

Universidade Federal do Rio Grande do Norte
Centro de Ciências Exatas e da Terra
Programa de Pós-Graduação em Matemática Aplicada e Estatística

Isaac Jales Costa Souza

Estimação Bayesiana no Modelo Potência Normal Bimodal Assimétrico

Natal, 2016

Isaac Jales Costa Souza

Estimação Bayesiana no Modelo Potência Normal Bimodal Assimétrico

Trabalho apresentado ao Programa de Pós-Graduação em Matemática Aplicada e Estatística da Universidade Federal do Rio Grande do Norte, em cumprimento com as exigências legais para obtenção do título de Mestre.

Área de Atuação: Probabilidade e Estatística

Orientador:

Prof. Dr. Bernardo Borba de Andrade

Co-orientador:

Prof. Dr. Fidel Ernesto Castro Morales

Natal, 2016

Catálogo da Publicação na Fonte. UFRN / SISBI / Biblioteca Setorial
Centro de Ciências Exatas e da Terra – CCET.

Souza, Isaac Jales Costa.

Estimação bayesiana no modelo potência normal bimodal assimétrico / Isaac
Jales Costa Souza. - Natal, 2016.
iv, 97f. : il.

Orientador: Prof. Dr. Bernardo Borba de Andrade.

Coorientador: Prof. Dr. Fidel Ernesto Castro Morales.

Dissertação (Mestrado) – Universidade Federal do Rio Grande do Norte. Centro
de Ciências Exatas e da Terra. Programa de Pós-Graduação em Matemática
Aplicada e Estatística.

1. Inferência bayesiana – Dissertação. 2. Assimetria – Dissertação. 3.
Bimodalidade – Dissertação. 4. DIC – Dissertação. 5. MCMC – Dissertação. 6.
Priori de Jeffreys – Dissertação. I. Andrade, Bernardo Borba de. II. Morales, Fidel
Ernesto Castro. III. Título.

RN/UF/BSE-CCET

CDU: 519.226

Isaac Jales Costa Souza

Estimação Bayesiana no Modelo Potência Normal Bimodal Assimétrico

Trabalho apresentado ao Programa de Pós-Graduação em Matemática Aplicada e Estatística da Universidade Federal do Rio Grande do Norte, em cumprimento com as exigências legais para obtenção do título de Mestre.

Aprovado em: 28 de janeiro de 2016

Banca Examinadora:

Prof. Dr. Bernardo Borba de Andrade
Departamento de Estatística - UNB
Orientador

Prof.^a Dr.^a Luz Milena Zea Fernández
Departamento de Estatística - UFRN
Examinador Interno

Prof. Dr. Fernando Ferraz do Nascimento
Departamento de Estatística - UFPI
Examinador Externo

Natal
Janeiro, 2016

Resumo

Neste trabalho, é apresentada uma abordagem bayesiana dos modelos potência normal bimodal (PNB) e potência normal bimodal assimétrico (PNBA). Primeiramente, apresentamos o modelo PNB e especificamos para este prioris não informativas e informativas do parâmetro que concentra a bimodalidade (α). Em seguida, obtemos a distribuição a posteriori pelo método MCMC, o qual testamos a viabilidade de seu uso a partir de um diagnóstico de convergência. Depois, fizemos uma análise de sensibilidade com o intuito de avaliar o efeito da variação dos hiperparâmetros da priori na distribuição a posteriori. Também foi feita uma simulação para avaliar o desempenho do estimador bayesiano utilizando prioris informativas. Constatamos que a estimativa da moda a posteriori apresentou em geral resultados melhores quanto ao erro quadratico médio (EQM) e viés percentual (VP) quando comparado ao estimador de máxima verossimilhança. Ajustamos o modelo PNB em dois conjunto de dados reais, estes foram o índice de desenvolvimento ambiental (EPI) dos países e o nível de RNA do gene MAOA de índios Pima. Por último, introduzimos o modelo de regressão linear com resíduos PNB. Quanto ao modelo PNBA, também especificamos prioris informativas e não informativas para os parâmetros de bimodalidade e assimetria. Fizemos o diagnóstico de convergência para o método MCMC, que também foi utilizado para obter a distribuição a posteriori. Fizemos uma análise de sensibilidade, aplicamos os dados do gêiser *Old Faithful* no modelo. Introduzimos o modelo de regressão linear com resíduos PNBA, o qual foi ajustado nos dados de pesca de camarões do RN.

Palavras-Chave: Assimetria. Bimodalidade. DIC. Inferência Bayesina. MCMC. Priori de Jeffreys.

Abstract

In this paper it is presented a Bayesian approach to the bimodal power-normal (BPN) models and the bimodal asymmetric power-normal (BAPN). First, we present the BPN model, specifying its non-informative and informative parameter α (bimodality). We obtain the posterior distribution by MCMC method, whose feasibility of use we tested from a convergence diagnose. After that, We use different informative priors for α and we do a sensitivity analysis in order to evaluate the effect of hyperparameters variation on the posterior distribution. Also, it is performed a simulation to evaluate the performance of the Bayesian estimator using informative priors. We noted that the Bayesian method shows more satisfactory results when compared to the maximum likelihood method. It is performed an application with bimodal data. Finally, we introduce the linear regression model with BPN error. As for the BAPN model we also specify informative and uninformative priors for bimodality and asymmetry parameters. We do the MCMC Convergence Diagnostics, which is also used to obtain the posterior distribution. We do a sensitivity analysis, applying actual data in the model and we introducing the linear regression model with PNBA error.

keywords: Assimetry. Bimodality. DIC. Bayesian Inference. MCMC. Jeffreys Prior.

Sumário

1	Introdução	5
1.1	Revisão Bibliográfica	6
1.1.1	Modelo Normal Assimétrico	6
1.1.2	Modelo Potência Normal	7
1.1.3	Extensões Bimodal do Modelo Normal Assimétrico	8
1.2	Conceitos Basicos de Inferência Bayesiana	8
1.3	Objetivos	9
1.4	Organização da Dissertação	9
2	O Modelo Potência Normal Bimodal	11
2.1	Modelo Potência Normal Bimodal	11
2.1.1	Função de Densidade	11
2.1.2	Média, Mediana, Moda e Variância	12
2.1.3	Extensão do Modelo com Parâmetros de Posição e Escala	13
2.2	Inferência	13
2.2.1	Função de Verossimilhança	13
2.2.2	Distribuições a Priori	13
2.2.3	Distribuições a Posteriori	18
2.3	Aspectos Computacionais	19
2.3.1	O Algoritmo MCMC	19
2.3.2	Diagnóstico de Convergência	25
2.4	Estudo de Simulação	28
2.4.1	Análise de Sensibilidade	28
2.4.2	Estudo do Estimador Bayesiano	39
2.5	Aplicações com Dados Bimodais	41
3	Regressão com resíduos PNB	46
3.1	Inferência	46
3.1.1	Função de Verossimilhança	46
3.1.2	Distribuição a Priori para o Vetor β	47
3.1.3	Distribuições a Posteriori	47

3.2	Aspectos Computacionais	48
4	O Modelo Potência Normal Bimodal Assimétrico	51
4.1	Modelo Potência Normal Bimodal Assimétrico	51
4.1.1	Função de Densidade	51
4.1.2	Média, Moda e Variância	52
4.1.3	Extensão para o Modelo com Parâmetro de Posição e Escala	53
4.2	Inferência	53
4.2.1	Função de Verossimilhança	53
4.2.2	Distribuição a Priori para o Vetor (α, λ)	53
4.2.3	Distribuições a Posteriori	58
4.3	Aspectos Computacionais	60
4.3.1	O Algoritmo MCMC	60
4.3.2	Diagnóstico de Convergência	61
4.4	Estudo de Simulação	64
4.4.1	Análise de Sensibilidade	64
4.5	Aplicações Com Dados Bimodais Assimétricos	74
5	Regressão com resíduos PNBA	77
5.1	Inferência	77
5.1.1	Função de Verossimilhança	77
5.1.2	Distribuições a Posteriori	78
5.2	Aspectos Computacionais	79
5.3	Aplicações Com Modelo de Regressão PNBA	80
6	Considerações Finais	84
7	Apêndice	85
7.1	Apêndice A	85
7.2	Apêndice B	91
	Referências	94

Capítulo 1

Introdução

Em aplicações, muitas vezes nos deparamos com dados cuja distribuição empírica apresenta um comportamento bimodal e, às vezes, assimétrico. Na prática, para modelagem de dados assimétricos, tem sido utilizado transformação de variáveis. Tal procedimento dificulta a interpretação dos resultados obtidos, uma vez que a análise é feita acerca dos dados transformados, e não dos originais. Para modelagem de dados bimodais, são considerados com frequência misturas de distribuições. No entanto, modelos de misturas de distribuições podem ser não identificáveis e não parcimoniosos. Como alternativa, alguns autores propuseram modelos assimétricos e bimodais. Os modelos mais recentes propostos com essas características foram os modelos potência normal bimodal (PNB) e potência normal bimodal assimétrico (PNBA) de Bolfarine (2013).

Dentre os modelos assimétricos propostos anteriormente aos de Bolfarine estão o modelo skew-normal (SN) (Azzalini, 1985) e potência normal (PN) (Durrans, 1992). Com base na distribuição SN, Kim (2005) propôs o modelo two-piece skew-normal (TN), o qual pode assumir formas bimodais simétricas. Mais adiante, Arnold (2009) propôs a generalização do modelo TN, o qual é denotado por extend two-piece skew-normal (ETN). Este pode assumir formas bimodais e assimétricas. Finalmente, baseado no modelo PN e TN, Bolfarine (2013) propôs o modelo PNB. E baseado no modelo PN e ETN, o mesmo propôs o modelo PNBA. Na seção de revisão bibliográfica deste trabalho, pode ser vista algumas propriedades dos modelos SN, PN, TN e ETN.

Por terem sido propostas recentemente, as distribuições PNB e PNBA não foram estudadas por muitos autores. O trabalho mais recente sobre tais distribuições foi o de Andrade et al. (2016), no qual foram estudadas a geração de números aleatórios e a estimação dos parâmetros destas distribuições. Porém, não foi feito um estudo a partir de uma perspectiva bayesiana para os modelos propostos por Bolfarine (2013). Neste trabalho, propomos um modelo bayesiano para as distribuições PNB e PNBA. Focamos mais o modelo PNB, no qual estudamos as propriedades frequentistas do estimador bayesiano e comparamos com o estimador de máxima verossimilhança. Além disso, desenvolvemos modelos de regressão linear com resíduos PNB e PNBA, o que até então não havia sido apresentado na literatura. Propomos para estes um

modelo bayesiano. Fizemos uma análise de sensibilidade afim de verificar o impacto na distribuição a posteriori ao mudar os parâmetros da priori. Além disso, ajustamos os modelos PNB e PNBA em dados reais. Comparamos estes ajustes com o de misturas de normais. Ainda empregamos os modelos de regressão com resíduos PNBA em uma das aplicações. Quatro conjunto de dados foram utilizados. Esses são:

- Dados do índice de desempenho ambiental (EPI) do ano de 2012.
- Dados do nível de RNA do gene monoamina oxidase A (MAOA) de índios pima do Arizona com predisposição a diabetes.
- Dados do tempo da erupção do gêiser *Old faithful*.
- Dados de pesca de camarões da espécie *farfantepenaeus brasiliensis* na costa do Rio Grande do Norte.

O índice dos dados EPI foi proposto por pesquisadores da universidade de Yale e de Columbia (Estes et al., 2005) e ele quantifica o quão eficiente são as políticas ambientais dos países pelo mundo. Os dados MAOA vieram do estudo de Mason et al. (2011), a MAOA é um gene cuja função é decompor importantes neurotransmissores do cérebro. O conjunto de dados do gêiser consta no software R. Gêisers são fontes termais de origem vulcânica que lançam jatos de água e vapor em intervalo regulares de tempo. O gêiser *Old faithful* é o mais famoso do mundo e está presente no parque nacional de Yellowstone, nos Estados Unidos da América. Os dados dos camarões se referem a medidas biométricas do crustáceo. Para os dados de pesca de camarões foram ajustados modelos de regressão PNBA.

1.1 Revisão Bibliográfica

1.1.1 Modelo Normal Assimétrico

O modelo normal assimétrico, também denotado por *skew-normal*, foi criado a partir dos trabalhos de Roberts (1966) e O'Hagan e Leonard (1976). Mais adiante, Azzalini (1985) estudou intensamente suas propriedades e criou uma estrutura geral para um modelo assimétrico, do qual a *skew-normal* é um caso particular. O modelo proposto por Azzalini é dado por:

$$f(z|\lambda) = 2f_0(z)G(\lambda z), \quad z \in \mathbb{R}, \lambda \in \mathbb{R}, \quad (1.1)$$

em que $f_0(z)$ é uma fdp (função densidade de probabilidade) simétrica em torno de zero e $G(z)$ uma fda (função de distribuição). Quando $f_0(z) = \phi(z)$ e $G(z) = \Phi(z)$, em que $\phi(z)$ e $\Phi(z)$ são a fdp e a fda da normal padrão, respectivamente, temos que $f(z|\lambda)$ é a fdp da *skew-normal*, a qual denotamos por $SN(\lambda)$, dada por:

$$f(z|\lambda) = 2\phi(z)\Phi(\lambda z), \quad z \in \mathbb{R}, \lambda \in \mathbb{R}.$$

O parâmetro λ controla a assimetria da distribuição. Se $\lambda > 0$, em (1.1), temos que $f(z|\lambda)$ é assimétrica à direita. Se $\lambda < 0$, $f(z|\lambda)$ é assimétrica à esquerda. Se $\lambda = 0$, então $f(z|\lambda) = \phi(z)$. Após Azzalini (1985) desenvolver a estrutura geral do modelo assimétrico, Henze (1986), Chiogna (1997) e Pewsey (2000) deram continuidade ao estudo do modelo. Um estudo a partir da perspectiva bayesiana do modelo *skew-normal* foi desenvolvido por Liseo e Loperfido (2006) cujos resultados obtidos através de simulações mostraram que a mediana da posteriori do modelo bayesiano proposto é um estimador menos tendencioso que o EMV. Além disso, o EMV apresenta problemas em algumas situações. Por exemplo, no caso em que todas as observações da amostra têm o mesmo sinal, o EMV obtido é $+\infty$ ou $-\infty$. Outro problema do EMV é que a matriz de informação de Fisher é singular quando $\lambda = 0$, conforme Azzalini (1985) mostrou. Posteriormente, Bayes e Branco (2007) propuseram outro modelo bayesiano, baseado na estimação de uma reparametrização de λ , que toma valores no intervalo $(-1, 1)$, cuja distribuição atribuída a priori foi uma uniforme no intervalo $(-1, 1)$, o qual denotamos de $U(-1, 1)$. O estimador de λ no modelo de Bayes e Branco (2007) apresentou um viés e um erro quadrático médio na estimação de λ menor do que o modelo proposto por Liseo e Loperfido (2006). A distribuição a priori utilizada por Liseo e Loperfido (2006) foi uma aproximação da distribuição de Jeffreys. Após o trabalho de Bayes e Branco (2007), Branco et al. (2012) propuseram um modelo bayesiano para estimar o parâmetro de assimetria no caso particular em que $f(z)$ e $G(z)$ são, respectivamente, a fdp e a fda de uma t de Student.

1.1.2 Modelo Potência Normal

O modelo potência normal foi desenvolvido por Durrans (1992), baseado no modelo proposto por Lehmen (1953). O modelo de Lehmen é dado por:

$$f(z|\alpha) = \alpha f_0(z) \{F_0(z)\}^{\alpha-1}, \quad z > 0, \alpha \in \mathbb{Q},$$

em que $f_0(z)$ é uma fdp, F_0 a função de distribuição de $f_0(z)$ e α um número racional. O parâmetro α é o parâmetro de forma de $f(z)$. Pode-se notar que, para um α inteiro, $f(z|\alpha)$ representa a fdp do máximo para uma amostra de tamanho α com fdp dada por $f_0(z)$. Durrans (1992) generalizou essa distribuição para o caso em que α é estritamente positivo, ou seja, $\alpha \in \mathbb{R}^+$, e denotou de gaussiana generalizada o modelo no qual $f_0(z) = \phi(z)$ e $F_0(z) = \Phi(z)$, conhecido como potência normal, cuja densidade é:

$$f(z|\alpha) = \alpha\phi(z) \{\Phi(z)\}^{\alpha-1}, \quad z > 0, \alpha \in \mathbb{R}^+.$$

Denotamos a fdp do modelo potência normal por $PN(\alpha)$. Note que quando $\alpha = 1$, a distribuição resultante é a normal padrão. As propriedades desse modelo foram estudadas adiante

por Gupta e Gupta (2008) e Pewsey (2012). Este último mostrou que a matriz de informação de Fisher é não singular no caso em que $\alpha = 1$, ou seja, quando $f(z|\alpha)$ é a fda da normal padrão.

1.1.3 Extensões Bimodal do Modelo Normal Assimétrico

Algumas extensões do modelo normal assimétrico surgiram com os trabalhos de Kim (2005) e Arnold (2009). Kim (2005) propôs o modelo *two-piece skew normal* (TN), uma extensão do modelo *skew normal*, que pode produzir densidades bimodais e é dado por:

$$f(z|\lambda) = c_\lambda \phi(z) \Phi(\lambda|z|), \quad z \in \mathbb{R}, \lambda \in \mathbb{R},$$

em que $c_\lambda = \frac{2}{1+2\tan^{-1}(\lambda/\pi)}$ e $|z|$ é o valor absoluto de z , ou seja, $|z| = z$ se $z \geq 0$ e $|z| = -z$ se $z < 0$. A densidade de $f(z|\lambda)$ é simétrica e bimodal para $\lambda > 0$. Mais adiante, Arnold (2009) propôs o modelo *extend two-piece skew-normal* (ETN), uma extensão do modelo *two-piece skew normal* dado por:

$$f(z|\lambda) = c_\lambda 2\phi(z) \Phi(\lambda|z|) \Phi(\beta z), \quad z \in \mathbb{R}, \lambda \in \mathbb{R}, \beta \in \mathbb{R},$$

em que c_λ é a mesma constante de normalização do modelo *two-piece skew-normal*. A densidade da *extend two-piece skew-normal* pode ser assimétrica e bimodal para $\lambda > 0$ e $\beta \neq 0$.

1.2 Conceitos Basicos de Inferência Bayesiana

A inferência bayesiana é um tipo de inferência que, ao contrário da clássica, trata um parâmetro θ de uma distribuição $f(x|\theta)$ como uma variável aleatória. A distribuição de θ , o qual denotamos neste trabalho por $\pi(\theta)$, representa o conhecimento que o pesquisador tem a priori sobre o parâmetro. A priori, no contexto bayesiano, significa antes da amostra. Logo, $\pi(\theta)$ representa de forma probabilística o conhecimento que o usuário tem sobre θ anterior às observações dos dados. Após as observações dos dados, a distribuição que reflete o conhecimento sobre θ é denotada de distribuição a posteriori e é dada por $\pi(\theta|\mathbf{x})$, sendo \mathbf{x} a amostra. A distribuição $\pi(\theta|\mathbf{x})$ é encontrada a partir da fórmula de bayes:

$$\pi(\theta|\mathbf{x}) = \frac{L(\theta|\mathbf{x})\pi(\theta)}{\int_{-\infty}^{\infty} L(\theta|\mathbf{x})\pi(\theta)d\theta}, \quad (1.2)$$

em que $L(\theta|\mathbf{x})$ é a função de verossimilhança da $f(x|\theta)$.

O denominador da expressão 1.2 é a distribuição preditiva a priori da amostra, a qual é denotado por $\pi(\mathbf{x})$ e não depende de θ . Assim, podemos rescrever a expressão 1.2 da seguinte maneira:

$$\pi(\theta|\mathbf{x}) \propto L(\theta|\mathbf{x})\pi(\theta),$$

em que \propto significa “proporcional a”.

A distribuição preditiva a priori representa a distribuição das observações anterior às observações dos dados. Após às observações \mathbf{x} , podemos obter a distribuição preditiva a posteriori de um novo valor x_n dada por:

$$\pi(x_n|\mathbf{x}) = \int_{-\infty}^{\infty} f(x_n|\theta)\pi(\theta|\mathbf{x})d\theta.$$

1.3 Objetivos

Esta dissertação tem como objetivo geral implementar métodos bayesianos de estimação para o modelo potência normal bimodal e potência normal bimodal assimétrico. Os objetivos específicos são:

- Propor distribuições a priori para os parâmetros do modelo potência normal bimodal e potência normal bimodal assimétrico e investigar se existem formas conjugadas simples, para que possa ser implementado o amostrador de Gibbs.
- Caso as formas conjugadas não sejam simples ou muito restritivas, investigar outras priors e implementar o método numérico de Metropolis-Hastings.
- Estudar a sensibilidade da distribuição a posteriori dos parâmetros dos modelos potência normal bimodal e potência normal bimodal assimétrico em relação a diferentes distribuições a priori.
- Estudar o comportamento do estimador bayesiano do modelo potência normal bimodal e compará-lo com o estimador de máxima verossimilhança.
- Aplicar o modelo bayesiano na estimação dos parâmetros das distribuições potência normal bimodal e potência normal bimodal assimétrico em dados reais.

1.4 Organização da Dissertação

A dissertação está organizada por capítulos. Em todo texto, utilizamos “.” como separador decimal. No Capítulo 2, falamos do modelo potência normal bimodal. Mostramos o modelo estendido na forma de posição e escala e a função de verossimilhança. Em seguida, apresentamos a distribuição a priori que foi utilizada no trabalho para o vetor de parâmetros do modelo potência normal bimodal. Obtemos as expressões proporcionais da distribuição a posteriori e das distribuições a posteriori condicionais completas. Depois de obtida essas expressões, expomos como utilizamos o método de monte carlo via cadeia de Markov (MCMC) proposto por Metropolis et al. (1953) e Hastings (1970) para a obtenção da distribuição a posteriori, uma

vez que não foi possível calcular a constante de normalização. Testamos a convergência do método MCMC utilizando o método de Gelman e Rubin. Fizemos uma análise de sensibilidade, a fim de verificar o impacto da mudança dos hiperparâmetros da distribuição a priori no ajuste bayesiano. Apresentamos um estudo de simulação para comparar as propriedades do estimador de máxima verossimilhança com os estimadores bayesianos para o modelo potência normal bimodal. Finalmente, uma aplicação com dados reais é realizada. No Capítulo 3, introduzimos o modelo de regressão com resíduos PNB e especificamos a priori utilizada e como ficaram as expressões da posteriori e das posteriores condicionais. Mostramos também como foi utilizado o método MCMC neste caso. Nos Capítulos 4 e 5 são feitos passos análogos aos Capítulos 2 e 3, respectivamente, utilizando a distribuição potência normal bimodal assimétrica. Porém, para esta distribuição, não estudamos as propriedades frequentistas do estimador bayesiano. Para concluir, no Capítulo 6 apresentamos as considerações finais.

Capítulo 2

O Modelo Potência Normal Bimodal

2.1 Modelo Potência Normal Bimodal

2.1.1 Função de Densidade

O Modelo potência normal bimodal (PNB) foi proposto por Bolfarine (2013), baseado nos modelos PN e TN propostos por Durrans (1992) e Kim (2005), respectivamente. O modelo potência normal bimodal pode assumir formas bimodais simétricas, e sua fdp é dada por:

$$f_Z(z|\alpha) = \alpha c_\alpha \phi(z) \{\Phi(|z|)\}^{\alpha-1}, \quad z \in \mathbb{R}, \quad \alpha > 0,$$

em que $c_\alpha = \frac{2^{\alpha-1}}{2^\alpha - 1}$ e c_α é a constante de normalização e α é o parâmetro de forma. Denotamos a variável aleatória Z com distribuição potência normal bimodal por $Z \sim PNB(\alpha)$.

Diferentes formas da distribuição PNB podem ser vistas na Figura 2.1. Notamos que a distribuição é simétrica em torno de zero. Para $\alpha \leq 1$ a distribuição é unimodal, e para $\alpha > 1$ a distribuição é bimodal. Quanto maior for α com relação a 1, mais fortemente bimodal é o modelo. No caso particular em que $\alpha = 1$, a potência normal bimodal é uma normal padrão. Perceba que para $\alpha \neq 1$, a função densidade tem um “bico” em $z = 0$. O que significa que não é diferenciável neste valor na condição restrita.

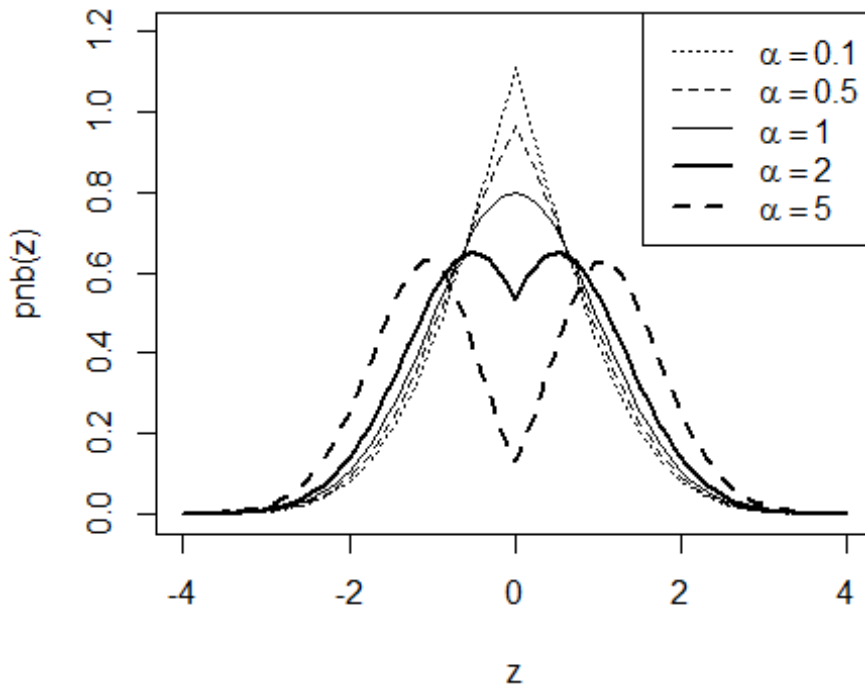


Figura 2.1: Densidade da potência normal bimodal para diferentes valores de α .

2.1.2 Média, Mediana, Moda e Variância

Pela simetria da distribuição potência normal bimodal, temos que:

$$E(Z) = Med(Z) = 0,$$

em que $Med(Z)$ é a mediana da variável aleatória Z .

Podemos obter numericamente os pontos críticos da distribuição por meio da equação:

$$\alpha - 1 = \frac{|z|\Phi(|z|)}{\phi(z)}.$$

Assim, podemos encontrar as duas modas da distribuição, caso $\alpha > 1$. Se $\alpha < 1$, a equação acima não terá solução, porém, a moda nessa situação é 0. Por fim, se $\alpha = 1$ a equação terá solução e a moda também será 0.

A expressão da variância é dada por:

$$Var(Z) = \alpha c_\alpha \left[\frac{(\alpha - 1)(\alpha - 2)}{4\pi} I + \alpha^{-1} - \left(\frac{1}{2}\right) \frac{\alpha^2 - \alpha - 4\pi}{4\pi\alpha} \right],$$

em que $I = \int_0^\infty \phi(\sqrt{3}z) \{\Phi(z)\}^{\alpha-3} dz$.

2.1.3 Extensão do Modelo com Parâmetros de Posição e Escala

O modelo potência normal bimodal estendido na forma de posição e escala é dado pela transformação $Y = \mu + \sigma Z$, em que $\mu \in \mathbb{R}$ é o parâmetro de posição e $\sigma > 0$ é o parâmetro de escala. A densidade de Y é dada por:

$$f_Y(y|\theta) = \frac{\alpha c_\alpha}{\sigma} \phi\left(\frac{y - \mu}{\sigma}\right) \left\{ \Phi\left(\frac{|y - \mu|}{\sigma}\right) \right\}^{\alpha-1},$$

em que $\theta = (\alpha, \mu, \sigma^2)$. Usamos a notação $Y \sim PNB(\alpha, \mu, \sigma^2)$. Notamos facilmente que $E(Y) = \mu$ e $Var(Y) = \sigma^2 Var(Z)$.

2.2 Inferência

2.2.1 Função de Verossimilhança

Seja $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ uma amostra independente e identicamente distribuída (*iid*) da distribuição $PNB(\alpha, \mu, \sigma^2)$ de tamanho n . Então, a função de verossimilhança para o vetor paramétrico θ é dada por:

$$L(\theta|\mathbf{y}) = \left(\frac{\alpha c_\alpha}{\sigma}\right)^n \prod_{i=1}^n \left[\phi\left(\frac{y_i - \mu}{\sigma}\right) \right] \left\{ \prod_{i=1}^n \Phi\left(\frac{|y_i - \mu|}{\sigma}\right) \right\}^{\alpha-1}, i = 1, \dots, n.$$

2.2.2 Distribuições a Priori

Na inferência bayesiana, é sempre necessário especificar as distribuições a priori dos parâmetros de interesse. A priori reflete o conhecimento que o pesquisador tem sobre os parâmetros antes de ter conhecimento sobre a amostra. Neste trabalho, estamos interessados no vetor de parâmetros $\theta = (\alpha, \mu, \sigma^2)$ da PNB, dando um maior enfoque a α . Por isso, nesta Seção discutimos diferentes formas de especificação da distribuição a priori do parâmetro α . Discutimos sobre as distribuições a priori conjugadas, Jeffreys, gama e log-normal para α . As prioris adotadas para μ e σ^2 foram normal e gama, respectivamente.

Distribuição a Priori Conjugada para α

Uma abordagem que facilita a análise bayesiana é a utilização de prioris conjugadas. Isso porque, nesse caso, as distribuições a priori e a posteriori pertencem à mesma classe de distribuições. Visando a simplificar a análise acerca do parâmetro α , procuramos obter uma família conjugada simples para este. O procedimento para encontrar a distribuição conjugada para α no caso mais simples em que $\mu = 0$ e $\sigma = 1$ é o seguinte:

- Identificar uma classe P de distribuição para α tal que $L(\alpha|\mathbf{y})$ seja proporcional ao membro desta classe.

- Verificar se P é fechada por amostragem, isto é, $\forall p_1, p_2 \in P, \exists k$ tal que $k p_1 p_2 \in P$.

Inicialmente, precisamos encontrar a classe P de distribuição para α :

$$\begin{aligned}
L(\alpha|\mathbf{y}) &\propto \alpha c_\alpha \left\{ \prod_{i=1}^n \Phi(|y_i|) \right\}^\alpha \\
&\propto \alpha^n \left(\frac{2^\alpha}{2^\alpha - 1} \right)^n \left\{ \prod_{i=1}^n \Phi(|y_i|) \right\}^\alpha \\
&\propto \left(\frac{\alpha}{2^\alpha - 1} \right)^n \left\{ \prod_{i=1}^n 2\Phi(|y_i|) \right\}^\alpha \\
&\propto \left(\frac{\alpha}{2^\alpha - 1} \right)^n \exp \left\{ \sum_{i=1}^n 2 \log \Phi(|y_i|) \alpha \right\}
\end{aligned} \tag{2.1}$$

Assim, a classe de distribuição P para α é dada por:

$$p(\alpha|r, t) = \kappa(r, t) \left(\frac{\alpha}{2^\alpha - 1} \right)^r e^{-t\alpha}, \quad \alpha > 0, \quad r > 0, \quad t > 0, \tag{2.2}$$

em que $\kappa(r, t)^{-1} = \int_0^\infty \left(\frac{\alpha}{2^\alpha - 1} \right)^r e^{-t\alpha} d\alpha$. Pode-se mostrar que $\int_0^\infty p(\alpha|r, t) d\alpha < \infty, \forall r > 0, t > 0$, o que confirma que $p(\alpha|r, t)$ é uma fdp. Tal demonstração pode ser vista no apêndice deste material. Perceba que a expressão (2.1) é um caso particular da expressão (2.2) quando $r = n$ e $t = -\sum_{i=1}^n 2 \log \Phi(|z_i|)$. Logo, $L(\theta|\mathbf{y})$ é proporcional a $p(\alpha|r = n, t = -\sum_{i=1}^n 2 \log \Phi(|z_i|))$.

Precisamos agora mostrar que p é fechada por amostragem. Sejam $p_1(\alpha|r_1, t_1)$ e $p_2(\alpha|r_2, t_2)$ duas distribuições da família P , então:

$$p_1 p_2 \propto \left(\frac{\alpha}{2^\alpha - 1} \right)^{r_1+r_2} e^{-(t_1+t_2)\alpha} \tag{2.3}$$

O lado direito da expressão (2.3) é proporcional a $p(\alpha|r_1+r_2, t_1+t_2)$. Portanto, P é fechada por amostragem. Logo, a distribuição $p(\alpha|r, t)$ é conjugada à família potência normal bimodal.

A família de distribuições P é uma classe que ainda não foi estudada anteriormente na literatura. Além disso, esta não possui uma forma simples, uma vez que não é possível obter analiticamente a constante de normalização. Dessa forma, a distribuição a priori conjugada não conseguiu atender aos nossos objetivos, uma vez que sua função seria simplificar a análise bayesiana.

Distribuição a Priori de Jeffreys para α

Em muitas situações, o conhecimento a priori do usuário quanto ao parâmetro é simplesmente nulo. Assim, é preciso que seja utilizada uma priori que reflita esta falta de informação.

Uma priori bastante conhecida e utilizada nesses casos é a priori de Jeffreys. Esta tem propriedades importantes como o fato de ela ser invariante a transformações 1 a 1. Isso quer dizer que a distribuição a priori de um parâmetro θ é a mesma de uma transformação $g(\theta)$, sendo g uma função injetiva. O que faz sentido, uma vez que, se não temos nenhuma informação com relação a θ , também não haverá informação sobre $g(\theta)$.

A distribuição a priori de Jeffreys para o parâmetro α é dada por:

$$J(\alpha) \propto \sqrt{I(\alpha)}.$$

Sendo:

$$I(\alpha) = -E \left[\frac{\partial^2 \log(f_Y(y|\alpha))}{\partial \alpha^2} \right] = n [\alpha^{-2} - \log^2(2)2^\alpha(2^\alpha - 1)^{-2}]. \quad (2.4)$$

A igualdade acima na expressão (2.4) foi demonstrada por Bolfarine et al. (2013). Pode-se mostrar também que a distribuição a priori de Jeffreys é imprópria, ou seja, $\int_0^\infty J(\alpha)d\alpha = \infty$. Logo, esta não pode ser caracterizada como uma fdp. A demonstração de que a priori de Jeffreys é imprópria pode ser encontrada no apêndice deste material.

Distribuições a Priori Informativa para α

Prioris informativas são prioris que refletem um conhecimento com relação ao parâmetro de interesse. Estamos interessados nestas, pois o pesquisador pode dispor de conhecimentos a priori sobre α . Queremos avaliar então como podemos refletir esse conhecimento do usuário em uma distribuição de probabilidade a priori.

O parâmetro α é o parâmetro de forma da potência normal bimodal. Seu suporte está no intervalo $(0, \infty)$. Assim, a família da distribuição em questão deve obedecer a essa restrição. Com relação aos hiperparâmetros, estamos interessados em escolhê-los de acordo com o conhecimento a priori do pesquisador sobre o fenômeno estudado. Vamos supor que o usuário possa ter a priori um conhecimento fraco, moderado, ou forte de que o parâmetro de interesse esteja dentro de um intervalo (x_1, x_2) , em que $x_1, x_2 \in [0, \infty)$. Então, dada a crença do usuário, estabelecemos uma restrição da priori da seguinte maneira:

$$P(x_1 < \alpha < x_2) = 0.4 \text{ (Crença fraca)} \quad (2.5)$$

$$P(x_1 < \alpha < x_2) = 0.7 \text{ (Crença moderada)} \quad (2.6)$$

$$P(x_1 < \alpha < x_2) = 0.95 \text{ (Crença forte)} \quad (2.7)$$

As probabilidades 0.4, 0.7 e 0.95 nas expressões 2.5, 2.6 e 2.7 foram atribuídas de forma arbitrária. Caso a família da distribuição de α possua dois parâmetros, devemos utilizar mais uma

restrição para encontrá-los. Sugerimos usar uma medida de tendência central (média, mediana ou moda) como sendo o ponto médio do intervalo (x_1, x_2) estabelecido. Isso porque faz mais sentido supor que o usuário tenha mais crença em valores próximos do centro do intervalo do que em valores extremos deste. A escolha mais conveniente entre média, mediana, ou moda como ponto médio do intervalo vai depender da família adotada para a distribuição de α . Por exemplo, se a família adotada tiver tendência a ter distribuições fortemente assimétricas, sugerimos optar pela mediana como ponto médio; caso contrário, a média é mais conveniente.

