac{(n+a-1)}{s}} z^{-(n+a-2)}}{-(n+a-2)} \right|_{\max |y_i|}^{\infty} \\ &= 2n^{-\frac{(n+a-1)}{s}} [(n+a-2)(\max |y_i|)^{n+a-2}]^{-1}. \end{aligned} \quad (\text{A.3})$$

Portanto

$$\int_{\mathfrak{R}} [n(\max |y_i| + |\mu|)^s]^{-\frac{(n+a-1)}{s}} d\mu = n^{-\frac{(n+a-1)}{s}} f_1(y), \quad (\text{A.4})$$

em que $f_1(y) = \frac{2}{(n+a-2)(\max |y_i|)^{n+a-2}}$ não depende de s .

Para a integral do lado direito, tem-se

$$\begin{aligned} \int_{\mathfrak{R}} [B(\mu, s)]^{-\frac{(n+a-1)}{s}} d\mu &= \int_{L1} [n|\bar{y} - \mu^*|^s]^{-\frac{(n+a-1)}{s}} d\mu + \int_{L2} [n|\bar{y} - \mu|^s]^{-\frac{(n+a-1)}{s}} d\mu \\ &= n^{-\frac{(n+a-1)}{s}} \left[\int_{L1} |\bar{y} - \mu^*|^{-(n+a-1)} d\mu + \int_{L2} |\bar{y} - \mu|^{-(n+a-1)} d\mu \right]. \end{aligned} \quad (\text{A.5})$$

Portanto

$$\int_{\mathfrak{R}} [B(\mu, s)]^{-\frac{(n+a-1)}{s}} d\mu = n^{-\frac{(n+a-1)}{s}} f_2(y), \quad (\text{A.6})$$

em que $f_2(y) = \int_{L1} |\bar{y} - \mu^*|^{-(n+a-1)} d\mu + \int_{L2} |\bar{y} - \mu|^{-(n+a-1)} d\mu$ não depende de s .

Considerando (A.4) e (A.6), tem-se que

$$f_1(y) \leq n^{-\frac{(n+a-1)}{s}} \int_{\mathfrak{R}} \left[\sum_{i=1}^n |y_i - \mu|^s \right]^{-\frac{(n+a-1)}{s}} d\mu \leq f_2(y) \quad (\text{A.7})$$

Portanto

$$\int_{\mathfrak{R}} \left[\sum_{i=1}^n |y_i - \mu|^s \right]^{-\frac{(n+a-1)}{s}} d\mu = O \left(n^{-\frac{(n+a-1)}{s}} \right) \quad (\text{A.8})$$

O resultado acima permitirá estudar o comportamento da verossimilhança integrada para s . Substituindo (A.8) em (A.2), tem-se

$$L(s|\mathbf{y}) \propto s^{-1} \Gamma \left(\frac{n+a-1}{s} \right) \Gamma \left(1 + \frac{1}{s} \right)^{-(n+a-1)} O \left(n^{-\frac{(n+a-1)}{s}} \right) \quad (\text{A.9})$$

Para estudar o comportamento da verossimilhança integrada para s , considera-se o seguinte resultado: a expansão da série $\frac{1}{\Gamma(z)} \approx z$ para z próximo de 0 (Abramowitz & Stegun, 1972). Portanto se $s \rightarrow \infty$, tem-se $\Gamma\left(\frac{1}{s}\right) \approx s$. Além disso, considerando a expansão da série de Taylor de 1ª ordem de $\log \Gamma(1+z)$ em torno de $z=0$, tem-se que $\log \Gamma(1+z) \approx z\Psi(1)$, onde $\Psi(1) \approx -0.57721$. Assim, $\Gamma\left(1+\frac{1}{s}\right) \approx e^{\frac{\Psi(1)}{s}}$ para s grande. Sendo assim, para $s \rightarrow \infty$, tem-se $\Gamma\left(\frac{n+a-1}{s}\right) \approx \frac{s}{n+a-1}$. Portanto

$$\begin{aligned} L(s|\mathbf{y}) &\approx s^{-1} \frac{s}{n+a-1} \left(e^{\frac{\Psi(1)}{s}}\right)^{-(n+a-1)} O\left(n^{-\frac{(n+a-1)}{s}}\right) \\ &\approx e^{\frac{-\Psi(1)(n+a-1)}{s}} O\left(n^{-\frac{(n+a-1)}{s}}\right) \\ &= O\left(e^{\frac{-(n+a-1)}{s}\{\Psi(1)+\log n\}}\right) \\ &= O(1). \end{aligned} \tag{A.10}$$

Completando a prova. □

Prova da Proposição 2.3: A prova será feita considerando-se a condição dada em (2.15) e os Corolários 2.1 e 2.2.

Considere a *a priori* de referência dada no Corolário 2.1, o resultado do Corolário 2.2 e a condição (2.15), tem-se que a *posteriori* de referência é própria se e somente

$$\int_1^\infty O(1)O\left(s^{-\frac{3}{2}}\right) ds < \infty.$$

Sendo assim,

$$\begin{aligned} \int_1^\infty O(1)O\left(s^{-\frac{3}{2}}\right) ds &= \int_1^\infty s^{-\frac{3}{2}} ds \\ &= 2 < \infty. \end{aligned} \tag{A.11}$$

Portanto a *a priori* de referência conduz a uma *posteriori* própria, completando a prova.

□.

Considere a *a priori* de Jeffreys dada no Corolário 2.1, o resultado do Corolário 2.2 e a condição (2.15), tem-se que a *posteriori* de Jeffreys é própria se e somente

$$\int_1^\infty O(1)O\left(s^{-1}\right) ds < \infty.$$

Sendo assim,

$$\begin{aligned} \int_1^\infty O(1)O\left(s^{-1}\right) ds &= \int_1^\infty \frac{ds}{s} \\ &= \infty. \end{aligned} \quad (\text{A.12})$$

Portanto a *a priori* de Jeffreys conduz a uma *posteriori* imprópria, completando a prova. \square

Prova da Proposição 4.2: Integrando em σ , obtém-se a verossimilhança para $(\boldsymbol{\beta}, s)$.

$$\begin{aligned} L(\boldsymbol{\beta}, s|\mathbf{y}) &= \int_0^\infty L(\boldsymbol{\beta}, s|\mathbf{y})\pi(\sigma)d\sigma \\ &= \int_0^\infty 2^{-n}\sigma^{-n} \exp\left\{-\sum_{i=1}^n \left(\frac{\Gamma\left(1+\frac{1}{s}\right)|y_i - \mathbf{x}_i^\top \boldsymbol{\beta}|}{\sigma}\right)^s\right\} \sigma^{-a} d\sigma \\ &\propto \int_0^\infty \sigma^{-(n+a)} \exp\left\{-\sum_{i=1}^n \left(\frac{\Gamma\left(1+\frac{1}{s}\right)|y_i - \mathbf{x}_i^\top \boldsymbol{\beta}|}{\sigma}\right)^s\right\} d\sigma. \end{aligned}$$

Considera-se a mudança de variável $z = \frac{1}{\sigma^s} \sum_{i=1}^n \left(\Gamma\left(1+\frac{1}{s}\right)|y_i - \mathbf{x}_i^\top \boldsymbol{\beta}\right)^s$. A função de verossimilhança para $(\boldsymbol{\beta}, s)$ reduz-se a

$$L(\boldsymbol{\beta}, s|\mathbf{y}) \propto \int_0^\infty z^{\frac{n+a-1}{s}-1} e^{-z} dz$$

Utiliza-se a definição da função gama, dada por $\Gamma(a) = \int_0^\infty y^{a-1} e^{-y} dy$. A função de verossimilhança para $(\boldsymbol{\beta}, s)$ pode ser reescrita como

$$L(\boldsymbol{\beta}, s|\mathbf{y}) \propto s^{-1} \Gamma\left(\frac{n+a-1}{s}\right) \left[\sum_{i=1}^n \left(\Gamma\left(1+\frac{1}{s}\right)|y_i - \mathbf{x}_i^\top \boldsymbol{\beta}\right)^s\right]^{\frac{-(n+a-1)}{s}}. \quad (\text{A.13})$$

