

Caroline Cavatti Vieira

**Estimação de Densidades via Mistura de Distribuições
Skew-normal por Processos de Dirichlet**

Belo Horizonte, Abril de 2011

Caroline Cavatti Vieira

**Estimação de Densidades via Mistura de Distribuições
Skew-normal por Processos de Dirichlet**

Orientador: Rosangela Helena Loschi

Co-Orientador: Denise Duarte

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA
INSTITUTO DE CIÊNCIAS EXATAS
UNIVERSIDADE FEDERAL DE MINAS GERAIS

Belo Horizonte, Abril de 2011

Agradecimentos

Agradeço a Deus por todas as oportunidades a mim concedidas, em especial por mais esta conquista.

Agradeço a minha mãe Heloisa pelo amor, carinho e dedicação incondicionais, que me deram força para chegar ao final desta jornada. Ao meu pai José pelo amor e incentivo moral e financeiro. Sinto não poder dividir esta conquista com ele, mas sei que se estivesse aqui, estaria extremamente orgulhoso de mim. Aos meus irmãos Di'Stephano e Isabela, o apoio e ombro amigo.

Ao grande amor da minha vida, Patrick, pela paciência, compreensão, suporte emocional e por incentivar meu crescimento profissional. Seu apoio e conselhos me ajudaram a superar os momentos difíceis e a seguir em frente.

A minha orientadora Rosangela Loschi e a minha co-orientadora Denise Duarte por todo empenho, sabedoria e compreensão concedidos e pela oportunidade de crescimento e aprendizado.

Aos membros da banca examinadora, Prof. Renato Assunção (UFMG) e Prof. Helio Migon (UFRJ), pela leitura, correções e sugestões da dissertação. Agradeço, também, ao CNPq pela bolsa de estudos concedida.

Agradeço a minha amiga Márcia pelos momentos de descontração, as palavras de incentivo e a acolhida em seu apartamento por tantos meses. A Jacqueline por ter me aconselhado a realizar a prova de admissão do mestrado da UFMG e pelas dúvidas esclarecidas. Aos amigos Cristiano, Zé Luiz e Fábio Demarqui pelas dicas de programação que contribuíram para que meu algoritmo finalmente funcionasse. E a todos que, de alguma forma, contribuíram para que a realização dessa dissertação fosse possível.

Resumo

Este trabalho analisa a estimação de densidades do ponto de vista Bayesiano não-paramétrico. Especificamente, utiliza o modelo hierárquico de misturas por processos de Dirichlet (MDP). O principal objetivo do trabalho é estender o modelo de mistura de distribuições normais por processos de Dirichlet (MNDP) para estimação de densidades, proposto por Escobar e West (1995). Nossa proposta consiste em estimar densidades utilizando um modelo de MDP cujo primeiro estágio é modelado segundo uma mistura de distribuições *skew-normal*. Como consequência, alguns resultados importantes referentes à inferência na família *skew-normal* de locação e escala são obtidos. Comparamos a qualidade das estimativas de densidade obtidas a partir do modelo de mistura de distribuições *skew-normal* por processos de Dirichlet (MSNDP) com àquelas obtidas pelo modelo MNDP. Essa comparação é realizada implementando-se o algoritmo de MacEachern e Müller (1998). Visto que a distribuição *skew-normal* é mais flexível que a distribuição normal, acomodando distribuições com diferentes graus de assimetria e também com caudas mais leves ou pesadas, o modelo proposto para estimação de densidades é também mais flexível que o modelo de Escobar e West (1995). Dessa forma, espera-se que o modelo MSNDP forneça melhores estimativas para densidades assimétricas e/ou estritamente positivas ou negativas que o modelo MNDP. Temos como objetivo secundário, comparar os algoritmos de Escobar e West (1995) e de MacEachern e Müller (1998) na estimação de densidades via modelo MNDP. A fim de atingir nossas metas, realizamos dois estudos envolvendo dados simulados e um estudo envolvendo dados reais.

Palavras-chaves: *Estimação de densidades; Processos de Dirichlet; Modelo de mistura skew-normal por processos de Dirichlet.*

Abstract

This work addresses the density estimation problem using non-parametric Bayesian approach. It is considered hierarchical mixture models where the uncertainty about the mixing measure is modeled using the Dirichlet process (for short, MDP). The main goal is to build a more flexible model for density estimation. The normal mixture model via Dirichlet process (MNDP) originally proposed by Escobar and West (1995) is extended by considering mixtures of skew-normal distributions (MSNDP), say, in the first stage of the hierarchical model, the normal distribution is replaced by the skew-normal one. As a by product, some important results related to Bayesian inference in the location-scale skew-normal family are introduced. The algorithm introduced by MacEachern and Müller (1998) to sample from the posteriors is used. Considering simulated data sets, the density estimates provided by MSNDP and MNDP are compared. The proposed model (MSNDP) provide much better estimates whenever the data sets comes from non-negative and skewed distributions as well as from mixture of them. If the data sets came from normal or symmetric distributions as well as mixtures of them, the results provided by both models are comparable. MacEachern and Müller (1998) and Escobar and West (1995)'s algorithms were also compared using the MNDP and simulated data sets. MacEachern and Müller (1998)'s algorithm usually provided better results. Finally, the Old Faithful Geyser data set taken from Silverman (1986) is analyzed using MSNDP and MNDP. The former model captured the data bimodality shown in the histogram.

Keywords: Density Estimation, Dirichlet Process, Skew-normal mixture model via Dirichlet process.

Sumário

Lista de Abreviaturas	vi
1 Introdução	1
2 Preliminares	5
2.1 Distribuição Skew-normal	5
2.2 Distribuição de Dirichlet	6
2.3 Processo de Dirichlet - DP	8
2.3.1 Resultado Probabilístico Relacionado com o Processo de Dirichlet	9
2.3.2 Representação de um Processo de Dirichlet via Esquema de Urna de Polya	10
3 Modelos de Misturas por Processos de Dirichlet (MDP)	13
3.1 Algoritmo de Escobar e West (1995)	14
3.2 Algoritmo de MacEachern e Müller (1998)	16
4 Estimação de Densidade Utilizando MDP	18
4.1 Modelo de Mistura Normal via Processos de Dirichlet (MNPDP) – Escobar e West, 1995	18
4.2 Modelo de Mistura Skew-normal via Processos de Dirichlet (MSNPD)	22
4.2.1 Mistura de Distribuições Skew-normal	23
4.2.2 Estimação de Densidades Utilizando Mistura Skew-normal via Processos de Dirichlet	24
5 Análise de Dados Simulados	27
5.1 Comparação dos Algoritmos de Escobar e West (1995) e MacEachern e Müller (1998)	28
5.2 Comparação dos Modelos MSNPD e MNDP	34
5.2.1 Análise da Distribuição a Posteriori do Parâmetro de Assimetria	45

SUMÁRIO

v

6	Estudo de Caso - Duração de Erupções do Géiser Old Faithful	48
7	Conclusões e Discussão	50
A	Demonstrações	52
B	Determinação das Representações Estocásticas	61
	Referências Bibliográficas	67

Lista de Abreviaturas

CRP	Chinese Restaurant Process.
DP	Processo de Dirichlet.
EQMI	Erro Quadrático Médio Integrado.
f.d.a	Função de Distribuição Acumulada.
f.d.p	Função Densidade de Probabilidade.
MCMC	Markov Chain Monte Carlo.
MDP	Mistura de Processos de Dirichlet.
MNPD	Modelo de Mistura Normal via Processos de Dirichlet
MNPD-EW	Modelo de Mistura Normal via Processos de Dirichlet via algoritmo de Escobar e West (1995)
MNPD-MM	Modelo de Mistura Normal via Processos de Dirichlet via algoritmo de MacEachern e Müller (1998)
MSNPD	Modelo de Mistura Skew-Normal via Processos de Dirichlet
SUN	Unified Skew-Normal.

Capítulo 1

Introdução

A estimação de densidade é um dos problemas centrais da Inferência Estatística. Ela se refere à construção de uma estimativa da função de densidade a partir de um conjunto de dados observados. Uma utilização muito natural da estimação de densidade é a investigação das propriedades de um determinado conjunto de dados. As estimativas de densidade podem fornecer indicações valiosas de certas características da distribuição, por exemplo, assimetria e multimodalidade (Silverman, 1986). Uma boa estimativa da densidade de probabilidade pode ser usada para resolver outros problemas, tais como problemas de regressão e de classificação.