Neste trabalho, utilizamos duas famílias distintas para a distribuição a priori de α : a gama, com parâmetros $a_0 > 0$ e $b_0 > 0$, e a log-normal, com parâmetros $\zeta \in \mathbb{R}$ e $\gamma > 0$. As densidades das distribuição estão dadas na Tabela 2.1 abaixo:

Distribuição	Densidade	$E(\alpha)$	$Var(\alpha)$
Gama	$\frac{b_0^{a_0}}{\Gamma(a_0)} \alpha^{a_0-1} e^{-\alpha b_0}$	$\frac{a_0}{b_0}$	$\frac{a_0}{b_0^2}$
Log-Normal	$\frac{1}{\alpha \gamma \sqrt{2\pi}} \exp\left(\frac{\log \alpha - \zeta}{\gamma}\right)^2$	$\exp\left(\zeta + \frac{\gamma^2}{2}\right)$	$\exp[2\zeta + \gamma^2] [\exp(\gamma^2) - 1]$

Tabela 2.1: Distribuições a priori informativas para α .

Denotamos de $\pi(\alpha)$ a fdp da priori de α . A escolha da distribuição gama é justificada pelo fato de ser uma distribuição bastante utilizada na análise de dados cuja variável resposta é maior que zero. A log-normal foi utilizada com outro intuito. Esta é classificada como uma distribuição de cauda pesada. A característica principal que define distribuições de cauda pesada é a de não apresentar função geradora de momentos. De acordo com Casella e Berger (2002), a distribuição gama tem função geradora de momentos e a log-normal não. Por ter cauda pesada, ao utilizar a priori log-normal, espera-se que o estimador bayesiano estime melhor α quando o usuário subestima o valor deste a priori.

Tanto a família gama como a log-normal têm dois parâmetros. Dessa forma, para obter os parâmetros de ambas, devemos utilizar de acordo com a crença do usuário, uma das equações (2.5), (2.6) e (2.7) e igualar uma medida estatística de tendência central ao ponto médio do intervalo (x_1, x_2) . Isto é, resolver o sistema:

$$\begin{cases} p = P(x_1 < \alpha < x_2) \\ E(\alpha) = m_0, Med(\alpha) = m_0 \text{ ou } Mod(\alpha) = m_0, \end{cases} \quad (2.8)$$

em que $Med(\alpha)$ e $Mod(\alpha)$ são a mediana e a moda de α , respectivamente, enquanto m_0 é o ponto médio do intervalo (x_1, x_2) , ou seja, $m_0 = \frac{x_1+x_2}{2}$.

Optamos por adotar $E(\alpha) = m_0$ no caso gama, pois a média é simples de ser calculada. Assim, seja $m_0 = E(\alpha)$, temos $a_0 = m_0 b_0$. No caso log-normal, que é uma distribuição que pertence a classe das que têm cauda pesada, optamos por esse motivo adotar $Med(\alpha) = m_0$. Assim, $\zeta = \log(m_0)$. Em algumas situações, pode ser que o sistema representado pela

equação (2.8) não apresente solução utilizando o m_0 que preferimos adotar em cada caso. Se isto acontecer, sugere-se utilizar uma das outras medidas. Caso nenhuma das outras medidas derem certo, sugere-se que o usuário controle a média e a variância no caso gama ou a mediana e a variância no caso log-normal. Neste trabalho, em todas as situações em que o sistema não teve solução com o m_0 sugerido em cada caso, o uso da moda como m_0 resolveu o problema para as duas famílias de distribuições utilizadas.

Distribuição a Priori para os parâmetros μ e σ^2

Como mencionado anteriormente, o enfoque é no parâmetro α . Por isso, utilizamos apenas uma distribuição a priori para μ e σ^2 . Poderíamos assumir para μ e σ^2 a priori de Jeffreys. Dessa forma, teríamos:

$$J_\mu(\mu) \propto k, k \in \mathbb{R}, \quad J_{\sigma^2}(\sigma^2) \propto \frac{1}{\sigma^2}.$$

A priori de Jeffreys em ambos os casos é imprópria. Apesar de que isso não impeça que a distribuição a posteriori seja própria, preferimos utilizar uma priori própria, pois teoricamente faz mais sentido trabalhar com prioris próprias. Optamos então assumir para μ uma distribuição a priori $N(\xi, \tau^2)$ e para σ^2 uma $Gama(a, b)$, cujas funções densidades são denotadas por $\pi(\mu)$ e $\pi(\sigma^2)$, respectivamente. Os hiperparâmetros dessas distribuições foram estabelecidos de modo que $\pi(\mu)$ e $\pi(\sigma^2)$ se aproximassem da distribuição a priori de jeffreys em cada caso. As funções densidades das prioris de μ e σ^2 são:

$$\pi(\mu) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \frac{(\mu - \xi)^2}{\tau^2} \right\}, \quad \mu \in \mathbb{R},$$

$$\pi(\sigma^2) = \frac{b^a}{\Gamma(a)} (\sigma^2)^{a-1} \exp \{ -\sigma^2 b \}.$$

Os hiperparâmetros adotados foram $\xi = 0$, $\tau^2 = 10^6$, $a = 10^{-3}$, $b = 10^{-6}$. Assim μ e σ^2 têm prioris não informativas e $\pi(\mu)$ e $\pi(\sigma^2)$ se aproximam de $J_\mu(\mu)$ e $J_{\sigma^2}(\sigma^2)$, respectivamente.

Distribuição a Priori Conjunta

Para fins de simplificação, a distribuição a priori considerada para o vetor de parâmetros θ é de tal forma que α , μ e σ^2 são independentes. Com isso, temos que:

$$\begin{aligned} \pi(\theta) &= \pi(\alpha, \mu, \sigma^2) \\ &= \pi(\alpha)\pi(\mu)\pi(\sigma^2) \\ &= \pi(\alpha) \frac{b^a}{\sqrt{2\pi}\Gamma(a)} (\sigma^2)^{a-1} \exp \left\{ -\frac{1}{2} \frac{(\mu - \xi)^2}{\tau^2} - b\sigma^2 \right\}. \end{aligned}$$

2.2.3 Distribuições a Posteriori

Posteriori Conjunta

Estabelecido a distribuição a priori conjunta proporcional do vetor θ , podemos obter a sua distribuição a posteriori proporcional:

$$\begin{aligned}\pi(\theta|\mathbf{y}) &= \frac{\pi(\alpha, \mu, \sigma^2)L(\alpha, \mu, \sigma^2|\mathbf{y})}{\pi(\mathbf{y})} \\ &\propto \pi(\alpha)\pi(\mu)\pi(\sigma^2)L(\alpha, \mu, \sigma|\mathbf{y}) \\ &\propto \pi(\alpha) \exp \left\{ -\frac{1}{2} \frac{(\mu - \xi)^2}{\tau^2} - \frac{1}{2} \frac{(y_i - \mu)^2}{\sigma^2} - b\sigma^2 \right\} \\ &\quad \alpha^{n-1} c_\alpha^n (\sigma^2)^{a - (\frac{n}{2} + 1)} \left\{ \prod_{i=1}^n \Phi \left(\frac{|y_i - \mu|}{\sigma} \right) \right\}^\alpha.\end{aligned}$$

A distribuição preditiva de \mathbf{y} é dada por:

$$\pi(\mathbf{y}) = \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{\infty} \pi(\alpha, \mu, \sigma^2) L(\alpha, \mu, \sigma|\mathbf{y}) d\alpha d\sigma^2 d\mu.$$

A resolução dessa integral só pode ser feita por métodos numéricos.

Distribuições a Posteriori Condicionais Completas

As distribuições a posteriori condicionais são muito utilizadas em alguns métodos de simulação estocástica para se obter as distribuições a posteriori. No método do amostrador de Gibbs é necessário conhecer as posterioris condicionais para gerar valores das posterioris. No algoritmo Métropolis Hastings, a implementação é facilitada com o conhecimento das distribuições condicionais, sem sequer ter a necessidade de conhecer as constantes de normalização destas.

A distribuição a posteriori condicional de um parâmetro é definida como sendo a distribuição deste, dado os demais parâmetros e a amostra. Abaixo, temos a densidade a posteriori condicional completa proporcional de α , μ e σ^2 :

$$\begin{aligned}\pi(\alpha|\mu, \sigma^2, \mathbf{y}) &= \frac{\pi(\alpha, \mu, \sigma^2)L(\alpha, \mu, \sigma^2|\mathbf{y})}{\pi(\mu, \sigma^2, \mathbf{y})} \\ &\propto \pi(\alpha)L(\alpha, \mu, \sigma^2|\mathbf{y}) \\ &\propto \pi(\alpha)\alpha^{n-1} c_\alpha^n \left\{ \prod_{i=1}^n \Phi \left(\frac{|y_i - \mu|}{\sigma} \right) \right\}^\alpha.\end{aligned}$$

$$\begin{aligned}\pi(\mu|\alpha, \sigma^2, \mathbf{y}) &= \frac{\pi(\alpha, \mu, \sigma^2)L(\alpha, \mu, \sigma^2|\mathbf{y})}{\pi(\alpha, \sigma^2, \mathbf{y})} \\ &\propto \pi(\mu)L(\alpha, \mu, \sigma^2|\mathbf{y}) \\ &\propto \exp \left\{ -\frac{1}{2} \frac{(\mu - \xi)^2}{\tau^2} - \frac{\sum_{i=1}^n (y_i - \mu)^2}{2\sigma^2} \right\} \left\{ \prod_{i=1}^n \Phi \left(\frac{|y_i - \mu|}{\sigma} \right) \right\}^{\alpha-1}.\end{aligned}$$

$$\begin{aligned}
\pi(\sigma^2|\alpha, \mu, \mathbf{y}) &= \frac{\pi(\alpha, \mu, \sigma^2)L(\alpha, \mu, \sigma^2|\mathbf{y})}{\pi(\alpha, \mu, \mathbf{y})} \\
&\propto \pi(\sigma^2)L(\alpha, \mu, \sigma^2|\mathbf{y}) \\
&\propto \exp\left\{-\frac{\sum_{i=1}^n (y_i - \mu)^2}{2\sigma^2} - b\sigma^2\right\} \sigma^{a-(\frac{n}{2}+1)} \left\{\prod_{i=1}^n \Phi\left(\frac{|y_i - \mu|}{\sigma}\right)\right\}^{\alpha-1}.
\end{aligned}$$

Como as distribuições a priori são independentes, o cálculo das distribuições a posteriori condicionais proporcionais é dado pelo produto da verossimilhança pela densidade a priori do parâmetro com o qual queremos obter a distribuição a posteriori condicional. Assim como não foi possível obter uma forma fechada para $\pi(\theta|\mathbf{y})$, também não foi possível encontrar uma expressão simplificada para as constantes de normalização das distribuições a posteriori condicionais.

2.3 Aspectos Computacionais

Diante do que vimos na seção anterior, não foi possível encontrar uma forma analítica para obter a distribuição preditiva da amostra. Com isso, as constantes de normalização envolvendo a distribuição a posteriori conjunta e as distribuições a posteriori condicionais não foram obtidas. A fim de solucionar esse problema, tivemos que recorrer ao método MCMC (*Markov Chain Monte Carlo*) para gerar amostras da distribuição a posteriori conjunta usando o algoritmo de Metropolis-Hastings (MH) proposto por Metropolis et al. (1953) e Hastings (1970). O algoritmo foi implementado no software estatístico R versão 3.2.1. Na seção seguinte damos uma noção introdutória do MCMC. Mais detalhes sobre o método pode ser obtido em Gamerman & Lopes (2006).

O esquema de amostragem utilizado para gerar aleatoriamente uma amostra da distribuição a posteriori conjunta foi amostrando valores das distribuições a posteriori condicionais via MH, atualizando cada parâmetro separadamente, conforme mostraremos mais adiante.

2.3.1 O Algoritmo MCMC

Métodos de Monte Carlo via Cadeias de Markov (MCMC) são métodos de simulação estocástica que consistem em gerar valores de uma cadeia de Markov que convirja para uma distribuição de equilíbrio. Essa distribuição de equilíbrio é a distribuição a posteriori no nosso caso. O método MCMC usado foi o MH. O procedimento para gerar a cadeia pelo MH consiste em escolher uma distribuição chamada de proposta cujo suporte é o mesmo da distribuição que pretendemos amostrar. A distribuição proposta deve seguir um processo de Markov, ou seja, o t -ésimo valor gerado desta dado todo o passado deve depender apenas do valor gerado na iteração $t - 1$. Cada valor amostrado da proposta pode ser aceito ou não como um valor gerado da

cadeia. Caso este não seja aceito, o novo valor amostrado da cadeia será o mesmo da iteração anterior.

Geralmente, utilizam-se métodos MCMC quando não é possível obter a constante de normalização da distribuição a posteriori. Esses métodos são capazes de obter uma amostra da posteriori mesmo quando esta não tiver uma forma fechada. Existem outros métodos que solucionam esse problema encontrado na posteriori, no entanto os métodos MCMC são mais simples de ser implementados que os demais. Para o MH precisamos escolher uma distribuição proposta relativamente simples com o suporte igual ao do parâmetro com o qual queremos amostrar.

Seja φ o parâmetro de interesse. Seja φ_t o t -ésimo valor da iteração no algoritmo MH, $q(\cdot|\varphi_t)$ a função densidade da proposta e $g(\varphi_t)$ a distribuição a posteriori de φ no ponto φ_t . Então, os passos para gerar uma amostra do Metropolis-Hastings são:

1. Inicialize o contador de iterações $t = 0$ e atribua um valor arbitrário φ_0 no suporte de φ .
2. Gere um valor φ_{t+1}^p da distribuição proposta $q(\cdot|\varphi_t)$. (φ_{t+1}^p é um valor proposto para φ_{t+1} que pode ser aceito ou rejeitado).
3. Calcule a probabilidade de aceitação $Q(\varphi_t, \varphi_{t+1}^p) = \min\left(1, \frac{g(\varphi_{t+1}^p)q(\varphi_t|\varphi_{t+1}^p)}{g(\varphi_t)q(\varphi_{t+1}^p|\varphi_t)}\right)$ e gere um valor u da distribuição $U(0, 1)$.
4. Se $u < Q(\varphi_t, \varphi_{t+1}^p)$, faça $\varphi_{t+1} = \varphi_{t+1}^p$. Caso contrário, faça $\varphi_{t+1} = \varphi_t$.
5. reinicie o procedimento a partir do segundo passo e adicione uma unidade no contador de t .

Perceba que não precisamos conhecer a constante de normalização da distribuição a posteriori para amostrar valores do algoritmo MH. A expressão da posteriori aparece apenas na razão $\frac{g(\varphi_{t+1})}{g(\varphi_t)}$. Nesse caso, a constante de normalização está no numerador e no denominador e será cancelada. Essa é uma importante propriedade do algoritmo MH. Podemos gerar valores de uma distribuição conhecendo apenas o núcleo da sua densidade.

Podemos generalizar os passos mencionados acima quando φ_t é um vetor de p dimensões. No entanto, existe uma maneira mais usual de se trabalhar com o algoritmo de MH quando queremos gerar amostras de um vetor. O procedimento consiste em gerar cada elemento do vetor separadamente. Assim, para cada elemento temos uma distribuição proposta associada. Ao invés de trabalhar com a distribuição a posteriori conjunta, trabalhamos com as posteriores condicionais. As posteriores condicionais de cada parâmetro serão atualizadas em cada iteração, de acordo com os últimos valores gerados pelo MH. O algoritmo que executa esse procedimento é conhecido como amostrador de Gibbs, passos de Metropolis.

Considere que φ é um vetor de dimensão p com elementos $\varphi_1, \varphi_2, \dots, \varphi_p$. Seja $\varphi_{t,i}$ o i -ésimo elemento do vetor gerado por MH na t -ésima iteração e $q_i(\cdot|\varphi_{t,i})$ a proposta de φ_i . Seja

também $g_i(\varphi_{t,i}|\varphi_{t,-i})$ a posteriori condicional de φ_i no ponto $\varphi_{t,i}$, em que $\varphi_{t,-i}$ é um vetor cujos elementos são os últimos valores amostrados do algoritmo MH para cada parâmetro excluindo o elemento $\varphi_{t,i}$, ou seja, $\varphi_{t,-i} = (\varphi_{t,1}, \dots, \varphi_{t,i-1}, \varphi_{t-1,i+1}, \dots, \varphi_{t-1,p})$. Assim, $g_i(\varphi_{t,i}|\varphi_{t,-i})$ é a posteriori de φ_i dado que conhecemos os demais parâmetros, cujos valores estão representados pelos últimos que foram gerados por MH.

Os passos do algoritmo agora são:

1. Inicialize o contador de iterações $t = 0$ e atribua um valor arbitrário para os $\varphi_{0,i}$ s. Faça $i = 1$.
2. Gere um valor $\varphi_{t+1,i}^p$ da distribuição proposta $q_i(\cdot|\varphi_{t,i})$.
3. Calcule a probabilidade de aceitação $Q_i(\varphi_{t,i}, \varphi_{t+1,i}^p) = \min\left(1, \frac{g_i(\varphi_{t+1,i}^p|\varphi_{t,-i})q_i(\varphi_{t,i}|\varphi_{t+1,i}^p)}{g_i(\varphi_{t,i}|\varphi_{t,-i})q_i(\varphi_{t+1,i}^p|\varphi_{t,i})}\right)$ e gere um valor u da distribuição $U(0, 1)$.
4. Se $u < Q_i(\varphi_{t,i}, \varphi_{t+1,i}^p)$, faça $\varphi_{t+1,i} = \varphi_{t+1,i}^p$. Caso contrário, faça $\varphi_{t+1,i} = \varphi_{t,i}$. Adicione uma unidade em i .
5. Reinicie o procedimento a partir do segundo passo. Se $i = p$, adicione uma unidade no contador de t e faça $i = 1$. Caso contrário, adicione uma unidade em i .

Distribuições Propostas

A distribuição proposta de um parâmetro φ no algoritmo MH deve ter o mesmo suporte de φ . Uma escolha adequada para uma distribuição proposta é aquela cuja distribuição é simples de amostrar valores e a taxa de aceitação seja em torno de 20% e 40%. Distribuições que são muito utilizadas quando $\varphi \in \mathbb{R}$ são a normal e a cauchy. Para $\varphi > 0$, a distribuição log-normal e gama são mais utilizadas. A parametrização das propostas geralmente é feita de tal modo que $E(\varphi_{t+1}^p) = \varphi_t$. Quanto à variância da proposta, devemos atribuir um valor de modo que temos uma taxa de aceitação razoável, que levará a cadeia referente à posteriori de φ a convergir mais rapidamente.

Neste trabalho, a distribuição proposta empregada para os parâmetros cujo suporte é o conjunto \mathbb{R}^+ é a distribuição gama. Para distribuições cujo suporte é o conjunto \mathbb{R} optamos por uma distribuição normal. Implementamos o algoritmo proposto por Roberts e Rosenthal (2006) para sintonizar a variância das distribuições propostas dos parâmetros de modo que a taxa de aceitação fique em torno de 20% a 40%.

Seja s_k a variância sintonizada a partir da k -ésima iteração do algoritmo de Roberts e Rosenthal (2006), temos então que este algoritmo segue os seguintes passos:

1. Inicialize o contador de iterações $k = 0$ e atribua um valor inicial s_0 para a variância da proposta.

2. Execute 50 iterações do algoritmo Metropolis-Hastings.
3. Se a taxa de aceitação dos 50 valores gerados de φ no passo anterior for maior que 40%, faça $s_{k+1} = s_k e^{\delta(k+1)}$; se não, faça $s_{k+1} = s_k e^{-\delta(k+1)}$.
4. Adicione uma unidade em k e volte para o passo 2.

Em que:

$$\delta(k+1) = \min(0.01, (k+1)^{-\frac{1}{2}}).$$

O algoritmo é encerrado quando terminar o período de aquecimento da cadeia de φ . Este período representa o tempo pelo qual a cadeia atinge a convergência. Perceba que, sempre que a taxa de aceitação de φ for maior do que a taxa desejada, aumentamos a variância da proposta; caso contrário, diminuimos. Isso acontece porque a taxa de aceitação tem uma relação inversa com a variância da proposta. Sempre que a taxa de aceitação for baixa, a variância da proposta estará maior do que deveria estar e vice-versa.

Abaixo, seguem as propostas consideradas para os parâmetros α , μ , σ^2 . A variância destas propostas foram sintonizadas usando o algoritmo de Roberts e Rosenthal (2006):

$$q_\alpha(\alpha_{t+1}^p | \alpha_t) \sim \text{Gama} \left(\frac{\alpha_t}{s_k^{(1)}}, \frac{1}{s_k^{(1)}} \right), \quad (2.9)$$

$$q_\mu(\mu_{t+1}^p | \mu_t) \sim N(\mu_t, s_k^{(2)}), \quad (2.10)$$

$$q_{\sigma^2}((\sigma^2)_{t+1}^p | \sigma_t^2) \sim \text{Gama} \left(\frac{\sigma_t^2}{s_k^{(3)}}, \frac{1}{s_k^{(3)}} \right), \quad (2.11)$$

em que:

- α_t^p , μ_t^p e $(\sigma^2)_t^p$ são os t -ésimos valores amostrados das distribuições propostas de α , μ e σ^2 , respectivamente.
- α_t , μ_t e σ_t^2 são os valores amostrados das distribuições a posteriori de α , μ , σ^2 , respectivamente, através do algoritmo MH na t -ésima iteração.
- $s_k^{(1)}$, $s_k^{(2)}$, $s_k^{(3)}$ são os sintonizadores da variância das propostas de α , μ e σ^2 , respectivamente, em que $k = \lceil \frac{t}{50} \rceil$, sendo $\lceil x \rceil$ a função maior inteiro, ou seja, $\lceil x \rceil = \max \{n \in \mathbb{Z}; n \leq x\}$. A quantidade t representa a iteração atual do algoritmo MH e k a iteração atual do algoritmo de Roberts e Rosenthal.

Adotamos para o índice k a função maior inteiro, pois só depois de 50 iterações do algoritmo MH, incrementamos um contador $k+1$ no algoritmo de Roberts e Rosenthal. Então, por exemplo, se $t = 75$, temos que $k = \lceil \frac{75}{50} \rceil = 2$, dado que o contador só foi atualizado uma vez

no momento em que t atingiu o valor 50. Perceba que o que a função maior inteiro faz nesse caso é justamente aumentar uma unidade em k toda vez que t atingir um múltiplo de 50.

Os parâmetros da proposta de α foram escolhidos de tal forma que $E(\alpha_{t+1}^p | \alpha_t) = \alpha_t$ e $Var(\alpha_{t+1}^p | \alpha_t) = \alpha_t s_k^{(1)}$. Utilizando a mesma lógica, definimos a parametrização da proposta de σ^2 . quanto a proposta μ_{t+1}^p , adotamos uma parametrização de tal modo que $E(\mu_{t+1}^p | \mu_t) = \mu_t$ e $Var(\mu_{t+1}^p | \mu_t) = s_t^{(2)}$. Na situação em que a posteriori concentra muita probabilidade em α pequeno, sugere-se parametrizar α_{t+1}^p de modo que α_t seja a moda da distribuição. Se isto for feito, a cadeia do MCMC terá a autocorrelação reduzida.

Valor Inicial para os Parâmetros e para a Variância Sintonizada

Para a cadeia de Markov gerada pelo algoritmo MH convergir rapidamente para a distribuição de equilíbrio (distribuição convergente da cadeia, que neste caso é a distribuição a posteriori) é preciso escolher bem os valores iniciais α_0 , μ_0 e σ_0^2 pelos quais a cadeia se inicializará. Além disso, precisamos também escolher bem os valores iniciais $s_0^{(1)}$, $s_0^{(2)}$ e $s_0^{(3)}$ para permitir que a variância das distribuições propostas sintonizem rapidamente. Com bons valores iniciais nos dois casos garantimos que o período de aquecimento da cadeia seja curto. Abaixo, temos a Tabela 2.2 que auxilia na escolha de um bom ponto inicial.

Parâmetro	valor inicial	$s_0^{(i)}$	valor inicial
α_0	$E(\alpha)$	$s_0^{(1)}$	$\sqrt{2}$
μ_0	$\hat{\mu}$	$s_0^{(2)}$	$\hat{\sigma}^2$
σ_0^2	$\hat{\sigma}^2$	$s_0^{(3)}$	$\frac{\hat{\sigma}^2}{(n-1)}$

Tabela 2.2: Valores iniciais para α_0 , μ_0 , σ_0^2 e os $s_0^{(i)}$ s.

Em que:

$$\hat{\mu} = \frac{\sum_{i=1}^n y_i}{n}, \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \mu)^2}{n-1}.$$

O valor inicial apresentado na Tabela 2.2 para α_0 foi baseado no conhecimento a priori que temos sobre este. Por isso, utilizamos o valor esperado da priori de α como valor inicial. No caso de μ_0 e σ_0^2 , foi pensado na estimativa de máxima verossilhança de um modelo normalmente distribuído. Com respeito aos $s_0^{(i)}$ s, pensamos em escolhê-los de tal forma que a variância de cada proposta fosse equivalente à variância de um estimador do parâmetro que se quer amostrar a partir daquela proposta. Por exemplo, um estimador para μ é o $\hat{\mu}$. Então, queremos escolher $s_0^{(2)}$ de maneira que $Var(\hat{\mu}) = Var(\mu_{t+1}^p | \mu_t)$. No entanto, não conhecemos $Var(\hat{\mu})$. Então, o estimamos por:

$$\widehat{Var}(\hat{\mu}) = \hat{\sigma}^2.$$

Assim, o valor inicial para $s_0^{(2)}$ será:

$$\widehat{Var}(\hat{\mu}) = Var(\mu_{t+1}^p | \mu_t) \Rightarrow s_0^{(2)} = \hat{\sigma}^2$$

Da mesma forma, o valor atribuído a $s_0^{(3)}$ deveria ser de tal modo que $\widehat{Var}(\hat{\sigma}^2) = Var((\sigma^2)_{t+1}^p | \sigma_t^2)$. Como não conhecemos $Var(\hat{\sigma}^2)$, o estimamos por $\widehat{Var}(\hat{\sigma}^2) = \frac{\hat{\sigma}^4}{n-1}$. Assim:

$$\widehat{Var}(\hat{\sigma}^2) = Var((\sigma^2)_{t+1}^p | \sigma_t^2) \Rightarrow \frac{\hat{\sigma}^4}{n-1} = \sigma_t^2 s_0^{(3)}.$$

Substituindo σ_t^2 pelo $\hat{\sigma}^2$ temos que:

$$s_0^{(3)} = \frac{\hat{\sigma}^2}{n-1}.$$

Obviamente, os estimadores de μ e σ^2 escolhidos têm algumas limitações, pois eles são estimadores adequados quando estamos trabalhando com uma distribuição normal, o que não é o caso. Porém, como estamos utilizando esses estimadores com um intuito de apenas propor um valor inicial, não há muita perda em considerá-los para esse propósito.

Para $s_0^{(1)}$, não o atribuímos usando a mesma lógica de $s_0^{(2)}$ e $s_0^{(3)}$. Testamos algumas situações através de simulações e vimos que o $s^{(1)}$ ótimo ($s_k^{(1)}$ que torna o valor esperado da taxa de aceitação 40%) varia geralmente em torno de 0.1 e 20 para valores de α entre 0 e 20 (α geralmente se encontra nesse intervalo na prática). Com isso, devemos atribuir a $s_0^{(1)}$ um valor entre 0.1 e 20. Pensamos então em encontrar o $s_0^{(1)}$ de tal modo que o tempo mínimo de convergência da taxa de aceitação seja o mesmo quando $s^{(1)}$ ótimo é 0.1 e quando $s^{(1)}$ ótimo é 20.

Seja o índice k_{\min} o menor tempo possível para que $s_k^{(1)}$ difira de $s^{(1)}$ em menos de ϵ unidades, sendo $\epsilon > 0$. Então, o $s_0^{(1)}$ desejado é aquele que resolve o sistema abaixo:

$$\begin{cases} 0.1 - \epsilon < s_0^{(1)} e^{\sum_{i=1}^{k_{\min}} \delta(i)} < 0.1 + \epsilon \\ 20 - \epsilon < s_0^{(1)} e^{-\sum_{i=1}^{k_{\min}} \delta(i)} < 20 + \epsilon \end{cases} \quad (2.12)$$

Multiplicando as duas inequações temos que:

$$(0.1 - \epsilon)(20 - \epsilon) < \left[s_0^{(1)} \right]^2 < (0.1 + \epsilon)(20 + \epsilon)$$

$$2 - 20.1\epsilon + \epsilon^2 < \left[s_0^{(1)} \right]^2 < 2 + 20.1\epsilon + \epsilon^2$$

Supondo que ϵ seja um valor bem próximo de 0, é fácil ver que $s_0^{(1)} = \sqrt{2}$ satisfaz a inequação acima. Logo, um bom valor inicial para $s_0^{(1)}$ é $\sqrt{2}$.

A expressão do meio da primeira inequação do sistema (2.12) representa a taxa de aceitação após k_{\min} iterações, caso todas as taxas nas k etapas sejam maiores do que 40%. Quanto à

expressão do meio da segunda inequação, temos a taxa de aceitação caso todas as taxas sejam menores que 40%. Obviamente, em algumas etapas a taxa pode ser maior ou menor que 40%. No entanto, estamos supondo que o tempo de convergência é o menor possível. Assim, $s^{(1)}$ ótimo estando no intervalo $(0.1, 20)$, $s_0^{(1)}$ deve também estar, e estando nele, caso $s^{(1)}$ ótimo seja 0.1, o tempo mínimo de convergência da taxa de aceitação é se todas as k_{\min} taxas forem maiores que 40%. Isso porque uma taxa maior que 40% na i -ésima etapa implica que $s_i^{(1)} < s_{i-1}^{(1)}$, o que vai ajudar na convergência de $s^{(1)}$, pois $s_{i-1}^{(1)} > 0.1$, caso a convergência ainda não tenha sido atingida. Usando o mesmo raciocínio, concluímos que caso o $s_k^{(1)}$ ótimo seja 20, o tempo mínimo de convergência da taxa de aceitação é se todas as k taxas forem menores que 40%.

2.3.2 Diagnóstico de Convergência

Em processos inferenciais via MCMC, só podemos considerar que os valores simulados foram gerados da distribuição a posteriori se a cadeia convergir. Uma forma simples para verificar se uma cadeia converge é gerar amostras do algoritmo MCMC inicializando em diferentes pontos e checar se as trajetórias das cadeias não diferem entre si. Há diversos critérios de diagnóstico de convergência. Neste trabalho, o critério utilizado foi o de Gelman e Rubin (1992).

Critério de Gelman e Rubin

O critério de Gelman e Rubin consiste em testar se m cadeias de Markov convergem para o mesmo ponto através de um valor escalar \hat{R} (redução potencial da escala). Esse critério usa uma idéia similar ao da ANOVA. Se a dispersão entre as cadeias for maior do que a dispersão dentro das cadeias, então pelo menos uma das cadeias se comporta diferente das demais. Se isso acontece é porque o histograma das cadeias como um todo apresenta uma maior variabilidade quando comparado com o histograma de cada cadeia individualmente.

Considere m cadeias de Markov de tamanho n cada uma das quais deveriam gerar valores da distribuição a posteriori de um parâmetro φ . Seja $\varphi_{i,j}$ o i -ésimo elemento da j -ésima cadeia, $\bar{\varphi}_i$ a média da i -ésima cadeia e $\bar{\varphi}$ a média geral de todas as cadeias. A variância dentro das cadeias (W) e a variância entre cadeias (B) são dadas por:

$$W = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1}^n (\varphi_{i,j} - \bar{\varphi}_i)^2, \quad B = \frac{n}{m-1} \sum_{i=1}^m (\bar{\varphi}_i - \bar{\varphi})^2.$$

Sob condições de convergência, as $m \times n$ observações das m cadeias serão gerados da distribuição a posteriori de φ . Sendo assim, um estimador não viesado da variância dessa distribuição é dado por:

$$\widehat{Var}(\varphi) = \left(1 - \frac{1}{n}\right) W + \frac{1}{n} B.$$

Se as cadeias estão convergindo, então $\widehat{Var}(\varphi)$ e W devem estar bem próximos, uma vez

que sob condição de convergência ambos estão estimando $Var(\varphi)$. No entanto, quando tal condição não é satisfeita, $\widehat{Var}(\varphi)$ tende a aumentar, pois o efeito da variabilidade entre as cadeias agora existe e contribui para o aumento da dispersão das observações. A quantidade \hat{R} mede o quanto maior é $\widehat{Var}(\varphi)$ com relação a W . Portanto, \hat{R} é uma boa medida para decidir se as cadeias convergem para o mesmo ponto ou não. A fórmula deste escalar é dada por:

$$\hat{R} = \sqrt{\frac{\widehat{Var}(\varphi)}{W}}.$$

Esse escalar sempre é maior do que 1 e a convergência da cadeia será garantida quando \hat{R} se aproxima de 1. Gelman e Rubin (1992) sugerem aceitar a convergência quando $\hat{R} \leq 1.2$.

Resultados da Análise de Convergência

A maior preocupação que temos quanto à convergência do algoritmo de MH é especialmente em dados cuja distribuição empírica nos leve a crer em uma distribuição unimodal. Nessa situação, às vezes a cadeia de Markov se comporta um pouco diferente do esperado. Em certos períodos de tempo, quando a cadeia atinge um valor bem próximo de zero, a mesma não se movimenta. O problema se agrava um pouco mais se a priori der muita informação para α pequeno. Com isso em mente, para o diagnóstico de convergência adotamos uma priori da família gama e log-normal com crença forte em $\alpha \in (0, 1)$. O ponto médio do intervalo é a média no caso gama e a mediana no caso log-normal. A parametrização dos modelos a priori é tal que $\alpha \sim \text{Gama}(0.018, 0.037)$ e $\alpha \sim \text{Log} - \text{Normal}(-0.693, 0.421)$. O procedimento utilizado para encontrar os hiperparâmetros das prioris pode ser encontrado na página 29. Adotamos também a priori de Jeffreys. Consideramos para a análise de convergência uma amostra de tamanho $n = 50$ da distribuição $PNB(0.1, 0, 1)$. Geramos, para cada priori, três amostras do algoritmo MH de tamanho $N = 50000$, inicializando, cada uma, em um ponto distinto conforme mostrado na Tabela 2.3 abaixo:

amostra	α_0	μ_0	σ_0^2
1	5	5	5
2	20	20	20
3	100	100	100

Tabela 2.3: Valores iniciais para o teste de convergência

O Diagnóstico de convergência de Gelman-Rubin apontou que todas as cadeias convergem, pois os valores de \hat{R} observados foram menores que 1.2, conforme mostrado na Tabela 2.4. Apesar disso, a Figura 2.3 ilustra que, no caso gama, após ser encontrada a região de convergência, ocorreu o problema mencionado anteriormente. Em um dado momento, a cadeia gera valores próximos de 1 aleatoriamente, e dado outro momento ela gera sucessivos valores iguais, que se aproximam de zero. No caso log-normal, a cadeia se comporta dentro do esperado, com

média e variância constante. Pensando no problema identificado, propomos outra priori gama para α . Ao invés de adotar a média como ponto médio do intervalo (0,1), adotamos a moda. A priori é de tal modo que $\alpha \sim \text{Gama}(6.768, 11.537)$. O comportamento da cadeia quando foi utilizado a priori gama com $Mod(\alpha) = 0.5$ ocorreu dentro do esperado. A Figura 2.2 ilustra as curvas das prioris. Podemos perceber que a distribuição gama cuja cadeia deu problemas com valores próximos de 0 concentra bastante probabilidade próxima a este valor.

Além da priori gama com $E(\alpha) = 0.5$ ter tido como consequência uma amostra do MCMC muito correlacionada, esta é muito assimétrica e não distribui bem os pesos no intervalo (0,1). Ela seria mais adequada quando o usuário tivesse muita crença de que α é muito próximo de 0. Para termos uma noção do quanto a assimetria é forte à direita, apesar de que $E(\alpha) = 0.5$, temos $P(\alpha > 0.5) = 0.93$. De qualquer maneira, esse exemplo nos mostrou que devemos estar atentos ao estabelecer um intervalo, o nível da crença e a medida de centralidade como ponto médio de x_1 e x_2 . Devemos analisar como a priori está distribuindo os pesos no seu suporte em diferentes intervalos, crenças e medidas de centralidade.