Considerando-se a verossimilhança $L(\boldsymbol{\beta}, s|\mathbf{y})$, integra-se em $\boldsymbol{\beta}$

$$\begin{aligned} L(s|\mathbf{y}) &= \int_{\mathbb{R}^p} L(\boldsymbol{\beta}, s|\mathbf{y})\pi(\boldsymbol{\beta})d\boldsymbol{\beta} \\ &\propto s^{-1} \Gamma\left(\frac{n+a-1}{s}\right) \Gamma\left(1+\frac{1}{s}\right)^{-(n+a-1)} \int_{\mathbb{R}^p} \left[\sum_{i=1}^n |y_i - \mathbf{x}_i^\top \boldsymbol{\beta}|^s\right]^{\frac{-(n+a-1)}{s}} d\boldsymbol{\beta}. \end{aligned} \quad (\text{A.14})$$

Além disso, pelo Lema 3.2 em [Salazar et al. \(2012\)](#), $\int_{\mathbb{R}^p} \left[\sum_{i=1}^n |y_i - \mathbf{x}_i^\top \boldsymbol{\beta}|^s\right]^{\frac{-(n+a-1)}{s}} d\boldsymbol{\beta}$ é limitada, de ordem $n^{\frac{-(n+a-1)}{s}}$ para $n > p + 1 - a$.

Portanto

$$L(s|\mathbf{y}) \propto s^{-1} \Gamma\left(\frac{n+a-1}{s}\right) \Gamma\left(1 + \frac{1}{s}\right)^{-(n+a-1)} O\left(n^{-\frac{(n+a-1)}{s}}\right). \quad (\text{A.15})$$

Para estudar o comportamento da verossimilhança integrada para s , consideram-se os resultados apresentados na prova do Corolário 2.2. Sendo assim, para $s \rightarrow \infty$, tem-se que $\Gamma\left(\frac{1}{s}\right) \approx s$, $\Gamma\left(1 + \frac{1}{s}\right) \approx e^{\frac{\Psi(1)}{s}}$ e $\Gamma\left(\frac{n+a-1}{s}\right) \approx \frac{s}{n+a-1}$.

Portanto

$$\begin{aligned} L(s|\mathbf{y}) &\approx s^{-1} \frac{s}{n+a-1} \left(e^{\frac{\Psi(1)}{s}}\right)^{-(n+a-1)} O\left(n^{-\frac{(n+a-1)}{s}}\right) \\ &\approx e^{-\frac{\Psi(1)(n+a-1)}{s}} O\left(n^{-\frac{(n+a-1)}{s}}\right) \\ &= O\left(e^{-\frac{(n+a-1)}{s}\{\Psi(1)+\log n\}}\right) \\ &= O(1). \end{aligned} \quad (\text{A.16})$$

Completando a prova. □

Prova da Proposição 4.3: A prova será feita considerando-se a condição dada em (4.19) e os Corolários 4.1 e 4.2.

Considerando-se a *a priori* de referência dada no Corolário 4.1, o resultado do Corolário 4.2 e a condição (4.19), tem-se que a *posteriori* de referência é própria se e somente

$$\int_1^\infty O(1)O\left(s^{-\frac{3}{2}}\right) ds < \infty.$$

Sendo assim,

$$\begin{aligned} \int_1^\infty O(1)O\left(s^{-\frac{3}{2}}\right) ds &= \int_1^\infty s^{-\frac{3}{2}} ds \\ &= 2 < \infty. \end{aligned} \quad (\text{A.17})$$

Portanto a *a priori* de referência conduz a uma *posteriori* própria.

Considerando-se a *a priori* de Jeffreys dada no Corolário 4.1, o resultado do Corolário 4.2 e a condição (4.19), tem-se que a *posteriori* de Jeffreys é própria se e somente

$$\int_1^\infty O(1)O\left(s^{\frac{p-2}{2}}\right) ds < \infty.$$

Sendo assim,

$$\int_1^{\infty} O(1)O\left(s^{\frac{p-2}{2}}\right) ds = \int_1^{\infty} s^{\frac{p-2}{2}} ds \quad (\text{A.18})$$

$$= \infty.$$

Portanto *a priori* de Jeffreys conduz a uma *posteriori* imprópria, completando a prova. \square

A.3 Função de sobrevivência

Se $t > \exp(\mu)$, a função de sobrevivência pode ser escrita por

$$S(t) = \int_t^{\infty} \frac{1}{2y\sigma} \exp\left\{-\left(\frac{\Gamma\left(1+\frac{1}{s}\right)(\log(y)-\mu)}{\sigma}\right)^s\right\} dy.$$

Considera-se a mudança de variável $z = \left(\frac{\Gamma\left(1+\frac{1}{s}\right)(\log(y)-\mu)}{\sigma}\right)^s$. A função de sobrevivência reduz-se a

$$S(t) = \frac{1}{2s\Gamma\left(1+\frac{1}{s}\right)} \int_{\left(\frac{\Gamma\left(1+\frac{1}{s}\right)(\log(t)-\mu)}{\sigma}\right)^s}^{\infty} z^{\frac{1}{s}-1} e^{-z} dz.$$

Utiliza-se a definição da função gama incompleta, dada por $\Gamma(a, t) = \int_t^{\infty} y^{a-1} \exp(-y) dy$.

A função de sobrevivência pode ser reescrita como

$$S(t) = \frac{\Gamma\left(\frac{1}{s}, \left(\frac{\Gamma\left(1+\frac{1}{s}\right)(\log(t)-\mu)}{\sigma}\right)^s\right)}{2s\Gamma\left(1+\frac{1}{s}\right)}.$$

De maneira semelhante, se $0 < t \leq \exp(\mu)$, pode-se mostrar que

$$S(t) = 1 - \frac{\Gamma\left(\frac{1}{s}, \left(\frac{\Gamma\left(1+\frac{1}{s}\right)(\mu-\log(t))}{\sigma}\right)^s\right)}{2s\Gamma\left(1+\frac{1}{s}\right)}.$$

Apêndice B

Metropolis-Hasting dentro do amostrador de Gibbs

Os detalhes do algoritmo Metropolis-Hasting dentro do amostrador de Gibbs, fazendo-se o uso do passeio aleatório, são apresentados a seguir.

O algoritmo Metropolis-Hasting dentro do amostrador de Gibbs é utilizado na geração dos valores de μ , σ e s . Para os parâmetros μ , σ e s , consideram-se as seguintes distribuições condicionais: $\pi^R(\mu|\sigma, s, \mathbf{y})$, $\pi^R(\sigma|\mu, s, \mathbf{y})$ e $\pi^R(s|\mu, \sigma, \mathbf{y})$, respectivamente. No entanto, sob o modelo dado, $\mu \in \mathfrak{R}$, $\sigma > 0$, $s > 1$, considera-se uma mudança de variáveis de (μ, σ, s) para $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) = (\mu, \log(\sigma), \log(s - 1))$. Isto transforma o espaço paramétrico em \mathfrak{R}^3 , e também ajuda a simetrizar a distribuição *a posteriori*. Considerando-se o Jacobiano dessa transformação, tem-se que a distribuição *a posteriori* é dada por

$$\pi^R(\boldsymbol{\theta}|\mathbf{y}) \propto \exp(\theta_2)^{-n} \exp(\theta_3) (1 + \exp(\theta_3))^{-\frac{3}{2}} \left[\left(1 + \frac{1}{(1 + \exp(\theta_3))} \right) \Psi' \left(1 + \frac{1}{(1 + \exp(\theta_3))} \right) \right]^{\frac{1}{2}} \\ \times \exp \left\{ - \sum_{i=1}^n \left(\frac{\Gamma \left(1 + \frac{1}{(1 + \exp(\theta_3))} \right) |y_i - \theta_1|}{\exp(\theta_2)} \right)^{(1 + \exp(\theta_3))} \right\}.$$

A implementação do algoritmo Metropolis-Hasting via passeio aleatório para o parâmetro θ_1 , por exemplo, é feita considerando-se núcleos de transição $q(\theta_1, \theta_1^*)$, de tal modo que θ_1^* é gerado pelo processo $\theta_1^* = \theta_1 + kz$, em que $z \sim N(0, \eta^2)$ e k é uma

constante de escala que controla a taxa de aceitação do algoritmo. Escolhe-se η^2 como o elemento diagonal correspondente da matriz de covariância da *posteriori* conjunta de θ , avaliada na moda a *posteriori* de $\pi^R(\theta|\mathbf{y})$ (Carlin & Louis, 2001b).