Duas abordagens muito utilizadas na estimação de densidade são as abordagens paramétrica e não-paramétrica. A diferença entre elas é que na abordagem paramétrica assume-se que os dados são modelados segundo uma distribuição paramétrica conhecida (*e.g.*, uniforme, normal, etc.), a qual pode ser descrita utilizando-se um número fixo e finito de parâmetros. Uma aproximação para a densidade desconhecida pode ser obtida estimando-se os parâmetros dessa distribuição. Já na abordagem não-paramétrica a estrutura da distribuição que originou os dados não é especificada *a priori*, e sim determinada a partir da amostra observada. O termo não-paramétrico, como utilizado, não sugere que os modelos desse tipo não possuem nenhum parâmetro, mas sim que o número e a natureza dos parâmetros desse tipo de modelo são flexíveis e não pré-determinados (Silverman, 1986).

Um dos métodos amplamente utilizados para estimar densidades faz uso de núcleo estimador. Esse método consiste em gerar uma função contínua (kernel) em torno de cada realização amostral, em geral utiliza-se um kernel Gaussiano. Uma estimativa da função densidade é então obtida a partir da soma, devidamente normalizada, dessas funções. Maiores detalhes a respeito de núcleo estimadores podem ser encontradas em Silverman (1986), Peters e Valafar (2003) e Dias (2002). Outro método muito utilizado faz uso de splines. Nesse método, o conjunto original $[a, b]$ é particionado em k subintervalos da forma $[x_0, x_1], \dots, [x_{k-1}, x_k]$, onde $a = x_0 < x_1 < \dots < x_{k-1} < x_k = b$. Uma aproximação da função densidade é obtida realizando-se, de forma mais suave possível, interpolações com polinômios de pequeno grau em cada subintervalo. O uso de splines na estimação de densidades encontra-se detalhado no trabalho de Dias (2002).

Em Inferência Bayesiana, o problema de estimação de densidades vem sendo tratado do ponto de vista não-paramétrico, utilizando misturas de distribuições simples e considerando a medida misturadora como um objeto aleatório. A incerteza sobre a medida misturadora é descrita utilizando-se processos de Dirichlet. Esta idéia, introduzida por Escobar e West (1995), traz mais flexibilidade para o problema de estimação de densidades, já que dispensa o estabelecimento prévio do número de elementos na mistura. Em particular, Escobar e West (1995) utilizam misturas de distribuições normais para estimar densidades.

O processo de Dirichlet (DP) foi introduzido por Ferguson (1973), o qual também discute algumas de suas propriedades básicas e prova sua existência. O DP é representado de diversas maneiras. Dentre elas, estão a representação em esquema de urna de Polya (Blackwell e MacQueen, 1973) e a representação como uma soma ponderada de massas pontuais (Sethuraman, 1994). A boa fundamentação teórica do DP, juntamente com o rápido desenvolvimento de métodos computacionais Bayesianos não-paramétricos, possibilitou a utilização do DP para resolver problemas aplicados, tais como a estimação de densidades. Do que temos conhecimento, Escobar (1994) foi o primeiro a utilizar o DP na solução de problemas aplicados, onde propõe uma abordagem Bayesiana não-paramétrica para estimar o vetor de médias de um modelo normal. Escobar (1994) também foi responsável pela introdução dos métodos de simulação MCMC, em particular o amostrador de Gibbs, em Inferência Bayesiana não-paramétrica. Desde então, diversos trabalhos vêm desenvolvendo eficientes algoritmos, baseados em métodos MCMC, para amostrar da distribuição *a posteriori* em um modelo de misturas envolvendo processos de Dirichlet. Dentre eles, estão os trabalhos de MacEachern (1994), MacEachern e Müller (1998), Neal (2000), entre outros.

Apesar da proposta de Escobar e West (1995) ser bastante eficiente na estimação de densidades, constatamos que ela tende a não fornecer boas estimativas para a densidade quando os dados provêm, por exemplo, de distribuições assimétricas ou de distribuições que põem massa apenas em valores não-negativos. Além disso, a utilização do algoritmo de Escobar e West (1995) só é interessante quando há conjugação entre a família de distribuições amostrais e a família a qual está associada a distribuição básica do DP. Quando não ocorre conjugação, métodos de integração numérica devem ser utilizados para aproximar algumas integrais. Isso torna a implementação do algoritmo de Escobar e West (1995) mais onerosa do ponto de vista computacional. MacEachern e Müller (1998) estendem a proposta de Escobar e West (1995) para situações mais gerais, criando um algoritmo cujo desempenho não é afetado pela falta de conjugação.

São dois os objetivos deste trabalho. O primeiro é implementar os algoritmos de Escobar e West (1995) e de MacEachern e Müller (1998) segundo a especificação de modelo proposta por Escobar e West (1995), ou seja, considerando misturas de distribuições normais para estimar densidades e uma normal-gama-invertida como distribuição básica do processo de Dirichlet. A meta é comparar

os dois algoritmos utilizando métodos MCMC e avaliar o desempenho do modelo proposto por Escobar e West (1995) na estimação de densidades assimétricas e/ou não-negativas. O segundo e mais importante objetivo é tornar a abordagem de Escobar e West (1995) ainda mais flexível, considerando misturas de distribuições *skew-normal* para a estimação de densidades. A versão da distribuição *skew-normal* que será considerada neste trabalho foi introduzida por Azzalini (1985).

A família de distribuições *skew-normal* introduzida por Azzalini (1985) inclui como caso particular a família normal e tem como casos limites as famílias *half-normal* positivas e negativas. Dessa forma, visamos obter melhores estimativas para densidades assimétricas e/ou densidades que põem massa apenas em valores não-negativos ou não-positivos.

Como demonstrado em Arellano-Valle *et al.* (2009) e Arellano-Valle *et al.* (2011), as distribuições *a priori* e *a posteriori* dos parâmetros da distribuição *skew-normal* pertencem a mesma classe de distribuições. Isso indica a existência de conjugação na família de distribuições *skew-normal*. Apesar disso, o cálculo das distribuições *a posteriori* é bastante intenso do ponto de vista computacional. Por essa razão e devido ao fato de o algoritmo de Escobar e West (1995) ter se mostrado menos eficiente na estimação de densidades que o algoritmo de MacEachern e Müller (1998), optamos por implementar este último.

Um estudo empregando dados simulados é realizado, através da implementação do algoritmo de MacEachern e Müller (1998), a fim de comparar as estimativas obtidas utilizando misturas de distribuições normais com as obtidas utilizando distribuições *skew-normal*. Realizamos também um estudo empregando dados reais, que dizem respeito à duração de erupções de um géiser no Parque Nacional de Yellowstone, nos EUA. Esses dados encontram-se disponíveis em Silverman (1986).

Misturas de distribuições *skew-normal* têm sido amplamente discutidas na literatura. Arellano-Valle *et al.* (2008) introduzem uma classe de misturas no parâmetro de assimetria da distribuição *skew-normal* e estudam suas principais propriedades. Arellano-Valle *et al.* (2009) consideram a classe de misturas introduzidas em Arellano-Valle *et al.* (2008) e discutem questões teóricas relativas à inferência Bayesiana na classe de distribuições considerada, como conjugação e robustez. Liu *et al.* (2007) consideram misturas finitas de distribuições *skew-normal* para a estimação de densidade. O trabalho de Liu *et al.* (2007) está focado no desenvolvimento do algoritmo EM para obtenção dos estimadores de máxima verossimilhança dos parâmetros da distribuição *skew-normal* e no desenvolvimento de um algoritmo MCMC para obter aproximações das distribuições *a posteriori* dos parâmetros do modelo de misturas finitas. A limitação do método proposto por Liu *et al.* (2007) está em considerar como conhecido o número de elementos envolvidos na mistura.

O conteúdo deste trabalho está organizado em 6 capítulos e 2 apêndices. No Capítulo 2 apresentamos, brevemente, conceitos que são fundamentais para o entendimento deste trabalho. Esse capítulo está dividido em três seções, denominadas respectivamente por Distribuição *skew-normal*, Distribuição Dirichlet e Processo de Dirichlet. Na última seção, começamos introduzindo a definição de processos de Dirichlet, em seguida apresentamos um importante resultado probabilístico envolvendo o mesmo e por fim descrevemos a representação de um DP em esquema de urna de Polya devido a Blackwell e MacQueen (1973). No Capítulo 3 introduzimos o conceito de modelo de Misturas de Processos de Dirichlet e apresentamos os algoritmos de Escobar e West (1995) e de MacEachern e Müller (1998), que fornecem aproximações das distribuições *a posteriori* dos parâmetros desse modelo. No Capítulo 4 apresentamos dois métodos de estimação de densidade que fazem uso de modelos de Misturas de Processos de Dirichlet, o método proposto por Escobar e West (1995) e o método proposto por este trabalho. No Capítulo 5 apresentamos os resultados obtidos a partir das análises de dados simulados. Já no Capítulo 6 apresentamos o resultado da análise dos dados reais. No sétimo e último capítulo apresentamos as principais conclusões, algumas discussões pertinentes e os trabalhos futuros a serem desenvolvidos. Nos Apêndices A e B encontram-se as demonstrações de alguns resultados apresentados neste trabalho.