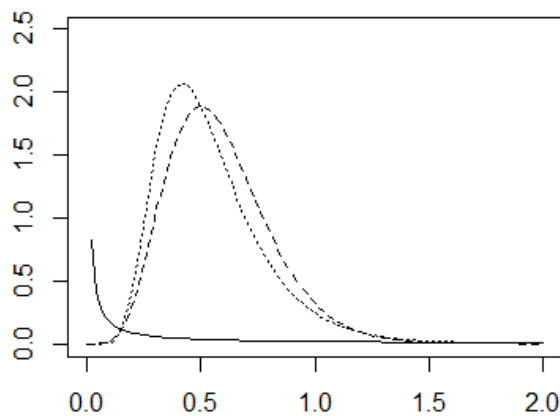


Figura 2.2: Distribuições a priori para α . $\alpha \sim \text{Gama}(0.018, 0.037)$ (linha sólida). $\alpha \sim \text{Gama}(6.768, 11.537)$ (linha tracejada). $\alpha \sim \text{Log} - \text{Normal}(-0.693, 0.421)$ (linha pontilhada)

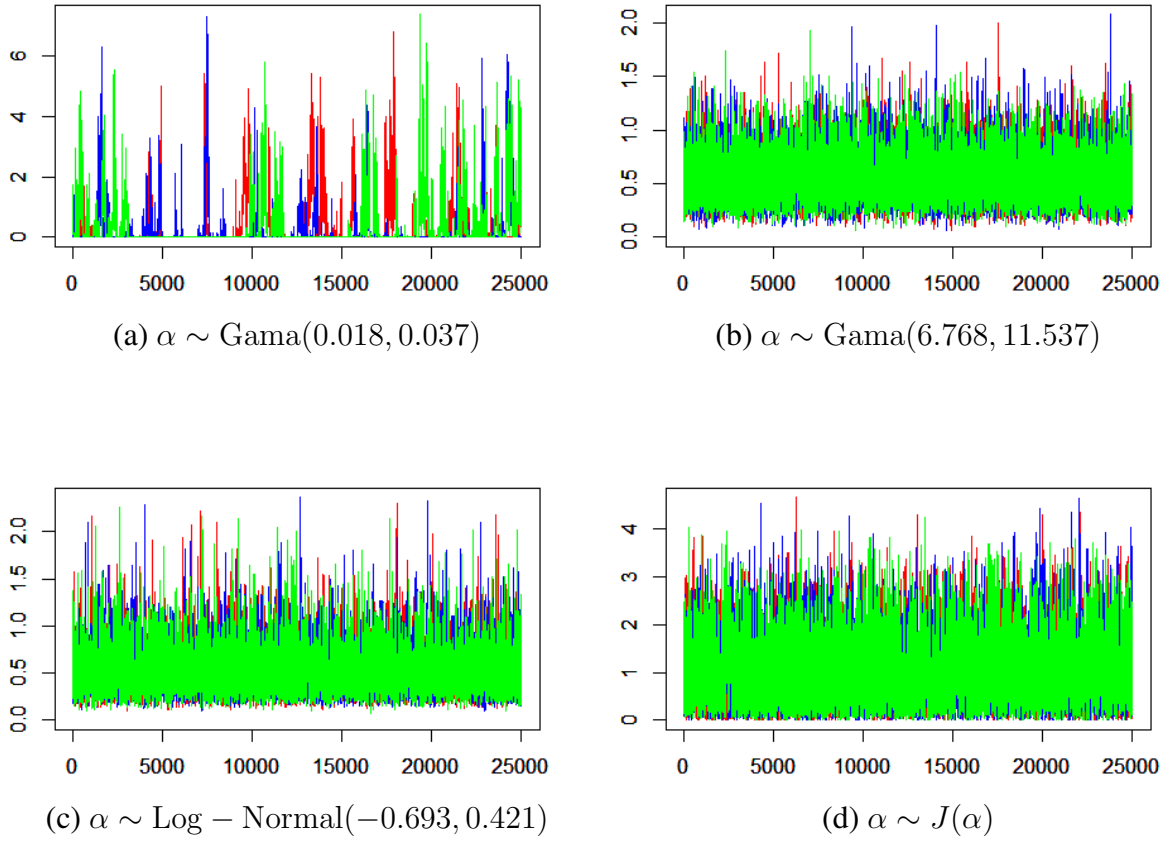


Figura 2.3: Cadeia gerada pelo MCMC da distribuição a posteriori de α a partir de uma amostra $PNB(0.1, 0, 1)$ de tamanho $n = 50$ iniciando em diferentes pontos.

Parâmetro	Priori utilizada			
	Gama(0.018, 0.037)	Gama(6.768, 11.537)	Log - Normal(-0.693, 0.421)	Jeffreys
α	1 (1.01)	1 (1)	1 (1)	1 (1.01)
μ	1 (1)	1 (1)	1 (1)	1 (1)
σ^2	1 (1)	1 (1)	1 (1)	1 (1)

Tabela 2.4: \hat{R} estimado em cada cenário. Estimativa pontual (fora dos parenteses). Limite superior do intervalo de confiança (entre parenteses).

2.4 Estudo de Simulação

2.4.1 Análise de Sensibilidade

Como mencionado na seção anterior, consideramos três possíveis crenças do usuário com respeito à informação a priori do parâmetro α dentro de um intervalo (x_1, x_2) dado. Nesta seção, utilizamos prioris favoráveis a modelos unimodais, prioris favoráveis a modelos bimodais e prioris que não favorecem nenhum dos dois modelos. Adotamos para cada uma dessas três crenças de modelo as crenças fraca, moderada e forte, tanto no caso gama como no caso

log-normal. Foi avaliado o DIC (*Deviance Information Criterion*) dos modelos bayesianos em amostras geradas a partir de um modelo unimodal e em amostras geradas a partir de um modelo bimodal. Esse critério foi proposto por Spiegelhalter et al. (2002) e utiliza uma medida bayesiana de ajuste e uma medida de penalidade pela complexidade do modelo para avaliá-lo. Dentre os modelos, aquele que apresentar o menor DIC é o mais adequado para se fazer inferência. Dessa forma, podemos avaliar o desempenho das posterioris em cenários favoráveis e desfavoráveis às prioris escolhidas para cada possibilidade do nível de certeza do usuário. Por exemplo, podemos analisar o estimador bayesiano quando a amostra é gerada a partir de uma distribuição $PNB(10, 0, 1)$ (bimodal), quando o usuário crer de fato que o modelo seja unimodal, ou vice-versa. Podemos analisar também o estimador bayesiano na situação em que o usuário cre a priori em um modelo que coincidentemente se aproxima do real, seja ele unimodal ou bimodal. No entanto, o foco principal do DIC nesta dissertação é avaliar o efeito da variação dos hiperparâmetros de cada distribuição separadamente na distribuição marginal a posteriori de α . Todavia, o que o DIC avaliou neste trabalho foi o efeito da variação dos hiperparâmetros na distribuição conjunta do vetor θ .

DIC (Deviance Information Criterion)

O DIC é uma medida bayesiana que avalia a qualidade do ajuste a partir dos desvios preditivos do modelo acrescentado de uma medida de penalidade, como já foi mencionado. Este foi proposto por Spiegelhalter et al. (2002) e é simples de se utilizar quando a distribuição a posteriori for obtida por algum método de simulação estocástico.

Seja y o vetor de observações da amostra, φ o parâmetro de interesse e $D(\varphi|y)$ o desvio do modelo, dada por $D(\varphi|y) = -2 \log(L(\varphi|y))$, sendo $L(\varphi|y)$ a função de verossimilhança. Então, a medida de qualidade do ajuste é dada por:

$$D(y) = E_{\varphi}(D(\varphi|y)).$$

A quantidade $D(y)$ nada mais é do que o desvio preditivo. Podemos estimar $D(y)$ utilizando simulações da posteriori de φ a partir de uma amostra $\varphi_1, \varphi_2, \dots, \varphi_m$:

$$\hat{D}(y) = \frac{1}{m} \sum_{i=1}^m D(y, \varphi_i).$$

O critério DIC também leva em conta uma medida de penalidade. Esta quantifica o efeito do número de parâmetros no modelo bayesiano e é dada por:

$$p_D = \hat{D}(y) - D(\hat{\varphi}|y),$$

em que $\hat{\varphi}$ é a média a posteriori de φ a qual pode ser obtida a partir da amostra do MCMC.

A quantidade p_D , grosseiramente falando, representa o número de parâmetros no modelo. No entanto, p_D é influenciado também pela informação a priori que temos sobre os parâmetros.

Quanto maior for a informação a priori, menor será p_D .

Tendo conhecimento de como se obtêm o desvio preditivo e a medida de complexidade do modelo p_D , temos então que o DIC é dado por:

$$DIC(\varphi) = \hat{D}(y) + p_D.$$

Deviance pequena e pouca complexidade do modelo são sinais de que o ajuste é bom. Portanto, quanto menor for o DIC do modelo, melhor é o ajuste bayesiano.

BIC (Critério de Informação Bayesiano)

O critério de informação bayesiano (BIC) foi proposto por Schwarz (1978) e avalia, assim como o DIC, a qualidade do ajuste. O cálculo do BIC é dada por:

$$BIC = \hat{D}(y) - q \log(n),$$

em que q é o número de parâmetros do modelo e n o tamanho da amostra. Note que a fórmula do BIC tem uma característica semelhante comparada a fórmula do DIC. Ela utiliza de uma medida de qualidade do ajuste, que no caso é a mesma utilizada no DIC, dada pela deviance $\hat{D}(y)$. E utiliza também uma medida de penalidade dada por $q \log(n)$.

O BIC pode ser utilizado ao invés do DIC quando o número efetivo de parâmetros calculado pelo DIC é negativo. Nestas situações não é recomendável utilizar o DIC. Este tipo de problema geralmente é encontrado quando a distribuição a posteriori dos parâmetros é bimodal ou assimétrica. No trabalho aqui apresentado, utilizamos o DIC para comparar modelos sempre que o p_D foi positivo e próximo do número de parâmetros. Quando isto não ocorreu, optamos por utilizar o BIC.

Prioris Utilizadas

As prioris adotadas na análise de sensibilidade estão representadas na Figura 2.4, e também na Tabela 2.6, onde na primeira e segunda colunas estão os intervalos em que o usuário crê que α esteja contido e o nível da crença, respectivamente. O intervalo $(0, 1)$ para α se refere a uma crença maior do usuário em modelos unimodais. O intervalo $(0, 2)$ mostra que o usuário tem uma crença indecisa entre um modelo unimodal ou fracamente bimodal. Já o intervalo $(2, 15)$ representa uma crença favorável a modelos bimodais. Ou seja, crer em $\alpha \in (0, 1)$ é uma crença unimodal, em $\alpha \in (0, 2)$ é uma crença indecisa, e em $\alpha \in (2, 15)$ é uma crença bimodal. Para cada intervalo e crença, temos uma priori gama e log-normal ajustada, suas propriedades estão apresentadas nas colunas 3 a 8 da Tabela 2.6. Nas colunas 3 e 6, que se referem à medida utilizada como ponto médio das famílias empregadas, podemos observar que utilizamos tanto a média, a mediana, como a moda. Na família gama, optamos por sempre tentar primeiramente utilizar a média. Entretanto, é impossível modelar a gama adotando como ponto médio a média

do intervalo $(0, 1)$ e $(0, 2)$ com crenças fracas e moderadas. A distribuição gama não é flexível o suficiente para atender a essas restrições. Para modelos unimodais ou indecisos com crença forte, a média pode ser o ponto médio. Todavia, nessas situações acontece algo semelhante ao que vimos na seção de análise de convergência. Teríamos uma priori que distribuiria os pesos no intervalo adotado de maneira muito desproporcional. Por esse motivo, preferimos nessa situação usar a moda como ponto médio. Com relação à família log-normal, a mediana foi a medida escolhida como prioridade a ser o ponto médio do intervalo. A razão disso é que a log-normal é uma distribuição que assume geralmente níveis de assimetria elevados. A mediana só não é possível de ser o ponto médio do intervalo quando a crença do usuário é fraca para modelos unimodais ou indecisos.

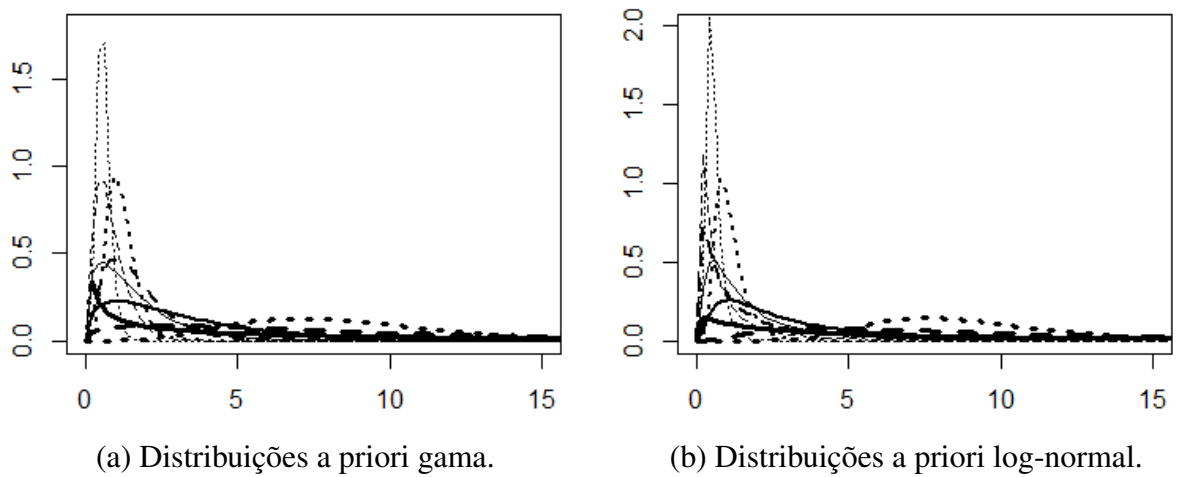


Figura 2.4: Distribuições a priori de α

Crença / Intervalo	(0,1)	(0,2)	(2,15)
Fraca	Sólida Fina	Sólida Média	Sólida Grossa
Moderada	Tracejada Fina	Tracejada Média	Tracejada Grossa
Forte	Pontilhada Fina	Pontilhada Média	Pontilhada Grossa

Tabela 2.5: Linhas das curvas das priors da Figura 2.4.

Intervalo	Crença	Ponto médio (Gama)	Hiperparâmetros (Gama)	Variância (Gama)	Ponto médio (Log-Normal)	Hiperparâmetros (Log-Normal)	Variância (Log-Normal)
(0, 1)	Fraca	Moda	$a_0 = 1.442$ $b_0 = 0.884$	1.845	Moda	$\zeta = 0.246$ $\gamma = 0.969$	6.511
(0, 1)	Moderada	Moda	$a_0 = 2.539$ $b_0 = 3.078$	0.268	Mediana	$\zeta = -0.693$ $\gamma = 1.322$	6.805
(0, 1)	Forte	Moda	$a_0 = 6.768$ $b_0 = 11.537$	0.051	Mediana	$\zeta = -0.693$ $\gamma = 0.421$	0.058
(0, 2)	Fraca	Moda	$a_0 = 1.442$ $b_0 = 0.442$	7.381	Moda	$\zeta = 0.939$ $\gamma = 0.969$	26.046
(0, 2)	Moderada	Moda	$a_0 = 2.539$ $b_0 = 1.539$	1.072	Mediana	$\zeta = 0$ $\gamma = 1.322$	27.221
(0, 2)	Forte	Moda	$a_0 = 6.768$ $b_0 = 5.768$	0.203	Mediana	$\zeta = 0$ $\gamma = 0.421$	0.232
(2, 15)	Fraca	Média	$a_0 = 0.417$ $b_0 = 0.049$	173.469	Mediana	$\zeta = 2.140$ $\gamma = 1.869$	75779.108
(2, 15)	Moderada	Média	$a_0 = 1.377$ $b_0 = 0.162$	52.469	Mediana	$\zeta = 2.140$ $\gamma = 0.858$	164.113
(2, 15)	Forte	Média	$a_0 = 6.086$ $b_0 = 0.716$	11.872	Mediana	$\zeta = 2.140$ $\gamma = 0.345$	10.287

Tabela 2.6: Prioris adotadas para a análise de sensibilidade.

Procedimento Numérico para Obter os Hiperparâmetros

Conforme foi visto na Seção 2.2.2, para encontrar os hiperparâmetros da distribuição a priori dada a crença do usuário, precisamos resolver o sistema:

$$\begin{cases} p = P(x_1 < \alpha < x_2) \\ E(\alpha) = m_0, Med(\alpha) = m_0 \text{ ou } Mod(\alpha) = m_0, \end{cases}$$

em que $Med(\alpha)$ e $Mod(\alpha)$ é a mediana e a moda de α , respectivamente, enquanto m_0 é o ponto médio do intervalo (x_1, x_2) , ou seja, $m_0 = \frac{x_1+x_2}{2}$.

Suponha que a distribuição a priori $\pi(\alpha)$ seja tal que $\alpha \sim \text{Gama}(a_0, b_0)$. Para encontrarmos os hiperparâmetros a_0 e b_0 , propomos um algoritmo que calcula a_0 a partir de uma função $f_1(b_0)$. Como visto anteriormente na Tabela 2.6, no caso gama usamos $E(\alpha) = m_0$ ou $Mod(\alpha) = m_0$. Assim, temos:

$$f_1(b_0) = \begin{cases} a_0 = m_0 b_0, \text{ se } E(\alpha) = m_0 \\ a_0 = m_0 b_0 + 1, \text{ se } Mod(\alpha) = m_0. \end{cases}$$

Propomos então uma função $h(x)$ tal que:

$$h_1(x) = P(\alpha < x_2 | a_0 = f_1(x), b_0 = x) - P(\alpha < x_1 | a_0 = f_1(x), b_0 = x) - p.$$

Note que em $h_1(x)$, o valor de x representa a escolha de b_0 e a função retorna o erro da

probabilidade de α estar contido no intervalo (x_1, x_2) em relação à probabilidade da crença do usuário. Obviamente, queremos $h_1(x) = 0$. Pensando nisso, inicializamos um contador c_1 a partir de $c_0 = 0.001$ com incremento de 0.005 , ou seja, $c_i = c_{i-1} + 0.005$. Decidimos encerrar o contador quando é encontrado um número inteiro j tal que $h_1(c_j) \times h_1(c_{j+1}) < 0$, o que significa que o valor de x que satisfaz $h_1(x) = 0$ está no intervalo (c_j, c_{j+1}) . Após encerrado o contador, escolhemos como b_0 o valor do vetor $(c_j, c_j + 0.001, \dots, c_j + 0.004, c_{j+1})$ que fornece o menor valor absoluto de $h_1(x)$, ou seja, o menor erro.

Considerando α de tal modo que $\alpha \sim \text{Log} - \text{Normal}(\zeta, \gamma)$, o procedimento é análogo. Podemos escrever ζ a partir de uma função $f_2(\zeta)$ da seguinte maneira:

$$f_2(\zeta) = \begin{cases} \zeta = \log(m_0), & \text{se } \text{Med}(\alpha) = m_0 \\ \zeta = \log(m_0) + \gamma^2, & \text{se } \text{Mod}(\alpha) = m_0, \end{cases}$$

e o erro associado à probabilidade da crença do usuário:

$$\begin{aligned} h_2(x) &= P(\alpha < x_2 | \zeta = f_2(x), \gamma = x) - P(\alpha < x_1 | \zeta = f_2(x), \gamma = x) - p \\ &= \phi\left(\frac{x_2 - f_2(x)}{x}\right) - \phi\left(\frac{x_1 - f_2(x)}{x}\right) - p. \end{aligned}$$

Simulação

Consideramos amostras geradas a partir da $PNB(1, 0, 1)$ (unimodal) e $PNB(10, 0, 1)$ (fortemente bimodal) com os tamanhos $n = 50$ (amostra pequena), $n = 100$ (amostra média), $n = 200$ (amostra grande), como pode ser observado na Tabela 2.7. Como já foi informado, foram utilizadas todas as distribuições a priori da Tabela 2.6 e calculado o DIC para todos os ajustes. Segundo Spiegelhalter (2002), se a diferença entre DICs de dois ajustes for maior que 5, isso indica que eles diferem substancialmente entre si. Assim, podemos avaliar se o efeito de mudar uma crença de um usuário altera de forma significativa o ajuste. No entanto, como já foi mencionado, o ajuste que queremos avaliar é o da posteriori marginal de α , e não da conjunta θ . Então, uma análise gráfica das distribuições a posteriori marginal de cada parâmetro foi feita, a fim de reforçar a análise acerca dos ajustes para α . O que desejamos observar nessa simulação é obviamente que todos os ajustes sejam semelhantes e produzam DICs baixos. Todavia, é de se esperar que os ajustes provenientes de uma crença “ruim” sobre α sejam diferentes.

n	Amostra	n	Amostra
50	$PNB(\alpha = 1, \mu = 0, \sigma^2 = 1)$	50	$PNB(\alpha = 10, \mu = 0, \sigma^2 = 1)$
100	$PNB(\alpha = 1, \mu = 0, \sigma^2 = 1)$	100	$PNB(\alpha = 10, \mu = 0, \sigma^2 = 1)$
200	$PNB(\alpha = 1, \mu = 0, \sigma^2 = 1)$	200	$PNB(\alpha = 10, \mu = 0, \sigma^2 = 1)$

Tabela 2.7: Cenários para a análise de sensibilidade.

Através da Tabela 2.9 e das Figuras 2.6 e 2.7, podemos perceber que modelos cujos DICs di-

ferem em mais de 5 unidades, realmente apresentam comportamentos bastante distintos quanto ao $\pi(\alpha|\mathbf{y})$. No entanto, em alguns casos em que a diferença do DIC de dois modelos é menor do que 5, as curvas das densidades marginais $\pi(\alpha|\mathbf{y})$ em ambos os modelos parecem ainda ser diferentes. Para entender esse fato, lembremos que o que afeta o ajuste é o fato de que a priori utilizada para α é distinta. Porém, as priors de μ e σ^2 são as mesmas. Isso contribui para que os modelos apresentem posterioris marginais de μ e σ^2 próximas. Assim, espera-se que ao estimar a diferença do DIC dos modelos marginais de α utilizando a diferença do DIC dos modelos globais, estamos subestimando a diferença do DIC do modelo marginal de α . Isso significa que quando afirmamos que dois modelos marginais de α diferem a partir do DIC do modelo global, é porque eles realmente diferem. Entretanto, pode ocorrer às vezes de encontrarmos DIC menores que 5, e decidirmos que os modelos marginais de α são parecidos, quando na realidade eles apresentam comportamentos bem distintos. Para esclarecer o que foi dito, observe a Figura 2.5: temos a marginal $\pi(\alpha|\mathbf{y})$, $\pi(\mu|\mathbf{y})$ e $\pi(\sigma^2|\mathbf{y})$ de duas posterioris conjuntas distintas. A diferença entre o DIC dos modelos globais foi 3.21. No entanto, acredita-se que a diferença do DIC dos modelos marginais de α seja maior que 3.21, pois o fato de as posterioris marginais $\pi(\mu|\mathbf{y})$ e $\pi(\sigma^2|\mathbf{y})$ serem mais próximas que as posterioris marginais $\pi(\alpha|\mathbf{y})$ faz com que as posterioris conjuntas $\pi(\alpha, \mu, \sigma^2|\mathbf{y})$ sejam mais semelhantes do que as posterioris $\pi(\alpha|\mathbf{y})$. A partir das Figuras A.1 e A.2 apresentadas no apêndice deste material, observa-se que, em todas as comparações de modelos que podemos notar uma diferença com relação à marginal $\pi(\alpha|\mathbf{y})$, essa diferença foi maior que comparada a diferença das marginais $\pi(\mu|\mathbf{y})$ e $\pi(\sigma^2|\mathbf{y})$. Ou seja, ocorreu sempre uma situação semelhante à da Figura 2.5. Inclusive, na maioria das situações nem existiu diferença das marginais $\pi(\mu|\mathbf{y})$ e $\pi(\sigma^2|\mathbf{y})$, ainda que notássemos diferença em $\pi(\alpha|\mathbf{y})$.

Após fazer essas conclusões, decidimos considerar que dois modelos marginais $\pi(\alpha|\mathbf{y})$ diferem substancialmente quando a diferença do DIC é maior que 5. Nos casos em que o DIC é menor que 5, mas a análise gráfica nos diz que as posterioris de α diferem, dizemos que essas distribuições potencialmente apresentam uma diferença importante quanto ao ajuste. Como mencionado, a priori de Jeffreys será utilizada como uma priori de referência. Isso quer dizer que as comparações dos ajustes baseados em priors informativas serão feitas em cima do ajuste baseado na priori de Jeffreys. Se um ajuste apresentar um DIC superior em 5 unidades com relação ao ajuste baseado na priori de Jeffreys, dizemos que este ajuste não é “bom” para os dados amostrados. Caso a diferença entre os DICs seja menor que 5, mas a análise gráfica nos remeta a outra conclusão, dizemos que o ajuste é potencialmente “ruim” para a amostra. Na Tabela 2.9, os números em negrito representam ajustes que foram considerados “ruins”, enquanto os números sublinhados representam ajustes considerados potencialmente “ruins”.

Ao analisar as Figuras 2.6 e 2.7, A.1 e A.2 (apresentadas no apêndice deste material) conjuntamente, notamos que quando temos crenças em modelos unimodais ou indecisos “erradas” (crença cujo intervalo adotado não contém α verdadeiro), a variância da distribuição $\pi(\mu|\mathbf{y})$ cresce à medida que o nível da crença aumenta. Reparamos ainda, que a crença “errada” com respeito a α nos leva a uma estimativa viesada de σ^2 . Quando acreditamos em α pequeno (caso

unimodal ou indeciso), mas na realidade α é grande (caso bimodal), a estimativa de σ^2 tende a ser superestimada. E quando cremos em α grande, mas α verdadeiro é pequeno, σ^2 estimado tende a ser subestimado.

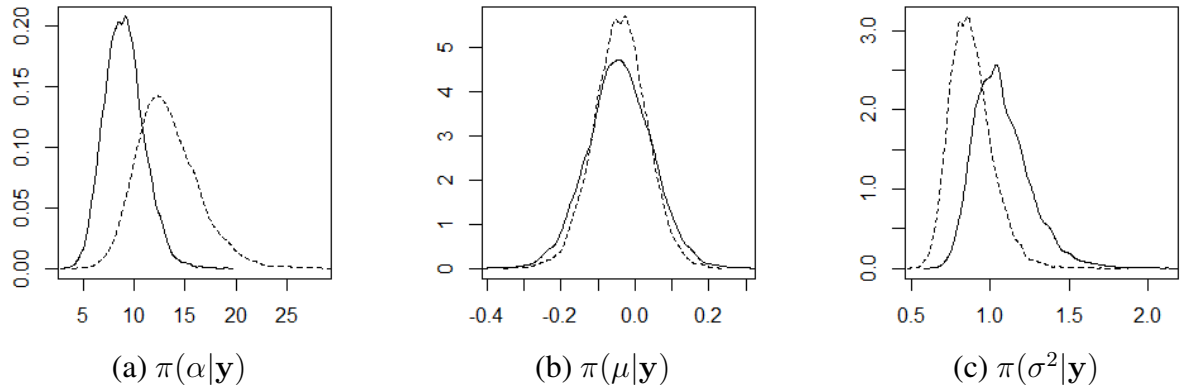


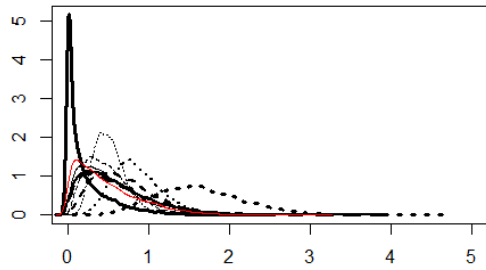
Figura 2.5: Distribuições a posteriori marginal utilizando uma priori gama (linha sólida) e Jeffreys (linha tracejada). Dados gerados a partir de uma $PNB(10, 0, 1)$ com $n = 50$.

A partir da Figura 2.9, notamos que os ajustes cujas crenças a priori eram “corretas” (crença cujo intervalo adotado contém α verdadeiro) não diferiram do ajuste baseado na priori de Jeffreys. Esses ajustes também não diferiram entre si. Isso quer dizer que os ajustes baseados em crença “correta” não foram influenciados pela mudança do nível desta no sentido de ela ser fraca, moderada, ou forte. No entanto, suspeita-se em duas situações que a crença unimodal “correta” promoveu um ajuste diferente desses demais ajustes baseados em crenças “corretas” quando $\alpha = 1$. Todavia, pode-se interpretar que a crença unimodal, isto é, em $\alpha \in (0, 1)$, não é uma crença “correta” quando $\alpha = 1$, uma vez que este parâmetro se encontra no extremo do intervalo.

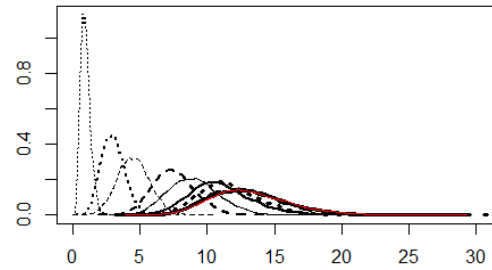
Analisando as prioris gama, vemos que as crenças muito distintas das observações não influenciaram na qualidade do ajuste se o nível destas for fraco. No entanto, a análise gráfica nos faz acreditar que há dois casos em que a crença fraca não promove um bom ajuste. Com relação às prioris log-normal, temos até o luxo de adotar uma crença moderada em modelos que não se identificam com a distribuição dos dados. Isso porque, apenas crenças fortes influenciaram de maneira negativa na qualidade de ajuste quando a distribuição dos dados não se comporta de acordo com a crença. No entanto, houve um caso em que a crença moderada que não corresponde com a distribuição dos dados foi potencialmente ruim.

Focando em amostras geradas a partir de distribuições bimodais, as posteriores cuja crença a priori não corresponde com a distribuição dos dados apresentaram DIC menores no caso gama comparado ao caso log-normal. Isso porque temos ao todo 10 modelos no caso gama que apresentaram DIC menor em mais de cinco unidades comparado ao caso Jeffreys. No caso log-normal, apenas com 3 modelos ocorreram tal situação. No cenário oposto, quando a distribuição

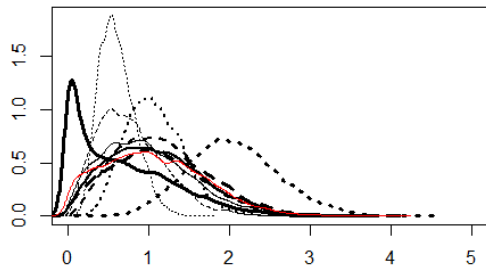
dos dados é unimodal, mas a crença é bimodal, foi observado a diferença no ajuste baseado na priori de Jeffreys em apenas 3 casos, 2 no caso gama e 1 no caso log-normal.



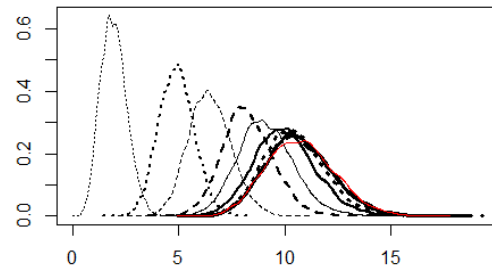
(a) $\alpha = 1; n = 50$



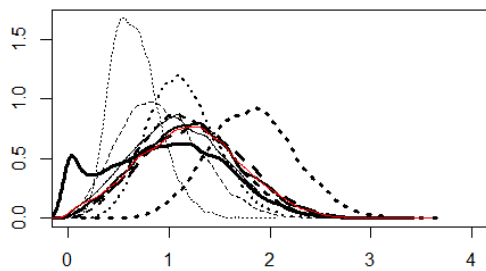
(b) $\alpha = 10; n = 50$



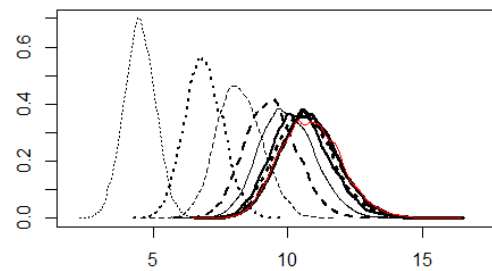
(c) $\alpha = 1; n = 100$



(d) $\alpha = 10; n = 100$



(e) $\alpha = 1; n = 200$



(f) $\alpha = 10; n = 200$

Figura 2.6: $\pi(\alpha|\mathbf{y})$ utilizando a priori gama.

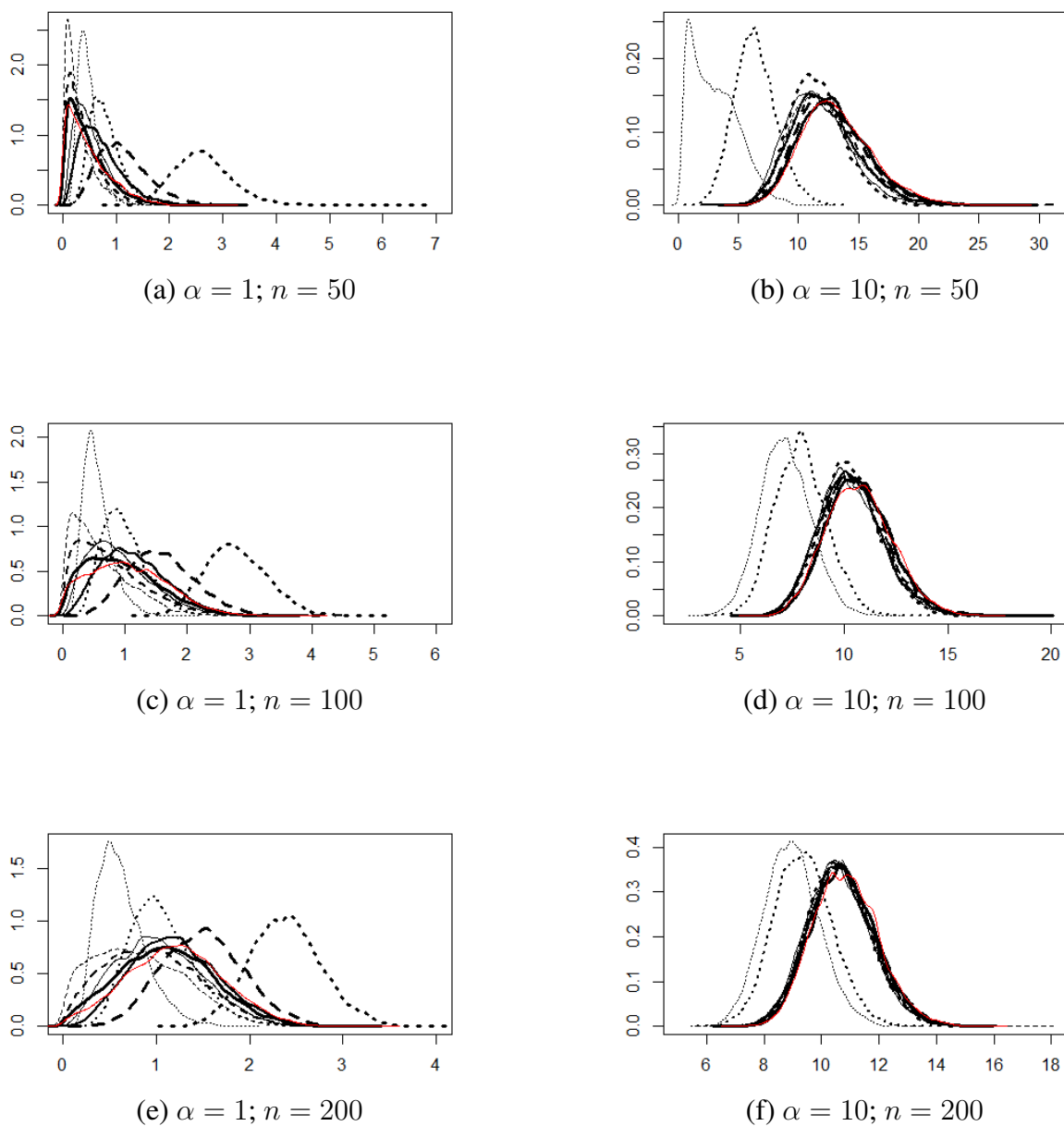


Figura 2.7: $\pi(\alpha|y)$ utilizando a priori de log-normal.