A estabilidade numérica do algoritmo é melhorada trabalhando-se na escala logarítmica. Os passos do algoritmo para gerar θ são os seguintes:

1. Inicie com $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}) = (\mu^{(0)}, \log(\sigma^{(0)}), \log(s^{(0)} - 1))$.
2. Gere θ_1^* de acordo com o núcleo transicional $q(\theta_1^{(0)}, \theta_1^*)$.
3. Gere um valor u da distribuição uniforme $U(0, 1)$.
4. Se $u \leq \min \left\{ 1, \exp \left[\log \pi^R \left(\theta_1^* | \theta_2^{(0)}, \theta_3^{(0)}, \mathbf{y} \right) - \log \pi^R \left(\theta_1^{(0)} | \theta_2^{(0)}, \theta_3^{(0)}, \mathbf{y} \right) \right] \right\}$, então atualize $\theta_1^{(1)}$ por θ_1^* . Caso contrário permaneça com $\theta_1^{(0)}$, ou seja, $\theta_1^{(1)} = \theta_1^{(0)}$.
5. Proceda analogamente para obter $\theta_2^{(1)}$ e $\theta_3^{(1)}$.
6. Repita os passos de 2 a 5 até obter uma amostra de uma distribuição estacionária.

Após a geração dos valores de θ , obtém-se $(\mu, \sigma = \exp(\theta_2), s = 1 + \exp(\theta_3))$. Assume-se para η o mesmo valor em todos os passos do algoritmo. O valor de k é definido de tal modo que a taxa de aceitação do algoritmo esteja entre 20% e 80% (Muller, 1991). Para verificar a convergência do algoritmo utiliza-se o método de Geweke (1992) juntamente com a análise gráfica.

Esse algoritmo também pode ser utilizado para gerar as amostras *a posteriori* dos parâmetros do modelo logNG. Para tanto, considera-se $\mathbf{y} = \log(\mathbf{t})$. Entretanto, quando a base de dados contém observações censuradas como descrito na Seção 3.2, o procedimento é semelhante, mas com algumas modificações. Na primeira iteração do algoritmo é necessário considerar as observações censuradas como não censuradas. Isto é feito com o objetivo de gerar os valores iniciais dos parâmetros utilizando-se todas as informações disponíveis. Depois de aplicar os passos de 2 a 5, um sexto passo adicional é gerar $n - r$ observações de uma distribuição logNG truncada no valor censurado. De posse destes valores, retorne para o passo 2 e continue o processo iterativo.

Apêndice C

Demais resultados

C.1 Distribuição normal generalizada

Os demais resultados, obtidos por meio do estudo de simulação e dos dados reais para o modelo normal generalizado, são apresentados a seguir.

Tabela C.1: Dados simulados. Médias das taxas de aceitação (%) no algoritmo Metropolis-Hastings para cada tamanho amostral e para cada parâmetro do modelo NG.

Parâmetros	<i>n</i>			
	30	100	300	500
$\mu = 0$	38,94	39,99	39,74	38,99
$\sigma = 1$	37,90	38,80	38,50	37,82
$s = 1,05$	37,78	38,82	38,58	37,94
$\mu = 0$	36,26	34,40	32,90	32,55
$\sigma = 1$	35,39	33,61	32,12	31,75
$s = 2$	35,00	33,47	31,94	31,60
$\mu = 0$	36,80	33,98	33,53	33,45
$\sigma = 1$	35,91	33,17	32,75	32,65
$s = 3$	35,10	32,95	32,57	32,47
$\mu = 0$	38,15	34,00	34,40	34,57
$\sigma = 1$	37,07	33,19	33,60	33,76
$s = 5$	35,91	32,98	33,41	33,59

Tabela C.2: Dados reais. Resultados do critério de Geweke para os modelos NG, normal e NG com mistura de escala uniforme.

Modelo	Parâmetros	Geweke
NG	μ	0,8889
	σ	-0,5580
	s	-0,9004
Normal	μ	-0,3637
	σ	0,1655
NG com mistura	μ	0,2134
	τ	-0,4003
	s	-1,2890

Tabela C.3: Dados reais. Autocorrelações das cadeias para o modelo NG com mistura de escala uniforme.

Modelo	Parâmetros	lag1	lag10	lag50	lag100	lag500	lag1000
NG com mistura	μ	0,9958	0,9560	0,7946	0,6417	0,1957	-0,0135
	τ	0,9988	0,9880	0,9461	0,9034	0,6871	0,5205
	s	0,9966	0,9673	0,8574	0,7598	0,5345	0,3818