Capítulo 2

Preliminares

Neste capítulo, apresentamos alguns resultados e definições que serão utilizados ao longo deste trabalho. Iniciamos, apresentando a distribuição normal assimétrica ou distribuição *skew-normal* introduzida por Azzalini (1985). Optamos por utilizar a denominação *skew-normal*.

2.1 Distribuição Skew-normal

A distribuição *skew-normal* é uma generalização da distribuição normal, que, como o próprio nome sugere, consiste de uma assimetria da distribuição normal. A função densidade de probabilidade (f.d.p) da distribuição *skew-normal* foi construída multiplicando-se a densidade da distribuição normal por sua função de distribuição acumulada (f.d.a), que é responsável por introduzir assimetria na f.d.p original. Outras maneiras de assimetrizar a distribuição normal foram propostas por O'Hagan e Leonhard (1976), Ferreira e Steel (2006), entre outros.

Azzalini (1985) define que, dados os parâmetros de locação $\mu \in \mathbb{R}$, de escala $V \in \mathbb{R}^+$ e de assimetria $\lambda \in \mathbb{R}$, uma variável aleatória (v.a) X tem distribuição *skew-normal*, denotada por $X|\mu, V, \lambda \sim SN(\mu, V, \lambda)$, se sua f.d.p é dada por:

$$f(x|\mu, V, \lambda) = \frac{2}{V^{\frac{1}{2}}} \phi\left(\frac{x - \mu}{V^{\frac{1}{2}}}\right) \Phi\left(\lambda \frac{x - \mu}{V^{\frac{1}{2}}}\right), \quad x \in \mathbb{R}, \quad (2.1)$$

onde ϕ e Φ denotam a f.d.p e a f.d.a de uma distribuição normal padrão, respectivamente.

A distribuição *skew-normal*, dada em (2.1), possui as seguintes propriedades:

1. Quando $\lambda = 0$, a densidade em (2.1) se reduz à densidade da distribuição normal, ou seja, não há assimetria.
2. Quanto maior o valor assumido por λ (em valor absoluto), maior será o grau de assimetria da distribuição.
3. Valores positivos (negativos) de λ indicam assimetria positiva (negativa) da distribuição.

4. Quando $|\lambda| \rightarrow \infty$, a densidade em (2.1) converge para uma função de densidade denominada “*half-normal*”.
5. Se $Z \sim SN(0, 1, \lambda)$, então $|Z|$ tem distribuição *half-normal*, para qualquer λ , e $Z^2 \sim \chi_1^2$.
6. A esperança e a variância de uma v.a com distribuição *skew-normal* são dadas, respectivamente, por:

$$E(X|\mu, V, \lambda) = \mu + V^{\frac{1}{2}}\delta\sqrt{\frac{2}{\pi}} \quad \text{e} \quad V(X|\mu, V, \lambda) = V \left(1 - \frac{2\delta^2}{\pi} \right),$$

onde $\delta = \lambda/\sqrt{1 + \lambda^2}$, $\delta \in [0, 1]$.

A Figura 2.1 mostra a f.d.p da distribuição *skew-normal* padrão para diferentes parâmetros de assimetria.

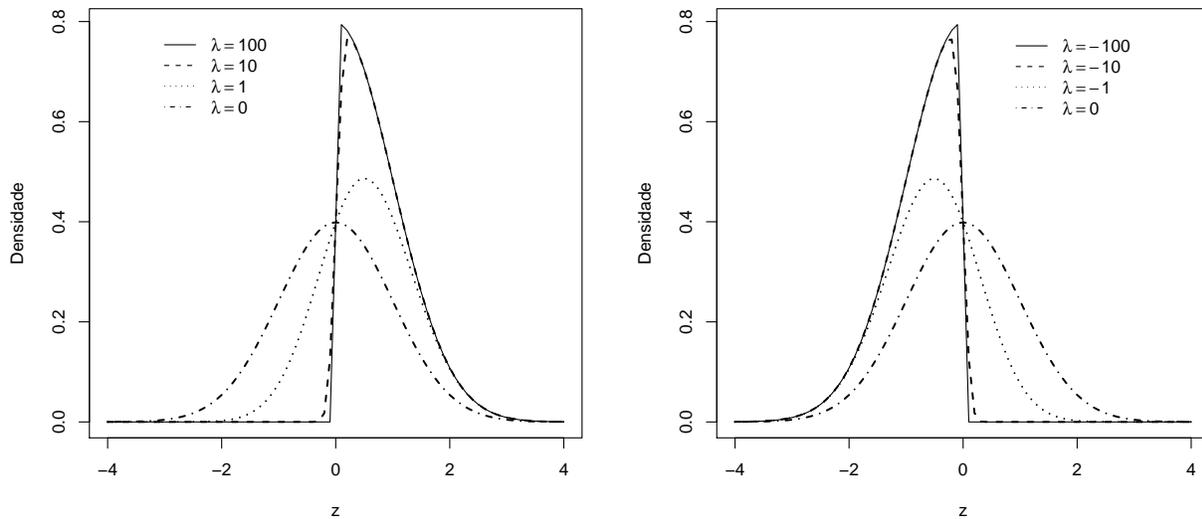


Figura 2.1: Função de densidade de probabilidade da *skew-normal* padrão considerando diversos valores para λ

Outro importante conceito utilizado ao longo deste trabalho é o de processo de Dirichlet. Antes de introduzirmos esse conceito, iremos apresentar a definição de distribuição de Dirichlet.

2.2 Distribuição de Dirichlet

A distribuição de Dirichlet é uma generalização multiparamétrica da distribuição Beta, assim como o modelo Multinomial é uma generalização multiparamétrica da distribuição Binomial (Zapata, 2007).

Seja $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_k)$ um vetor k dimensional tal que $0 \leq \pi_i \leq 1$, para todo $i = 1, \dots, k$, e $\sum_{i=1}^k \pi_i = 1$. Dizemos que $\boldsymbol{\pi}$ tem distribuição Dirichlet com parâmetros $\alpha_i > 0$, $i = 1, \dots, k$, denotada por $\boldsymbol{\pi} \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_k)$, se sua f.d.p é da forma

$$f(\boldsymbol{\pi} | \alpha_1, \dots, \alpha_k) = \frac{\Gamma(\sum_j \alpha_j)}{\prod_j \Gamma(\alpha_j)} \prod_{j=1}^k \pi_j^{\alpha_j - 1}. \quad (2.2)$$

Se em (2.2) k for igual a dois, obtemos a distribuição Beta com parâmetros α_1 e α_2 , uma vez que a f.d.p de $\boldsymbol{\pi}$ torna-se

$$f(\pi_1, \pi_2 | \alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \pi_1^{\alpha_1 - 1} (1 - \pi_1)^{\alpha_2 - 1}.$$

O exemplo a seguir ilustra o comportamento da distribuição de Dirichlet em função de seus parâmetros.

Exemplo: Seja $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)$. Suponha que $\boldsymbol{\pi} \sim \text{Dirichlet}(\alpha_1, \alpha_2, \alpha_3)$, $\alpha_i > 0$, $i = 1, 2, 3$. O domínio desta distribuição é o prisma definido pelos pontos $(0, 0, 1)$, $(0, 1, 0)$ e $(1, 0, 0)$, dado na Figura 2.2.

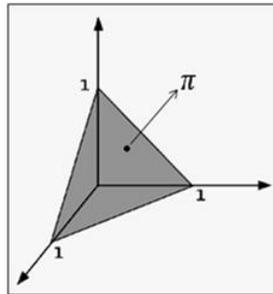


Figura 2.2: Domínio da distribuição de Dirichlet, $k = 3$

A probabilidade de selecionar as diferentes combinações de $\boldsymbol{\pi}$ vai depender dos valores de $(\alpha_1, \alpha_2, \alpha_3)$. A Figura 2.3 mostra as curvas de nível para diferentes valores de α_1, α_2 e α_3 . As regiões com cores mais intensas são aquelas onde há maior massa de probabilidade. Essa figura foi extraída de Teh (2007).

A Figura 2.3 demonstra que, nos casos onde os parâmetros α_1, α_2 e α_3 são todos iguais, a distribuição de Dirichlet é simétrica. Especificamente, no caso em que $\alpha_1 = \alpha_2 = \alpha_3 = 1$, tem-se uma distribuição uniforme no simplex. Já no caso em que todos os parâmetros são iguais e menores que um, a distribuição de Dirichlet tende a colocar mais massa nas extremidades do simplex.