Crença / Intervalo	(0,1)	(0,2)	(2,15)	(0, ∞)
Fraca	Sólida Fina	Sólida Média	Sólida Grossa	-
Moderada	Tracejada Fina	Tracejada Média	Tracejada Grossa	-
Forte	Pontilhada Fina	Pontilhada Média	Pontilhada Grossa	-
Não Informativa	-	-	-	Vermelha

Tabela 2.8: Linhas das curvas das prioris da Figura 2.6 e Figura 2.7.

n	Amostra	Família	Crença fraca em $\alpha \in (0, 1)$	Crença mode-rada em $\alpha \in (0, 1)$	Crença forte em $\alpha \in (0, 1)$	Crença fraca em $\alpha \in (0, 2)$	Crença mode-rada em $\alpha \in (0, 2)$	Crença forte em $\alpha \in (0, 2)$	Crença fraca em $\alpha \in (2, 15)$	Crença mode-rada em $\alpha \in (2, 15)$	Crença forte em $\alpha \in (2, 15)$	Jeffreys
50	$PNB(1, 0, 1)$	Gama	126.59	126.56	126.35	126.90	127.23	127.76	125.66	127.28	132.05	126.70
50	$PNB(1, 0, 1)$	LN	126.71	125.85	126.31	127.36	126.33	127.51	126.54	129.32	138.85	
100	$PNB(1, 0, 1)$	Gama	220.10	219.80	219.60	220.26	220.06	219.62	220.39	220.38	222.14	220.41
100	$PNB(1, 0, 1)$	LN	219.99	220.07	219.72	220.15	220.24	219.61	220.30	220.77	225.48	
200	$PNB(1, 0, 1)$	Gama	467.61	467.46	467.54	467.73	467.47	466.88	468.48	467.76	468.58	467.82
200	$PNB(1, 0, 1)$	LN	467.67	468.20	467.73	467.50	467.93	467.03	467.80	467.73	471.83	
50	$PNB(10, 0, 1)$	Gama	130.50	145.81	173.23	128.10	133.60	155.30	127.29	127.24	127.42	127.29
50	$PNB(10, 0, 1)$	LN	128.06	127.92	158.09	127.66	127.73	137.99	127.23	127.47	127.49	
100	$PNB(10, 0, 1)$	Gama	256.70	266.10	316.27	255.65	258.51	277.63	255.41	255.46	255.28	255.43
100	$PNB(10, 0, 1)$	LN	255.55	255.67	262.94	255.52	255.53	260.02	255.34	255.28	255.24	
200	$PNB(10, 0, 1)$	Gama	522.07	528.35	574.35	521.30	523.38	538.33	521.30	521.34	521.30	521.34
200	$PNB(10, 0, 1)$	LN	521.36	521.38	524.92	521.43	521.36	523.60	521.29	521.44	521.29	

Tabela 2.9: DIC das distribuições a posteriori em cada amostra.

2.4.2 Estudo do Estimador Bayesiano

Nesta dissertação, foi feita uma simulação para avaliar o método de estimação bayesiano para os parâmetros da distribuição $PNB(\alpha, \mu, \sigma^2)$. Os cenários considerados foram amostras de tamanho $n = 50$, $n = 100$ e $n = 200$, com o parâmetro de forma $\alpha = 0.1$, $\alpha = 1$ e $\alpha = 10$, permanecendo fixo os parâmetros de posição e de escala, dados por $\mu = 0$ e $\sigma^2 = 1$, respectivamente. Simulamos 1000 amostras em todos cenários, e para cada amostra, geramos uma cadeia do algoritmo MCMC de tamanho $N = 5000$. As 20000 primeiras observações da cadeia foram descartadas e as 5000 restantes foram utilizadas para estimar a média, a mediana e a moda a posteriori dos parâmetros, além de fornecer o intervalo HPD. Utilizamos também a rotina OPTIM do R para obter o estimador de máxima verossimilhança (EMV). Após simular todos os cenários, estimamos o Viés Percentual (VP) e o Erro Quadrático Médio (EQM) da média a posteriori, da mediana a posteriori, da moda a posteriori e do EMV. As estimativas do VP e do EQM para um parâmetro φ com m amostras são dados por $\widehat{VP} = \frac{1}{m} \sum_{i=1}^m \left(\frac{\hat{\varphi}_i - \varphi}{\varphi} \right)$ e $\widehat{EQM} = \frac{1}{m} \sum_{i=1}^m (\hat{\varphi}_i - \varphi)^2$, em que $\hat{\varphi}_i$ é a estimativa de φ da amostra i .

A distribuição a priori de α utilizada foi a distribuição gama com parâmetros $a_0 = 1.056$ e $b_0 = 0.056$. A parametrização foi adotada de tal modo que a crença é fraca com respeito a α estar no intervalo $(0, 10)$. Ou seja, $P(\alpha < 10) = 0.4$. Isso quer dizer que cremos fracamente a priori que o modelo pode ser unimodal até bimodal tão fortemente quanto $\alpha = 10$. Essa crença foi adotada com base nos modelos bimodais que geralmente encontramos em aplicações, normalmente eles não são mais fortemente bimodais que um modelo $PNB(10, 0, 1)$. Diferentemente do que foi proposto na Seção 2.2.2, o ponto médio do intervalo não foi escolhido para ser igual a uma medida de tendência central. Para a simulação, escolhemos uma priori que fornecesse um peso maior para $\alpha = 1$ (modelo normal), por este ser mais simples que os demais. Portanto, a moda de $\pi(\alpha)$ é tal que $Mod(\alpha) = 1$. Dando um peso maior para $\alpha = 1$, $\pi(\alpha)$ é crescente no intervalo $(0, 1)$ e decrescente no intervalo $(1, \infty)$. Isso quer dizer que, quanto mais bimodal é o modelo, menos probabilidade a priori estamos fornecendo a ele. E, dentre todos os modelos unimodais, estamos dando maior probabilidade àqueles que se aproximam da distribuição normal. O gráfico da distribuição a priori pode ser visto na Figura 2.8.

A priori utilizada foi a gama, pois, na análise de sensibilidade, esta mostrou ter desempenho semelhante ao da distribuição log-normal quando a crença é fraca. Como a distribuição gama é mais simples, preferimos utilizá-la no estudo do estimador bayesiano.

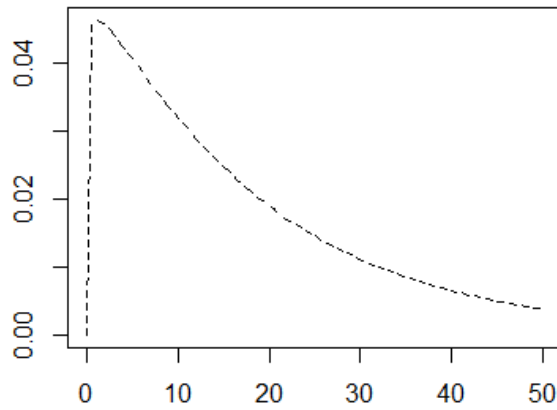
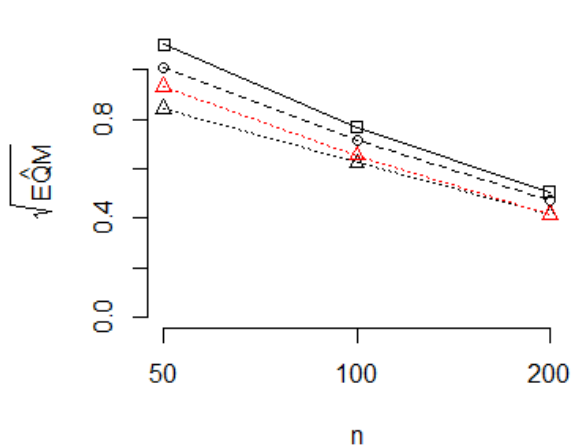
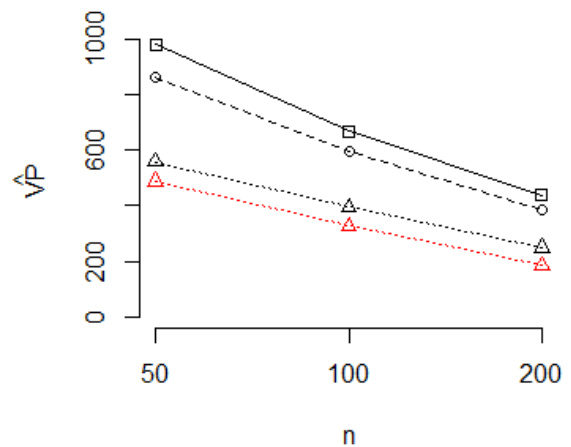


Figura 2.8: Distribuição a priori de α .

A partir da Figuras 2.9 e com o auxílio da Tabela A.1, que pode ser encontrada no apêndice, notamos que a moda a posteriori apresentou em geral um desempenho melhor do que a média a posteriori e a mediana a posteriori quanto ao VP e o EQM, e, portanto, é aconselhável utilizá-la para se fazer estimativas pontuais. Podemos perceber também que, na maioria dos cenários, a moda a posteriori teve um desempenho melhor do que o EMV quanto ao EQM.



(a) $\alpha = 0.1$



(b) $\alpha = 0.1$

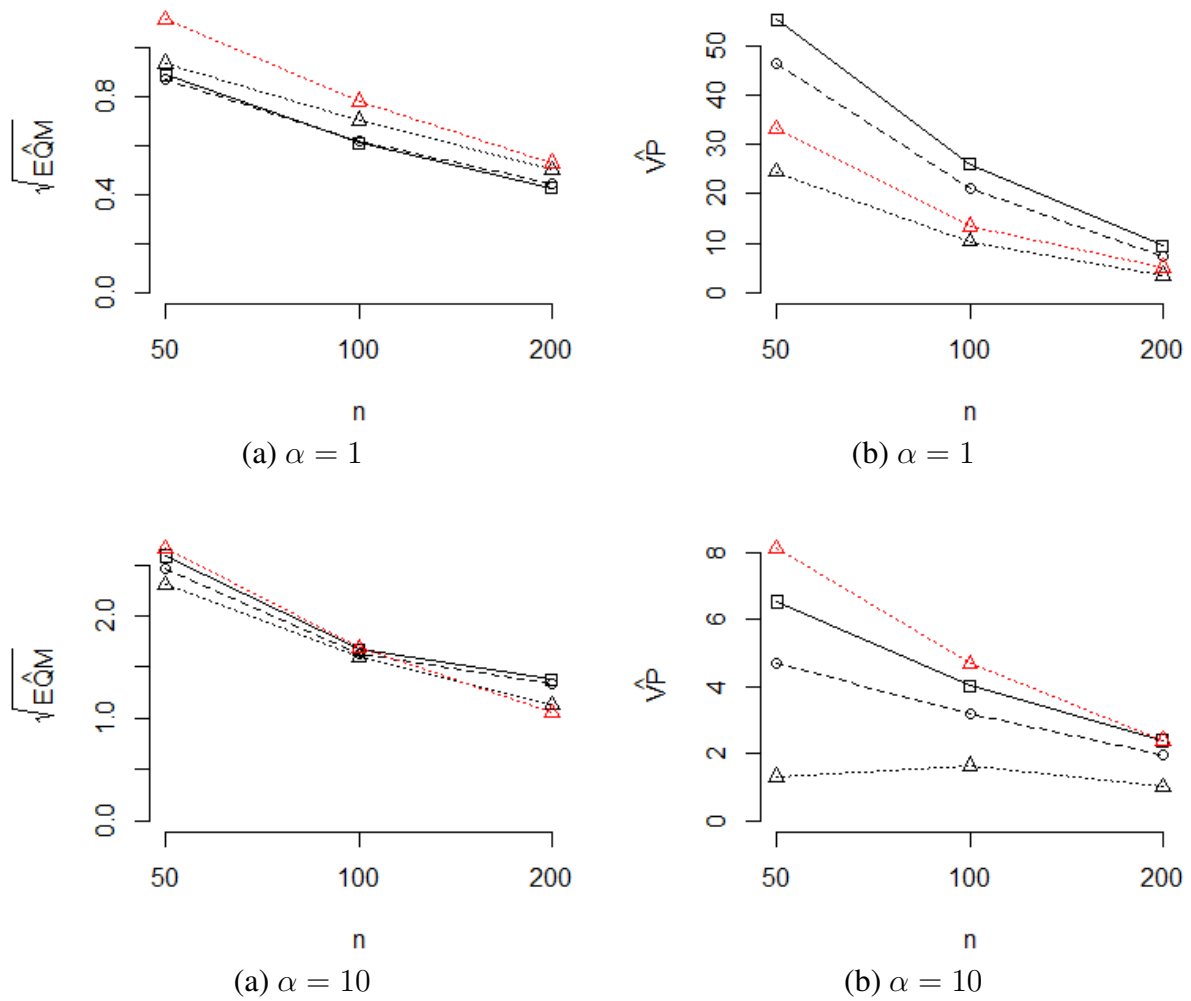


Figura 2.9: \hat{VP} e \sqrt{EQM} dos estimadores bayesianos e EMV. Média a posteriori (linha preta contínua). Mediana a posteriori (linha preta tracejada). Moda a posteriori (linha preta pontilhada). EMV (linha vermelha tracejada).

2.5 Aplicações com Dados Bimodais

Consideramos dois conjuntos de dados para a aplicação. O primeiro, cujo comportamento nos leva a crer em um modelo simétrico fracamente bimodal, é o Índice de Desempenho Ambiental (EPI) do ano de 2012. Esse índice foi proposto por pesquisadores da universidade de Yale e de Columbia (Estes et al., 2005) e tem como objetivo quantificar e qualificar numericamente as políticas ambientais dos países. O segundo, cujo comportamento é fortemente bimodal, surgiu dos estudos de Mason et al. (2011). São dados que representam o nível de RNA do gene da monoamina oxidase A (MAOA), coletados a partir de uma amostra de 227 índios Pima saudáveis, no entanto com pré-disposição a terem diabetes. As estatísticas descritivas dos dados podem ser vista na Tabela 2.10. Esta monoamina é responsável por decompor importantes neurotransmissores do cérebro. O histograma de ambos os conjuntos de dados pode ser visto na

Tabela 2.10 abaixo:

Dados	Tamanho amostral	Média	Desvio padrão	Coefficiente de assimetria
EPI	132	53.06	9.83	-0.18
MAOA	227	5.36	1.47	0.26

Tabela 2.10: Estatísticas descritivas dos dados EPI e MAOA

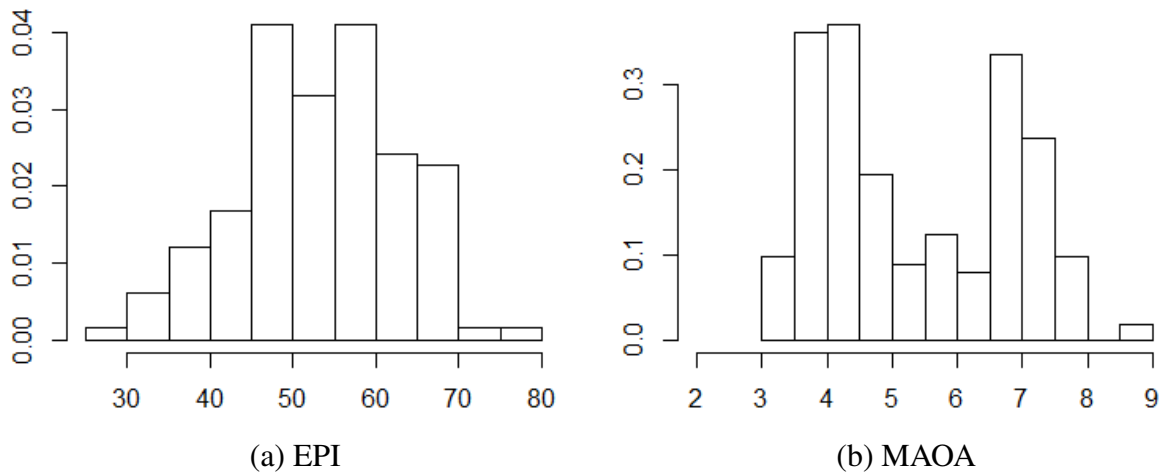


Figura 2.10: Histograma dos dados EPI e MAOA.

Para a modelagem bayesiana do modelo PNB, consideramos para ambos conjuntos de dados a priori de Jeffreys e a priori gamma. A priori gamma é a mesma empregada na Seção 2.3.3, visto que não temos conhecimento sobre EPI e genética. Dessa forma, $\alpha \sim \text{Gama}(1.056, 0.056)$. Consideramos também o modelo normal para os dados EPI, uma vez que pelo histograma não é simples identificar se a bimodalidade existe. Nos dados MAOA, também utilizamos o modelo de mistura dado por:

$$f(x|\mu_1, \sigma^2) = p\phi(x, \mu_1, \sigma^2) + (1 - p)\phi(x, \mu_2, \sigma^2).$$

Em que $\phi(x, \mu, \sigma^2)$ é a densidade de uma distribuição normal com média μ e variância σ^2 .

As prioris consideradas para os parâmetros da normal foram as mesmas dos parâmetros de posição e escala adotada no modelo PNB. No modelo de mistura temos que as prioris são:

$$p \sim \text{Beta}(\epsilon_{0,1}, \epsilon_{0,2})$$

$$\sigma^2 \sim \text{InvGamma}\left(\frac{\nu_0}{2}, \frac{\nu_0 S_0}{2}\right)$$

$$\mu \sim Normal(b_0, B_0)$$

Distribuição	Densidade	Valor Esperado	Variância
Beta	$\frac{\Gamma(\epsilon_{0,1} + \epsilon_{0,2})}{\Gamma(\epsilon_{0,1})\Gamma(\epsilon_{0,2})} p^{\epsilon_{0,1}-1} (1-p)^{\epsilon_{0,2}-1}$	$\frac{\epsilon_{0,1}}{\epsilon_{0,1} + \epsilon_{0,2}}$	$\frac{\epsilon_{0,1}\epsilon_{0,2}}{(\epsilon_{0,1} + \epsilon_{0,2} + 1)(\epsilon_{0,1} + \epsilon_{0,2})^2}$
InvGamma	$\frac{(\nu_0 S_0)^{\frac{\nu_0}{2}}}{2^{\frac{\nu_0}{2}} \Gamma(\frac{\nu_0}{2})} (\sigma^2)^{-(\frac{\nu_0}{2} + 1)} \exp\left\{-\frac{\nu_0 S_0}{2\sigma^2}\right\}$	$\frac{\nu_0 S_0}{2}$ $\frac{\nu_0}{2} - 1$	$\frac{\nu_0 S_0}{2}$ $\left(\frac{\nu_0}{2} - 1\right)^2 \left(\frac{\nu_0}{2} - 2\right)$

Tabela 2.11: Distribuições a priori informativa para α .

A priori adotada no modelo de mistura foi uma não informativa. Com isso, temos que $\epsilon_{0,1} = \epsilon_{0,2} = 1$, $\nu_0 = S_0 = e^{-16}$, $b_0 = \hat{y}$ e $B_0 = e^{16}$.

Utilizamos o pacote *bayesmix* do R para gerar a cadeia do MCMC do modelo de mistura. Escolhemos o *burn-in* de 1000 para um total de 6000 iterações geradas. Com relação as cadeias dos modelos normal e PNB o *burn-in* foi de 20000 com um total de 5000 iterações geradas. Assim, foi possível obter a posteriori como pode ser visto na Figura 2.11. Adotamos a moda a posteriori para estimar o parâmetro α do modelo PNB. Para os demais parâmetros empregamos a média. No modelo normal e no de mistura adotamos a média a posteriori para estimar os parâmetros. Com isso, foram ajustados os seguintes modelos:

Modelo	EPI	
	Bayes	EMV
Normal	N(53.04,97.78)	N(53.06,95.86)
PNB (jef.)	PNB(1.39,52.97,91.26)	PNB(1.50,53.04,86.84)
PNB (gam.)	PNB(1.52,52.96,91,63)	
Modelo	MAOA	
	Bayes	EMV
PNB (jef.)	PNB(7.21,5.49,0.97)	PNB(7.09,5.50,0.98)
PNB (gam.)	PNB(7.07,5.50,0.98)	
Mistura	0.56 N(4.18,0.35) + 0.44 N(6.88,0.35)	

Tabela 2.12: Ajustes dos modelos para os dados EPI e MAOA.

Como era de se esperar, analisando a Tabela 2.12 as estimativas de α no caso PNB e μ_1 , μ_2 e σ^2 no caso do modelo de mistura, vemos que os dados MAOA apresentaram uma bimodalidade forte. Percebemos que a falta de evidência sobre a bimodalidade dos dados EPI implicou no modelo normal como o melhor segundo o BIC. Analisando a Tabela 2.13, podemos ver que o limite inferior (LI) do intervalo de credibilidade HPD de α para os dados EPI é menor que 1, e o limite superior (LS) é maior que 1, para as duas prioris consideradas. Isso significa que o valor 1 pertence ao intervalo HPD. Logo, não temos fortes evidências para saber se os dados são unimodais ou bimodais. Quanto ao intervalo de credibilidade referente aos dados MAOA, através da Tabela 2.14, este mostra fortes evidências de bimodalidade acentuada. O melhor ajuste de acordo com o DIC para os dados MAOA foi o do modelo de mistura. O estimador

bayesiano (EB) e o estimador de máxima verossimilhança (EMV) praticamente não diferiram nas estimativas.

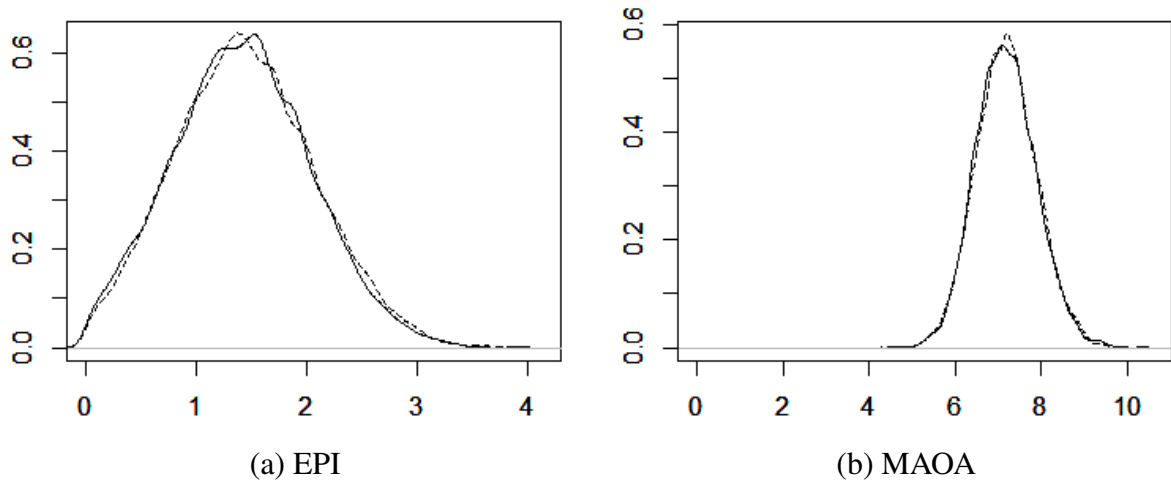


Figura 2.11: Distribuições a posteriori do parâmetro α do modelo PNB para os dados EPI e MAOA. Priori Gamma (linha contínua). Priori de Jeffreys (linha tracejada).

Modelo	BIC	Parâmetro	Estimador Bayesiano					
			Média	Mediana	Moda	LI	LS	EMV
Normal	988.65	μ	53.04	53.02	52.94	51.37	54.64	53.06
		σ^2	97.78	96.72	93.23	72.38	121.46	95.86
PNB (jef.)	993.93	α	1.42	1.40	1.39	0.21	2.65	1.50
		μ	52.97	53.00	53.21	51.18	54.52	53.04
		σ^2	91.26	89.48	87.10	62.64	122.62	86.84
PNB (gam.)	993.90	α	1.40	1.40	1.52	0.16	2.54	1.50
		μ	52.96	52.98	53.05	51.20	54.59	53.04
		σ^2	91.63	90.06	91.56	65.16	124.38	86.84

Tabela 2.13: Estimativas dos parâmetros do modelo potência normal bimodal para os dados EPI.

Para uma análise mais aprofundada dos dados, podemos também estar interessado em explicá-los em função de covariáveis. Se a relação dos dados com tais covariáveis é linear, podemos utilizar um modelo de regressão linear. Nessa situação, esperamos que os resíduos apresentem uma distribuição potência normal bimodal, uma vez que os dados também têm essa distribuição. No próximo capítulo, mostramos como analisar dados na presença de covariáveis a partir de uma perspectiva bayesiana utilizando regressão com resíduos cuja distribuição é potência normal bimodal.

Modelo	BIC	Parâmetro	Estimador Bayesiano					EMV
			Média	Mediana	Moda	LI	LS	
PNB (jef.)	727.82	α	7.18	7.15	7.09	5.87	8.61	7.21
		μ	5.50	5.50	5.48	5.41	5.60	5.49
		σ^2	0.98	0.97	0.96	0.84	1.13	0.97
PNB (gam.)	727.80	α	7.19	7.17	7.07	5.84	8.49	7.21
		μ	5.50	5.50	5.49	5.41	5.60	5.49
		σ^2	0.98	0.97	0.97	0.84	1.12	0.97
Mistura	718.70	p	0.56	0.56	0.56	5.84	8.49	0.56
		μ_1	4.18	4.18	4.18	0.00	0.00	4.18
		μ_2	6.88	6.88	6.89	5.41	5.60	6.88
		σ^2	0.35	0.34	0.34	0.84	1.12	0.34

Tabela 2.14: Estimativas dos parâmetros do modelo potência normal bimodal para os dados MAOA.

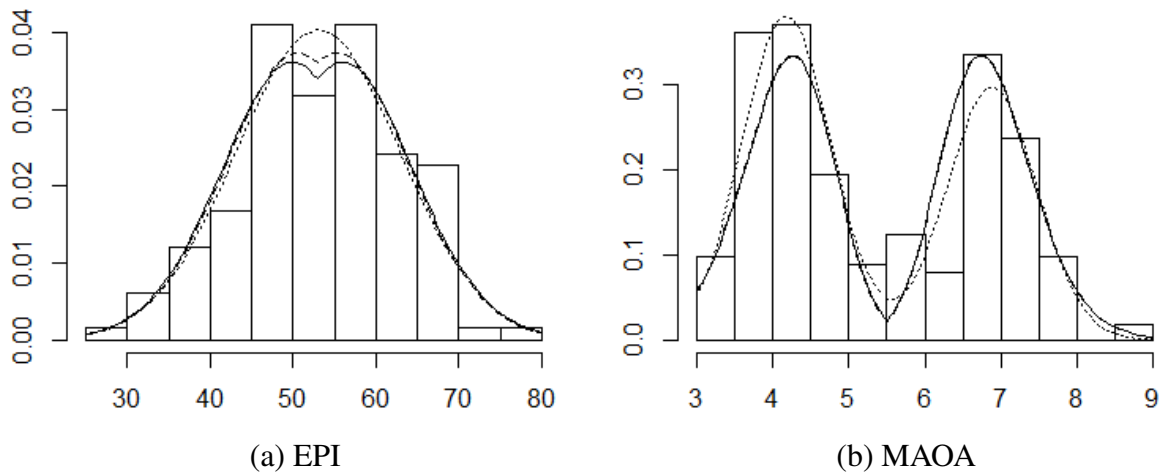


Figura 2.12: Histograma dos dados EPI e MAOA sobreposto ao modelo bayesiano. Dados EPI: PNB com priori gamma (linha contínua), PNB com priori de jeffreys (linha tracejada), normal (linha pontilhada). Dados MAOA: PNB com priori gamma (linha contínua), PNB com priori de jeffreys (linha tracejada), mistura (linha pontilhada).

Capítulo 3

Regressão com resíduos PNB

3.1 Inferência

O modelo de regressão com resíduos PNB é dado por:

$$y_i = X_i' \beta + \eta_i$$

em que:

- β' : vetor de coeficientes de dimensão $k + 1$ das covariáveis observadas. O primeiro elemento é o intercepto.
- X_i' : vetor de dimensão $k + 1$ das covariáveis observadas na i -ésima observação, sendo o primeiro elemento o número 1.
- η_i : Resíduos do modelo. $\eta_i \sim PNB(\alpha, 0, \sigma^2)$.

3.1.1 Função de Verossimilhança

O modelo de regressão linear com resíduos PNB tem função de verossimilhança dada por:

$$L(\theta | \underline{X}, \mathbf{y}) = \left(\frac{2\alpha c_\alpha}{\sigma} \right)^n \prod_{i=1}^n \left[\phi \left(\frac{y_i - X_i' \beta}{\sigma} \right) \right] \left\{ \prod_{i=1}^n \Phi \left(\frac{|y_i - X_i' \beta|}{\sigma} \right) \right\}^{\alpha-1}, \quad i = 1, \dots, n, \quad (3.1)$$

em que \underline{X} é a matriz de delineamento com dimensão $n \times (k + 1)$ cuja i -ésima linha é dada pelo vetor X_i' .

O vetor θ agora será $\theta = (\alpha, \beta, \sigma^2)$. As suposições que devem ser estabelecidas para a veracidade da expressão 3.1 é que os vetores X_1', X_2', \dots, X_n' e as observações y_1, y_2, \dots, y_n devem ser independentes. A quantidade $X_i' \beta$ representa a média da variável y dado o vetor de covariáveis X_i' . Quanto aos β_i s, podemos dizer que, para cada unidade a mais da i -ésima covariável do vetor X_i' , a média de Y aumenta β_i unidades, mantendo as demais covariáveis fixas.

3.1.2 Distribuição a Priori para o Vetor β

Os elementos do vetor de parâmetros β são os coeficientes do modelo de regressão. Assumimos para β uma distribuição a priori $N_{k+1}(\xi, \Sigma)$, em que ξ é o vetor de médias e Σ a matriz de covariância. Então, a fdp de β é dada por:

$$\pi(\beta) = |\Sigma|^{-1/2} (2\pi)^{-(k+1)/2} \exp \left\{ -\frac{(\beta - \xi)' \Sigma^{-1} (\beta - \xi)}{2} \right\}, \quad \beta \in \mathbb{R}^{k+1}$$

Para os hiperparâmetros, escolhemos $\xi = 0$ e $\Sigma = I10^6$, em que I é a matriz identidade de tamanho $(k + 1) \times (k + 1)$. Ou seja, assumimos independência entre cada β_i e consideramos que todos eles têm distribuição normal padrão com média 0 e variância 10^6 . A média 0 foi escolhida de modo que não tivéssemos nenhum conhecimento a priori com relação ao sinal dos β_i s e a variância é tão grande para que a distribuição a priori seja não informativa. Dessa forma, as distribuições a priori dos β_i s se aproximam de uma constante, que é a distribuição de Jeffreys para um parâmetro de posição, que é o caso dos β_i s.

Distribuição a Priori Conjunta

A distribuição a priori considerada para o vetor de parâmetros θ é de tal forma que α , β e σ^2 são independentes. Com isso, temos que:

$$\begin{aligned} \pi(\theta) &= \pi(\alpha, \beta, \sigma^2) \\ &= \pi(\alpha)\pi(\beta)\pi(\sigma^2) \\ &= \pi(\alpha) |\Sigma|^{-1/2} (2\pi)^{-\frac{p+1}{2}} \frac{b^a}{\Gamma(a)} (\sigma^2)^{a-1} \\ &\quad \exp \left\{ -\frac{(\beta - \xi)' \Sigma^{-1} (\beta - \xi)}{2} - b\sigma^2 \right\}. \end{aligned}$$

3.1.3 Distribuições a Posteriori

Posteriori Conjunta

Obtida a distribuição a priori conjunta proporcional de $(\alpha, \beta, \sigma^2)$, podemos obter a sua distribuição a posteriori proporcional:

$$\begin{aligned} \pi(\alpha, \beta, \sigma^2 | \mathcal{X}, \mathbf{y}) &= \frac{\pi(\alpha, \beta, \sigma^2) L(\alpha, \beta, \sigma^2 | \mathcal{X}, \mathbf{y})}{\pi(\mathbf{y})} \\ &\propto \pi(\alpha)\pi(\beta)\pi(\sigma^2) L(\alpha, \beta, \sigma | \mathcal{X}, \mathbf{y}) \\ &\propto \pi(\alpha) \exp \left\{ -\frac{(\beta - \xi)' \Sigma^{-1} (\beta - \xi)}{2} - \frac{(y_i - X_i' \beta)' (y_i - X_i' \beta)}{2\sigma^2} - b\sigma^2 \right\} \\ &\quad \alpha^{n-1} c_\alpha^n (\sigma^2)^{a - (\frac{n}{2} + 1)} \left\{ \prod_{i=1}^n \Phi \left(\frac{|y_i - X_i' \beta|}{\sigma} \right) \right\}^\alpha. \end{aligned}$$

A distribuição preditiva de \mathbf{y} é dada por:

$$\pi(y) = \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{\infty} \pi(\theta) L(\theta | \underline{X}, \mathbf{y}) d\alpha d\sigma^2 d\beta.$$

A resolução dessa integral só pode ser feita por métodos numéricos, portanto, não podemos obter uma forma fechada para $\pi(\theta | \mathbf{y})$.

Distribuições a Posteriori Condicionais

A distribuição a posteriori condicional é definida como sendo a distribuição de um parâmetro dados os demais parâmetros e a amostra. Abaixo, temos a densidade a posteriori condicional proporcional de α , β e σ^2 :

$$\begin{aligned} \pi(\alpha | \beta, \sigma^2, \underline{X}, \mathbf{y}) &= \frac{\pi(\alpha, \beta, \sigma^2) L(\alpha, \beta, \sigma^2 | \underline{X}, \mathbf{y})}{\pi(\beta, \sigma^2, Y)} \\ &\propto \pi(\alpha) L(\alpha, \beta, \sigma^2 | \underline{X}, \mathbf{y}) \\ &\propto \pi(\alpha) \alpha^{n-1} c_{\alpha}^n \left\{ \prod_{i=1}^n \Phi \left(\frac{|y_i - X_i' \beta|}{\sigma} \right) \right\}^{\alpha}, \end{aligned}$$

$$\begin{aligned} \pi(\beta | \alpha, \sigma^2, \underline{X}, \mathbf{y}) &= \frac{\pi(\alpha, \beta, \sigma^2) L(\alpha, \beta, \sigma^2 | \underline{X}, \mathbf{y})}{\pi(\alpha, \sigma^2, \mathbf{y})} \\ &\propto \pi(\beta) L(\alpha, \beta, \sigma^2 | \underline{X}, \mathbf{y}) \\ &\propto \exp \left\{ \frac{(\beta - \xi)' (\Sigma^2)^{-1} (\beta - \xi)}{2} - \frac{\sum_{i=1}^n (y_i - X_i' \beta)^2}{2\sigma^2} \right\} \\ &\quad \left\{ \prod_{i=1}^n \Phi \left(\frac{|y_i - X_i' \beta|}{\sigma} \right) \right\}^{\alpha-1}, \end{aligned}$$

$$\begin{aligned} \pi(\sigma^2 | \alpha, \beta, \mathbf{y}) &= \frac{\pi(\alpha, \beta, \sigma^2) L(\alpha, \beta, \sigma^2 | \underline{X}, \mathbf{y})}{\pi(\alpha, \beta, \mathbf{y})} \\ &\propto \pi(\sigma^2) L(\alpha, \beta, \sigma^2 | \underline{X}, \mathbf{y}) \\ &\propto \exp \left\{ -\frac{\sum_{i=1}^n (y_i - X_i' \beta)^2}{2\sigma^2} - b\sigma^2 \right\} \sigma^{a - (\frac{n}{2} + 1)} \\ &\quad \left\{ \prod_{i=1}^n \Phi \left(\frac{|y_i - X_i' \beta|}{\sigma} \right) \right\}^{\alpha-1}. \end{aligned}$$

3.2 Aspectos Computacionais

Distribuição Proposta para β

No caso univariado, para o parâmetro de locação da distribuição potência normal bimodal empregamos a proposta de tal modo que $q_{\mu}(\mu_{t+1}^p | \mu_t) \sim N(\mu_t, s_t^{(2)})$. Assim, para o caso multivariado, a distribuição proposta escolhida foi uma normal multivariada dada pela expressão

abaixo:

$$q_{\beta}(\beta_{t+1}^p | \beta_t) \sim N_{k+1}(\beta_t, (\underline{X}' \underline{X})^{-1} s_t^{(2)})$$

A estrutura de covariância estabelecida se deve ao fato de que a ausência de correlação entre os elementos de β_t^p implica uma amostra de β_i s gerada pelo MCMC bastante auto-correlacionada. Então, optamos por trabalhar com a matriz de covariância dada por $s_t^{(2)} (\underline{X}' \underline{X})^{-1}$, a fim de decidir por uma proposta com uma estrutura de correlação análoga ao do EMV dos coeficientes de um modelo de regressão linear com resíduos normalmente distribuídos. Nesse caso, $s^{(2)}$ é o sintonizador da variância da proposta de β .