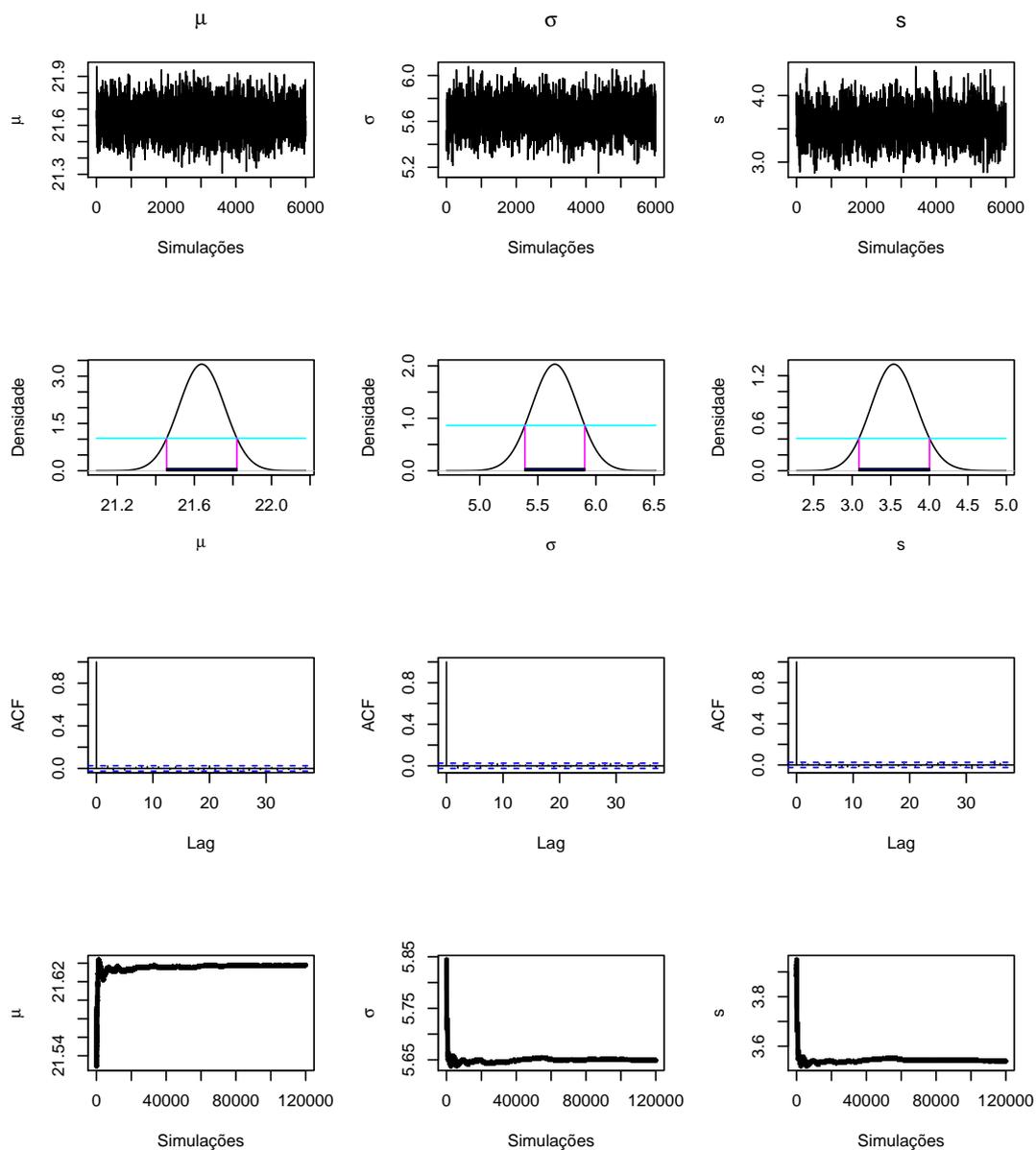


Figura C.1: Dados reais. Históricos das cadeias, densidades *a posteriori* marginais dos parâmetros do modelo NG com HPD a 95%, autocorrelações amostrais e médias ergódicas.

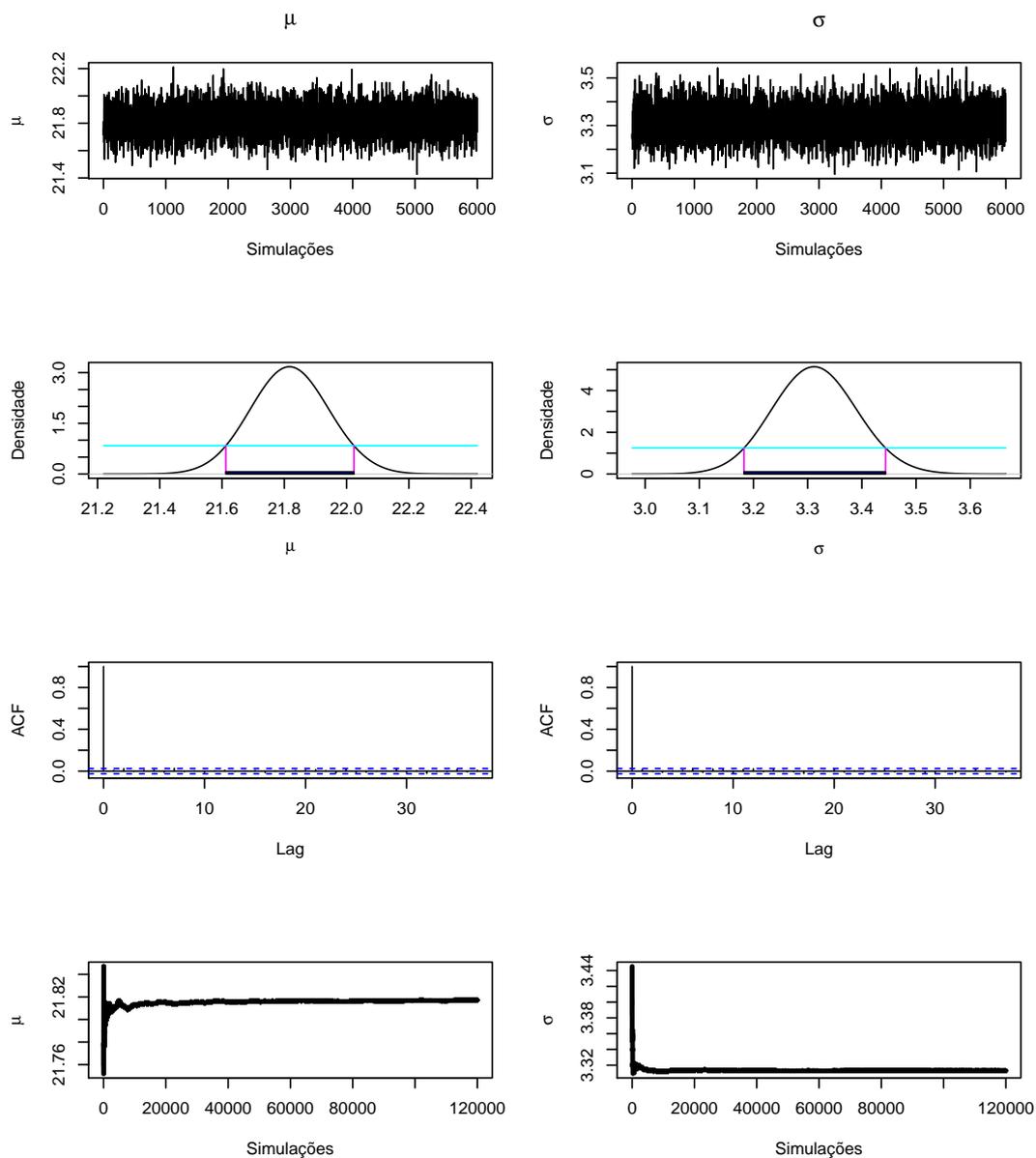


Figura C.2: Dados reais. Históricos das cadeias, densidades *a posteriori* marginais dos parâmetros do modelo normal com HPD a 95%, autocorrelações amostrais e médias ergódicas.