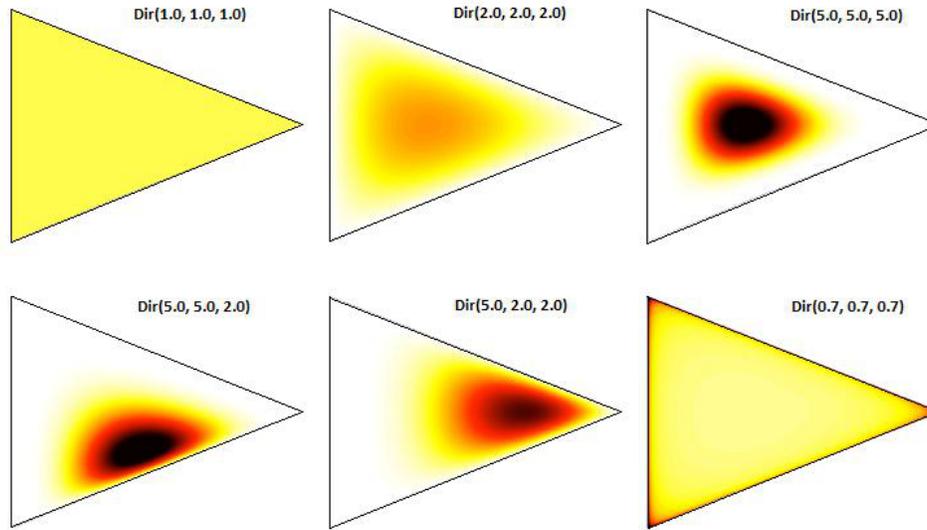


Figura 2.3: Exemplos de distribuição de Dirichlet.

2.3 Processo de Dirichlet - DP

O processo de Dirichlet (DP) foi introduzido por Ferguson (1973) como uma distribuição *a priori* sobre o espaço de medidas de probabilidade definidas sobre um espaço mensurável $(\Theta, \mathcal{B}(\Theta))$, onde Θ é um conjunto não enumerável (*e.g.* \mathbb{R} , $[0, 1]$, \mathbb{R}^k , \dots) e $\mathcal{B}(\Theta)$ é uma σ -álgebra de subconjuntos de Θ . Tal processo pode ser entendido como uma distribuição sobre medidas de probabilidade, isto é, cada realização de um processo de Dirichlet é, por si só, uma distribuição de probabilidade. Distribuições obtidas a partir de um processo de Dirichlet são discretas com probabilidade um (Sethuraman, 1994).

Seja G uma medida de probabilidade aleatória sobre $(\Theta, \mathcal{B}(\Theta))$. Essa medida é vista como um “parâmetro” em Estatística não-paramétrica e assume valores em \mathcal{G} , o conjunto de todas as possíveis medidas de probabilidade que podem ser definidas sobre $(\Theta, \mathcal{B}(\Theta))$ (Sethuraman, 1994). Formalmente, tem-se:

Definição: Uma medida de probabilidade aleatória G é distribuída segundo um processo de Dirichlet se, para toda partição finita A_1, \dots, A_k de Θ , o vetor aleatório de probabilidades $(G(A_1), \dots, G(A_k))$ tem distribuição de Dirichlet com parâmetros $\alpha G_0(A_i)$, $i = 1, \dots, k$, denotada por

$$(G(A_1), \dots, G(A_k)) \sim \text{Dirichlet}(\alpha G_0(A_1), \dots, \alpha G_0(A_k)),$$

onde G_0 é a distribuição basal e α é um número real positivo.

A notação $G \sim DP(\alpha, G_0)$ será utilizada para indicar que a distribuição de probabilidade aleatória G é definida por um processo de Dirichlet com parâmetros $\alpha > 0$ e G_0 .

A distribuição base G_0 pode ser interpretada como uma distribuição média a ser subjetivamente escolhida de tal forma que $E[G(A)] = G_0(A)$, para qualquer subconjunto mensurável $A \subset \Theta$. O parâmetro α , por sua vez, controla a dispersão das distribuições obtidas via DP em torno da distribuição básica G_0 . Se $G \sim DP(\alpha, G_0)$, segue que $V[G(A)] = G_0(A)(1 - G_0(A))/(1 - \alpha)$. Se α for pequeno, significa que há muita variação em torno de G_0 . Por outro lado, se α for grande as distribuições obtidas via DP são similares a G_0 .

2.3.1 Resultado Probabilístico Relacionado com o Processo de Dirichlet

A conjugação é uma característica interessante do processo de Dirichlet. Essa característica nos permite facilmente obter a distribuição *a posteriori* da medida de probabilidade aleatória G .

Suponha que $G \sim DP(\alpha, G_0)$. Sejam $\theta_1, \dots, \theta_n$ variáveis aleatórias tais que, dado G , são i.i.d. com distribuição G , isto é, $\theta_1, \dots, \theta_n | G \stackrel{iid}{\sim} G$. Note que os θ_i 's assumem valores em Θ , uma vez que G é uma distribuição definida sobre este espaço. Estamos interessados na distribuição de G dado os valores observados de $\theta_1, \dots, \theta_n$.

Seja A_1, \dots, A_k uma partição qualquer de Θ , finita e mensurável. Assim, por definição, tem-se que $(G(A_1), \dots, G(A_k)) \sim Dirichlet(\alpha G_0(A_1), \dots, \alpha G_0(A_k))$. Tem-se ainda que, para todo $j = 1, \dots, n$,

$$P(\theta_j \in A_i | G) = G(A_1)^{\delta_{\theta_j}(A_1)} \dots G(A_k)^{\delta_{\theta_j}(A_k)}; \quad i = 1, \dots, k,$$

$$\text{onde } \delta_{\theta_j}(A_i) = \begin{cases} 1 & \text{se } \theta_j \in A_i, \\ 0 & \text{caso contrário.} \end{cases}$$

Dessa forma, segue que

$$\begin{aligned} P(\theta_1, \dots, \theta_n | G) &= \prod_{j=1}^n G(A_1)^{\delta_{\theta_j}(A_1)} \dots G(A_k)^{\delta_{\theta_j}(A_k)} \\ &= G(A_1)^{\sum_{j=1}^n \delta_{\theta_j}(A_1)} \dots G(A_k)^{\sum_{j=1}^n \delta_{\theta_j}(A_k)}, \end{aligned}$$

que é um produto de distribuições multinomiais independentes com parâmetros $(1, G(A_1), \dots, G(A_k))$.

Utilizando resultados de conjugação entre as famílias Dirichlet e Multinomial, tem-se que

$$\begin{aligned} P(G(A_1), \dots, G(A_k) | \theta_1, \dots, \theta_n) &\propto P(\theta_1, \dots, \theta_n | G) P(G(A_1), \dots, G(A_k)) \\ &\propto \left(G(A_1)^{\sum_{j=1}^n \delta_{\theta_j(A_1)}} \dots G(A_k)^{\sum_{j=1}^n \delta_{\theta_j(A_k)}} \right) \left(G(A_1)^{\alpha G_0(A_1)-1} \dots G(A_k)^{\alpha G_0(A_k)-1} \right) \\ &\propto \left(G(A_1)^{\alpha G_0(A_1) + \sum_{j=1}^n \delta_{\theta_j(A_1)} - 1} \dots G(A_k)^{\alpha G_0(A_k) + \sum_{j=1}^n \delta_{\theta_j(A_k)} - 1} \right), \end{aligned}$$

ou seja, segue que

$$(G(A_1), \dots, G(A_k)) | \theta_1, \dots, \theta_n \sim \text{Dirichlet}(\alpha G_0(A_1) + n_1, \dots, \alpha G_0(A_k) + n_k), \quad (2.3)$$

onde $n_k = \#\{i : \theta_i \in A_k\}$ denota o número valores observados em A_k , ou seja, $n_k = \sum_{i=1}^n \delta_{\theta_i(A_k)}$.

Uma vez que (2.3) é válida para qualquer partição finita mensurável de Θ , a distribuição de G condicional nos valores observados $\theta_1, \dots, \theta_n$ segue um processo de Dirichlet com parâmetros $\alpha^* = \alpha + n$ e $G_0^* = \frac{\alpha}{\alpha+n} G_0 + \frac{n}{\alpha+n} \frac{\sum_{i=1}^n \delta_{\theta_i}}{n}$, ou seja, a distribuição *a posteriori* de G é dada por

$$G | \theta_1, \dots, \theta_n \sim DP \left(\alpha + n, \frac{\alpha}{\alpha+n} G_0 + \frac{n}{\alpha+n} \frac{\sum_{i=1}^n \delta_{\theta_i}}{n} \right). \quad (2.4)$$

2.3.2 Representação de um Processo de Dirichlet via Esquema de Urna de Polya

Existem três formas usuais de se representar o processo de Dirichlet, a representação em esquema de urna de Polya (Blackwell e MacQueen, 1973), a representação como uma soma ponderada de massas pontuais (Sethuraman, 1994) e a representação denominada Chinese Restaurant Process (CRP)¹. Cada uma dessas representações fornece um método para se gerar realizações de um DP.