Valor inicial para os Parâmetros e para os $s_t^{(i)}$ s

Usando o mesmo raciocínio utilizado para propor valores iniciais para os parâmetros e os $s_t^{(i)}$ s no capítulo anterior, temos que:

Parâmetro	valor inicial	$s_0^{(i)}$	valor inicial
α_0	Qualquer valor entre 0 e 20	$s_0^{(1)}$	$\sqrt{2}$
β_0	$\hat{\beta}$	$s_0^{(2)}$	QME
σ_0^2	QME	$s_0^{(3)}$	$\frac{QME}{(n-k-1)}$

Tabela 3.1: Valores iniciais para α_0 , β_0 , σ_0^2 e os $s_0^{(i)}$ s.

Em que:

$$\hat{\beta} = (\underline{X}' \underline{X})^{-1} \underline{X}' \mathbf{y}, \quad QME = \frac{(\mathbf{y} - \underline{X} \hat{\beta})' (\mathbf{y} - \underline{X} \hat{\beta})}{n - k - 1}.$$

Os estimadores usados como referência para o valor inicial de β_0 e σ_0^2 , foram os estimadores de máxima verossilhança de um modelo de regressão com resíduos normalmente distribuídos, dados por $\hat{\beta}$ e $\hat{\sigma}^2$, respectivamente. O ideal seria escolher $s_0^{(2)}$ de maneira que $Var(\hat{\beta}) = Var(\beta_{t+1}^p | \beta_t)$. No entanto, não conhecemos $Var(\hat{\beta})$. Então, o estimamos por:

$$\widehat{Var}(\hat{\beta}) = (X'X)^{-1} \hat{\sigma}^2, \quad \text{em que } \hat{\sigma}^2 = QME.$$

Assim, o valor inicial para $s_0^{(2)}$ será:

$$\widehat{Var}(\hat{\beta}) = Var(\beta_{t+1}^p | \beta_t) \Rightarrow (X'X)^{-1} QME = (X'X)^{-1} s_0^{(2)} \Rightarrow s_0^{(2)} = QME.$$

Da mesma forma, o valor atribuído a $s_0^{(3)}$ deveria ser de tal forma que $Var(\hat{\sigma}^2) = Var((\sigma^2)_{t+1}^p | \sigma_t^2)$. Como não conhecemos $Var(\hat{\sigma}^2)$, o estimamos por $\frac{QME^2}{n-k-1}$. Assim:

$$\frac{QME^2}{n - k - 1} = \text{Var}((\sigma^2)_{t+1}^p | \sigma_t^2) = \sigma_t^2 s_0^{(3)}.$$

Substituindo σ_t^2 pelo QME temos que:

$$s_0^{(3)} = \frac{QME}{n - k - 1}.$$

Capítulo 4

O Modelo Potência Normal Bimodal Assimétrico

4.1 Modelo Potência Normal Bimodal Assimétrico

4.1.1 Função de Densidade

Bolfarine et al. (2013) propôs uma extensão do modelo PNB o qual foi denominado de Potência Normal Bimodal Assimétrico (PNBA). Este teve enfoque no modelo *extended two-piece skew normal* (ETN) proposto por Arnold et al. (2009). O modelo PNBA pode assumir formas bimodais e assimétricas. Sua fdp é dada por:

$$f_Z(z|\alpha, \lambda) = 2\alpha c_\alpha \phi(z) \{\Phi(|z|)\}^{\alpha-1} \Phi(\lambda z), \quad z \in \mathbb{R}, \quad \alpha > 0, \quad \lambda \in \mathbb{R}$$

em que c_α é a mesma constante de normalização do modelo PNB(α) e λ é o parâmetro de assimetria. Assim, denotamos o modelo por $Z \sim PNBA(\alpha, \lambda)$. A Figura 4.1 ilustra o gráfico das fdp's do modelo potência normal bimodal assimétrico para diferentes valores de α e λ .

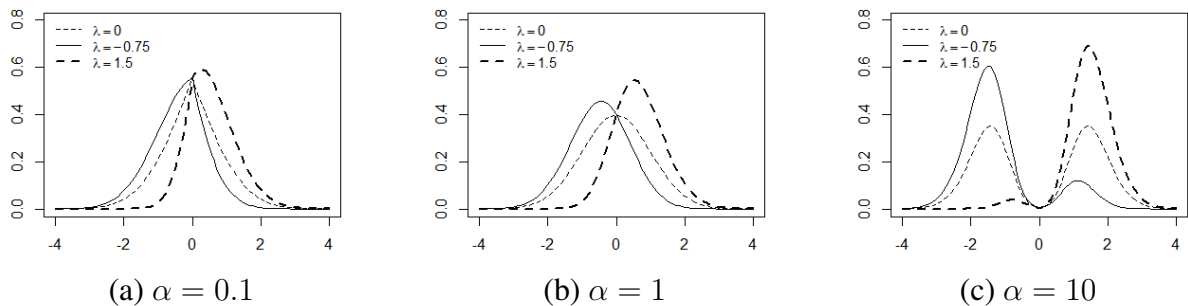


Figura 4.1: Densidades da potência normal bimodal assimétrica.

O modelo $PNBA(\alpha, \lambda)$ é bastante flexível e pode ser visto como a generalização de vários

modelos estudados na literatura:

Distribuição	Densidade
PNB(α)	$f_Z(z \alpha, 0) = \alpha c_\alpha \phi(z) \{\Phi(z)\}^{\alpha-1}, \quad z \in \mathbb{R}, \quad \alpha > 0$
N(0,1)	$f_Z(z 1, 0) = \phi(z), \quad z \in \mathbb{R}$
SN(λ)	$f_Z(z 1, \lambda) = 2\phi(z)\Phi(\lambda z), \quad z \in \mathbb{R}, \lambda \in \mathbb{R}$
ETN(1, λ)	$f_Z(z 2, \lambda) = \frac{4}{3}\phi(z)\Phi(z)\Phi(\lambda z), \quad z \in \mathbb{R}, \lambda \in \mathbb{R}$

Tabela 4.1: Casos particulares do modelo potência normal bimodal assimétrico.

A vantagem do modelo PNBA com relação ao *extend two-piece skew-normal*, cuja distribuição também pode assumir formatos assimétricos e bimodais, é a flexibilidade. Quando o ETN é bimodal, os dois picos do modelo são sempre muito próximos com relação ao eixo das abscissas. Já o modelo PNBA, por exemplo, pode assumir formatos em que os dois picos se distanciam bastante. Andrade et al. (2016) estudou as propriedades dos modelos PNB(α) e PNBA(α, λ) e modelou um conjunto de dados cujos picos não eram tão próximos e, como era de se esperar, o modelo *extend two-piece skew-normal* não se ajustou bem aos dados, enquanto o modelo PNBA teve um bom ajuste. Além de estudar as propriedades, Andrade et al. (2016) estudou os estimadores de máxima verossimilhança e de momentos através de simulações.

4.1.2 Média, Moda e Variância

Seja $\mu_r(\alpha, 0)$ e $\mu_r(\alpha, \lambda)$ os momentos incompletos de ordem r das distribuições potência normal e potência normal bimodal assimétrica, respectivamente, dados por:

$$\mu_r(\alpha, 0) = \alpha \int_0^\infty z^r \phi(z) \{\Phi(z)\}^{\alpha-1} dz,$$

$$\mu_r(\alpha, \lambda) = 2\alpha c_\alpha \int_0^\infty z^r \phi(z) \{\Phi(z)\}^{\alpha-1} \Phi(\lambda z) dz.$$

Então, o valor esperado e a variância de $Z \sim \text{PNBA}(\alpha, \lambda)$ é dado por:

$$E(Z) = 2[\mu_1(\alpha, \lambda) - c_\alpha \mu_1(\alpha, 0)],$$

$$\text{Var}(Z) = \mu_2(\alpha, 0) - 4[\mu_1(\alpha, \lambda) - c_\alpha \mu_1(\alpha, 0)]^2.$$

Os pontos críticos da distribuição PNBA(α, λ) podem ser encontrados obtendo a solução da equação abaixo:

$$\alpha - 1 = |z| \frac{\Phi(|z|)}{\phi(z)} \left[1 - \frac{\lambda \Phi(\lambda z)}{z \phi(z)} \right].$$

Dessa maneira, o modelo é bimodal se $\alpha > 1$ e $\left[1 - \frac{\lambda \Phi(\lambda z)}{z \phi(z)} \right] > 0$.

4.1.3 Extensão para o Modelo com Parâmetro de Posição e Escala

A família de posição e escala do modelo potência normal bimodal assimétrico é dada por:

$$f_Y(y|\alpha, \mu, \sigma^2) = 2\alpha c_\alpha \phi\left(\frac{y-\mu}{\sigma}\right) \left\{ \Phi\left(\frac{|y-\mu|}{\sigma}\right) \right\}^{\alpha-1} \Phi\left(\lambda \frac{y-\mu}{\sigma}\right)$$

A notação que usamos é $Y \sim \text{PNBA}(\alpha, \lambda, \mu, \sigma^2)$. Notamos facilmente que $E(Y) = \mu + \sigma E(Z)$ e $\text{Var}(Y) = \sigma^2 \text{Var}(Z)$.

4.2 Inferência

4.2.1 Função de Verossimilhança

A função de verossimilhança do modelo potência normal bimodal assimétrico é dada por:

$$L(\theta|\mathbf{y}) = \left(\frac{2\alpha c_\alpha^n}{\sigma}\right)^n \prod_{i=1}^n \left[\phi\left(\frac{y_i-\mu}{\sigma}\right)\right] \left\{ \prod_{i=1}^n \Phi\left(\frac{|y_i-\mu|}{\sigma}\right) \right\}^{\alpha-1} \prod_{i=1}^n \Phi\left(\lambda \frac{y_i-\mu}{\sigma}\right), i = 1, \dots, n,$$

em que $\theta = (\alpha, \lambda, \mu, \sigma^2)$.

4.2.2 Distribuição a Priori para o Vetor (α, λ)

No modelo PNBA, propomos algumas prioris informativas e não informativas para o vetor de parâmetros (α, λ) . Apresentamos duas priori não informativas: uma baseada na Jeffreys e outra baseada em uma reparametrização de λ . Quanto às prioris informativas, analisaremos formas bem simples.

Distribuição a Priori de Jeffreys

Bolfarine et al. (2013) mostrou que a matriz de informação de fisher do vetor (α, λ) é dado por:

$$IF(\alpha, \lambda) = \begin{pmatrix} n [\alpha^{-2} - \log^2(2)2^\alpha(2^\alpha - 1)^{-2}] & 0 \\ 0 & n \left\{ \lambda E \left[Z^3 \frac{\phi(\lambda Z)}{\Phi(\lambda|Z|)} \right] + E \left[Z \frac{\phi(\lambda Z)}{\Phi(\lambda|Z|)} \right]^2 \right\} \end{pmatrix}$$

em que Z é tal que $Z \sim \text{PNBA}(\alpha, \lambda)$. Dessa forma, temos que a priori de Jeffreys do vetor (α, λ) é:

$$\begin{aligned}\pi_J(\alpha, \lambda) &\propto \sqrt{\det(IF(\alpha, \lambda))} \\ &\propto [\alpha^{-2} - \log^2(2)2^\alpha(2^\alpha - 1)^{-2}] \left\{ \lambda E \left[Z^3 \frac{\phi(\lambda Z)}{\Phi(\lambda|Z|)} \right] + E \left[Z \frac{\phi(\lambda Z)}{\Phi(\lambda|Z|)} \right]^2 \right\}\end{aligned}\quad (4.1)$$

Os valores esperados apresentados na expressão acima não podem ser calculados analiticamente.

Distribuição a Priori Não Informativa Uniforme

Consideramos um resultado obtido por Andrade et al. (2016) em que foi mostrado que a distribuição PNBA pode ser gerada a partir da transformação:

$$Z = SV,$$

em que:

- $V \sim |PNB(\alpha)|$.
- S é tal que $P(S = 1|V = v) = \Phi(\lambda v)$ e $P(S = -1|V = v) = 1 - \Phi(\lambda v)$.

Seja $p = P(S = 1) = \Phi(\lambda V)$ e suponha $p \sim U(0, 1)$ a distribuição a priori não informativa de p . Então, temos que:

$$\Phi^{-1}(p) = \lambda V \Rightarrow \lambda = \frac{N(0, 1)}{|PNB(\alpha)|}.$$

No apêndice deste trabalho, verificamos que a distribuição de λ dado α é escrita da seguinte maneira:

$$f_{\lambda|\alpha}(\lambda|\alpha) = \sqrt{\frac{2}{\pi}} \alpha c_\alpha \int_0^\infty v \phi(\sqrt{\lambda^2 + 1}v) \{\Phi(v)\}^{\alpha-1} dv. \quad (4.2)$$

A distribuição a priori não informativa de α utilizada é a de Jeffreys. Temos, então, que a distribuição conjunta a priori $\pi(\alpha, \lambda)$ é dada por:

$$\begin{aligned}\pi(\alpha, \lambda) &= J(\alpha) f_{\lambda|\alpha}(\lambda|\alpha) \\ &\propto \sqrt{\alpha^{-2} - \log^2(2)2^\alpha(2^\alpha - 1)^{-2}} \alpha c_\alpha \int_0^\infty v \phi(\sqrt{\lambda^2 + 1}v) \{\Phi(v)\}^{\alpha-1} dv\end{aligned}\quad (4.3)$$

A priori da expressão (4.2) é própria, a constante de normalização é conhecida e sabemos como gerar valores da distribuição condicional $f_{\lambda|\alpha}(\lambda|\alpha)$. Se atribuirmos uma priori simples para α , sabemos como gerar valores aleatórios da distribuição conjunta de (α, λ) . Estas são as vantagens desta priori comparada a de jeffreys. Ambas prioris tiveram o proposito de serem

utilizadas apenas para comparar com prioris informativas, do mesmo modo como foi feito na Seção 2.

Distribuição a Priori Informativa

Atribuímos distribuições a priori informativas para o vetor (α, λ) de maneira que estas fossem bem simples. Consideramos que α e λ são independentes, o que significa que $\pi(\alpha, \lambda) = \pi(\alpha)\pi(\lambda)$. Usamos a priori log-normal (ζ, γ) para o parâmetro α . Quanto ao parâmetro λ , atribuímos uma distribuição normal com média δ e variância η^2 , ou seja:

$$\lambda \sim N(\delta, \eta^2)$$

Denotamos $\pi(\alpha)$ e $\pi(\lambda)$ como as distribuições marginais de α e λ , respectivamente. A distribuição conjunta é denotada por $\pi(\alpha, \lambda)$ e é dada por:

$$\pi(\alpha, \lambda) = \frac{1}{2\pi\alpha\gamma\eta} \exp \left\{ -\frac{1}{2} \left[\frac{(\log \alpha - \zeta)^2}{\gamma^2} - \frac{(\lambda - \delta)^2}{\eta^2} \right] \right\}$$

Nesta seção usamos um procedimento um pouco diferente para encontrar os hiperparâmetros das distribuições de α e λ . O usuário fornece a probabilidade do modelo ser unimodal e a probabilidade do modelo ser assimétrico à direita ou à esquerda denotada por p_u e p_a , respectivamente, ou seja, é dada $p_u = P(\alpha < 1)$ e $p_a = P(\lambda < 0)$. Essa informação é importante, pois, em muitas situações, um modelo PNBA unimodal se ajustará bem aos dados assim como um modelo PNBA bimodal com o sinal de λ oposto ao PNBA unimodal. Um exemplo de como a situação descrita acontece é ilustrado na Figura 4.2 abaixo. Dois ajustes, um unimodal assimétrico à esquerda ($\alpha < 1, \lambda < 0$) e um bimodal com a curva à direita mais concentrada ($\alpha > 1, \lambda > 0$) é apresentado:

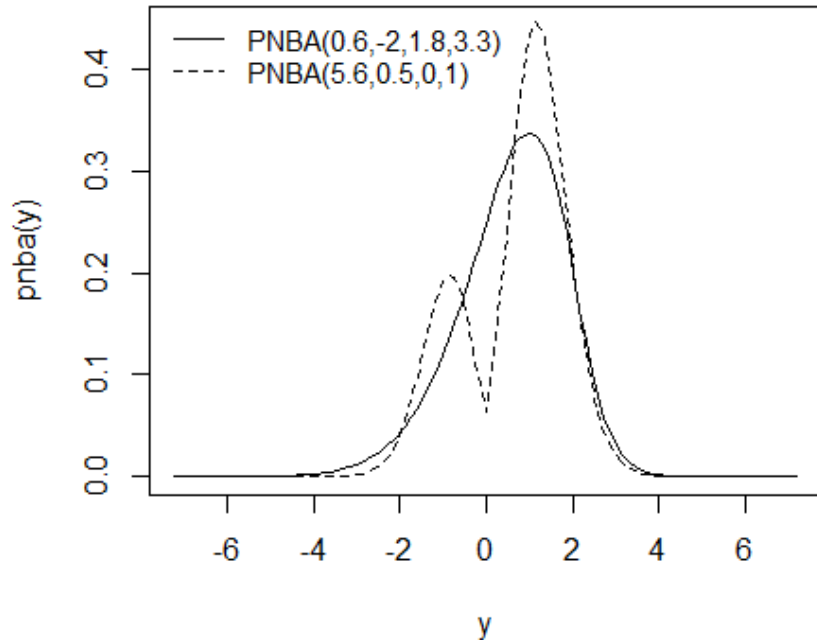


Figura 4.2: Densidade das distribuições PNBA(0.6,-2,1.7,3.27) e PNBA(5.6,0.5,0,1).

Além de fornecer essa probabilidade, o usuário deve fornecer também a probabilidade $p_\alpha = P(\alpha < q_\alpha | \alpha > 1)$, em que q_α é um número maior que 1. Esta probabilidade é interpretada como um intervalo a priori para α dado que o modelo é bimodal. Ou seja, q_α é o valor pelo qual o usuário crê fortemente que α não o ultrapassará dado a bimodalidade. Por isso, p_α deve ser um valor bem elevado. Para λ , o usuário deve fornecer a probabilidade p_λ da seguinte maneira:

$$p_\lambda = \begin{cases} P(\lambda > -q_\lambda | \lambda < 0), & \text{se } P(\lambda < 0) > 0.5 \\ P(\lambda < q_\lambda | \lambda > 0), & \text{se } P(\lambda > 0) > 0.5 \\ 1 - P(|\lambda| < q_\lambda), & \text{se } P(\lambda > 0) = 0.5 \end{cases} \quad (4.4)$$

em que $q_\lambda > 0$. Essa probabilidade p_λ é interpretada da mesma forma que interpretamos p_α . Se o usuário crê por exemplo que a distribuição é assimétrica à esquerda ($\lambda < 0$), ele deve estipular um valor crítico para λ menor que 0, afirmando que λ não será menor que esse valor caso a crença dele esteja correta. Ou seja, a distribuição não terá uma assimetria tão forte quanto a assimetria de uma PNBA cujo λ é $-q_\lambda$ com probabilidade p_λ . Na situação da terceira equação da expressão (4.4), a crença quanto ao sinal da assimetria é não informativa. Por isso que nesse caso utilizaram-se dois pontos extremos (q_λ e $-q_\lambda$).

Os valores mais recomendados para q_α são aqueles tais que $q_\alpha \geq 10$, pois, como já foi visto no modelo PNB, valores de $\alpha > 10$ implicam uma distribuição cujos picos são muito afastados, o que não acontece frequentemente na prática. Já as quantidades sugeridas para q_λ são tais que $q_\lambda \geq 2$, pois a flexibilidade do modelo é muito grande no caso em que $-2 < \lambda < 2$. Para um modelo com $\alpha = 1$ (*skew-normal*), o coeficiente de assimetria da distribuição denotado por

CA para $-2 < \lambda < 2$ varia entre $(-0.9, 0.9)$, sendo que a classe de distribuição skew-normal compreende níveis de assimetria entre $(-0.99, 0.99)$. Isso significa que a escolha de $\pi(\lambda)$ não será boa nos casos em que $|CV| > 0.9$, mesmo que a crença na direção de λ seja correta. No caso em que a crença do usuário favorece modelos bimodais, recomenda-se utilizar $q_\lambda \geq 1.5$. Sendo bem conservador e supondo um modelo bimodal fraco com $\alpha = 3$, temos que $P(Y > \mu)$ varia entre $(0.15, 0.85)$ se $-1.5 < \lambda < 1.5$. Para valores de α maiores, $P(Y > \mu)$ tem uma maior variação no intervalo de λ estabelecido. Se $\alpha = 9$ (bimodalidade forte), $0.05 < P(Y > \mu) < 0.95$. A quantidade $P(Y > \mu)$ representa o quanto de probabilidade se concentra na curva da direita do modelo bimodal. Além de $P(Y > \mu)$ apresentar uma boa variação no intervalo de λ no qual se crê fortemente, o intervalo $(-1.5, 1.5)$ para λ compreende todos os valores do CA caso $\alpha > 4.5$. Andrade et al. (2016) obteve o CA do modelo para diversos valores de α e λ . Na Figura 4.3 e 4.4 são ilustradas curvas do modelo PNBA com α e λ variando dentro do que é recomendado para eles.

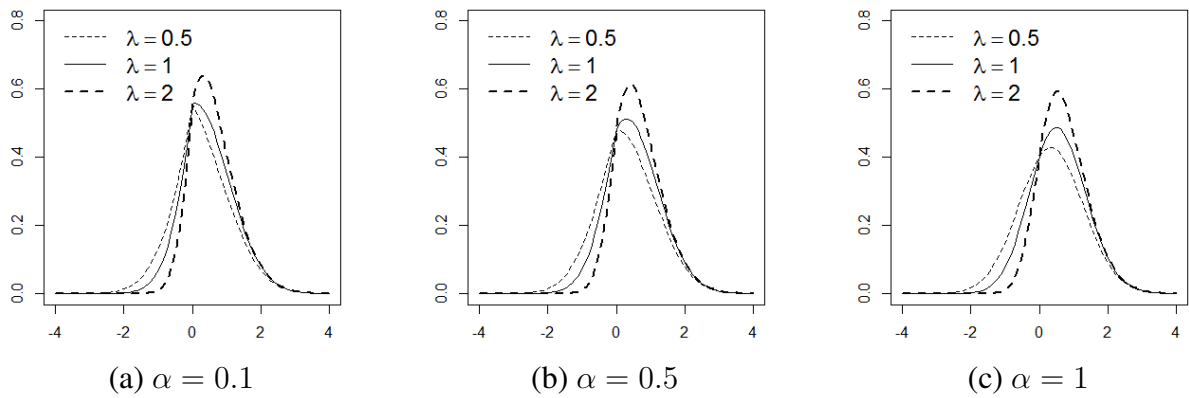


Figura 4.3: Densidade da potência normal bimodal assimétrica em casos unimodais.

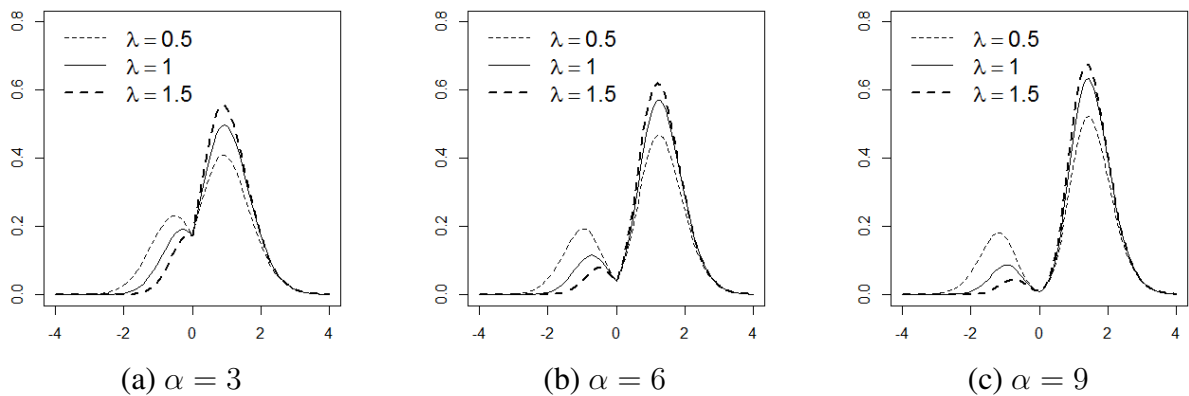


Figura 4.4: Densidade da potência normal bimodal assimétrica em casos bimodais.

Para encontrar os hiperparâmetros de α (ζ e γ) e λ (δ e τ), precisamos resolver os sistemas:

$$\begin{cases} p_u = P(\alpha < 1) \\ p_\alpha = P(\alpha < q_\alpha | \alpha > 1), \end{cases}$$

$$\begin{cases} p_a = P(\lambda < 0) \\ p_\lambda = \begin{cases} P(\lambda > -q_\lambda | \lambda < 0), \text{ se } P(\lambda < 0) > 0.5 \\ P(\lambda < q_\lambda | \lambda > 0), \text{ se } P(\lambda > 0) > 0.5 \\ 1 - P(|\lambda| < q_\lambda), \text{ se } P(\lambda > 0) = 0.5 \end{cases} \end{cases}$$

Nesta Seção, também usamos a força da crença do usuário para estabelecer a probabilidade p_a e p_u . A Tabela 4.2 ilustra as crenças com as probabilidades:

Crença (α)	p_u	Crença (λ)	p_a
Forte Bimodal	0.1	Forte Assimetrica à Direita	0.1
Moderada Bimodal	0.3	Moderada Assimetrica à Direita	0.3
Não Informativa	0.5	Não Informativa	0.5
Moderada Unimodal	0.7	Moderada Assimetrica à Esquerda	0.7
Moderada Unimodal	0.9	Moderada Assimetrica à Esquerda	0.9

Tabela 4.2: Distribuições a priori informativa para α .

Distribuição a Priori Conjunta

Considerando que o vetor (α, λ) é independente de μ e σ^2 e que as prioris de μ e σ^2 são as mesmas da Seção 2. Então, temos:

$$\begin{aligned} \pi(\theta) &= \pi(\alpha, \lambda, \mu, \sigma^2) \\ &= \pi(\alpha, \lambda)\pi(\mu)\pi(\sigma^2) \\ &= \pi(\alpha, \lambda) \frac{b^a}{\sqrt{2\pi}\Gamma(a)} (\sigma^2)^{a-1} \exp \left\{ -\frac{1}{2} \frac{(\mu - \xi)^2}{\tau^2} - b\sigma^2 \right\}. \end{aligned}$$

4.2.3 Distribuições a Posteriori

Posteriori Conjunta

As distribuições a posteriori conjunta dos parâmetros α , λ , μ , σ^2 são dadas por:

$$\begin{aligned}
\pi(\theta|\mathbf{y}) &= \frac{\pi(\alpha, \lambda, \mu, \sigma^2)L(\alpha, \lambda, \mu, \sigma^2|\mathbf{y})}{\pi(\mathbf{y})} \\
&\propto \pi(\alpha, \lambda)\pi(\mu)\pi(\sigma^2)L(\theta|\mathbf{y}) \\
&\propto \pi(\alpha, \lambda) \exp \left\{ -\frac{1}{2} \frac{(\mu - \xi)^2}{\tau^2} - \frac{1}{2} \frac{(y_i - \mu)^2}{\sigma^2} - b\sigma^2 \right\} \\
&\quad \alpha^{n-1} c_\alpha^n (\sigma^2)^{a - (\frac{n}{2} + 1)} \left\{ \prod_{i=1}^n \Phi \left(\frac{|y_i - \mu|}{\sigma} \right) \right\}^\alpha.
\end{aligned}$$

A distribuição preditiva de \mathbf{y} é dada por:

$$\pi(\mathbf{y}) = \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \int_0^\infty \pi(\alpha, \lambda, \mu, \sigma^2)L(\alpha, \lambda, \mu, \sigma^2|\mathbf{y}) d\alpha d\lambda d\mu d\sigma^2.$$

A resolução da integral acima só pode ser obtida via métodos numéricos.

Posteriores Condicionais

Uma vez que não foi possível obter analiticamente a distribuição preditiva e com o fato de que a expressão da posteriori não é simples, métodos numéricos devem ser implementados para gerar amostras da posteriori conjunta. Algo que facilita a implementação do método que vamos utilizar é a obtenção das distribuições a posteriori condicionais, dadas por:

$$\begin{aligned}
\pi(\alpha|\lambda, \mu, \sigma^2, \mathbf{y}) &= \frac{\pi(\alpha, \lambda, \mu, \sigma^2)L(\alpha, \lambda, \mu, \sigma^2|\mathbf{y})}{\pi(\mu, \sigma^2, \mathbf{y})} \\
&\propto \pi(\alpha, \lambda)L(\alpha, \mu, \sigma^2|\mathbf{y}) \\
&\propto \pi(\alpha, \lambda) \alpha^{n-1} c_\alpha^n \left\{ \prod_{i=1}^n \Phi \left(\frac{|y_i - \mu|}{\sigma} \right) \right\}^\alpha.
\end{aligned}$$

$$\begin{aligned}
\pi(\lambda|\alpha, \mu, \sigma^2, \mathbf{y}) &= \frac{\pi(\alpha, \lambda, \mu, \sigma^2)L(\alpha, \lambda, \mu, \sigma^2|\mathbf{y})}{\pi(\mu, \sigma^2, \mathbf{y})} \\
&\propto \pi(\alpha, \lambda)L(\alpha, \mu, \sigma^2|\mathbf{y}) \\
&\propto \pi(\alpha, \lambda) \left\{ \prod_{i=1}^n \Phi \left(\lambda \frac{y_i - \mu}{\sigma} \right) \right\}.
\end{aligned}$$

$$\begin{aligned}
\pi(\mu|\alpha, \lambda, \sigma^2, \mathbf{y}) &= \frac{\pi(\alpha, \lambda, \mu, \sigma^2)L(\alpha, \lambda, \mu, \sigma^2|\mathbf{y})}{\pi(\alpha, \sigma^2, \mathbf{y})} \\
&\propto \pi(\mu)L(\alpha, \lambda, \mu, \sigma^2|\mathbf{y}) \\
&\propto \exp \left\{ -\frac{1}{2} \frac{(\mu - \xi)^2}{\tau^2} - \frac{\sum_{i=1}^n (y_i - \mu)^2}{2\sigma^2} \right\} \left\{ \prod_{i=1}^n \Phi \left(\frac{|y_i - \mu|}{\sigma} \right) \right\}^{\alpha-1} \\
&\quad \left\{ \prod_{i=1}^n \Phi \left(\lambda \frac{y_i - \mu}{\sigma} \right) \right\}
\end{aligned}$$

$$\begin{aligned}
\pi(\sigma^2|\alpha, \lambda, \mu, \mathbf{y}) &= \frac{\pi(\alpha, \lambda, \mu, \sigma^2)L(\alpha, \lambda, \mu, \sigma^2|\mathbf{y})}{\pi(\alpha, \mu, \mathbf{y})} \\
&\propto \pi(\sigma^2)L(\alpha, \lambda, \mu, \sigma^2|\mathbf{y}) \\
&\propto \exp\left\{-\frac{\sum_{i=1}^n (y_i - \mu)^2}{2\sigma^2} - b\sigma^2\right\} \sigma^{a-(\frac{n}{2}+1)} \left\{\prod_{i=1}^n \Phi\left(\frac{|y_i - \mu|}{\sigma}\right)\right\}^{\alpha-1} \\
&\quad \left\{\prod_{i=1}^n \Phi\left(\lambda \frac{y_i - \mu}{\sigma}\right)\right\}.
\end{aligned}$$

4.3 Aspectos Computacionais

4.3.1 O Algoritmo MCMC

Uma vez que não foi possível obter uma forma fechada para a distribuição a posteriori dos parâmetros da PNBA, métodos de simulação estocásticos foram utilizados para a inferência. O método empregado foi o MCMC, com o auxílio do algoritmo de MH.

Distribuições Propostas

As distribuições propostas utilizadas no algoritmo de MH foram as mesmas utilizadas na Seção 2 quando nos referimos às propostas de α e σ^2 . A distribuição proposta de μ não foi empregada no algoritmo. Ao invés disso, usamos a distribuição conjunta proposta do vetor (λ, μ) , ao qual denotamos de $q_{\lambda, \mu}(\lambda_{t+1}^p, \mu_{t+1}^p | \lambda_t, \mu_t)$. A decisão de optar pela conjunta se deve ao fato de que existe um método simples de implementar que estabelece uma estrutura de covariância entre distribuições propostas cujo suporte é o conjunto dos \mathbb{R} . Esse método foi proposto por Haario et al. (2001) e se utiliza da matriz de covariância amostral da posteriori na estruturação da matriz de covariância da distribuição proposta. Dessa forma, a proposta fica mais adequada, uma vez que passa a se assemelhar mais à posteriori conjunta de μ e λ .

A implementação do algoritmo de Haario consiste em primeiramente escolher uma matriz de covariância inicial para a proposta, que receberá sua primeira atualização após um certo número n_0 de iterações do MH. Após n_0 iterações, o algoritmo deve ser atualizado em todas as iterações seguintes. Em cada atualização, utilizamos a matriz de covariância da distribuição a posteriori adquirida através do algoritmo MH até o certo momento. A expressão da matriz de covariância da proposta é dada por:

$$\begin{cases} C_t = COV_t \times s + sI\epsilon, \text{ se } t > n_0 \\ C_t = COV_0, \text{ se } t \leq n_0, \end{cases}$$

em que:

- COV_t é a matriz de covariância amostral do algoritmo MCMC com t iterações.
- COV_0 é a matriz inicial atribuída à matriz de covariância.

- s é um fator multiplicativo da covariância amostral.
- I é a matriz identidade de dimensão d .
- ϵ é um valor suficientemente pequeno maior que 0 para que a matriz C_t seja não singular.

Baseada no algoritmo de Haario, a distribuição proposta conjunta de λ e μ é dada por:

$$q_{\lambda,\mu}(\lambda_{t+1}^p, \mu_{t+1}^p | \lambda_t, \mu_t) \sim N \left(\begin{bmatrix} \lambda_t^p \\ \mu_t^p \end{bmatrix}, C_t \right),$$

em que C_t é a matriz de covariância de dimensões 2×2 do algoritmo de Haario no tempo t . O valor de s escolhido foi a variância sintonizada do algoritmo de Roberts e Rosenthal. Isso significa que abordamos s de uma maneira um pouco diferente, não o tratamos como uma constante, mas sim como uma variável aleatória no tempo que converge para um valor quando t é suficientemente grande. Quanto ao valor de ϵ , atribuímos a ele o valor 0, uma vez que ao desprezar ϵ a matriz C_t não teve problema de singularidade.

Outro ponto que devemos considerar é o valor inicial que atribuímos à matriz C_0 . Com uma boa escolha para este, garantimos que as n_0 primeiras observações da cadeia serão não tão autocorrelacionadas de modo que se possa estimar razoavelmente bem a covariância a posteriori através de COV_1 . Estimando bem COV_1 , C_t convergirá mais rápido. A matriz que atribuímos a C_0 foi de tal modo que:

$$C_0 = \begin{bmatrix} 1 & 0 \\ 0 & \frac{\hat{\sigma}^2}{n} \end{bmatrix}$$

A matriz C_0 representa a crença que temos a respeito da matriz de covariância da distribuição a posteriori. O primeiro e segundo elemento da diagonal principal representam respectivamente a crença sobre a variância de $\pi(\lambda|\mathbf{y})$ e $\pi(\mu|\mathbf{y})$. Atribuímos 1 para a variância de μ , pois segundo Andrade et al. (2016), α geralmente está entre (-2,2). Assim, se a distribuição a priori de μ é uma normal padrão, em que sabemos que tem variância 1, então $P(-2 < \mu < 2) = 0.95$. A variância que cremos da posteriori de μ é a variância estimada da média amostral dada por $\frac{\hat{\sigma}^2}{n}$. Quanto à covariância entre μ e λ , estabelecemos uma crença sem nenhuma informação. Atribuímos 0 para a covariância a fim de não favorecer sua direção.

4.3.2 Diagnóstico de Convergência

Com o intuito de averiguar se a cadeia gerada por MH converge, utilizamos o método de Gelman e Rubin (1992). Consideramos duas amostras para esse propósito. Amostras geradas de um modelo PNBA(0.6,-2,1.8,3.3) e PNBA(5.6,0.5,0,1), com tamanho $n = 50$ cada. Esses modelos foram escolhidos porque ambos apresentam parametrizações bem distintas e curvas bem semelhantes conforme já ilustramos na Figura 4.2. Dessa forma, o método de Gelman

e Rubin pode constatar que as cadeias não convergem para o mesmo ponto. Por exemplo, se iniciarmos a cadeia no ponto $(\alpha = 0.6, \lambda = -2, \mu = 1.8, \sigma^2 = 3.3)$, não é absurdo acreditar que esta gerará apenas para valores próximos desse ponto. Do mesmo modo, se inicializarmos a cadeia no ponto $(\alpha = 5.6, \lambda = 0.5, \mu = 0, \sigma^2 = 1)$, podemos crer que esta desprezará pontos próximos de $(\alpha = 0.6, \lambda = -2, \mu = 1.8, \sigma^2 = 3.3)$.