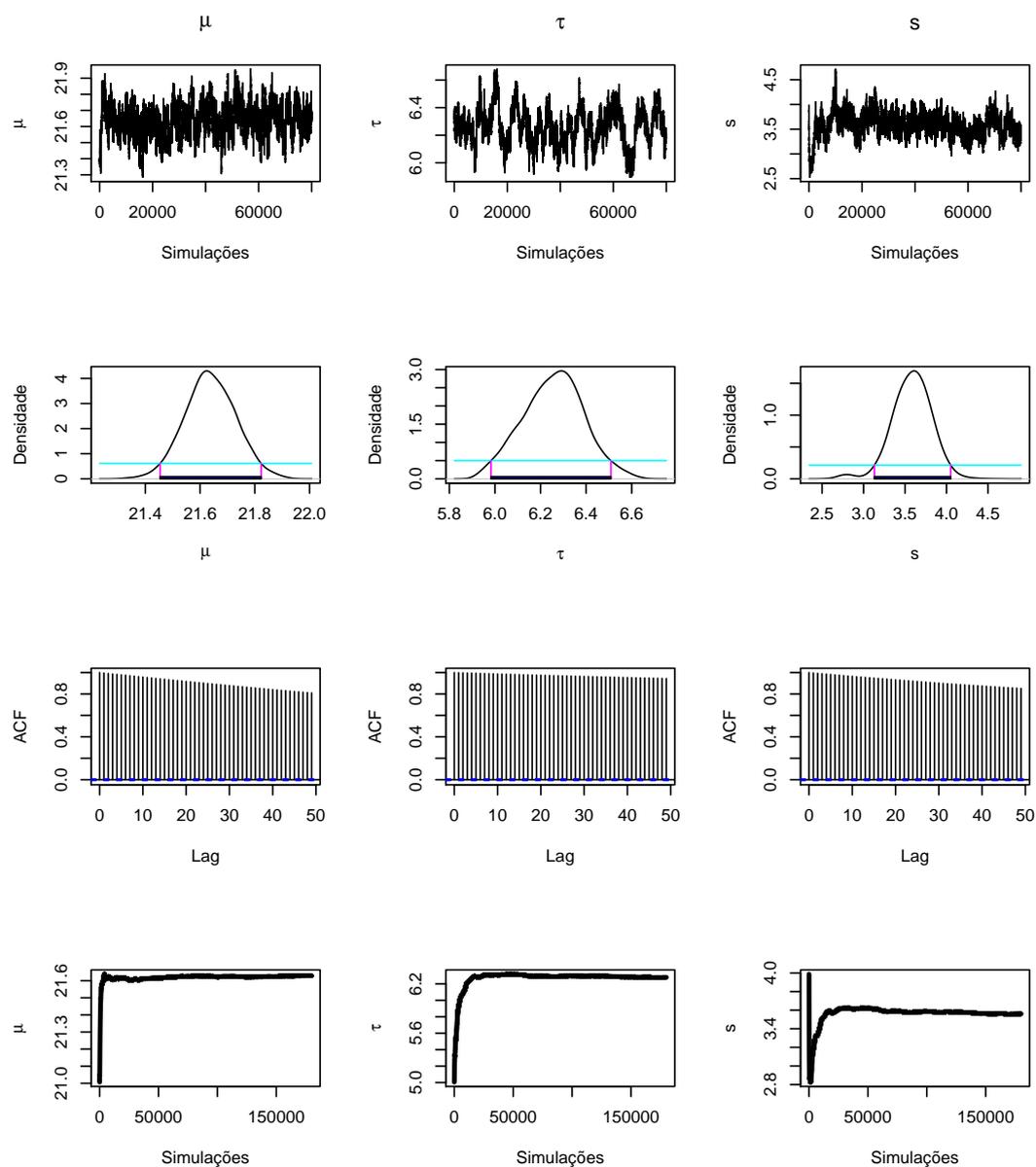


Figura C.3: Dados reais. Históricos das cadeias, densidades *a posteriori* marginais dos parâmetros do modelo NG com mistura de escala uniforme, considerando-se HPD a 95%, autocorrelações amostrais e médias ergódicas.

C.2 Distribuição lognormal generalizada

Os demais resultados, obtidos por meio do estudo de dados artificiais e dos dados reais para o modelo lognormal generalizado, são apresentados a seguir.

Tabela C.4: Dados artificiais. Resultados do critério de Geweke para o modelo logNG.

Modelo	Parâmetros	Geweke
logNG	μ	-0,7568
	σ	-0,1053
	s	-0,2300

Tabela C.5: Dados reais. Resultados do critério de Geweke para os modelos logNG, lognormal e logNG com mistura de escala uniforme.

Modelo	Parâmetros	Geweke
logNG	μ	-0,9059
	σ	-0,8406
	s	-0,1495
lognormal	μ	0,1550
	σ	-1,8860
logNG com mistura	μ	0,4633
	τ	-0,3198
	s	-0,6374

Tabela C.6: Dados reais. Autocorrelações das cadeias para o modelo logNG com mistura de escala uniforme.

Modelo	Parâmetros	lag1	lag10	lag50	lag100	lag500	lag1000
logNG com mistura	μ	0,9819	0,8357	0,4247	0,1913	0,0311	0,0445
	τ	0,9959	0,9604	0,8222	0,6835	0,1810	0,1112
	s	0,9809	0,8516	0,6319	0,5170	0,1349	0,0694

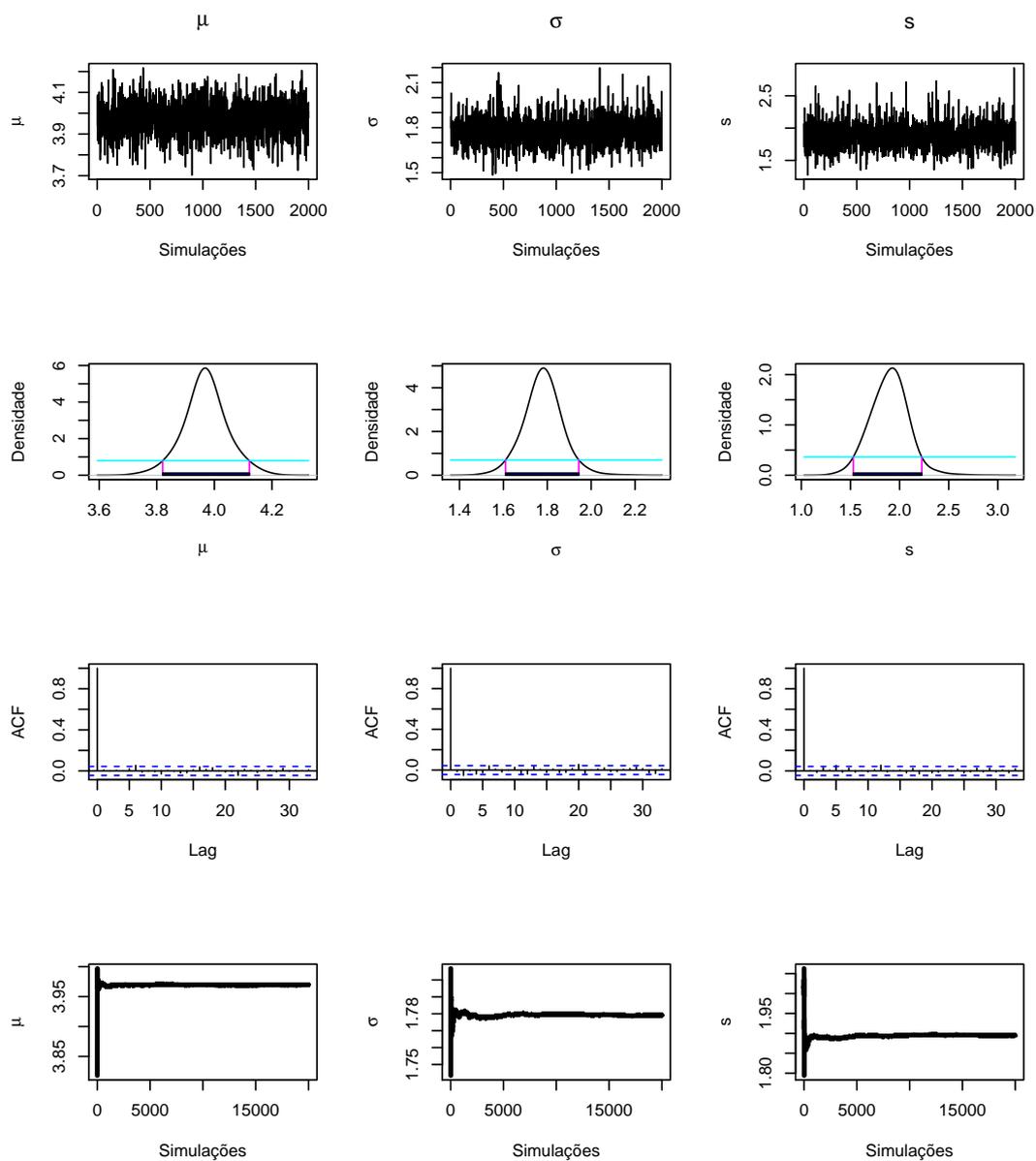


Figura C.4: Dados artificiais. Históricos das cadeias, densidades a *posteriori* marginais dos parâmetros do modelo logNG com HPD a 95%, autocorrelações amostrais e médias ergódicas.