Nesta seção discutiremos a representação em esquema de urna de Polya. Essa representação é utilizada no cálculo das distribuições condicionais completas, necessárias para a implementação dos algoritmos MCMC considerados neste trabalho, e na obtenção de uma aproximação da distribuição preditiva *a posteriori*, que serve como uma estimativa da densidade desconhecida.

Suponha que $G \sim DP(\alpha, G_0)$ e que $\theta_1, \dots, \theta_n | G \stackrel{iid}{\sim} G$. Seja A um subconjunto qualquer de Θ . Seja θ_{n+1} um valor futuro a ser observado. Assuma que, dado G , $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ e θ_{n+1} são independentes e tais que $\theta_{n+1} | G, \boldsymbol{\theta} \sim G$. Segue então que

¹Para maiores detalhes a respeito do CRP acesse os sites <http://www.cs.princeton.edu/courses/archive/fall07/cos597C/scribe/20070921.pdf> e http://videlectures.net/icml05_jordan.dpcrp/.

$$\begin{aligned}
P(\theta_{n+1} \in A | \theta_1, \dots, \theta_n) &= \int P(\theta_{n+1} \in A | G) P(G | \boldsymbol{\theta}) dG \\
&= \int G(A) P(G | \boldsymbol{\theta}) dG = E(G(A) | \theta_1, \dots, \theta_n) \\
&= \frac{1}{\alpha + n} \left(\alpha G_0 + \sum_{j=1}^n \delta_{\theta_j} \right).
\end{aligned}$$

Dessa forma, a distribuição preditiva para θ_{n+1} condicionada ao vetor $\theta_1, \dots, \theta_n$ é dada por:

$$\theta_{n+1} | \theta_1, \dots, \theta_n \sim \frac{\alpha}{\alpha + n} G_0 + \frac{n}{\alpha + n} \frac{\sum_{i=1}^n \delta_{\theta_i}}{n}. \quad (2.5)$$

Pode-se notar que a distribuição base do processo de Dirichlet em (2.4) é também a distribuição preditiva de θ_{n+1} , dados $\theta_1, \dots, \theta_n$, exibida em (2.5).

A sequência de distribuições preditivas em (2.5) para $\theta_1, \theta_2, \dots$ é denominada esquema de Urna de Polya de Blackwell-MacQueen (Teh, 2010 apud Blackwell e MacQueen, 1973).

Uma das vantagens deste tipo de representação é que se pode construir um algoritmo para se amostrar do processo de Dirichlet. A partir de (2.5) tem-se que os θ_i 's podem ser gerados da seguinte forma:

Passo 1: Gere $\theta_1 \sim G_0$;

Passo 2: Para $i = 2, \dots, n$, gere

$$\theta_i | \theta_1, \dots, \theta_{i-1} = \begin{cases} \theta_1 & \text{com prob. } 1/(\alpha + i - 1) \\ \vdots & \\ \theta_{i-1} & \text{com prob. } 1/(\alpha + i - 1) \\ \text{Nova retirada de } G_0 & \text{com prob. } \alpha/(\alpha + i - 1) \end{cases}$$

A representação de Blackwell e MacQueen (1973) demonstra que, com probabilidade positiva, alguns θ_i 's possuirão valores coincidentes. Dessa forma, com probabilidade positiva, uma sequência de n valores $(\theta_1, \dots, \theta_n)$ será reduzida a uma sequência de $k \leq n$ valores distintos, os quais serão denotados por $\theta_1^*, \dots, \theta_k^*$.

A sequência $\theta_1^*, \dots, \theta_k^*$ induz uma partição do conjunto $[n] = \{1, \dots, n\}$ em k clusters, tal que se $i, i = 1, \dots, n$, estiver no cluster j , então $\theta_i = \theta_j^*$. Uma vez que os θ_i 's são aleatórios, tem-se que a partição induzida por eles também será aleatória.

Dessa forma, a distribuição preditiva em (2.5) pode ser equivalentemente representada por:

$$\theta_{n+1}|\theta_1, \dots, \theta_n \sim \frac{\alpha}{\alpha + n} G_0 + \frac{1}{\alpha + n} \sum_{j=1}^k n_j \delta_{\theta_j^*} \quad (2.6)$$

Note que o valor de θ_{n+1} será igual ao valor θ_j^* com probabilidade proporcional a n_j , que denota o número de observações no *cluster* j . Assim, quanto maior n_j , maior a probabilidade de que θ_{n+1} pertença ao j -ésimo *cluster*. Este fenômeno é denotado por “*rich-gets-richer*”.

Note ainda que para $i \geq 2$, a observação θ_i toma um novo valor (incrementando assim k em 1 unidade) com probabilidade $\alpha/((\alpha + i - 1))$. Portanto, α controla o número de *clusters* de uma maneira bem direta, com valores grandes de α implicando em um grande número de *clusters*.

Esta é a idéia por trás da modelagem Bayesiana para estimação de densidades que será vista no Capítulo 4.

Capítulo 3

Modelos de Misturas por Processos de Dirichlet (MDP)

Apesar da idéia de agrupamento, necessária à estimação de densidades, ser inerente ao processo de Dirichlet, vimos que distribuições obtidas a partir de um DP são discretas com probabilidade um, tornando tal processo inadequado para estimar densidades. Uma extensão do DP, que busca remover a restrição à medidas discretas, consiste em realizar uma mistura entre uma distribuição contínua com uma medida de probabilidade aleatória que segue um processo de Dirichlet ¹. Ou seja, consiste em construir um modelo de mistura via processos de Dirichlet (MDP). Seja \mathcal{Y} o espaço amostral e Θ o espaço paramétrico. Para cada $\theta \in \Theta$, considere $f(y|\theta)$ uma família de distribuições (densidades) amostrais. Seja G uma medida de probabilidade aleatória sobre Θ e suponha que $G \sim DP(\alpha, G_0)$. Então, define-se uma densidade aleatória para $y \in \mathcal{Y}$ como se segue

$$f(y) = \int f(y|\theta)G(\theta) d\theta. \quad (3.1)$$

A mistura em (3.1) pode ser hierarquicamente representada como segue:

$$\begin{aligned} Y_i|\theta_i &\stackrel{iid}{\sim} f(\cdot|\theta_i), \quad i = 1, \dots, n; \\ \theta_1, \dots, \theta_n|G &\stackrel{iid}{\sim} G; \\ G|\alpha, G_0 &\sim DP(\alpha, G_0). \end{aligned} \quad (3.2)$$

O modelo em (3.2) representa uma mistura de densidades que dependem de um parâmetro θ_i , $i = 1, \dots, n$, onde a medida misturadora é a distribuição G , que é desconhecida e, portanto, aleatória. O processo de Dirichlet é utilizado para descrever o comportamento *a priori* dessa distribuição aleatória.

Uma vez que utilizamos o processo de Dirichlet para descrever a incerteza sobre G , alguns valores de θ_i poderão ser iguais. Porém, não sabemos quantos θ_i 's são distintos. Dessa forma, o modelo anterior pode ser visto com um modelo de mistura onde Y_i 's com o mesmo valor de θ_i pertencem a um mesmo *cluster*, ou seja, possuem a mesma distribuição (Teh, 2010).

¹Para maiores detalhes ver Zapata (2007) e Vallejos (2008), Capítulo 1.

A fim de realizar inferências sobre $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ é necessário obter a distribuição *a posteriori* desse vetor de parâmetros. Uma aproximação da distribuição *a posteriori* pode ser obtida utilizando-se métodos computacionais comumente considerados em Estatística Bayesiana.

Neste capítulo apresentamos os algoritmos de Escobar e West (1995) e de MacEachern e Müller (1998) para se obter aproximações das distribuições *a posteriori* em um modelo de MDP. Esses algoritmos são baseados em esquemas MCMC, os quais foram trazidos para a Inferência Bayesiana por Gelfand e Smith (1990). Um dos métodos MCMC que vem sendo amplamente utilizado é o amostrador de Gibbs, o qual é descrito a seguir.