Quanto às prioris $\pi(\alpha, \lambda)$ adotadas na análise de convergência, consideramos duas para cada amostra: a priori não informativa uniforme para p e a priori informativa com crença não informativa em $\alpha < 1$ e $\lambda < 0$. Atribuímos os valores $q_\alpha = 20, p_\alpha = 0.95, q_\lambda = 3, p_\lambda = 0.95$. Isso significa que encontramos os hiperparâmetros da distribuição $\pi(\alpha, \lambda)$ usando as restrições: $P(\alpha < 1) = P(\lambda < 0) = 0.5$ e $P(\alpha < 20 | \alpha > 1) = P(|\lambda| > 3) = 0.95$. Dessa forma, as distribuições de α e λ são tais que $\alpha \sim \text{Log} - \text{Normal}(0, 1.528)$ e $\lambda \sim \text{Normal}(0, 1.531)$. Geramos três cadeias que se inicializam em pontos distintos conforme ilustra a Tabela 4.3. As duas primeiras inicializam nos reais valores dos parâmetros das distribuições PNBA geradas. A terceira inicializa em um ponto bem distinto de ambos. Cada cadeia tem tamanho $N = 250000$ e o *burn-in* estipulado (período de aquecimento) é de 50000 iterações.

amostra	α_0	λ_0	μ_0	σ_0^2
1	0.6	-2	1.8	3.3
2	5.6	0.5	0	1
3	15	3	10	10

Tabela 4.3: Valores iniciais para o teste de convergência.

A Figura 4.6 ilustra o comportamento de uma das cadeias de Markov gerada pelo MCMC para os parâmetros α e λ . Analisando as Figura 4.6 (c) e 4.6 (d), notamos que em diferentes períodos de tempo as cadeias de α e λ convergem para um ponto diferente. Tal comportamento indica que a distribuição a posteriori conjunta tem duas modas locais. Todavia, ao analisar a cadeia inteira de α a partir da Figura 4.6 (a), observamos que há um terceiro ponto em que α atinge valores bastante elevados, indicando que há mais uma moda local na distribuição. Uma análise conjunta das cadeias mostram que a correlação existente entre os parâmetros é muito forte. Na Tabela 4.4 podemos verificar essa correlação. Ao observar a cadeia completa para valores de α que não são elevados na Figura 4.6 (e), não é possível notar o comportamento visto na Figura 4.6 (c), que mostra a forte autocorrelação e a presença de modas locais. Isso aconteceu por conta de que o número de iterações da cadeia completa é muito maior do que o número de iterações das cadeias da Figura 4.6 (a). Por isso que foi necessário utilizar N muito grande.

As Figuras 4.7, 4.8 e 4.9 ilustram gráficos das cadeias se inicializando em diferentes pontos. Através das Figuras 4.7 e 4.8 não é possível averiguar o comportamento das cadeias de α para valores pequenos deste parâmetro. No entanto, temos a Figura 4.9 que nos auxilia nisso. Através da análise conjunta dos gráficos representados pelas três figuras, podemos perceber que a trajetória das três cadeias são mais semelhantes nos casos em que a priori utilizada é infor-

Parâmetro	α	λ	μ	σ^2
α	1.00	0.66	-0.71	-0.74
λ	0.66	1.00	-0.95	-0.83
μ	-0.71	-0.95	1.00	0.83
σ^2	-0.74	-0.83	0.83	1.00

Tabela 4.4: Correlação das distribuições a posteriori dos parâmetros.

mativa. A estimativa da redução de escala potencial \hat{R} na Tabela 4.5 indica que há cadeias não convergem para o mesmo ponto na amostra unimodal para as duas prioris consideradas. Todavia, ao analisar os gráficos do MCMC quando a priori utilizada é a informativa, temos a impressão de que as cadeias seguem percursos semelhantes. Suspeitamos então que o método de Gelman e Rubin não identificou a convergência por causa da forte autocorrelação existente entre as cadeias. Para utilizar o método com mais precisão, geramos o MCMC novamente para a mesma amostra unimodal aumentando o número de iterações para 500000. Desta vez, o método de Gelman e Rubin apontou que a convergência existe. O \hat{R} estimado para a cadeia geradora de α foi de 1.01. O gráfico da cadeia pode ser visto na Figura 4.10.

Para solucionar o problema de identificabilidade e de autorrelação entre as cadeias sugerimos utilizar prioris informativas. Na seção seguinte, mostramos através de uma análise de sensibilidade prioris informativas que torna o modelo mais identificável. Se o problema de identificabilidade do modelo não existir ou for resolvido com prioris informativas, pode ser que a cadeia gere alguns pontos elevados de α . Este α está associado a um λ um pouco distante de zero, exemplos disso podem ser observados nas cadeias que geramos para a análise de convergência. Uma alternativa que podemos ter para a cadeia não visitar estes lugares é concentrar menos probabilidade ainda em pontos elevados de α e $|\lambda|$. Sugerimos mudar q_α para 15 ao invés de 20 e / ou mudar q_λ para 2 ao invés de 3. Assim, dificilmente estes pontos serão visitados. Todavia, tal decisão pode diminuir o poder preditivo do modelo, uma vez que a informação a priori é maior. Caso o problema de identificabilidade não for resolvido, a estimativa dos parâmetros fica comprometida e então, neste caso, é mais conveniente utilizar a distribuição preditiva para modelar os dados. Na Figura 4.5 abaixo, temos a distribuição preditiva dos dados gerados a partir do modelo PNBA(5.6,0.5,0,1) de tamanho $n = 50$.

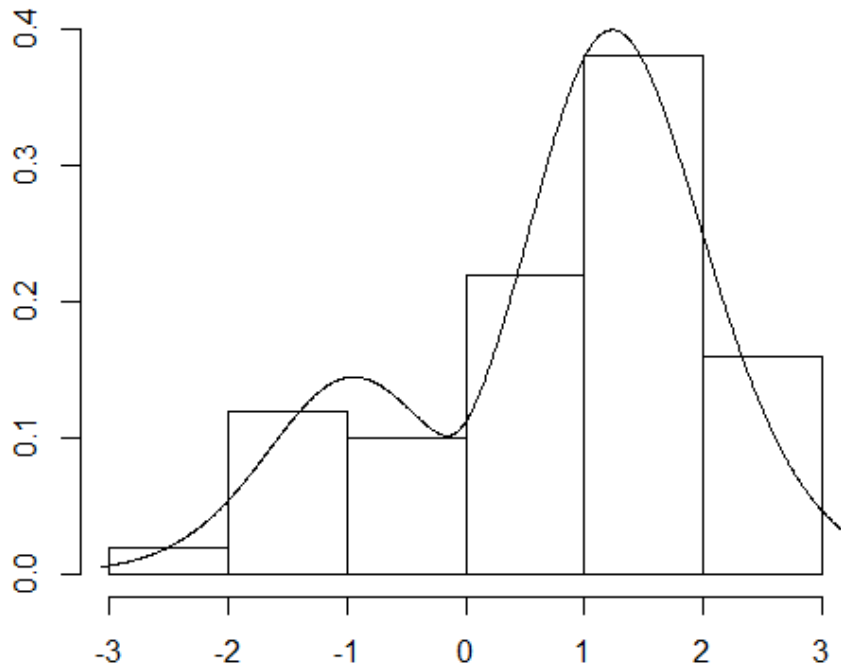


Figura 4.5: Distribuição preditiva dos dados. Amostra gerada de uma $PNBA(5.6, 0.5, 0, 1)$ de tamanho $n = 50$.

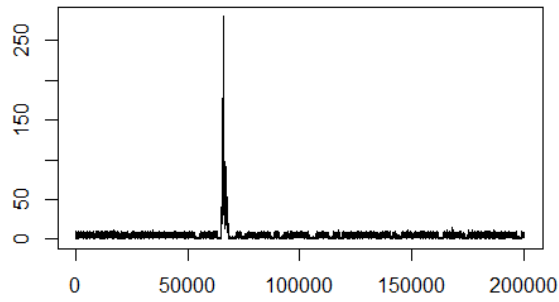
Amostra	Parâmetro	Priori utilizada	
		Não Informativa	Informativa
PNBA(0.6, -2, 1.8, 3.3)	α	1.18 (1.54)	1.25 (1.56)
	λ	1.85 (3.12)	1.01 (1.01)
	μ	1.39 (2.17)	1.07 (1.16)
	σ^2	1.27 (1.8)	1.02 (1.04)
PNBA(5.6, 0.5, 0, 1)	α	1 (1)	1 (1)
	λ	1 (1)	1 (1.01)
	μ	1 (1)	1 (1)
	σ^2	1 (1)	1 (1)

Tabela 4.5: \hat{R} estimado em cada cenário. Estimativa pontual (fora dos parênteses). Limite superior do intervalo de confiança (entre parênteses).

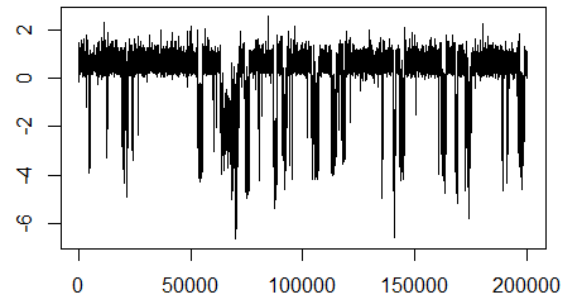
4.4 Estudo de Simulação

4.4.1 Análise de Sensibilidade

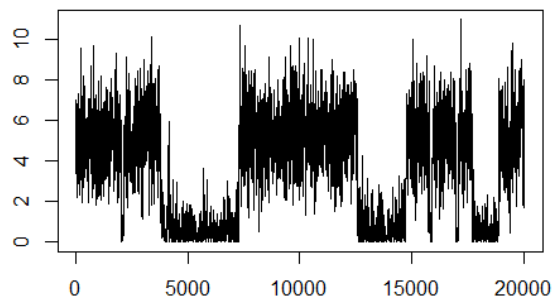
Para a análise de sensibilidade do modelo PNBA, geramos aleatoriamente amostras de tamanho $n = 50$, $n = 100$ e $n = 200$ dos modelos $PNBA(0.6, -2, 1.8, 3.3)$ (unimodal assimétrico à esquerda) e $PNBA(5.6, 0.5, 0, 1)$ (bimodal com a curva à direita mais concentrada). Para cada amostra foram avaliadas as posteriores baseadas em prioris com os tipos de crenças já mencionados. Cada priori $\pi(\alpha, \lambda)$ considerada tem a mesma força de crença para $\alpha < 1$ e $\lambda < 0$. O que muda entre elas é o nível da crença. Os valores atribuídos para p_α , p_λ , q_α e q_λ



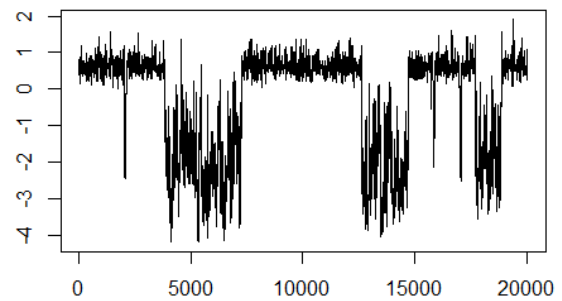
(a) Cadeia geradora de $\pi(\alpha|y)$.



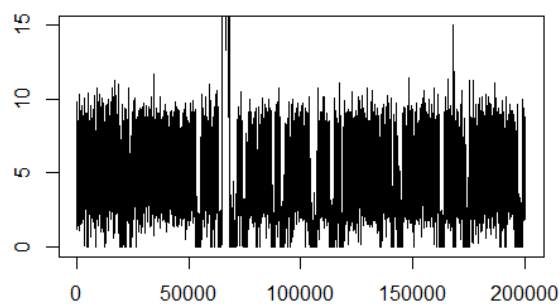
(b) Cadeia geradora de $\pi(\lambda|y)$.



(c) Sequência de 20000 iterações da cadeia geradora de $\pi(\alpha|y)$.



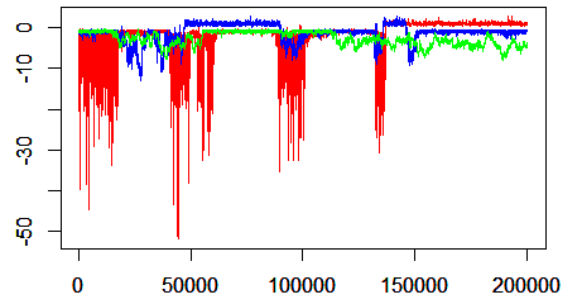
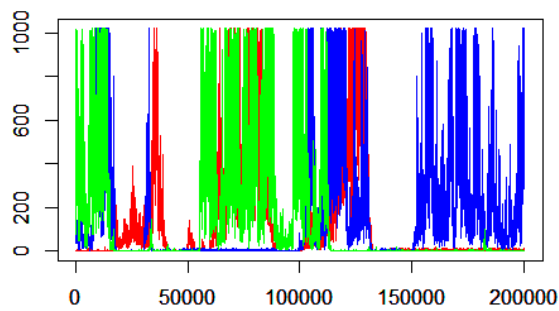
(d) Sequência de 20000 iterações da cadeia geradora de $\pi(\lambda|y)$.



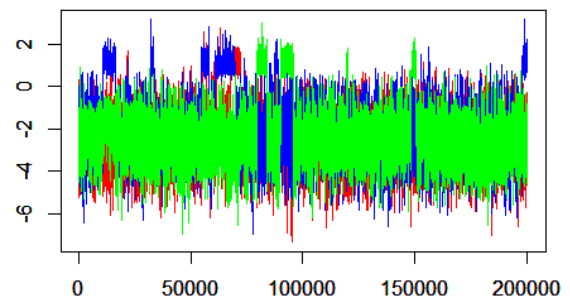
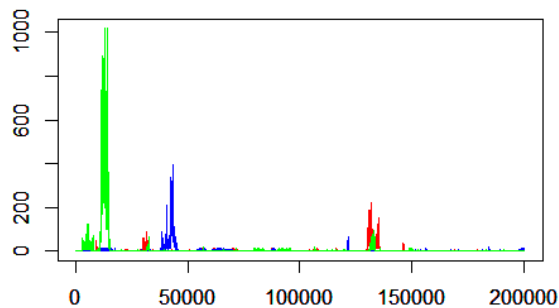
(e) Cadeia completa geradora de $\pi(\lambda|y)$. Eixo y limitado no intervalo $(0,15)$.

Figura 4.6: Cadeia gerada pelo MCMC das distribuições $\pi(\alpha|y)$ e $\pi(\lambda|y)$ utilizando a priori informativa. Dados gerados a partir de uma amostra PNBA(5.6,0.5,0,1) de tamanho $n = 50$.

foram tais que $p_\alpha = 0.95$, $p_\lambda = 0.95$, $q_\alpha = 20$ e $q_\lambda = 3$ para crenças que favorecem a unimodalidade e $q_\lambda = 2$ para as que favorecem a bimodalidade. Isso quer dizer que:



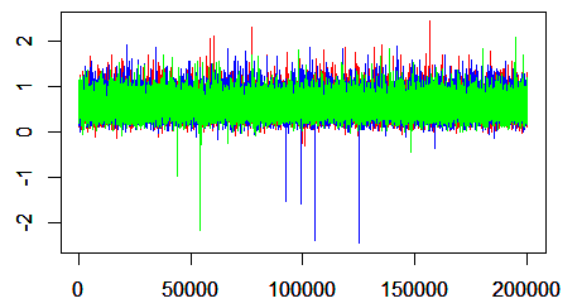
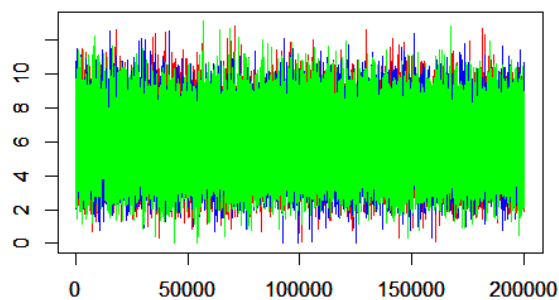
(a) $\pi(\alpha|\mathbf{y})$ utilizando a priori não informativa. (b) $\pi(\lambda|\mathbf{y})$ utilizando a priori não informativa.



(c) $\pi(\alpha|\mathbf{y})$ utilizando a priori informativa.

(d) $\pi(\lambda|\mathbf{y})$ utilizando a priori informativa.

Figura 4.7: Cadeia gerada pelo MCMC das distribuições $\pi(\alpha|\mathbf{y})$. Dados gerados a partir de uma amostra PNBA(0.6,-2,1.8,3.3) de tamanho $n = 50$.

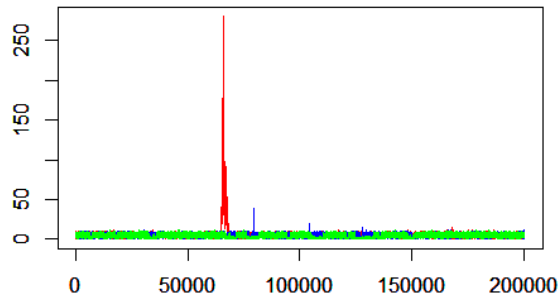


(a) $\pi(\alpha|\mathbf{y})$ utilizando a priori não informativa.

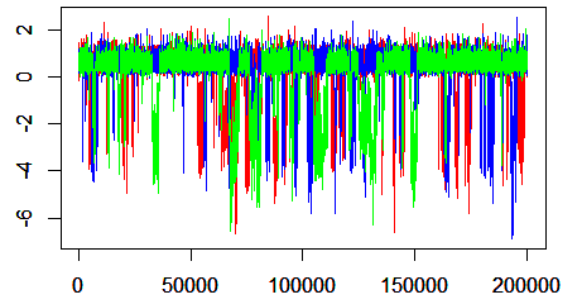
(b) $\pi(\lambda|\mathbf{y})$ utilizando a priori não informativa.

$$P(\alpha < 20 | \alpha > 1) = 0.95$$

$$P(\lambda > -3 | \lambda < 0) = 0.95, \text{ se } P(\lambda < 0) > 0.5 \text{ e } P(\alpha < 1) > 0.5$$

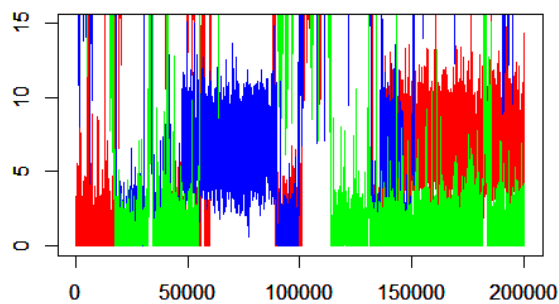


(c) $\pi(\alpha|y)$ utilizando a priori informativa.

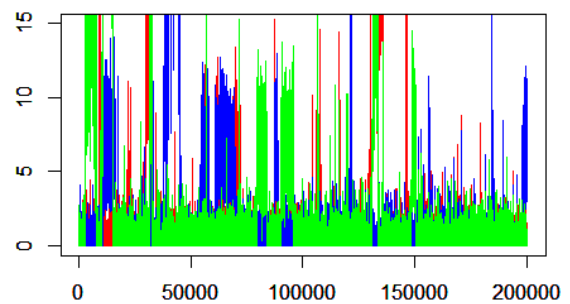


(d) $\pi(\lambda|y)$ utilizando a priori informativa.

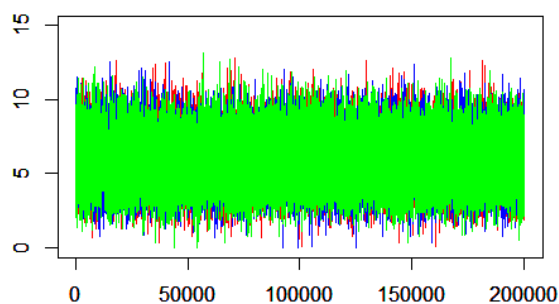
Figura 4.8: Cadeia gerada pelo MCMC das distribuições $\pi(\alpha|y)$ e $\pi(\lambda|y)$. Dados gerados a partir de uma amostra PNBA(5.6,0.5,0,1) de tamanho $n = 50$.



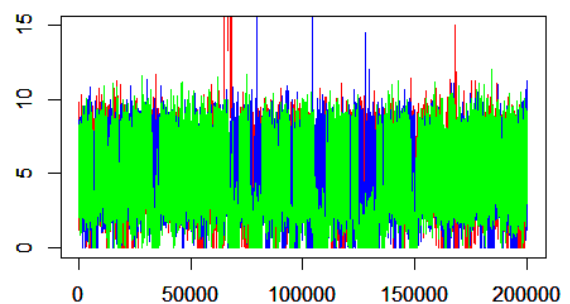
(a) $\pi(\alpha|y)$ utilizando a priori não informativa. Dados gerados da PNBA(0.6,-2,1.8,3.3).



(b) $\pi(\alpha|y)$ utilizando a priori informativa. Dados gerados da PNBA(0.6,-2,1.8,3.3).



(c) $\pi(\alpha|y)$ utilizando a priori não informativa. Dados gerados da PNBA(5.6,0.5,0,1).



(d) $\pi(\alpha|y)$ utilizando a priori informativa. Dados gerados da PNBA(5.6,0.5,0,1).

Figura 4.9: Cadeia gerada pelo MCMC das distribuições $\pi(\alpha|y)$.

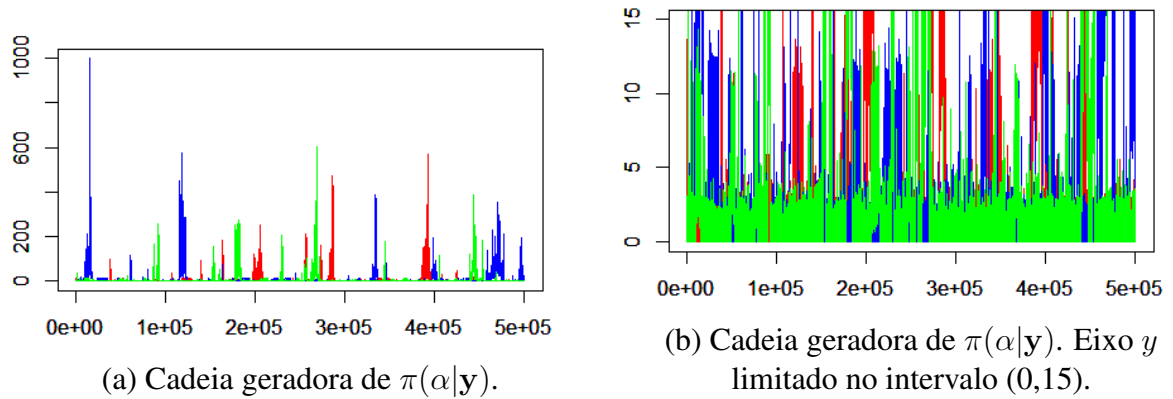


Figura 4.10: Cadeia gerada pelo MCMC da distribuição $\pi(\alpha|y)$ com 500000 iterações. Dados gerados a partir de uma amostra PNBA(0.6,-2,1.8,3.3) de tamanho $n = 50$.

$$P(\lambda < 2 | \lambda > 0) = 0.95, \text{ se } P(\lambda < 0) < 0.5 \text{ e } P(\alpha < 1) < 0.5$$

$$P(|\lambda| < 3) = 0.95, \text{ se } P(\lambda < 0) = 0.5 \text{ e } P(\alpha < 1) = 0.5$$

A Tabela 4.6 ilustra as crenças utilizadas junto com os hiperparâmetros de $\pi(\alpha, \lambda)$ associados e a Figura 4.11 os gráficos das distribuições a priori.

Crença	p_α / p_λ	q_α / q_λ	Hiperparâmetros	$E(\alpha) / E(\lambda)$	$Var(\alpha) / Var(\lambda)$
BF e AEF	0.1 / 0.1	20 / 2	$\zeta = 1.290$ $\gamma = 1.006$ $\delta = 0.861$ $\tau = 0.672$	6.025 / 0.861	36.767 / 0.451
BM e AEM	0.3 / 0.3	20 / 2	$\zeta = 0.672$ $\gamma = 1.282$ $\delta = 0.449$ $\tau = 0.856$	4.457 / 0.449	37.832 / 0.733
NI	0.5 / 0.5	20 / 3	$\zeta = 0$ $\gamma = 1.528$ $\delta = 0$ $\tau = 1.531$	3.878 / 0	39.346 / 2.343
UM e ADM	0.7 / 0.7	20 / 3	$\zeta = -0.955$ $\gamma = 1.820$ $\delta = -0.584$ $\tau = 1.113$	2.018 / -0.584	41.189 / 1.240
UF e ADF	0.9 / 0.9	20 / 3	$\zeta = -2.967$ $\gamma = 2.315$ $\delta = -0.997$ $\tau = 0.778$	0.750 / -0.997	43.911 / 0.605

Tabela 4.6: Prioris adotadas para a análise de sensibilidade. BF (bimodal forte). AEF (assimétrica à esquerda forte). BM (bimodal moderada). AEM (assimétrica à esquerda moderada). NI (não informativa). UM (unimodal moderada). ADM (Assimétrica à direita moderada). UF (unimodal forte). ADF (assimétrica à direita forte).

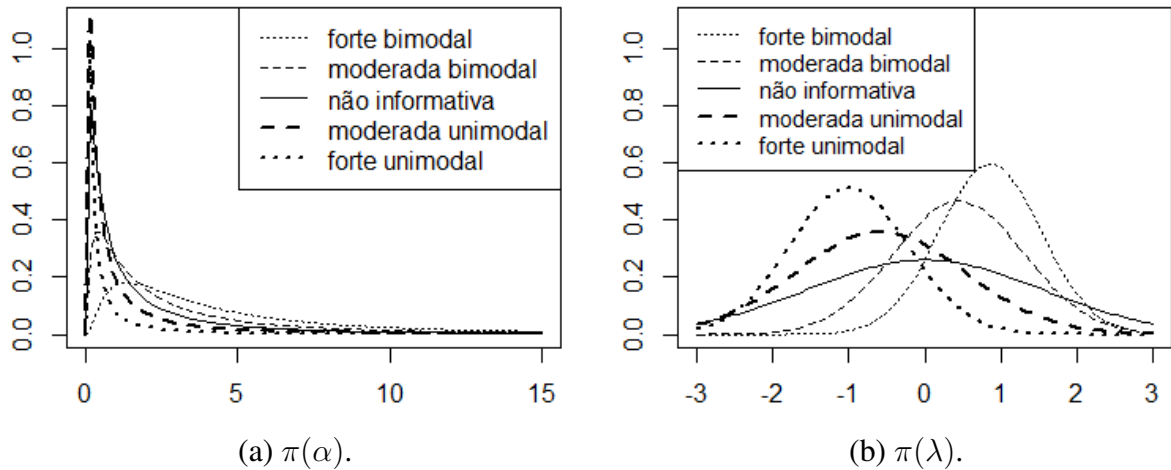


Figura 4.11: Distribuições a priori de α e λ consideradas na análise de sensibilidade.

Resultados da Análise de Sensibilidade

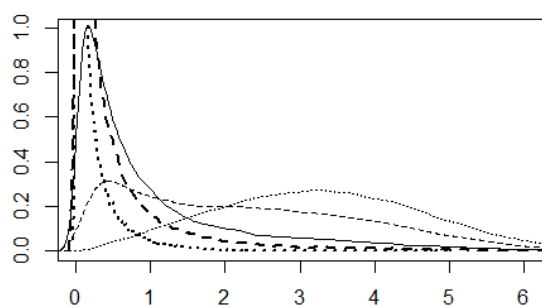
Através das Figuras 4.12 e 4.13, podemos perceber em geral um comportamento bimodal das distribuições a posteriori $\pi(\alpha|\mathbf{y})$ e $\pi(\lambda|\mathbf{y})$. Uma moda representa um ajuste unimodal e a outra um ajuste bimodal. Além disso, percebemos que $\pi(\alpha|\mathbf{y})$ e $\pi(\lambda|\mathbf{y})$ são bem sensíveis quanto à crença à priori. Observando os gráficos, vemos que a diferença das posteriores em geral se resume na probabilidade em modelos unimodais comparado a bimodais e na probabilidade em modelos assimétricos à esquerda comparado a assimétricos à direita. Por exemplo, se cremos moderadamente em $\lambda < 0$, a posteriori $\pi(\lambda|\mathbf{y})$ difere da posteriori baseada na crença moderada em $\lambda > 0$ no que diz respeito a $P(\lambda < 0|\mathbf{y})$. Em geral, as curvas em ambos os casos seguem o mesmo padrão e têm modas locais parecidas. Elas se diferenciam mais quanto à quantidade de probabilidade concentrada na curva à esquerda e na curva à direita. A sensibilidade da posteriori quanto à crença à priori se deve ao fato de que a verossimilhança não é eficiente para acusar se o modelo é unimodal ou bimodal. Logo, a informação a priori influencia de maneira significativa em $\pi(\alpha|\mathbf{y})$ e $\pi(\lambda|\mathbf{y})$. Apenas em uma das seis amostras as posteriores não diferiram.

Notamos também que em todos os cenários as posteriores baseadas em prioris não informativas ou informativas que acusam corretamente a unimodalidade ou bimodalidade e assimetria do modelo se comportam como o esperado. Essas posteriores concentram mais probabilidade no lugar “certo”. Isso significa que quando o modelo “real” é bimodal, a posteriori $\pi(\alpha|\mathbf{y})$ concentra mais probabilidade na curva à direita, ou seja, na curva cujos valores de α apontam para um modelo bimodal. Quando é unimodal, $\pi(\alpha|\mathbf{y})$ concentra mais probabilidade na curva à esquerda, ou seja, curva cujos valores de α apontam para um modelo unimodal. Da mesma forma, quando o modelo “real” é unimodal assimétrico à esquerda ou bimodal com a distribuição à esquerda mais concentrada, ou seja, quando λ verdadeiro é menor que zero, a posteriori $\pi(\lambda|\mathbf{y})$ é mais concentrada na curva à esquerda, aquela que acusa valores de $\lambda < 0$. Quando o

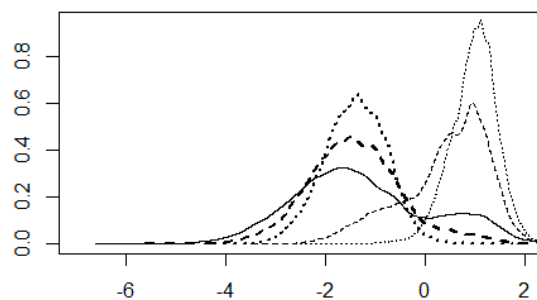
oposto acontece, ou seja, $\lambda > 0$, a posteriori $\pi(\lambda|\mathbf{y})$ é mais concentrada na curva à direita.

Quanto a influência da priori na identificabilidade, percebemos que uma crença moderada “correta” com relação a assimetria e o número de modas do modelo é suficiente para tornar a posteriori unimodal. Também vimos que para amostras com $n = 200$, até mesmo uma priori não informativa com relação a bimodalidade e assimetria torna o modelo identificável. Além disso, evidenciamos que quando o modelo verdadeiro é bimodal, o problema de identificabilidade é menos frequente comparado ao caso em que a distribuição teórica dos dados é unimodal.

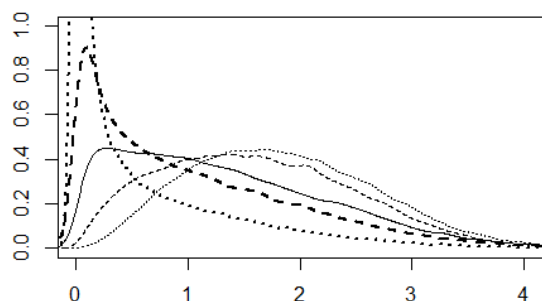
Outro ponto a se observar através das Figuras 4.12 e 4.13 e da Tabela 4.7 é que os modelos, apesar de alguns possuírem posteriors bem distintas, apresentam BIC semelhantes. A justificativa disso é que o que difere nessas posteriors é que uma acusa o modelo unimodal assimétrico à esquerda como sendo o melhor ajuste, e a outra, o modelo bimodal assimétrico à direita. Uma foi baseada em uma crença a priori unimodal e a outra baseada em uma crença a priori bimodal. Sabemos que o que o BIC mede é a qualidade do ajuste bayesiano. Ora, se modelos unimodais têm verossimilhança parecida com modelos bimodais, é de se esperar que uma posteriori que concentra altas probabilidades em modelos unimodais tenha o mesmo desempenho de posteriors que concentre altas probabilidades em modelos bimodais. Com isso, o BIC neste caso não é eficiente para apontar diferenças nas curvas das posteriors.



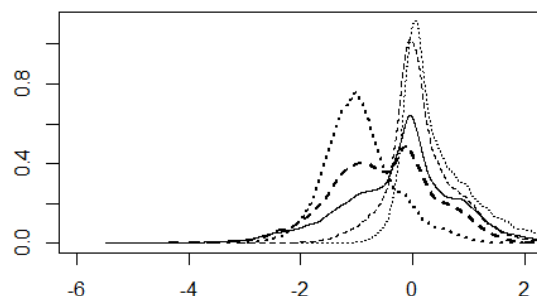
(a) $\pi(\alpha|\mathbf{y})$, sendo \mathbf{y} uma amostra da $PNBA(0.6, -2, 1.8, 3.3)$ com $n = 50$.



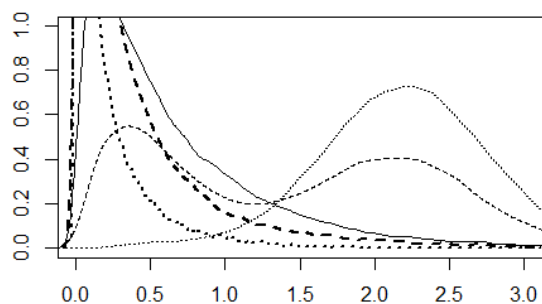
(b) $\pi(\lambda|\mathbf{y})$, sendo \mathbf{y} uma amostra da $PNBA(0.6, -2, 1.8, 3.3)$ com $n = 50$.



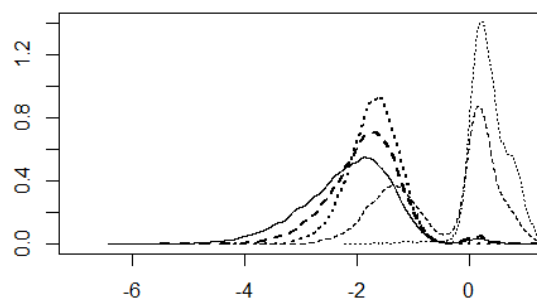
(c) $\pi(\alpha|\mathbf{y})$, sendo \mathbf{y} uma amostra da $PNBA(0.6, -2, 1.8, 3.3)$ com $n = 100$.



(d) $\pi(\lambda|\mathbf{y})$, sendo \mathbf{y} uma amostra da $PNBA(0.6, -2, 1.8, 3.3)$ com $n = 100$.

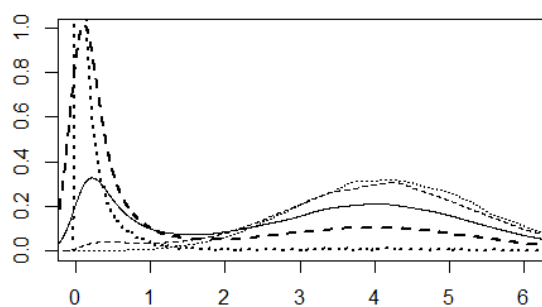


(e) $\pi(\alpha|\mathbf{y})$, sendo \mathbf{y} uma amostra da $PNBA(0.6, -2, 1.8, 3.3)$ com $n = 200$.

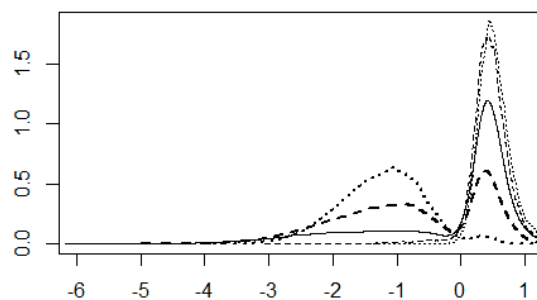


(f) $\pi(\lambda|\mathbf{y})$, sendo \mathbf{y} uma amostra da $PNBA(0.6, -2, 1.8, 3.3)$ com $n = 200$.