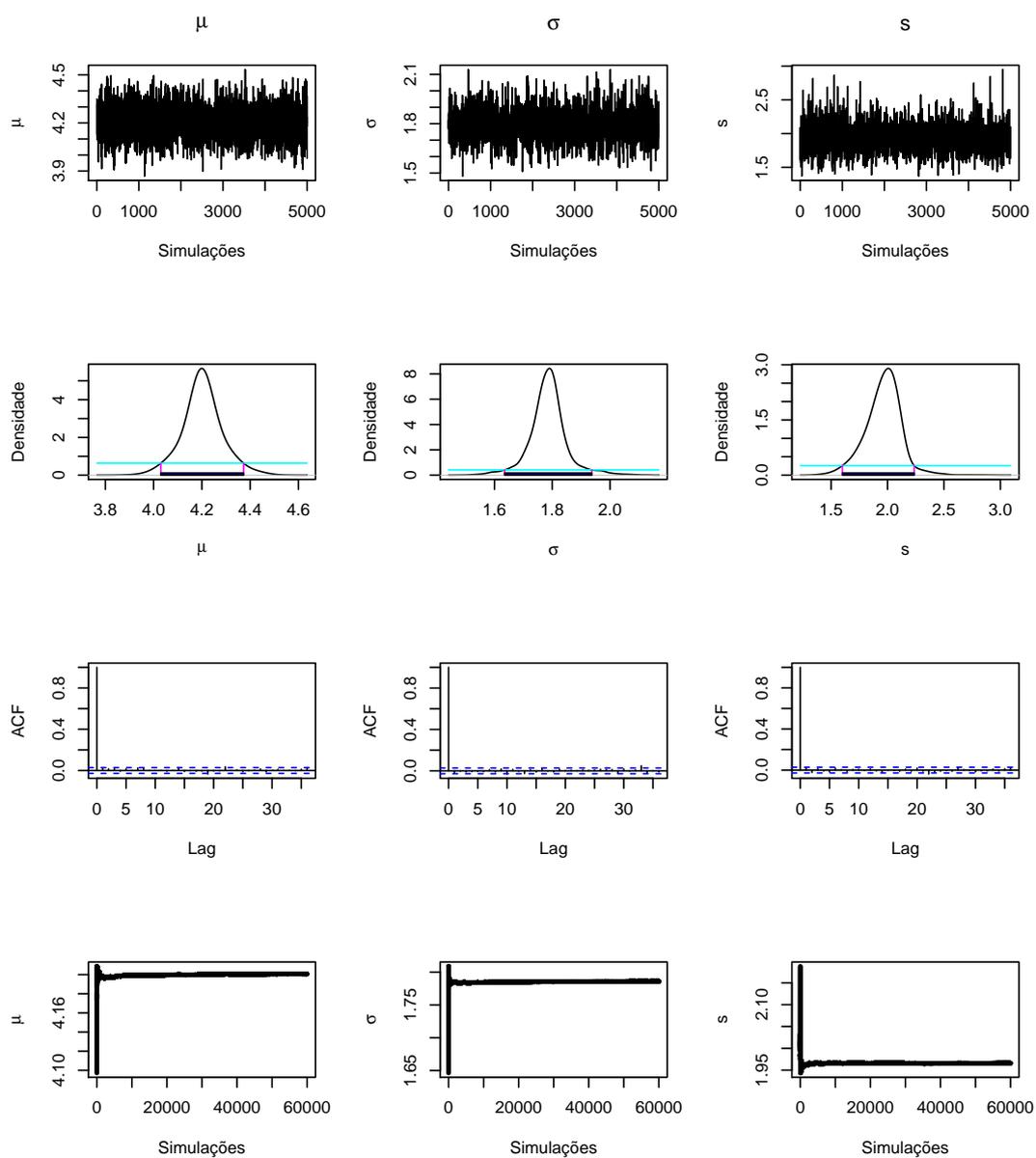


Figura C.5: Dados reais. Históricos das cadeias, densidades *a posteriori* marginais dos parâmetros do modelo logNG com HPD a 95%, autocorrelações amostrais e médias ergódicas.

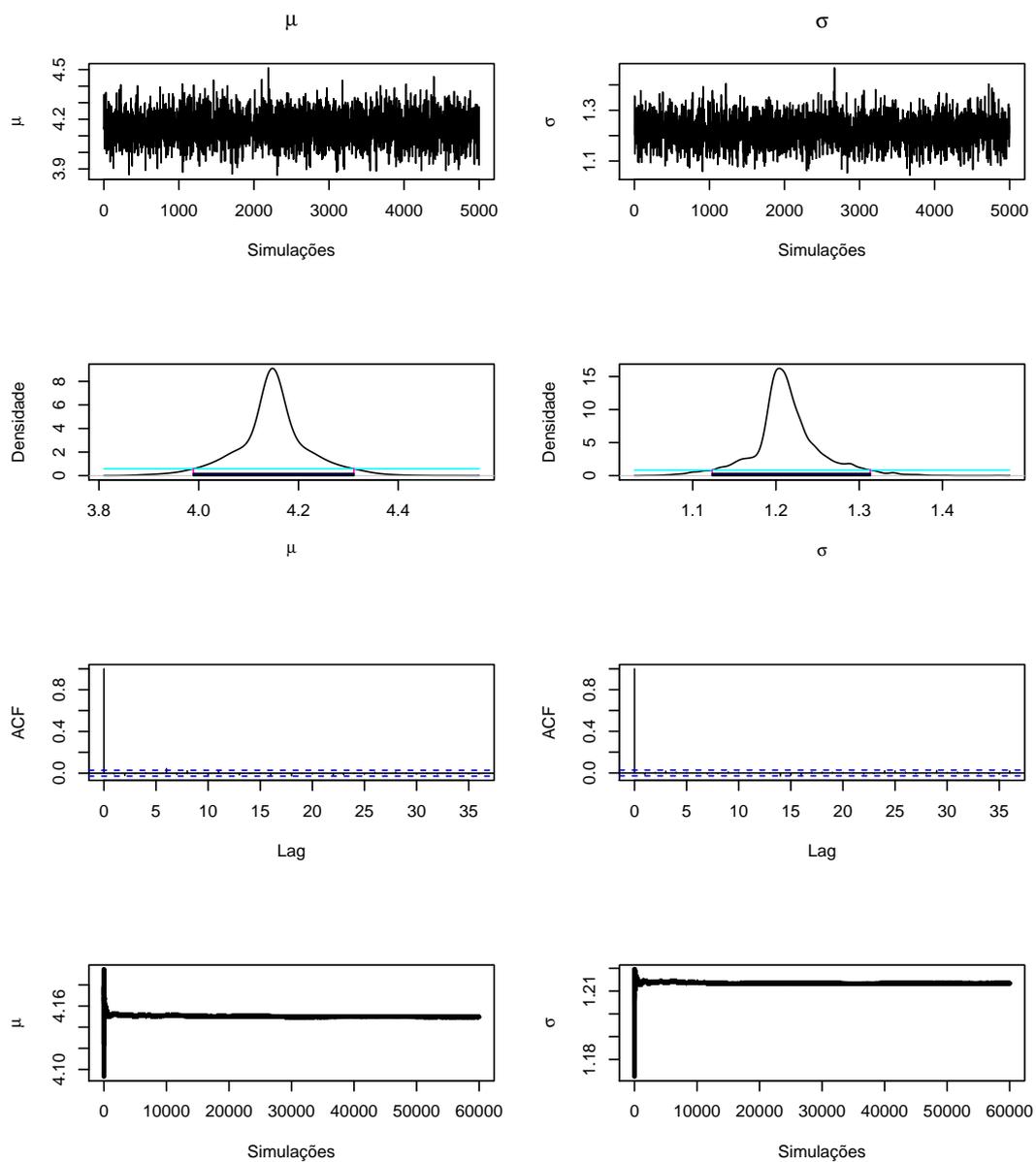


Figura C.6: Dados reais. Históricos das cadeias, densidades *a posteriori* marginais dos parâmetros do modelo lognormal com HPD a 95%, autocorrelações amostrais e médias ergódicas.