Suponha que se queira obter uma amostra da distribuição *a posteriori* de $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$, a qual será denotada por $f(\boldsymbol{\theta}|\mathbf{y})$. Uma condição necessária para a implementação do amostrador de Gibbs, é que as distribuições condicionais completas *a posteriori* $f(\theta_i|\boldsymbol{\theta}_{(-i)}, \mathbf{y})$ sejam todas conhecidas, onde $\boldsymbol{\theta}_{(-i)} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n)$. O algoritmo para amostrar da distribuição *a posteriori* é:

Passo 1: Inicialize o vetor $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n)$ com os valores $(\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_n^{(0)})$.

Passo 2: Para $r = 1, \dots, R$ faça:

- Gere $\theta_1^{(r)}$ da distribuição $f(\theta_1|\theta_2^{(r-1)}, \dots, \theta_n^{(r-1)}, \mathbf{y})$;
- Gere $\theta_2^{(r)}$ da distribuição $f(\theta_2|\theta_1^{(r)}, \theta_3^{(r-1)}, \dots, \theta_n^{(r-1)}, \mathbf{y})$;
- Gere $\theta_3^{(r)}$ da distribuição $f(\theta_3|\theta_1^{(r)}, \theta_2^{(r)}, \theta_4^{(r-1)}, \dots, \theta_n^{(r-1)}, \mathbf{y})$
- ⋮
- Gere $\theta_n^{(r)}$ da distribuição $f(\theta_n|\theta_1^{(r)}, \dots, \theta_{n-1}^{(r)}, \mathbf{y})$;

onde $\mathbf{y} = (y_1, \dots, y_n)$.

O Passo 2 é repetido até que se obtenha uma amostra do tamanho desejado. A amostra é formada após a convergência ter sido atingida, digamos no passo M . Após descartar os M primeiros valores gerados, seleciona-se um valor a cada l iterações para pertencer à amostra, evitando assim que os valores amostrais sejam correlacionados. Toda inferência é realizada a partir dessa amostra obtida. Mais detalhes podem ser vistos em Gamerman e Lopes (2006), Casella e Roberts (1999), entre outros.

3.1 Algoritmo de Escobar e West (1995)

O método computacional proposto por Escobar e West (1995) para se amostrar de um modelo de MDP é uma extensão direta do método proposto por Escobar (1994). Ele é particularmente

útil quando o modelo de MDP é construído assumindo-se que a família de distribuições amostrais $f(y_i|\theta_i)$ é conjugada natural da família de distribuições básicas G_0 do DP. Quando não há conjugação, este algoritmo passa a demandar muito esforço computacional, visto que o cálculo de algumas distribuições tem que ser feitos de forma numérica. Isso torna esse algoritmo menos eficiente.

Escobar (1994) prova que num processo de Dirichlet as distribuições condicionais completas *a posteriori* são dadas por:

$$f(\theta_i|\boldsymbol{\theta}_{(-i)}, \mathbf{y}) \propto q_0 f(\theta_i|y_i) + \sum_{j=1, j \neq i}^n q_j \delta_{\theta_j}(\theta_i), \quad (3.3)$$

ou, equivalentemente,

$$(\theta_i|\boldsymbol{\theta}_{-i}, \mathbf{y}) = \left\{ \begin{array}{ll} \theta_1 & \text{com prob. proporcional a } q_1 \\ \vdots & \\ \theta_{i-1} & \text{com prob. proporcional a } q_{i-1} \\ \theta_{i+1} & \text{com prob. proporcional a } q_{i+1} \\ \vdots & \\ \theta_n & \text{com prob. proporcional a } q_n \\ \text{Retirada de uma } f(\theta_i|y_i) & \text{com prob. proporcional a } q_0, \end{array} \right.$$

onde

$$f(\theta_i|y_i) = \frac{\alpha}{q_0} f(y_i|\theta_i) g_0(\theta_i) = \frac{f(y_i|\theta_i) g_0(\theta_i)}{\int f(y_i|\theta_i) g_0(\theta_i) d\theta_i},$$

$$q_0 = \alpha \int f(y_i|\theta_i) g_0(\theta_i) d\theta_i,$$

$q_j = f(y_i|\theta_j)$ é a densidade de Y no ponto y_i , dado θ_j , $g_0(\cdot)$ é a densidade obtida da distribuição base G_0 e α é o parâmetro do DP. Uma amostra da distribuição *a posteriori* de $\boldsymbol{\theta}$ é obtida implementando-se o amostrador de Gibbs descrito na introdução deste capítulo.

Note que este algoritmo deve ser implementado, preferencialmente, no caso em que haja a possibilidade de se calcular analiticamente a integral necessária para o cálculo de q_0 . Isso é possível, por exemplo, quando a distribuição *a priori* $g_0(\cdot)$ é conjugada da família amostral à qual $f(y_i|\theta_i)$ pertence. Quando uma distribuição *a priori* não-conjugada é utilizada, torna-se, muitas vezes, necessário realizar integração numérica, o que torna o algoritmo menos eficiente e sua implementação mais difícil.

MacEachern e Müller (1998) apresentaram um algoritmo mais eficiente para amostrarmos da distribuição *a posteriori* em modelos de MPD no caso em que q_0 não pode ser obtido analiticamente.

Esse tipo de situação pode ocorrer, particularmente, quando não há conjugação entre a família de distribuições amostrais $f(y_i|\theta_i)$ e a família de distribuições basais G_0 . Tal algoritmo é descrito a seguir.

3.2 Algoritmo de MacEachern e Müller (1998)

Iniciamos a apresentação do algoritmo proposto por MacEachern e Müller (1998) estabelecendo a notação que será utilizada. Como em (3.2), assumamos que $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$. Seja $\boldsymbol{\theta}^* = (\theta_1^*, \dots, \theta_k^*)$, $k \leq n$, o vetor formado pelos valores de $\boldsymbol{\theta}$ distintos. Seja $\mathbf{c} = (c_1, \dots, c_n)$ o vetor de indicadores de grupo ou *cluster* ao qual cada θ_i pertence, ou seja, $c_i = j$ se $\theta_i = \theta_j^*$, $i = 1, \dots, n$ e $j = 1, \dots, k$. Seja n_j o número de elementos pertencentes ao j -ésimo *cluster*. Aqui, o termo “*cluster*” está sendo utilizado para se referir ao conjunto de todas as observações y_i , ou dos parâmetros θ_i correspondentes, que possuem o mesmo indicador de grupo c_i .

Considere o vetor $\boldsymbol{\theta}_a^* = (\underbrace{\theta_1^*, \dots, \theta_k^*}_{\boldsymbol{\theta}_F^*}, \underbrace{\theta_{k+1}^*, \dots, \theta_n^*}_{\boldsymbol{\theta}_E^*})$, onde $\boldsymbol{\theta}_E^*$ corresponde ao vetor de potenciais *clusters* que ainda não foram formados, denominado vetor de *clusters* vazios, e $\boldsymbol{\theta}_F^* = \boldsymbol{\theta}^*$ é o vetor de *clusters* que contém pelo menos um θ_i , denominado vetor de *clusters* cheios.

O algoritmo de MacEachern e Müller (1998) pode ser descrito da seguinte forma:

Passo 1: Gere o vetor $(\theta_1^{*(0)}, \dots, \theta_k^{*(0)}, \theta_{k+1}^{*(0)}, \dots, \theta_n^{*(0)})$ tal que $\theta_j^{*(0)} \sim G_0$, $j = 1, \dots, n$.

Passo 2: Atribua ao número de *clusters* k um valor inicial $k_0 \leq n$.

Passo 3: Gere o número de elementos dentro de cada um dos k_0 *clusters* a partir de uma distribuição *Multinomial* (n, p_1, \dots, p_{k_0}) , onde $p_j = 1/k_0$, $\forall j, j = 1, \dots, k_0$.

Passo 4: Com base nos valores obtidos no Passo 3, construa o vetor $\mathbf{c}^{(0)}$ (indicador de *cluster*).

Passo 5: Para $r = 1, \dots, R$ faça:

Passo 5.1: Atualização do vetor $\mathbf{c}^{(r-1)}$ de forma a obter o vetor $\mathbf{c}^{(r)}$. Para $i = 1, \dots, n$ faça:

- Determine o valor de $n_{c_i}^{[r-1; i-1]}$, o número de componentes no vetor $\mathbf{c}^{[r-1; i-1]} = (c_1^{(r)}, \dots, c_{i-1}^{(r)}, c_i^{(r-1)}, c_{i+1}^{(r-1)}, \dots, c_n^{(r-1)})$ iguais a $c_i^{(r-1)}$.