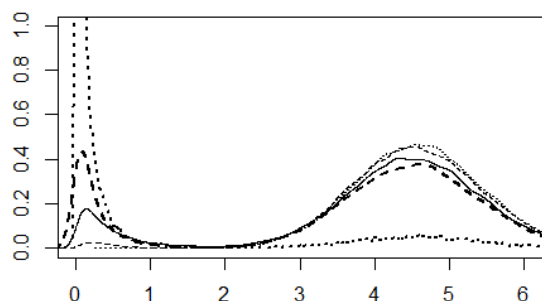
Figura 4.12: Distribuições a posteriori $\pi(\alpha|\mathbf{y})$ e $\pi(\lambda|\mathbf{y})$ em amostras da $PNBA(0.6, -2, 1.8, 3.3)$. Forte bimodal (linha fina pontilhada). Moderada bimodal (linha fina tracejada). Não informativa (linha contínua). Moderada unimodal (linha grossa tracejada). Forte unimodal (linha grossa pontilhada).



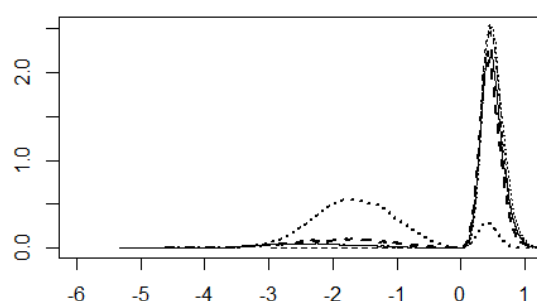
(a) $\pi(\alpha|\mathbf{y})$, sendo \mathbf{y} uma amostra da $PNBA(5.6, 0.5, 0, 1)$ com $n = 50$.



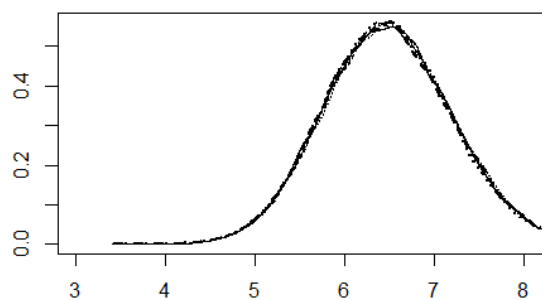
(b) $\pi(\lambda|\mathbf{y})$, sendo \mathbf{y} uma amostra da $PNBA(5.6, 0.5, 0, 1)$ com $n = 50$.



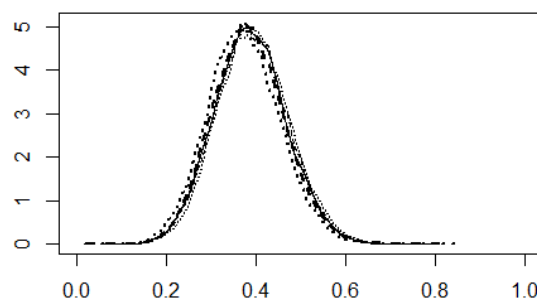
(c) $\pi(\alpha|\mathbf{y})$, sendo \mathbf{y} uma amostra da $PNBA(5.6, 0.5, 0, 1)$ com $n = 100$.



(d) $\pi(\lambda|\mathbf{y})$, sendo \mathbf{y} uma amostra da $PNBA(5.6, 0.5, 0, 1)$ com $n = 100$.



(e) $\pi(\alpha|\mathbf{y})$, sendo \mathbf{y} uma amostra da $PNBA(5.6, 0.5, 0, 1)$ com $n = 200$.



(f) $\pi(\lambda|\mathbf{y})$, sendo \mathbf{y} uma amostra da $PNBA(5.6, 0.5, 0, 1)$ com $n = 200$.

Figura 4.13: Distribuições a posteriori $\pi(\alpha|\mathbf{y})$ e $\pi(\lambda|\mathbf{y})$ em amostras da $PNBA(5.6, 0.5, 0, 1)$. Forte bimodal (linha fina pontilhada). Moderada bimodal (linha fina tracejada). Não informativa (linha contínua). Moderada unimodal (linha grossa tracejada). Forte unimodal (linha grossa pontilhada).

n	Amostra	Crença BF e AEF	Crença BM e AEM	Crença NI	Crença UM e ADM	UF e ADF
50	PNBA(0.6,-2,1.8,3.3)	178.59	179.15	178.03	177.66	177.31
100	PNBA(0.6,-2,1.8,3.3)	345.77	345.71	346.64	346.96	348.00
200	PNBA(0.6,-2,1.8,3.3)	690.13	688.70	685.54	685.56	685.45
50	PNBA(5.6,0.5,0,1)	162.76	163.10	164.81	166.75	168.08
100	PNBA(5.6,0.5,0,1)	327.98	327.99	328.23	327.92	336.00
200	PNBA(5.6,0.5,0,1)	622.58	622.60	622.62	622.60	622.61

Tabela 4.7: BIC das distribuições a posteriori em cada amostra.

4.5 Aplicações Com Dados Bimodais Assimétricos

Consideramos um conjunto de dados para a aplicação do modelo PNBA. Este apresenta um comportamento bimodal assimétrico e se refere ao tempo de espera entre as erupções do gêiser *Old Faithful*, que se apresenta no parque nacional de Yellowstone, nos Estados Unidos da América. Estes dados foram coletados do software estatístico R (R development core team, 2015). Na Tabela abaixo, podemos observar as estatísticas descritiva dos dados. E na Figura 4.14, a distribuição empírica da amostra:

Tamanho amostral	Média	Desvio padrão	Coefficiente de assimetria
272	70.90	13.60	-0.41

Tabela 4.8: Estatísticas descritivas dos dados Faithful

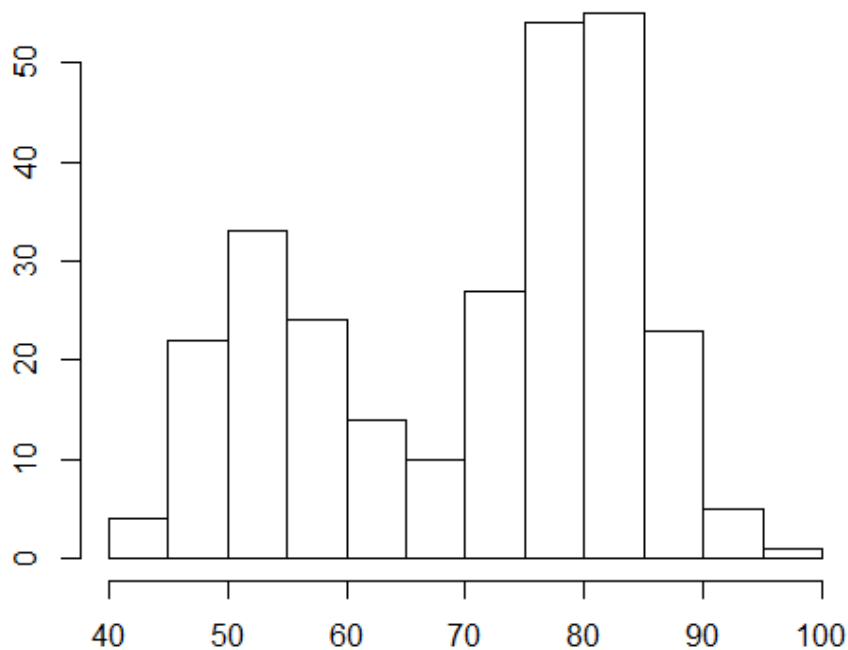


Figura 4.14: Histograma dos dados Faithful.

Para a modelagem bayesiana, consideramos o modelo PNBA e o modelo de mistura. No modelo PNBA, utilizamos a priori informativa com crença não informativa com respeito à unimodalidade ou bimodalidade dos dados e com respeito à assimetria. Ou seja, a priori está restrita a $P(\alpha < 1) = P(\lambda < 0) = 0.5$. Os valores q_α e q_λ escolhidos foram 20 e 3, respectivamente. Atribuímos a p_α e p_λ o valor 0.95. Tais informações implicam que $P(\alpha < 20 | \alpha > 1) = P(\lambda < 3 | \lambda > 0) = 0.95$. Dessa forma, as prioris são tais que $\alpha \sim \log - \text{normal}(0, 1.528)$ e $\lambda \sim \text{Normal}(0, 1.531)$. As prioris consideradas no modelo de mistura foram as mesmas da aplicação na seção 2.5.

Utilizamos o algoritmo de MH para gerar valores da posteriori. Como os dados apresentam um comportamento claramente bimodal, não foi necessário gerar tantas iterações do algoritmo MCMC quando trabalhamos com o modelo PNBA. Geramos aleatoriamente 50000 iterações e as 20000 primeiras foram descartadas. As distribuições a priori e a posteriori de α e λ estão representadas na Figura 4.15 abaixo:

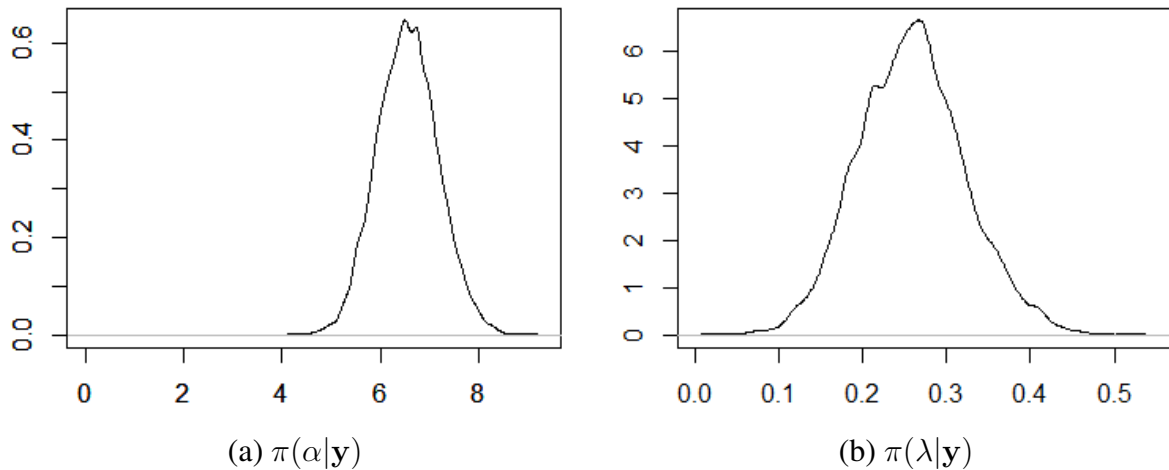


Figura 4.15: Distribuições a posteriori dos parâmetros α e λ para os dados Faithful.

Adotamos os estimadores da média, mediana e moda da distribuição a posteriori marginal para cada parâmetro. Obtemos também o EMV para comparar com os estimadores bayesianos. A partir da Tabela 4.9 podemos perceber que as estimativas foram muito próximas. A Figura 4.16 mostra que as curvas ajustadas pelos diferentes estimadores praticamente não diferiram. As estimativas e o intervalo HPD para λ , onde LI e LS representam o limite inferior e superior do intervalo, respectivamente, mostram que os dados são assimétricos e que a curva ajustada à esquerda é mais concentrada do que a da direita.

Modelo	BIC	Parâmetro	Estimador Bayesiano					EMV
			Média	Mediana	Moda	LI	LS	
PNBA	2102.45	α	6.57	6.54	6.45	5.36	7.70	6.69
		λ	0.26	0.26	0.26	0.14	0.38	0.26
		μ	66.78	66.79	66.84	65.83	67.70	66.75
		σ^2	95.04	94.61	94.36	81.38	109.32	93.49
Mistura	2094.43	p	0.36	0.36	0.36	5.84	8.49	0.36
		μ_1	54.60	54.60	54.52	0.00	0.00	54.61
		μ_2	80.07	80.08	80.09	5.41	5.60	80.09
		σ^2	35.21	35.03	34.86	0.84	1.12	34.44

Tabela 4.9: Estimativas dos parâmetros do modelo PNBA e de mistura para os dados Faithful.

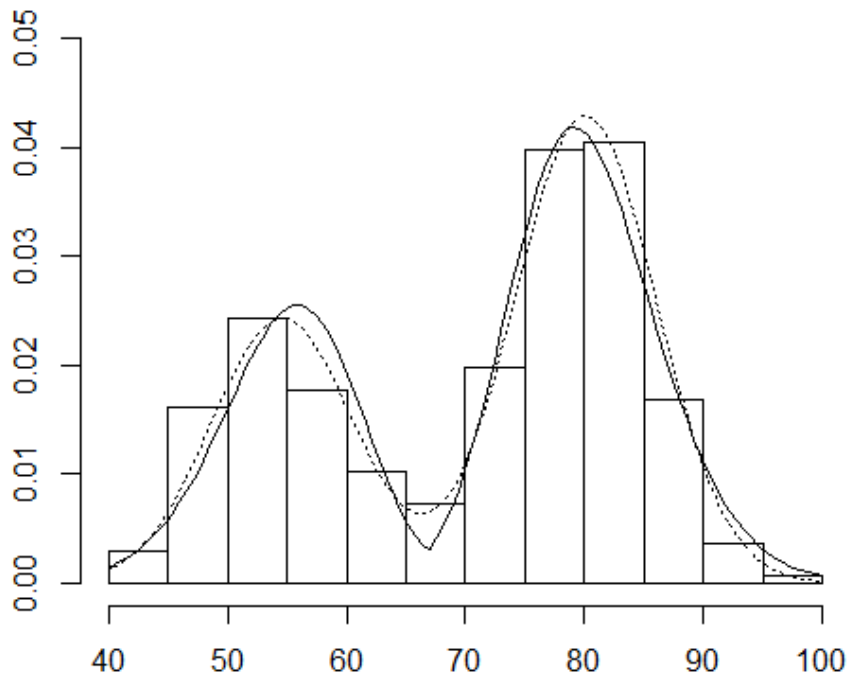


Figura 4.16: Histograma dos dados Faithful sobrepostos às curvas ajustadas pelos estimadores bayesianos. Ajuste do modelo PNBA (linha contínua). Ajuste do modelo de mistura (linha tracejada).

Para uma análise mais aprofundada dos dados, podemos também estar interessados em explicá-los em função de covariáveis. Se a relação dos dados com tais covariáveis é linear, podemos utilizar um modelo de regressão linear. Nessa situação, esperamos que os resíduos apresentem uma distribuição potencial normal bimodal, uma vez que os dados também têm essa distribuição. No próximo capítulo, mostramos como analisar dados na presença de covariáveis a partir de uma perspectiva bayesiana utilizando regressão com resíduos cuja distribuição é potencial normal bimodal.

Capítulo 5

Regressão com resíduos PNBA

5.1 Inferência

5.1.1 Função de Verossimilhança

O modelo de regressão linear com resíduos PNBA tem função de verossimilhança dada por:

$$L(\theta|X, \mathbf{y}) = \left(\frac{2\alpha c_\alpha}{\sigma}\right)^n \prod_{i=1}^n \left[\phi\left(\frac{y_i - X_i'\beta}{\sigma}\right)\right] \left\{ \prod_{i=1}^n \Phi\left(\frac{|y_i - X_i'\beta|}{\sigma}\right) \right\}^{\alpha-1} \left\{ \Phi\left(\lambda \frac{y_i - X_i'\beta}{\sigma}\right) \right\}^{\alpha-1}, i = 1, \dots, n. \quad (5.1)$$

O vetor de parâmetros θ agora é dado por $\theta = (\alpha, \lambda, \beta, \sigma^2)$. As suposições para a veracidade da expressão (5.1) são as mesmas para a verossimilhança do modelo de regressão PNB. Os dados têm uma distribuição de modo que $(y_i|\theta, X_i') \sim \text{PNBA}(\alpha, \lambda, X_i'\beta, \sigma^2)$. Podemos dizer que, para cada unidade a mais da i -ésima covariável do vetor X_i' , a média de Y aumenta β_i unidades, mantendo as demais covariáveis fixas.

Distribuição a Priori Conjunta

A distribuição a priori considerada para o vetor θ é a mesma considerada no modelo com resíduos PNB. Assumindo que a priori $\pi(\lambda)$ independa do vetor β , então a distribuição a priori conjunta é dada por:

$$\begin{aligned} \pi(\theta) &= \pi(\alpha, \lambda, \beta, \sigma^2) \\ &= \pi(\alpha, \lambda)\pi(\beta)\pi(\sigma^2) \\ &= \pi(\alpha, \lambda)|\Sigma|^{-1/2}(2\pi)^{-\frac{p+1}{2}} \frac{b^a}{\Gamma(a)} (\sigma^2)^{a-1} \\ &\quad \exp\left\{-\frac{(\beta - \xi)'\Sigma^{-1}(\beta - \xi)}{2} - b\sigma^2\right\} \end{aligned}$$

5.1.2 Distribuições a Posteriori

Posteriori Conjunta

A distribuição a posteriori proporcional para o modelo PNBA é dada por:

$$\begin{aligned}
 \pi(\theta|\underline{X}, \mathbf{y}) &= \frac{\pi(\alpha, \beta, \sigma^2)L(\alpha, \beta, \sigma^2|\underline{X}, \mathbf{y})}{\pi(\mathbf{y})} \\
 &\propto \pi(\alpha, \lambda)\pi(\beta)\pi(\sigma^2)L(\alpha, \lambda, \beta, \sigma|\underline{X}, \mathbf{y}) \\
 &\propto \pi(\alpha, \lambda) \exp \left\{ -\frac{(\beta - \xi)' \Sigma^{-1} (\beta - \xi)}{2} - \frac{(y_i - X_i' \beta)' (y_i - X_i' \beta)}{2\sigma^2} - b\sigma^2 \right\} \\
 &\quad \alpha^{n-1} c_\alpha^n (\sigma^2)^{a - (\frac{n}{2} + 1)} \left\{ \prod_{i=1}^n \Phi \left(\frac{|y_i - X_i' \beta|}{\sigma} \right) \right\}^\alpha \left\{ \Phi \left(\lambda \frac{y_i - X_i' \beta}{\sigma} \right) \right\}^{\alpha-1}
 \end{aligned}$$

A distribuição preditiva de \mathbf{y} é dada por:

$$\pi(\mathbf{y}) = \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{\infty} \pi(\alpha, \lambda, \beta, \sigma^2) L(\alpha, \lambda, \beta, \sigma|\underline{X}, \mathbf{y}) d\alpha d\sigma^2 d\beta$$

A resolução da integral acima só pode ser feita por métodos numéricos, portanto, não podemos obter uma forma fechada para $\pi(\alpha, \lambda, \beta, \sigma^2|\mathbf{y})$.

Distribuições a Posteriori Condicionais

As distribuições a posteriori condicionais de cada parâmetro são dadas por:

$$\begin{aligned}
 \pi(\alpha|\lambda, \beta, \sigma^2, \underline{X}, \mathbf{y}) &= \frac{\pi(\alpha, \lambda, \beta, \sigma^2)L(\alpha, \beta, \sigma^2|\underline{X}, \mathbf{y})}{\pi(\beta, \sigma^2, \mathbf{y})} \\
 &\propto \pi(\alpha)L(\alpha, \beta, \sigma^2|\underline{X}, \mathbf{y}) \\
 &\propto \pi(\alpha)\alpha^{n-1} c_\alpha^n \left\{ \prod_{i=1}^n \Phi \left(\frac{|y_i - X_i' \beta|}{\sigma} \right) \right\}^\alpha \left\{ \prod_{i=1}^n \left(\lambda \frac{y_i - X_i' \beta}{\sigma} \right) \right\}.
 \end{aligned}$$

$$\begin{aligned}
 \pi(\lambda|\alpha, \beta, \sigma^2, \mathbf{y}) &= \frac{\pi(\alpha, \lambda, \beta, \sigma^2)L(\alpha, \lambda, \beta, \sigma^2|\mathbf{y})}{\pi(\mu, \sigma^2, \mathbf{y})} \\
 &\propto \pi(\alpha, \lambda)L(\alpha, \beta, \sigma^2|\mathbf{y}) \\
 &\propto \pi(\alpha, \lambda)\Phi \left\{ \prod_{i=1}^n \left(\lambda \frac{y_i - X_i' \beta}{\sigma} \right) \right\}.
 \end{aligned}$$

$$\begin{aligned}
\pi(\beta|\alpha, \lambda, \sigma^2, \mathbf{X}, \mathbf{y}) &= \frac{\pi(\alpha, \lambda, \beta, \sigma^2)L(\alpha, \beta, \sigma^2|\mathbf{X}, \mathbf{y})}{\pi(\alpha, \sigma^2, \mathbf{y})} \\
&\propto \pi(\beta)L(\alpha, \beta, \sigma^2|\mathbf{X}, \mathbf{y}) \\
&\propto \exp \left\{ \frac{(\beta - \xi)'(\Sigma^2)^{-1}(\beta - \xi)}{2} - \frac{\sum_{i=1}^n (y_i - X_i'\beta)^2}{2\sigma^2} \right\} \\
&\quad \left\{ \prod_{i=1}^n \Phi \left(\frac{|y_i - X_i'\beta|}{\sigma} \right) \right\}^{\alpha-1} \left\{ \prod_{i=1}^n \Phi \left(\lambda \frac{y_i - X_i'\beta}{\sigma} \right) \right\}.
\end{aligned}$$

$$\begin{aligned}
\pi(\sigma^2|\alpha, \lambda, \beta, \mathbf{y}) &= \frac{\pi(\alpha, \lambda, \beta, \sigma^2)L(\alpha, \beta, \sigma^2|\mathbf{X}, \mathbf{y})}{\pi(\alpha, \beta, \mathbf{y})} \\
&\propto \pi(\sigma^2)L(\alpha, \beta, \sigma^2|\mathbf{X}, \mathbf{y}) \\
&\propto \exp \left\{ -\frac{\sum_{i=1}^n (y_i - X_i'\beta)^2}{2\sigma^2} - b\sigma^2 \right\} \sigma^{a-(\frac{n}{2}+1)} \\
&\quad \left\{ \prod_{i=1}^n \Phi \left(\frac{|y_i - X_i'\beta|}{\sigma} \right) \right\}^{\alpha-1} \left\{ \prod_{i=1}^n \Phi \left(\lambda \frac{y_i - X_i'\beta}{\sigma} \right) \right\}.
\end{aligned}$$

5.2 Aspectos Computacionais

Distribuição Proposta para o Vetor (λ, β)

Na seção anterior, não atribuímos uma proposta para a marginal de μ , mas sim uma proposta conjunta para o vetor (λ, μ) . Neste capítulo, não será diferente com relação ao parâmetro β . Consideramos uma distribuição proposta para o vetor $(\lambda, \beta_0, \beta_1, \dots, \beta_k)$, o qual pode ser escrito na forma particionada por $(\lambda, \beta)'$ de modo que:

$$q_{\lambda, \beta}(\lambda_{t+1}^p, \beta_{t+1}^p | \lambda_t, \beta_t) \sim N \left(\begin{bmatrix} \lambda_t^p \\ \beta_t^p \end{bmatrix}, C_t \right),$$

em que C_t é a matriz de covariância do vetor $(\lambda, \beta)'$ de dimensões $(k+2) \times (k+2)$ do algoritmo de Hario visto na seção anterior. Não houve mudanças no cálculo de C_t comparado com o que foi feito na seção anterior. A quantidade s no tempo foi calculada da mesma maneira e $\epsilon_0 = 0$ também. O valor inicial atribuído para a matriz C_0 foi de tal modo que:

$$C_0 = \begin{bmatrix} 1 & 0'_{k+1} \\ 0_{k+1} & (X'X)^{-1}QME, \end{bmatrix}$$

em que $0'_{k+1}$ é um vetor de zeros de dimensão $k+1$. A matriz $(X'X)^{-1}QME$ representa a crença que temos a priori sobre a matriz de covariância da distribuição a posteriori de β . Neste caso, estamos acreditando a priori que a matriz de covariância de $\pi(\beta|\mathbf{X}, \mathbf{y})$ é a matriz

de covariância do EMV.

5.3 Aplicações Com Modelo de Regressão PNBA

Consideramos um conjunto de dados para a aplicação do modelo de regressão PNBA. Os dados proveram do estudo de Pinheiro (2008) e se refere a características genéticas e biométricas de camarões da espécie *farfantepenaeus brasiliensis*. A amostra foi coletada na costa do Rio Grande do Norte em três regiões distintas: Baía Formosa, Diego Lopes e Touros. Três variáveis foram observadas: comprimento, peso e sexo. Encontramos resíduos com distribuição assimétrica e bimodal quando adotamos a variável peso como variável dependente e sexo como variável independente. Abaixo, segue o histograma e as estatísticas descritivas dos dados.

Comprimento	Tamanho amostral	Média	Desvio padrão	Coefficiente de assimetria
Todos os camarões	120	102.70	19.56	-0.42
camarões macho	35	89.82	10.69	-1.56
camarões fêmea	85	107.8	19.95	-0.93

Tabela 5.1: Estatísticas descritivas dos dados EPI e MAOA

Ajustamos o modelo PNBA sem a covariável sexo e com a variável sexo. Assumimos uma priori tal que $q_\alpha = 20$, $q_\lambda = 3$ e $p_\alpha = p_\lambda = 0.95$. Assim, $\alpha \sim \log - \text{normal}(0, 1.528)$ e $\lambda \sim \text{Normal}(0, 1.531)$.

O algoritmo de MH foi utilizado para obter uma amostra da distribuição a posteriori. Geramos 50000 iterações desprezando as 20000 primeiras. Na Figura 5.2 pode ser observada as distribuições a posteriori de α e λ nos dois modelos utilizados.

Calculamos os estimadores da média, mediana e moda da distribuição a posteriori marginal para cada parâmetro. Obtemos também o EMV para comparar com os estimadores bayesianos. A partir da Tabela 5.2 podemos perceber que as estimativas foram muito próximas. O BIC mostrou que o ajuste com a covariável sexo foi melhor do que o ajuste sem covariáveis.

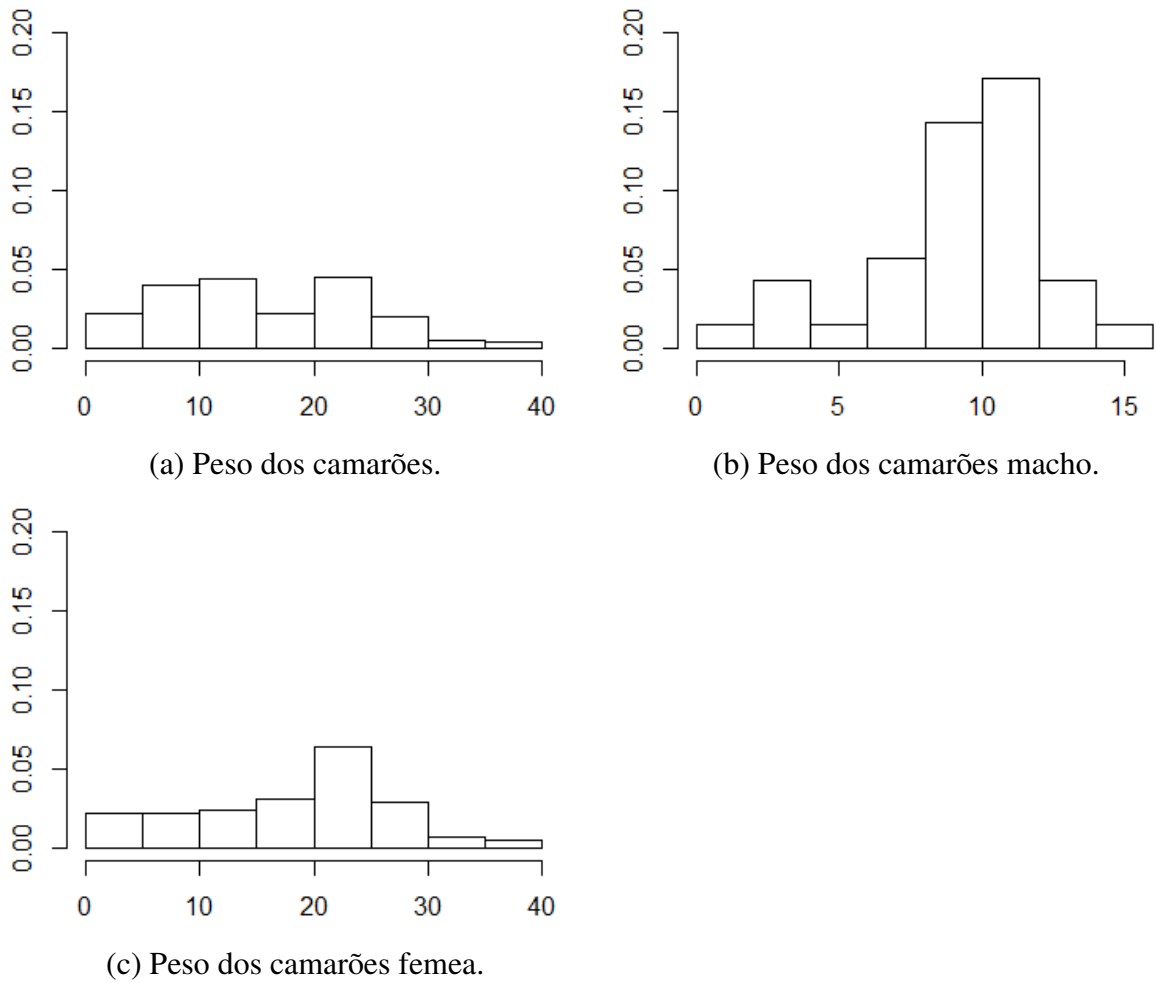
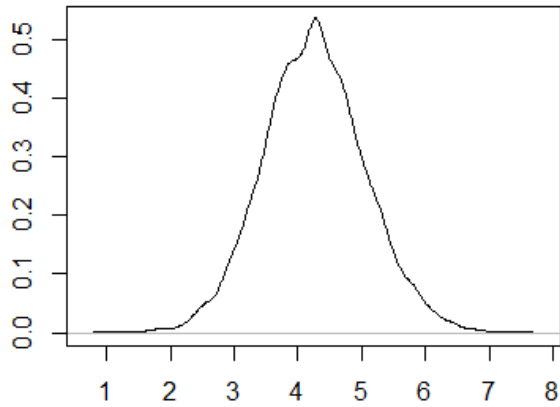


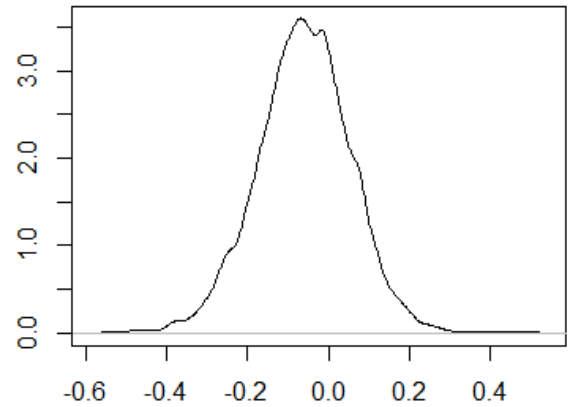
Figura 5.1: Histograma dos dados dos camarões fêmea.

Modelo	BIC	Parâmetro	Estimador Bayesiano					EMV
			Média	Mediana	Moda	LI	LS	
PNBA	849.32	α	4.26	4.25	4.27	2.75	5.88	4.61
		λ	-0.06	-0.06	-0.07	-0.30	0.15	-0.07
		μ	16.29	16.30	16.34	15.15	17.54	16.42
		σ^2	42.24	41.51	40.48	31.50	54.00	39.45
Reg. PNBA	815.11	α	5.19	5.19	5.20	3.42	7.07	5.70
		λ	0.96	0.95	0.91	0.56	1.43	0.96
		β_0	12.03	11.84	11.76	10.20	14.61	11.74
		β_1	-11.93	-11.91	-11.84	-14.15	-9.86	-11.95
		σ^2	57.15	56.38	55.96	39.15	74.74	54.97

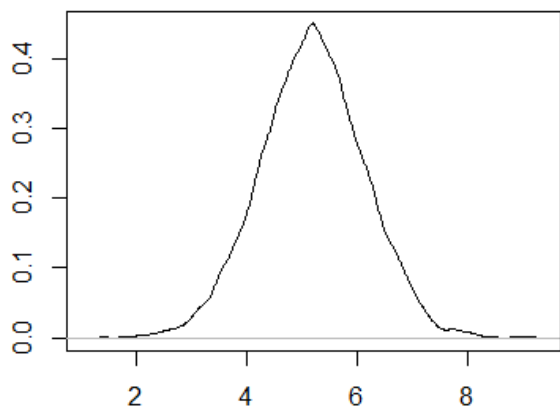
Tabela 5.2: Estimativas dos parâmetros do modelo PNBA dos dados camarões.



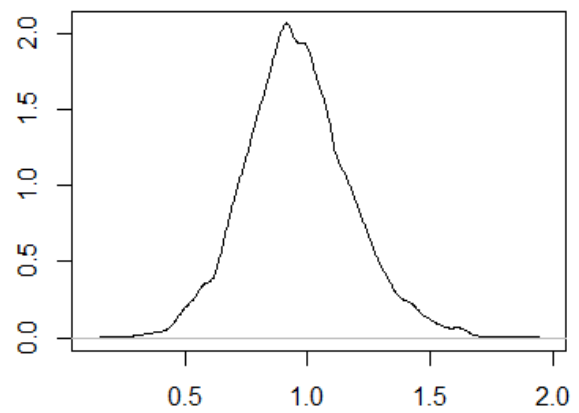
(a) $\pi(\alpha|\mathbf{y})$. Modelo PNBA sem covariáveis.



(b) $\pi(\lambda|\mathbf{y})$. Modelo PNBA sem covariáveis.



(c) $\pi(\alpha|\mathbf{y})$. Modelo PNBA com a covariável sexo.



(d) $\pi(\lambda|\mathbf{y})$. Modelo PNBA com a covariável sexo.

Figura 5.2: Distribuições a posteriori dos parâmetros α e λ para os dados dos camarões.

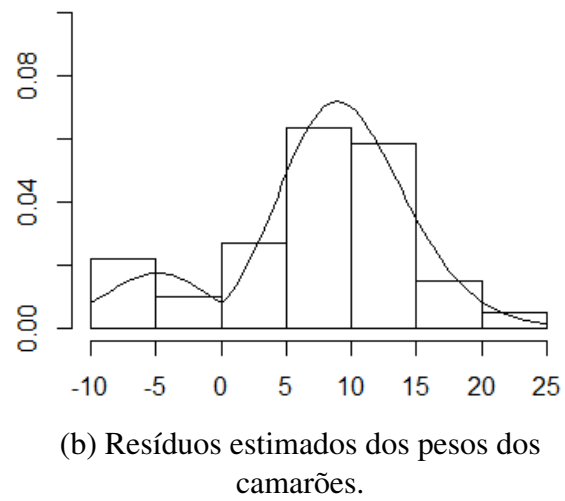
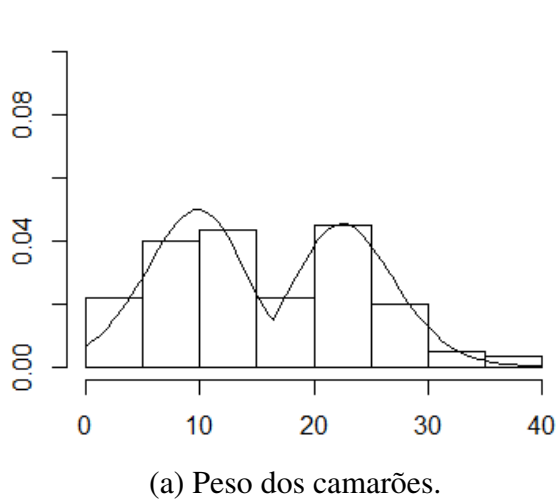


Figura 5.3: Histograma dos dados camarões sobrepostos às curvas ajustadas pelo estimador bayesiano da média a posteriori.