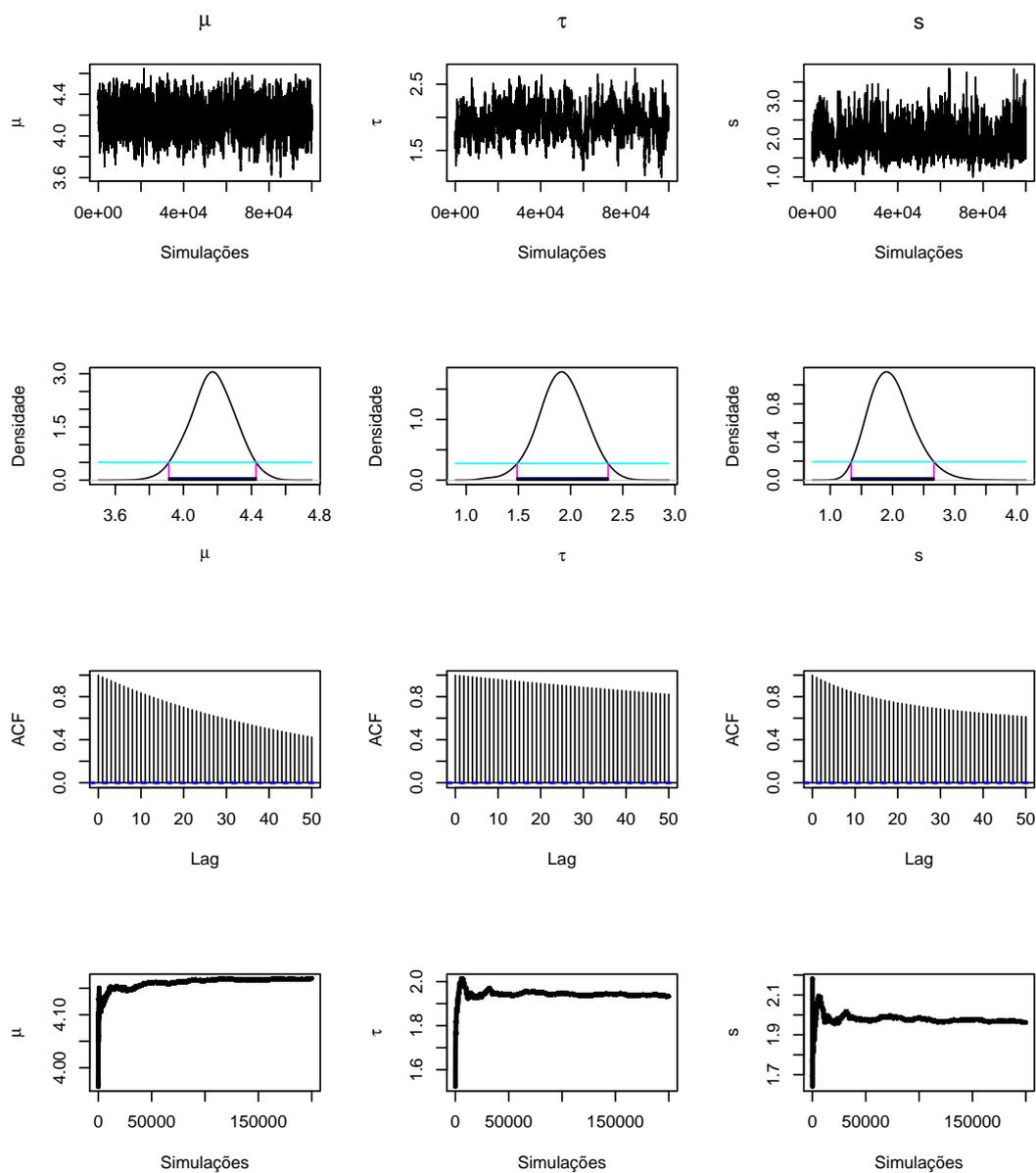


Figura C.7: Dados reais. Históricos das cadeias, densidades *a posteriori* marginais dos parâmetros do modelo logNG com mistura de escala uniforme, considerando-se HPD a 95%, autocorrelações amostrais e médias ergódicas.

C.3 Modelo de regressão normal generalizado

Os demais resultados, obtidos por meio do estudo de dados artificiais e dos dados reais para o modelo de regressão normal generalizado, são apresentados a seguir.

Tabela C.7: Dados artificiais. Resultados do critério de Geweke para o modelo de regressão NG.

Modelo	Parâmetros	Geweke
NG	β_1	-0,3594
	β_2	0,2483
	σ	0,0814
	s	0,2122

Tabela C.8: Dados reais. Resultados do critério de Geweke para os modelos de regressão NG e normal.

Modelo	Parâmetros	Geweke
NG	β_1	0,3758
	β_2	-0,3906
	σ	0,3756
	s	0,4790
Normal	β_1	1,1100
	β_2	-1,0720
	σ	-0,8476

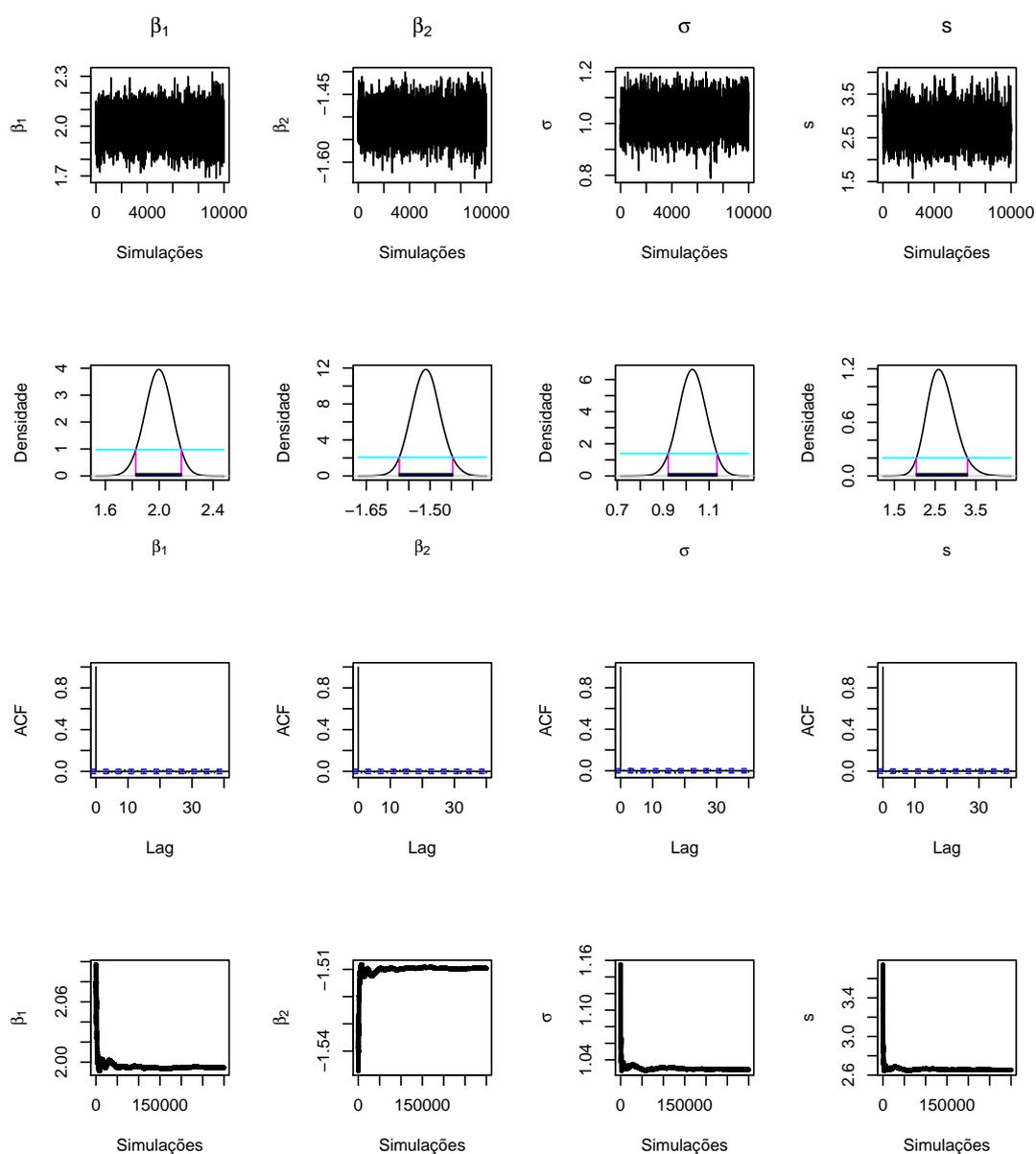


Figura C.8: Dados artificiais. Históricos das cadeias, densidades a *posteriori* marginais dos parâmetros do modelo de regressão NG com HPD a 95%, autocorrelações amostrais e médias ergódicas.