- Se $n_{c_i}^{[r-1; i-1]} > 1$, então gere $c_i^{(r)}$ de acordo com a seguinte distribuição

$$(c_i | \mathbf{c}_{(-i)}^{[r-1; i-1]}, \boldsymbol{\theta}_a^{*(r-1)}, \mathbf{y}) \propto \frac{\alpha}{k_{i-1} + 1} f(y_i | \theta_{k+1}^{*(r-1)}) + \sum_{j=1}^{k_{i-1}} n_{j(-i)} f(y_i | \theta_j^{*(r-1)}), \quad (3.4)$$

onde $n_{j(-i)}$ é o número de observações, com exceção da i -ésima observação, que pertencem ao j -ésimo *cluster*; k_{i-1} é o número de *clusters* em $\mathbf{c}^{[r-1; i-1]}$; e $\mathbf{c}_{(-i)}^{[r-1; i-1]} = (c_1^{(r)}, \dots, c_{i-1}^{(r)}, c_{i+1}^{(r-1)}, \dots, c_n^{(r-1)})$.

- Se a i -ésima observação está sozinha em seu *cluster*, isto é, se $n_{c_i}^{[r-1; i-1]} = 1$, então gere $c_i^{(r)}$ da seguinte forma: (i) $c_i^{(r)} = c_i^{(r-1)}$ com probabilidade $(k_{i-1} - 1)/k_{i-1}$; e (ii) $c_i^{(r)}$ assume um novo valor gerado da seguinte distribuição

$$(c_i | \mathbf{c}_{(-i)}^{[r-1; i-1]}, \boldsymbol{\theta}_a^{*(r-1)}, \mathbf{y}) \propto \frac{\alpha}{k_{i-1}} f(y_i | \theta_{k+1}^{*(r-1)}) + \sum_{j=1}^{k_{i-1}} n_{j(-i)} f(y_i | \theta_j^{*(r-1)}), \quad (3.5)$$

com probabilidade $1/k_{i-1}$.

- Determine o valor de k_i . Para $i = 1, \dots, n-1$, k_i denota o número de *clusters* em $\mathbf{c}^{[r-1; i]}$ e para $i = n$, k_i denota o número de *clusters* em $\mathbf{c}^{(r)}$.

Passo 5.2: Gere o vetor $(\theta_1^{*(r)}, \dots, \theta_k^{*(r)}, \theta_{k+1}^{*(r)}, \dots, \theta_n^{*(r)})$. Para $j = k+1, \dots, n$ isto significa gerar um novo valor da distribuição base G_0 , ou seja, $\theta_j^* \sim G_0$. Para $j = 1, \dots, k$, isto significa gerar um valor da distribuição condicional completa *a posteriori* de θ_j^* .

Em Inferência Bayesiana, o modelo de Mistura via Processos de Dirichlet vem sendo utilizado, entre outras coisas, para estimar densidades. No próximo capítulo discutiremos dois modelos, utilizados na estimação de densidades, que fazem uso de modelos de MDP.

Capítulo 4

Estimação de Densidade Utilizando MDP

Neste capítulo, revemos o modelo de MDP proposto por Escobar e West (1995) para a estimação de densidades e introduzimos um modelo que é uma extensão do primeiro. Ambos os modelos possuem a mesma estrutura hierárquica apresentada na expressão (3.2). A diferença entre eles está na distribuição paramétrica assumida no primeiro estágio do modelo hierárquico em (3.2). Na primeira proposta, assume-se uma distribuição normal no primeiro estágio do modelo, enquanto que na segunda, assume-se uma distribuição *skew-normal*. Visto que a distribuição *skew-normal* é mais flexível que a distribuição normal, acomodando distribuições com diferentes graus de assimetria e também com caudas mais leves ou pesadas, o modelo proposto para estimação de densidades é também mais flexível que o modelo de Escobar e West (1995). Dessa forma, espera-se que o modelo proposto neste trabalho forneça melhores estimativas para densidades assimétricas que o modelo proposto por Escobar e West (1995).

Qualquer que seja o modelo adotado, a estimação de densidades via modelos de MDP é realizada de acordo com o procedimento geral abaixo:

- 1) Gere uma amostra da distribuição *a posteriori* do vetor de parâmetros $\theta = (\theta_1, \dots, \theta_n)$;
- 2) Com base na amostra gerada, obtenha uma aproximação da distribuição preditiva *a posteriori*. Essa aproximação é uma estimativa da densidade desconhecida.

O primeiro passo desse procedimento pode ser executado através da implementação dos métodos computacionais Bayesianos descritos no Capítulo 3. Enquanto que o segundo passo é realizado utilizando-se a técnica de aproximação de Monte Carlo, que será descrita posteriormente.

4.1 Modelo de Mistura Normal via Processos de Dirichlet (MNPD) – Escobar e West, 1995

Escobar e West (1995) introduzem um método de estimação de densidades utilizando uma abordagem Bayesiana não-paramétrica. Assumem que a distribuição dos dados pode ser bem aproximada por misturas de distribuições normal com diferentes parâmetros e utilizam Processos de Dirichlet para descrever a incerteza sobre a distribuição dos parâmetros que indexam a

família de distribuições normal. Formalmente, suponha que $\theta_i = (\mu_i, V_i)$, $i = 1, \dots, n$, onde $\mu_i \in \mathbb{R}$ e $V_i \in \mathbb{R}^+$. O modelo considerado por Escobar e West (1995) é dado por:

$$\begin{aligned} Y_i | \theta_i &\stackrel{ind}{\sim} N(\theta_i), \quad i = 1, \dots, n; \\ \theta_i | G &\stackrel{iid}{\sim} G; \\ G | \alpha, G_0 &\sim DP(\alpha, G_0). \end{aligned} \quad (4.1)$$

Para completar a especificação do modelo, Escobar e West (1995) consideram que *a priori* a distribuição basal do DP $G_0(\cdot)$ é uma distribuição normal-gama-invertida, que é a família conjugada natural da distribuição normal com média e variância desconhecidas. Dessa forma, sob $G_0(\cdot)$, assumem que

$$\begin{cases} V_i^{-1} \sim Gama(s/2, S/2), \\ (\mu_i | V_i) \sim N(m, \tau V), \end{cases} \quad (4.2)$$

onde $\tau > 0$ é um fator de escala, s e S são números reais positivos e $m \in \mathbb{R}$.

Se $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ é definido como em (4.1), segue de resultados vistos na Seção 2.3.2 que um novo parâmetro θ_{n+1} , dado $\boldsymbol{\theta}$, tem distribuição dada pela seguinte mistura

$$\theta_{n+1} | \boldsymbol{\theta} \sim \frac{\alpha}{\alpha + n} G_0(\theta_{n+1}) + \frac{1}{\alpha + n} \sum_{i=1}^n \delta_{\theta_i}(\theta_{n+1}), \quad (4.3)$$

ou, de forma equivalente,

$$\theta_{n+1} | \boldsymbol{\theta} \sim \frac{\alpha}{\alpha + n} G_0(\theta_{n+1}) + \frac{1}{\alpha + n} \sum_{j=1}^k n_j \delta_{\theta_j^*}(\theta_{n+1}), \quad (4.4)$$

onde $\boldsymbol{\theta}^* = (\theta_1^*, \dots, \theta_k^*)$ é o vetor formado pelos θ_i 's distintos, sendo $k \leq n$ o número de elementos distintos do vetor $\boldsymbol{\theta}$; e n_j é o número de ocorrências do valor θ_j^* , $j = 1, \dots, k$, ou seja, é o número de elementos pertencentes ao j -ésimo *cluster*.

Tem-se ainda que a distribuição de $\theta_i | \boldsymbol{\theta}_{(-i)}$ é dada por:

$$\theta_i | \boldsymbol{\theta}_{(-i)} \sim \frac{\alpha}{\alpha + n - 1} G_0(\theta_i) + \frac{1}{\alpha + n - 1} \sum_{j=1, j \neq i}^n \delta_{\theta_j}(\theta_i),$$

onde $\boldsymbol{\theta}_{(-i)}$ é igual ao vetor $\boldsymbol{\theta}$ sem a coordenada i , ou seja, $\boldsymbol{\theta}_{(-i)} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n)$, $i = 1, \dots, n$.

Escobar e West (1995) utilizam o método iterativo proposto por Escobar (1998, 1994) a fim de se obter uma aproximação para a distribuição *a posteriori* $f(\boldsymbol{\theta}|\mathbf{y})$. Esse método foi descrito na Seção 3.1. Uma vez que o algoritmo de Escobar e West (1995) se baseia no método de amostragem de Gibbs, faz-se necessário encontrar a distribuição condicional completa $(\theta_i|\boldsymbol{\theta}_{(-i)}, \mathbf{y})$, $i = 1, \dots, n$. Da equação (3.3) tem-se que, para cada i ,

$$(\theta_i|\boldsymbol{\theta}_{(-i)}, \mathbf{y}) \sim q_0 G_i(\theta_i) + \sum_{j=1, j \neq i}^n q_j \delta_{\theta_j}(\theta_i), \quad (4.5)$$

onde $G_i(\theta_i)$ denota a distribuição *a posteriori* de θ_i quando $Y_i|\theta_i \sim N(\mu_i, V_i)$ e (μ_i, V_i) segue a distribuição normal-gama-invertida em (4.2). A distribuição $G_i(\theta_i)$ é, portanto, uma normal-gama-invertida tal que $V_i^{-1} \sim G((1+s)/2, S_i/2)$ e $(\mu_i|V_i) \sim N(m_i, \tau(\tau+1)^{-1}V_i)$, onde $S_i = S + (y_i - m)^2/(\tau+1)$ e $m_i = (\tau y_i + m)/(\tau+1)$.