Capítulo 6

Considerações Finais

Nesta dissertação, estudamos o modelo potência normal bimodal a partir de uma perspectiva bayesiana. Desenvolvemos a distribuição a priori de Jeffreys para o parâmetro α e obtivemos também a sua distribuição conjugada, que não é uma distribuição simples e conhecida. Propusemos prioris informativas com distribuição gama e log-normal. Expressamos possíveis crenças que o usuário pode ter a respeito de α e, a partir delas, encontramos os hiperparâmetros das distribuições a priori informativa. Fizemos um estudo de diagnóstico de convergência das cadeias geradas pelo MCMC e concluímos que a cadeia é convergente. No entanto, se utilizarmos prioris com probabilidade alta em intervalos próximos de 0, podemos ter problemas de autocorrelação forte da cadeia gerada pelo MCMC. Foi feita uma análise de sensibilidade na qual constatamos que a escolha da família log-normal como priori de α é vantajosa no caso em que crermos a priori em $\alpha \in (0, 1)$ (unimodal) ou crermos em $\alpha \in (0, 2)$ (unimodal ou fracamente bimodal) quando os dados são fortemente bimodais. Vimos também que crenças fracas em modelos que não condizem com a realidade dos dados, em geral, não comprometem a qualidade do ajuste no caso gama. No caso log-normal, temos até o luxo de atribuir uma crença moderada em prioris “erradas”. Fizemos um estudo de simulações para analisar o desempenho do estimador bayesiano e compará-lo com o EMV. Constatamos que a moda a posteriori, no geral, tem propriedades frequentistas melhores do que a média a posteriori, mediana a posteriori e o EMV quanto ao EQM e VP. Apresentamos uma aplicação com dados bimodais em que utilizamos o modelo bayesiano para se fazer inferências. Para finalizar o estudo do modelo potência normal bimodal, apresentamos o modelo de regressão linear com resíduos com distribuição potência normal bimodal.

Estudamos também o modelo potência normal bimodal assimétrico. Neste, desenvolvemos uma distribuição a priori não informativa baseada em uma reparametrização de λ . Propusemos uma priori informativa simples para α e λ . Também expressamos a priori em função da crença do usuário. O estudo de convergência constatou que a priori não informativa não converge, mas a informativa sim. Constatamos através da análise de sensibilidade que a distribuição a posteriori é sensível à priori escolhida para α e λ . Por fim, apresentamos uma aplicação com dados bimodais assimétricos.

Capítulo 7

Apêndice

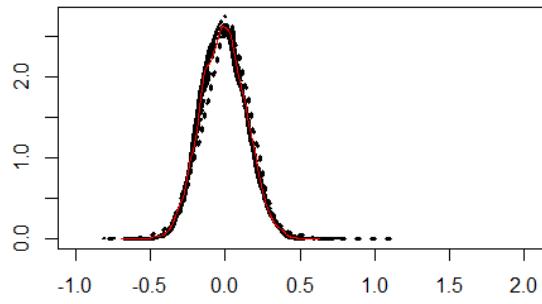
7.1 Apêndice A

Tabela A.1: Viés percentual do modelo potência normal bimodal (fora dos parênteses). Desvio padrão do VP (entre parênteses).

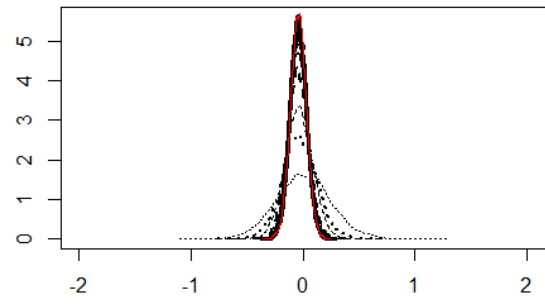
		HPD	Media (EB)	Mediana (EB)	Moda (EB)	EMV
$n = 50$	α	0.95	980.06 (15.95)	862.36 (16.66)	558.26 (20.04)	487.74 (25.18)
$\alpha = 0.1$	μ	0.95	0.02 (0.04)	0.02 (0.04)	0.02 (0.04)	0.04 (0.04)
	σ^2	0.84	-13.48 (0.64)	-16.79 (0.62)	-22.26 (0.60)	-10.62 (0.74)
$n = 100$	α	0.94	668.71 (11.95)	594.69 (12.58)	396.15 (15.28)	326.89 (17.91)
$\alpha = 0.1$	μ	0.94	0.06 (0.03)	0.06 (0.03)	0.06 (0.03)	0.07 (0.03)
	σ^2	0.87	-9.98 (0.48)	-11.75 (0.48)	-15.03 (0.48)	-6.41 (0.57)
$n = 200$	α	0.96	438.40 (8.04)	386.57 (8.47)	248.99 (10.37)	184.51 (11.76)
$\alpha = 0.1$	μ	0.95	0.03 (0.02)	0.03 (0.02)	0.03 (0.02)	0.03 (0.02)
	σ^2	0.89	-8.22 (0.35)	-9.08 (0.35)	-10.54 (0.36)	-4.62 (0.41)
$n = 50$	α	0.97	55.25 (2.19)	46.31 (2.32)	24.44 (2.85)	33.18 (3.36)
$\alpha = 1$	μ	0.92	-0.09 (0.05)	-0.09 (0.05)	-0.09 (0.05)	-0.11 (0.05)
	σ^2	0.93	-3.06 (0.72)	-7.19 (0.69)	-14.12 (0.67)	-5.05 (0.85)
$n = 100$	α	0.97	25.98 (1.74)	21.21 (1.84)	10.34 (2.20)	13.47 (2.43)
$\alpha = 1$	μ	0.95	0.03 (0.03)	0.03 (0.03)	0.03 (0.03)	0.03 (0.03)
	σ^2	0.94	-0.44 (0.56)	-2.76 (0.55)	-6.84 (0.54)	-1.40 (0.66)
$n = 200$	α	0.95	9.39 (1.32)	7.39 (1.38)	3.49 (1.58)	5.12 (1.67)
$\alpha = 1$	μ	0.94	0.00 (0.02)	0.00 (0.02)	0.00 (0.02)	0.00 (0.02)
	σ^2	0.95	0.11 (0.41)	-1.21 (0.41)	-3.65 (0.42)	-1.06 (0.47)
$n = 50$	α	0.95	6.52 (0.79)	4.69 (0.77)	1.31 (0.73)	8.10 (0.80)
$\alpha = 10$	μ	0.95	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.01 (0.03)
	σ^2	0.94	1.90 (0.50)	-0.07 (0.48)	-3.36 (0.47)	-2.13 (0.46)
$n = 100$	α	0.95	4.02 (0.51)	3.17 (0.51)	1.65 (0.50)	4.68 (0.51)
$\alpha = 10$	μ	0.96	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)	0.00 (0.02)
	σ^2	0.94	0.65 (0.34)	-0.28 (0.34)	-1.98 (0.33)	-1.25 (0.33)
$n = 200$	α	0.96	2.39 (0.43)	1.95 (0.42)	1.02 (0.36)	2.39 (0.33)
$\alpha = 10$	μ	0.95	-0.06 (0.07)	-0.06 (0.08)	-0.06 (0.08)	-0.13 (0.15)
	σ^2	0.96	1.14 (1.00)	0.66 (1.04)	-0.05 (1.02)	1.20 (2.03)

Tabela A.2: EQM do modelo potência normal bimodal (fora dos parênteses). Desvio padrão do EQM (entre parênteses).

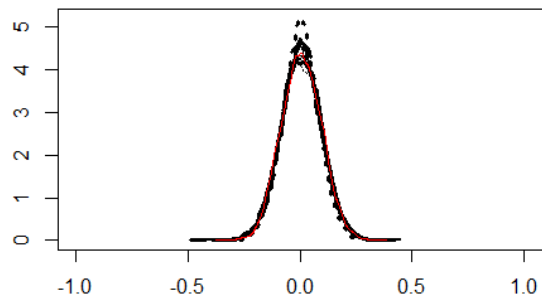
		HPD	Media (EB)	Mediana (EB)	Moda (EB)	EMV
$n = 50$	α	0.95	1.21 (0.05)	1.02 (0.05)	0.71 (0.05)	0.87 (0.06)
$\alpha = 0.1$	μ	0.95	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)
	σ^2	0.84	0.06 (0.00)	0.07 (0.00)	0.09 (0.00)	0.07 (0.00)
$n = 100$	α	0.94	0.59 (0.02)	0.51 (0.02)	0.39 (0.03)	0.43 (0.03)
$\alpha = 0.1$	μ	0.94	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)
	σ^2	0.87	0.03 (0.00)	0.04 (0.00)	0.05 (0.00)	0.04 (0.00)
$n = 200$	α	0.96	0.26 (0.01)	0.22 (0.01)	0.17 (0.01)	0.17 (0.01)
$\alpha = 0.1$	μ	0.95	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
	σ^2	0.89	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)
$n = 50$	α	0.97	0.79 (0.05)	0.75 (0.05)	0.87 (0.05)	1.24 (0.06)
$\alpha = 1$	μ	0.92	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.03 (0.00)
	σ^2	0.93	0.05 (0.00)	0.05 (0.00)	0.06 (0.00)	0.07 (0.00)
$n = 100$	α	0.97	0.37 (0.02)	0.38 (0.02)	0.50 (0.02)	0.61 (0.03)
$\alpha = 1$	μ	0.95	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)
	σ^2	0.94	0.03 (0.00)	0.03 (0.00)	0.03 (0.00)	0.04 (0.00)
$n = 200$	α	0.95	0.18 (0.01)	0.19 (0.01)	0.25 (0.01)	0.28 (0.01)
$\alpha = 1$	μ	0.94	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)
	σ^2	0.95	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)
$n = 50$	α	0.95	6.69 (0.46)	6.09 (0.41)	5.32 (0.36)	7.06 (0.50)
$\alpha = 10$	μ	0.95	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)
	σ^2	0.94	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)
$n = 100$	α	0.95	2.81 (0.16)	2.66 (0.15)	2.56 (0.14)	2.85 (0.16)
$\alpha = 10$	μ	0.96	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
	σ^2	0.94	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)
$n = 200$	α	0.96	1.91 (0.80)	1.77 (0.70)	1.28 (0.23)	1.12 (0.06)
$\alpha = 10$	μ	0.95	0.05 (0.05)	0.06 (0.06)	0.06 (0.06)	0.22 (0.22)
	σ^2	0.96	0.10 (0.09)	0.11 (0.10)	0.10 (0.10)	0.41 (0.41)



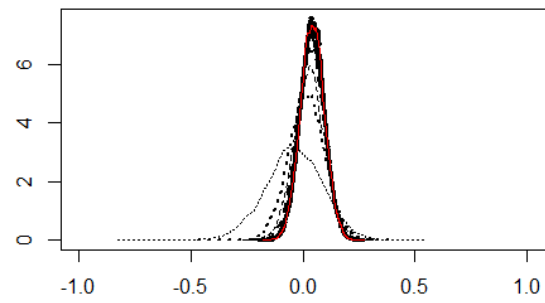
(a) $\alpha = 1; n = 50$



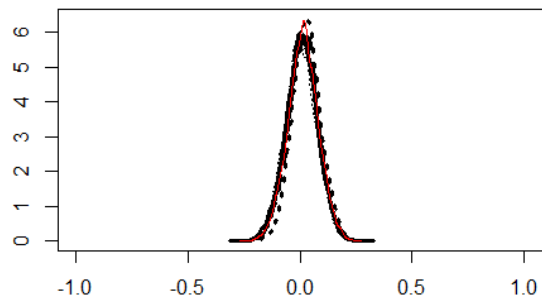
(b) $\alpha = 10; n = 50$



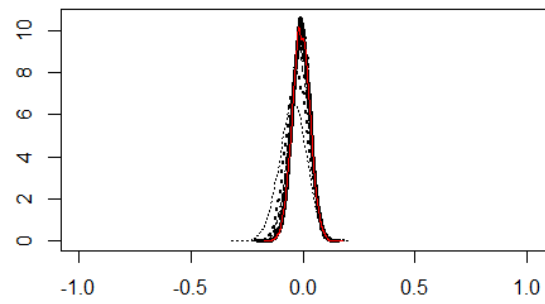
(c) $\alpha = 1; n = 100$



(d) $\alpha = 10; n = 100$

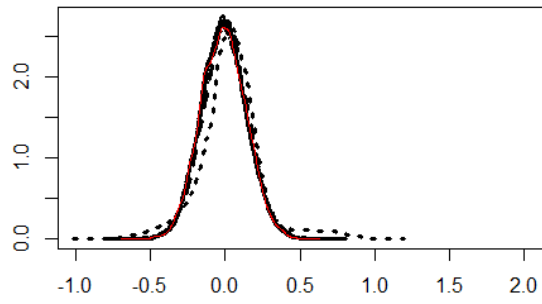


(e) $\alpha = 1; n = 200$

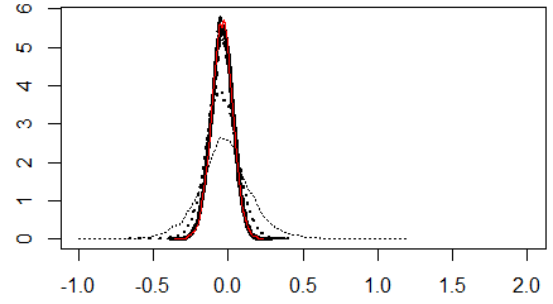


(f) $\alpha = 10; n = 200$

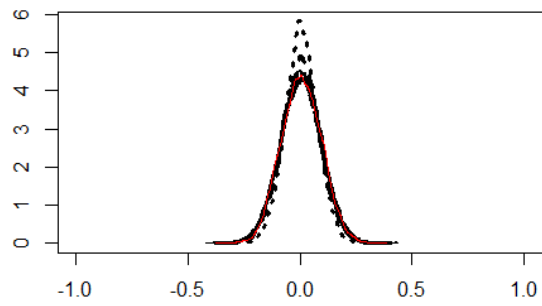
Figura A.1: $\pi(\mu|y)$ utilizando a priori gama para $\pi(\alpha)$.



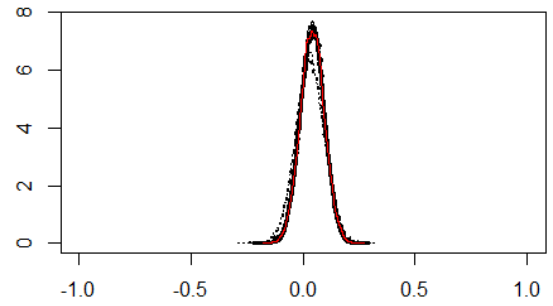
(a) $\alpha = 1; n = 50$



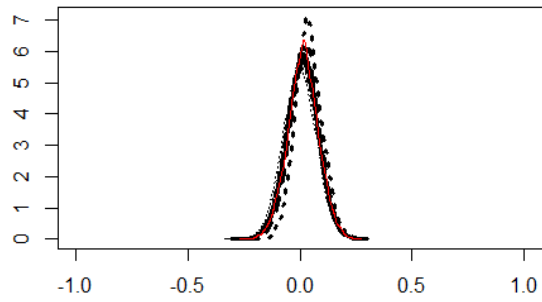
(b) $\alpha = 10; n = 50$



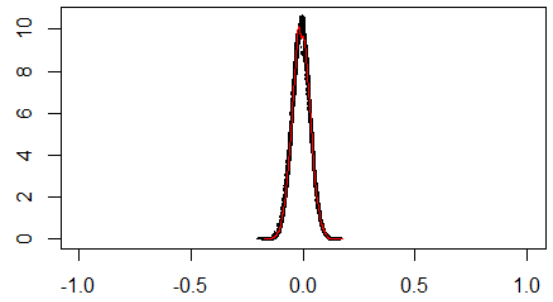
(c) $\alpha = 1; n = 100$



(d) $\alpha = 10; n = 100$

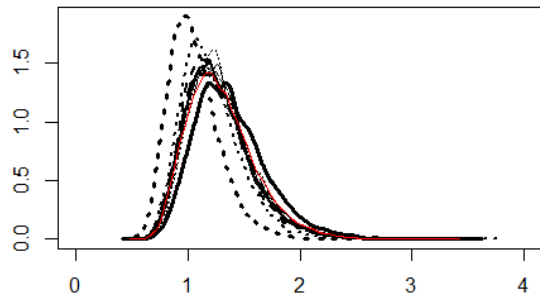


(e) $\alpha = 1; n = 200$

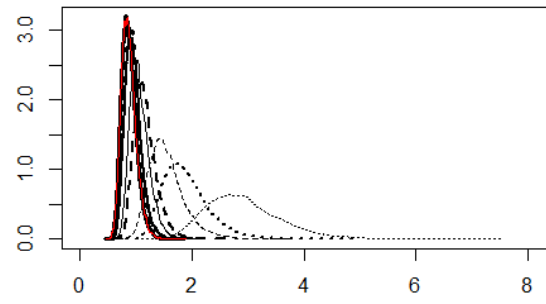


(f) $\alpha = 10; n = 200$

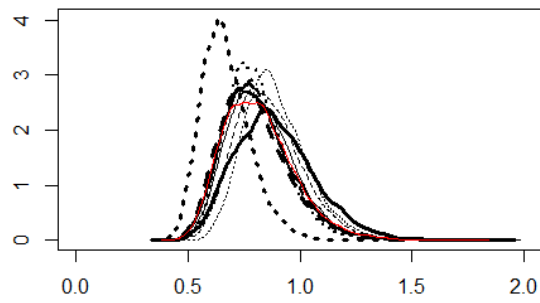
Figura A.2: $\pi(\mu|y)$ utilizando a priori log-normal para $\pi(\alpha)$.



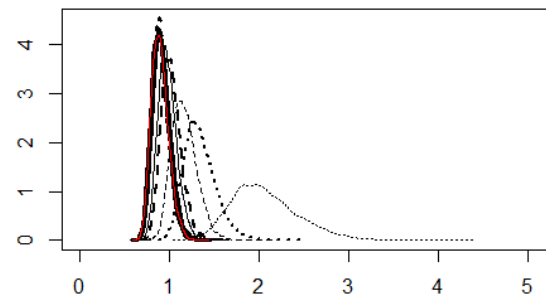
(a) $\alpha = 1; n = 20$



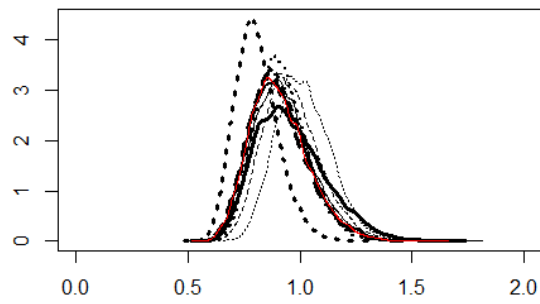
(b) $\alpha = 10; n = 20$



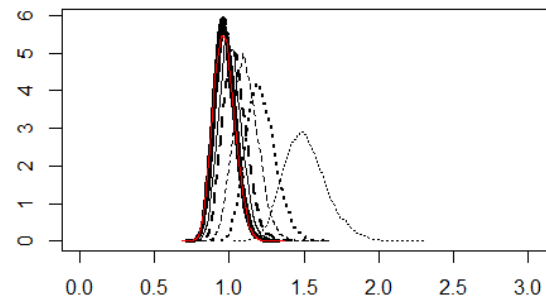
(c) $\alpha = 1; n = 100$



(d) $\alpha = 10; n = 100$



(e) $\alpha = 1; n = 200$



(f) $\alpha = 10; n = 200$

Figura A.3: $\pi(\sigma^2|\mathbf{y})$ utilizando a priori gama para $\pi(\alpha)$.

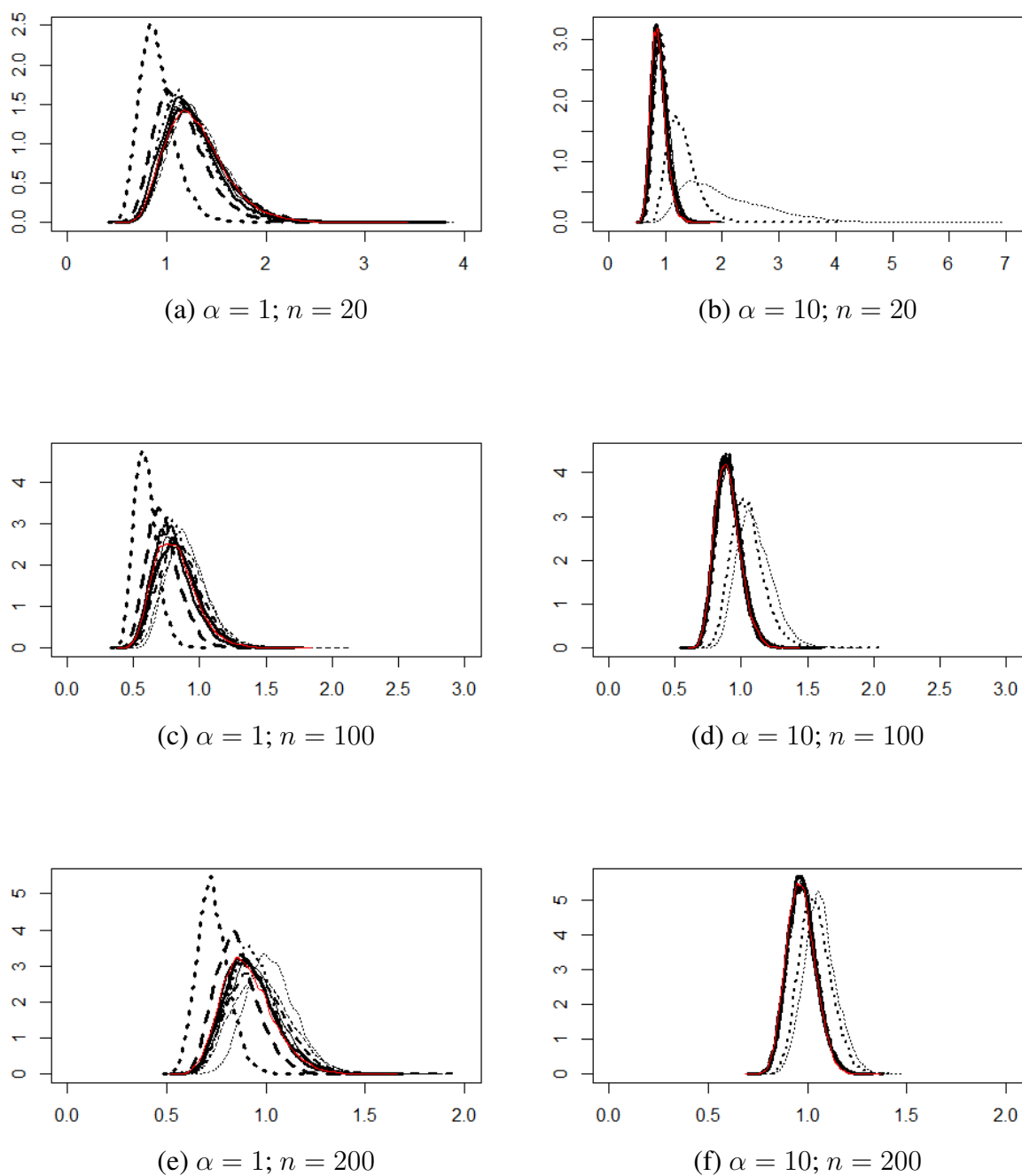


Figura A.4: $\pi(\sigma^2|\mathbf{y})$ utilizando a priori log-normal para $\pi(\alpha)$.

Tabela A.2: Linhas das curvas das prioris das Figuras A.1, A.2 e A.3.

Crença / Intervalo	(0,1)	(0,2)	(2,15)	(0,∞)
Fraca	Sólida Fina	Sólida Média	Sólida Grossa	-
Moderada	Tracejada Fina	Tracejada Média	Tracejada Grossa	-
Forte	Pontilhada Fina	Pontilhada Média	Pontilhada Grossa	-
Não Informativa	-	-	-	Vermelha

7.2 Apêndice B

Proposição 1: $I = \int_0^\infty \sqrt{I(\alpha)} d\alpha = \int_0^\infty \sqrt{\frac{1}{\alpha^2} - \log^2(2)2^\alpha(2^\alpha - 1)^{-2}} d\alpha > \infty$:

Prova: Seja k, c números do conjunto dos \mathbb{R}^+ . Sabemos que $\int_k^\infty \frac{1}{c\alpha} d\alpha > \infty$. Se mostrarmos que $\exists \alpha_0$ t.q $\forall \alpha > \alpha_0, \sqrt{I(\alpha)} > \frac{1}{c\alpha}$, então:

$$\int_{\alpha_0}^\infty \sqrt{I(\alpha)} d\alpha > \int_{\alpha_0}^\infty \frac{1}{c\alpha} > \infty \quad \text{B.1}$$

Para mostrar que tal α_0 que satisfaça B.1 existe, mostramos primeiro que:

$$\begin{aligned} \lim_{\alpha \rightarrow \infty} \alpha^2 \log^2(2)2^\alpha(2^\alpha - 1)^{-2} &= \log^2(2) \lim_{\alpha \rightarrow \infty} \frac{\alpha^2}{2^\alpha - 1} \lim_{\alpha \rightarrow \infty} \frac{2^\alpha}{2^\alpha - 1} \\ &= \log^2(2) \times 0 \times 1 = 0 \end{aligned} \quad \text{B.2}$$

Com o auxílio da expressão B.2, podemos ver que tomando $c > 1$ e α suficientemente grande temos que:

$$1 - \alpha^2 \log^2(2)2^\alpha(2^\alpha - 1)^{-2} > \frac{1}{c} \quad \text{B.3}$$

Desenvolvendo a expressão B.3:

$$\begin{aligned} \frac{1}{\alpha^2} - \log^2(2)2^\alpha(2^\alpha - 1)^{-2} &> \frac{1}{c\alpha^2} \\ \sqrt{\frac{1}{\alpha^2} - \log^2(2)2^\alpha(2^\alpha - 1)^{-2}} &> \frac{1}{\sqrt{c}\alpha} \end{aligned}$$

Logo, a expressão B.1 é verdadeira e pelo teste da comparação a integral a esquerda em B.1 diverge. Como $\sqrt{I(\alpha)} > 0$ é uma fdp temos que $\sqrt{I(\alpha)} > 0 \forall \alpha > 0$. Então, $\int_0^{\alpha_0} \sqrt{I(\alpha)} d\alpha > 0$. Isso implica que:

$$\int_0^\infty \sqrt{I(\alpha)} d\alpha = \int_0^{\alpha_0} \sqrt{I(\alpha)} d\alpha + \int_{\alpha_0}^\infty \sqrt{I(\alpha)} d\alpha > \infty$$

Portanto, a priori de Jeffreys é imprópria.

Proposição 2: $I = \int_0^\infty p(\alpha|r, t) d\alpha = \int_0^\infty \left(\frac{\alpha}{2^\alpha - 1}\right)^r e^{-t\alpha} d\alpha > \infty$:

Prova: Primeiramente, analisamos o comportamento de $p(\alpha|r, t)$ e mostramos que tal função não tem uma assintota vertical em $\alpha = 0$:

$$\lim_{\alpha \rightarrow \infty} \frac{\alpha}{2^\alpha - 1} = \lim_{\alpha \rightarrow \infty} \frac{1}{2^\alpha \log(2)} = \frac{1}{\log(2)}$$

Assim, temos:

$$\lim_{\alpha \rightarrow \infty} \left(\frac{\alpha}{2^\alpha - 1}\right)^r e^{-t\alpha} = \left(\lim_{\alpha \rightarrow \infty} \frac{\alpha}{2^\alpha - 1}\right)^r \lim_{\alpha \rightarrow \infty} e^{-t\alpha} = \left[\frac{1}{\log(2)}\right]^r \times 0 = 0$$

Logo, como $\lim_{\alpha \rightarrow 0} p(\alpha|r, t) = 0$ e $p(\alpha|r, t)$ é contínua no conjunto dos \mathbb{R}^+ , então $\int_0^a p(\alpha|r, t) d\alpha < \infty, \forall a > 0$. Para demonstrar que $\int_0^\infty p(\alpha|r, t) d\alpha$ é convergente, basta tomar um $a > 0$ e mostrar que $\int_a^\infty p(\alpha|r, t) d\alpha$ converge. Escolhendo $a = 1$ e usando o teste da comparação, temos que:

$$\int_1^\infty \frac{\alpha}{(2^\alpha - 1)^r} \alpha^r e^{-t\alpha} d\alpha > \int_1^\infty \alpha^r e^{-t\alpha} d\alpha > \infty \quad \text{B.4}$$

A desigualdade em B.4 é verdadeira pois a função integrando da esquerda é a função integrando da direita multiplicada por uma função que é menor que 1 $\forall \alpha > 1$. A função integrando da direita é o núcleo de uma distribuição *gamma*(r, t), portanto pelo teste da comparação a função da esquerda que é menor converge. Assim, como $\int_0^1 p(\alpha|r, t) d\alpha$ e $\int_1^\infty p(\alpha|r, t) d\alpha$ convergem, logo $\int_0^\infty p(\alpha|r, t) d\alpha$ também converge. Por isso, $p(\alpha|r, t)$ é uma fdp própria.

Proposição 3: Seja X_1, X_2 variáveis aleatórias independentes tal que $X_1 \sim N(0, 1)$ e $X_2 \sim |PNB(\alpha)|$. Então, a fdp da transformação $\lambda = X_1/X_2$ é dada por:

$$f_{\lambda|\alpha}(\lambda|\alpha) = \sqrt{\frac{2}{\pi}} \alpha c_\alpha \int_0^\infty v \phi(\sqrt{\lambda^2 + 1}v) \{\Phi(v)\}^{\alpha-1} dv$$

Prova: Considere as transformações $Y_1 = X_1$ e $Y_2 = X_1/X_2$. Dado isso, temos que as inversas dessas transformações podem ser escritas como:

$$X_1 = g_1^{-1}(y_1, y_2) = y_1$$

$$X_2 = g_2^{-1}(y_1, y_2) = \frac{y_1}{y_2}$$

Dessa forma, temos o modulo do jacobiano dado por:

$$|J| = \begin{vmatrix} \frac{\partial X_1}{\partial y_1} & \frac{\partial X_1}{\partial y_2} \\ \frac{\partial X_2}{\partial y_1} & \frac{\partial X_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} \frac{\partial X_1}{\partial y_1} & \frac{\partial X_1}{\partial y_2} \\ \frac{\partial X_2}{\partial y_1} & \frac{\partial X_2}{\partial y_2} \end{vmatrix} = \frac{|y_1|}{y_2}$$

As densidades de X_1, X_2 e a conjunta delas são dadas por:

$$f_{X_1}(x_1) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x_1^2}{2}\right\}, x_1 \in \mathbb{R}$$

$$f_{X_2}(x_2) = 2\alpha c_\alpha \phi(x_2) \{\Phi(|x_2|)\}^{\alpha-1}, x_2 \in \mathbb{R}^+$$

$$f_{X_1, X_2}(x_1, x_2) = \frac{\alpha c_\alpha}{\pi} \exp\left\{-\frac{1}{2}(x_1^2 + x_2^2)\right\} \{\Phi(|x_2|)\}^{\alpha-1}, x_1 \in \mathbb{R}$$

Assim, obtemos a conjunta de Y_1 e Y_2 dada por:

$$\begin{aligned} f_{Y_1, Y_2}(y_1, y_2) &= |J| f_{X_1, X_2}(g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2)), y_1 y_2 > 0 \\ &= \frac{\alpha c_\alpha |y_1|}{\pi y_2^2} \exp\left\{-\frac{(1 + y_2^2)}{2} y_1^2\right\} \left\{\Phi\left(\frac{y_1}{y_2}\right)\right\}^{\alpha-1}, y_1 y_2 > 0 \end{aligned}$$

O suporte de (Y_1, Y_2) é definido na região tal que $y_1 y_2 > 0$, pois o sinal de Y_1 será sempre o mesmo de Y_2 , uma vez que $Y_2 = Y_1/X_2$, em que $X_2 > 0$.

Obtida a conjunta de Y_1 e Y_2 , podemos calcular a fdp de λ dado α :

$$\begin{aligned}
f_\lambda(\lambda|\alpha) &= f_{Y_2}(y_2) \\
&= \int_{y_1} f_{Y_1, Y_2}(y_1, y_2) dy_1 \\
&= \begin{cases} \frac{\alpha c_\alpha}{\pi} \frac{1}{\lambda^2} \int_{-\infty}^0 |y_1| \exp \left\{ -\frac{(1+\lambda^2)}{2} y_1^2 \right\} \left\{ \Phi \left(\frac{y_1}{\lambda} \right) \right\}^{\alpha-1} dy_1, & \text{se } \lambda < 0 \\ \frac{\alpha c_\alpha}{\pi} \frac{1}{\lambda^2} \int_0^{\infty} |y_1| \exp \left\{ -\frac{(1+\lambda^2)}{2} y_1^2 \right\} \left\{ \Phi \left(\frac{y_1}{\lambda} \right) \right\}^{\alpha-1} dy_1, & \text{se } \lambda > 0 \end{cases}
\end{aligned}$$

Pela simetria da integral em B.7:

$$f_\lambda(\lambda|\alpha) = \frac{\alpha c_\alpha}{\pi} \frac{1}{\lambda^2} \int_0^{\infty} |v| \exp \left\{ -\frac{(1+\lambda^2)}{2} v^2 \right\} \left\{ \Phi \left(\frac{v}{\lambda} \right) \right\}^{\alpha-1} dv, \text{ se } \lambda > 0$$

Referências Bibliográficas

- [1] Andrade, B. B., Bolfarine, H. and Siroky, A.N. (2016). Random number generation and estimation with the bimodal asymmetric power-normal distribution. *Journal of Statistical Computation and Simulation*, 86:3, 460-476.
- [2] Arnold, B. C., Gómez, H. W., Salinas, H. S. (2009). On multiple constraint skewed models. *Statistics: A Journal of Theoretical and Applied Statistics*, 43:3, 279-293.
- [3] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171-178.
- [4] Bayes, C. L., Branco, M. D. (2007). Bayesian inference for the skewness parameter of the scalar skew-normal distribution. *Brazilian Journal of Probability and Statistics*, 21, 141-163.
- [5] Bolfarine, H., Martínez-Flórez, G. and Salinas, H.S. (2013). Bimodal symmetricasymmetric power-normal families. *Communications in Statistics: Theory and Methods*. Access options DOI:10.1080/03610926.2013.765475.
- [6] Branco, M. D, Genton, M. G., Liseo, B. (2012). Objective Bayesian Analysis of Skew-t Distributions. *Scandinavian Journal of Statistics*, 40, 63-85.
- [7] Casella, G., Berger, R.L., 2002. *Statistical Inference*. Thonson Learning, Buxbury.
- [8] Chiogna, M. (1997). Notes on estimation problems with scalar skew-normal distributions. Technical Report 15, University of Padua, Dept Statistical Sciences.
- [9] Durrans, S. R. (1992). Distributions of fractional order statistics in hydrology. *Water Resources Research* 28, 1649-1655.
- [10] Estes, R. et al. (2005) *environmental sustainability index: benchmarking national environmental stewardship*. New Haven: Yale Center for Environemntal Law and Policy.
- [11] Gelman, A., Rubin, D. R. (1992). Inference from interactive simulation using multiple sequences. *Stat. Sci.*, v.7, p.457-511.

- [12] Gupta, D. and Gupta, R. C. (2008). Analyzing skewed data by power normal model. *Test* 17, 197-210.
- [13] Haario H., Saksman E., Tamminen J. (2001). An Adaptive Metropolis Algorithm. *Bernoulli* 7, 223-242.
- [14] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97-109.
- [15] Henze, N. (1986) A probabilistic representation of the skew-normal distribution. *Scandinavian Journal of Statistics*, 13, 271-275.
- [16] Kim, H. J. (2005). On a class of two-piece skew-normal distributions. *Statistics: A Journal of Theoretical and Applied Statistics* 39:6, 537-553
- [17] Lehmann, E. L. (1953). The power of rank tests. *Annals of Mathematical Statistics* 24, 23-43.
- [18] Liseo, B., Loperfido, N. (2006). A note on reference priors for the scalar skew-normal distribution. *Journal of Statistical Planning and Inference*, 136, 373-389.
- [19] Mason et al. (2011). Bimodal distribution of RNA expression levels in human skeletal muscle tissue. *BMC Genomics*, 12:98, 1471-2164.
- [20] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087-1092.
- [21] O'Hagan, A and Leonard, T (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, 63, 201-203.
- [22] Pewsey, A (2000). Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics*, 27, 859-870.
- [23] Pewsey, A, Gómez, H. W. and Bolfarine, H. (2012). Likelihood-based inference for power distributions. *Test* 21(4), 775-789.
- [24] R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.Rproject.org/>.
- [25] Roberts, C. D (1966). A correlation model useful in the study of twins. *Journal of the American Statistical Association*, 61, 1184-1190.
- [26] Roberts, G. O., Rosenthal, J. S. (2006). Examples of adaptive MCMC, Technical report, Lancaster University.

- [27] Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics*, 6(2), 461-464.
- [28] Spiegelhalter D. J. et al (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, 64:583-640.