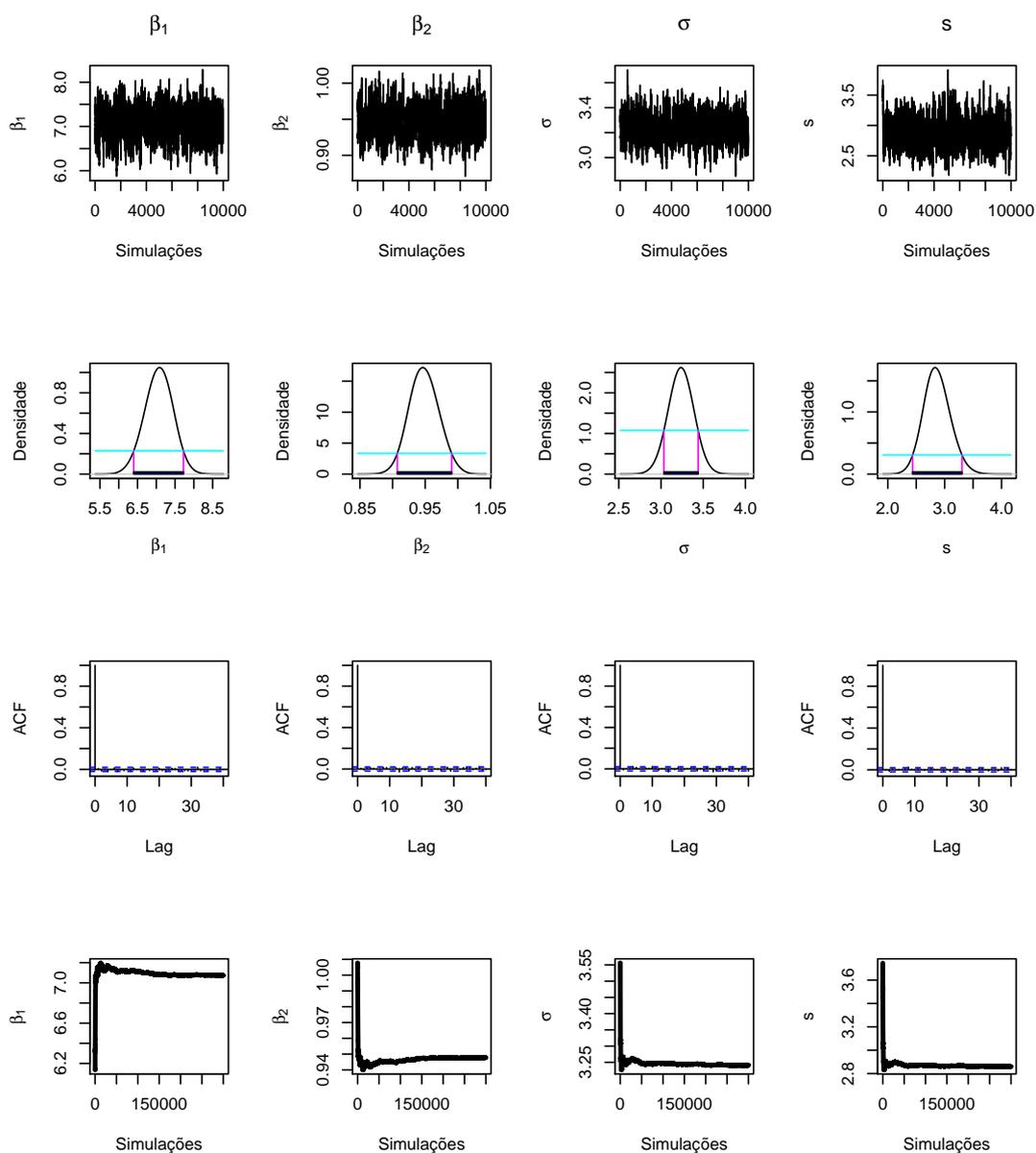


Figura C.9: Dados reais. Históricos das cadeias, densidades *a posteriori* marginais dos parâmetros do modelo de regressão NG com HPD a 95%, autocorrelações amostrais e médias ergódicas.

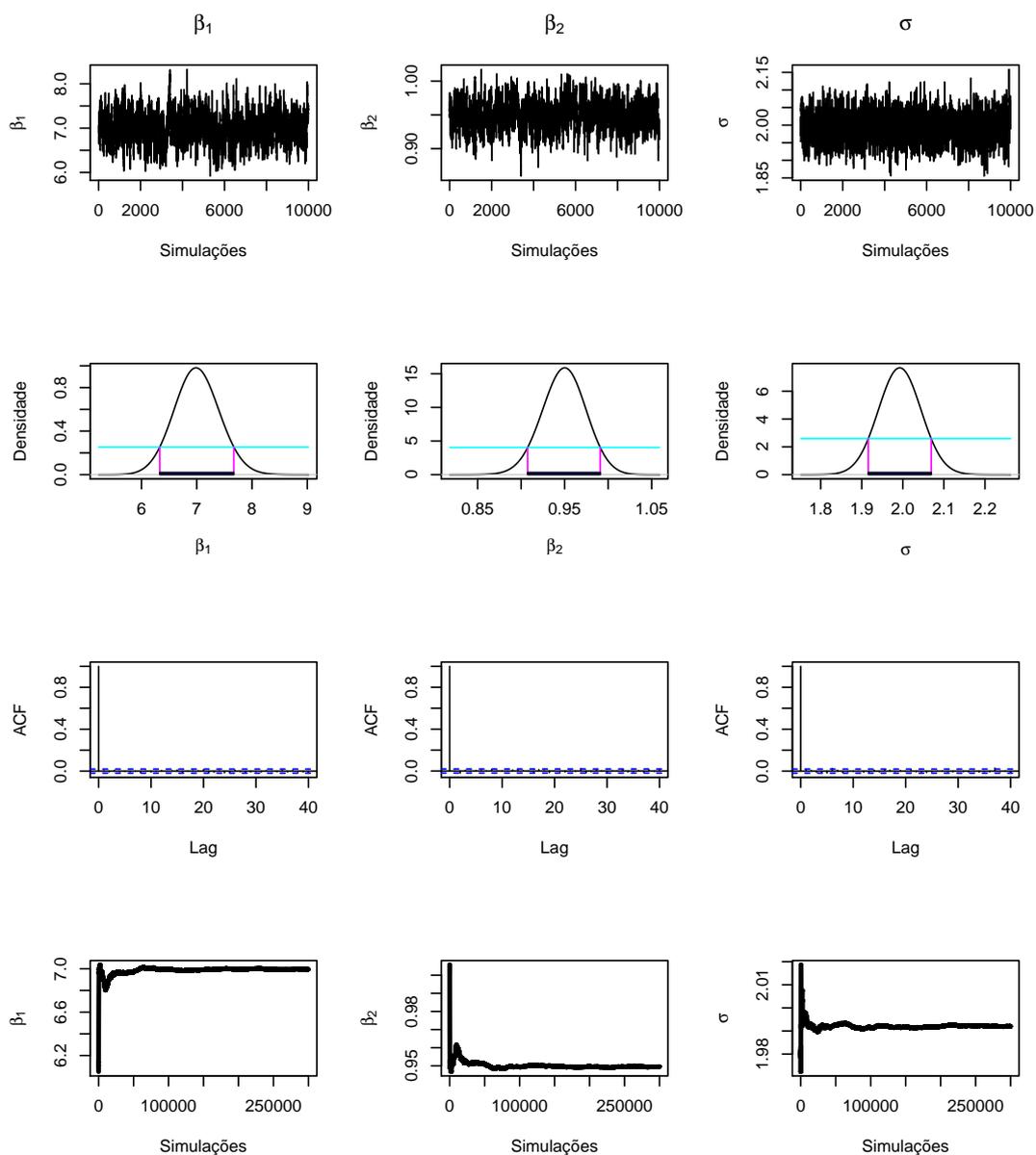


Figura C.10: Dados reais. Históricos das cadeias, densidades *a posteriori* marginais dos parâmetros do modelo de regressão normal com HPD a 95%, autocorrelações amostrais e médias ergódicas.