Os pesos q_0 e q_j na equação (4.5) são dados, respectivamente, por

$$q_0 \propto \alpha c(s) [1 + (y_i - m)^2/(sM)]^{-\frac{s+1}{2}} M^{-\frac{1}{2}},$$

$$q_j \propto \frac{1}{(2V_j)^{\frac{1}{2}}} \exp\left\{-\frac{(y_i - \mu_j)^2}{2V_j}\right\}, \quad j = 1, \dots, n; \quad j \neq i,$$

sujeitos ao fato de que $q_1 + \dots + q_{i-1} + q_{i+1} + \dots + q_n = 1$, onde $M = (\tau+1)S/s$ e $c(s) = \Gamma((s+1)/2) \Gamma(s/2)^{-1} s^{-1/2}$. Assim, o peso q_0 é proporcional a α vezes a densidade de uma distribuição t-Student com s graus de liberdade, parâmetro de locação m e parâmetro de escala $M^{1/2}$, denotada $T_s(m, M)$, avaliada no ponto y_i . Enquanto que o peso q_j é proporcional a densidade da distribuição normal $N(\mu_j, V_j)$ avaliada no ponto y_i . A demonstração do resultado em (4.5) pode ser encontrada no Apêndice A.

Uma amostra da distribuição $f(\boldsymbol{\theta}|\mathbf{y})$ é obtida gerando-se sucessivos valores de $\boldsymbol{\theta}$ a partir das condicionais completas em (4.5). De posse dessa amostra, o próximo passo do procedimento geral de estimação de densidade é obter uma estimativa da distribuição preditiva *a posteriori*, que é dada pela seguinte expressão

$$\begin{aligned} P(y_{n+1}|\mathbf{y}) &= \int P(y_{n+1}|\boldsymbol{\theta}) P(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \\ &= \iint P(y_{n+1}|\theta_{n+1}) P(\theta_{n+1}|\boldsymbol{\theta}) d\theta_{n+1} P(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}. \end{aligned} \quad (4.6)$$

Escobar e West (1995) obtêm uma estimativa da distribuição preditiva (4.6) utilizando a técnica de aproximação de Monte Carlo. Eles utilizam o seguinte procedimento para realizar essa aproximação

Procedimento 1:

Passo 1: Particione o contradomínio de Y_{n+1} em L valores, $y_{n+1}^{(1)}, \dots, y_{n+1}^{(L)}$;

Passo 2: Para $i = 1, \dots, L$ e $r = 1, \dots, N$, calcule $P(y_{n+1}^{(i)}|\boldsymbol{\theta}^{(r)})$;

Passo 3: Para $i = 1, \dots, L$, obtenha

$$P(y_{n+1}^{(i)}|\mathbf{y}) \approx N^{-1} \sum_{r=1}^N P(y_{n+1}^{(i)}|\boldsymbol{\theta}^{(r)}), \quad (4.7)$$

onde N é o tamanho da amostra da distribuição *a posteriori* de $\boldsymbol{\theta}$ após o descarte do período de *burn-in*.

Uma vez que estamos considerando que $Y_{n+1}|\boldsymbol{\theta}_{n+1} \sim N(\mu_{n+1}, V_{n+1})$ e a partir de (4.3), segue que a distribuição $P(y_{n+1}^{(i)}|\boldsymbol{\theta}^{(r)})$ em (4.7) é dada pela seguinte mistura

$$Y_{n+1}|\boldsymbol{\theta} \sim \frac{\alpha}{\alpha+n} T_s(m, M) + \frac{1}{\alpha+n} \sum_{i=1}^n N(\mu_i, V_i), \quad (4.8)$$

ou, de forma equivalente

$$Y_{n+1}|\boldsymbol{\theta} \sim \frac{\alpha}{\alpha+n} T_s(m, M) + \frac{1}{\alpha+n} \sum_{j=1}^k n_j N(\mu_j^*, V_j^*). \quad (4.9)$$

O resultado em (4.8) decorre de alguns resultados de cálculo de probabilidades e sua demonstração encontra-se no Apêndice A.

Com o objetivo de comparar os algoritmos de Escobar e West (1995) e de MacEachern e Müller (1998), implementou-se neste trabalho o algoritmo de MacEachern e Müller (1998)¹ considerando o mesmo modelo especificado por Escobar e West (1995). Pretende-se com essa comparação avaliar qual o algoritmo que fornece estimativas de densidade mais próximas da densidade teórica.

A fim de implementar o algoritmo de MacEachern e Müller (1998) é necessário determinar a distribuição condicional completa *a posteriori* de $\theta_j^* = (\mu_j^*, V_j^*)$, $j = 1, \dots, k$, onde k é o número de *clusters*. Uma vez que estamos considerando o modelo em (4.1), tem-se que a distribuição condicional completa *a posteriori* de θ_j^* é a seguinte distribuição normal-gama-invertida:

¹Para maiores detalhes a respeito do algoritmo de MacEachern e Müller (1998) veja a Seção 3.2.

$$\begin{cases} V_j^{*-1} | \mathbf{y}_{(j)} \sim G \left(\frac{n_j + s}{2}; \frac{1}{2} \left[S + \sum_{i=1}^{n_j} (y_i - \bar{y}_{(j)})^2 + \frac{an_j}{2(a+n_j)} (\bar{y}_{(j)} - m)^2 \right] \right), \\ (\mu_j^* | V_j^*, \mathbf{y}_{(j)}) \sim N \left(\frac{am + n_j \bar{y}_{(j)}}{a + n_j}; \frac{V_j^*}{a + n_j} \right), \end{cases} \quad (4.10)$$

onde n_j é o número de observações pertencentes ao j -ésimo *cluster*, $\mathbf{y}_{(j)} = (y_1, \dots, y_{n_j})$ é o vetor de observações pertencentes ao j -ésimo *cluster*, $\bar{y}_{(j)}$ é a média de $\mathbf{y}_{(j)}$ e $a = 1/\tau$. Veja a demonstração do resultado em (4.10) no Apêndice A.

A implementação desse algoritmo fornece uma amostra da distribuição *a posteriori* do vetor de parâmetros $\boldsymbol{\theta}$. De posse dessa amostra, uma estimativa da distribuição preditiva (4.6) é obtida utilizando-se a técnica de aproximação de Monte Carlo. MacEachern e Müller (1998) propõem o seguinte procedimento para realizar tal aproximação

Procedimento 2:

Passo 1: Para $r = 1, \dots, N$, gere um novo parâmetro $\theta_{n+1}^{(r)}$, dado $\boldsymbol{\theta}^{(r)}$, a partir da seguinte distribuição

$$\theta_{n+1}^{(r)} | \boldsymbol{\theta}^{(r)} \sim \left[\frac{\alpha}{\alpha + n} g_0(\theta_{n+1}^{(r)}) + \frac{1}{\alpha + n} \sum_{j=1}^k n_j \delta_{\theta_j^{*(r)}}(\theta_{n+1}^{(r)}) \right];$$

Passo 2: Particione o contradomínio de Y_{n+1} em L valores, $y_{n+1}^{(1)}, \dots, y_{n+1}^{(L)}$;

Passo 3: Para $i = 1, \dots, L$ e $r = 1, \dots, N$, calcule $P(y_{n+1}^{(i)} | \theta_{n+1}^{(r)})$;

Passo 4: Para $i = 1, \dots, L$, obtenha

$$P(y_{n+1}^{(i)} | \mathbf{y}) \approx N^{-1} \sum_{r=1}^N P(y_{n+1}^{(i)} | \theta_{n+1}^{(r)}). \quad (4.11)$$

4.2 Modelo de Mistura Skew-normal via Processos de Dirichlet (MSNPD)

Nesta seção, apresentamos uma extensão do trabalho de Escobar e West (1995), que considera distribuições *skew-normal* no primeiro estágio do modelo de misturas via DP. Alguns resultados importantes referentes à inferência na família *skew-normal* de locação e escala são obtidos como consequência. Esses resultados são utilizados na estimação de densidades via modelos de MDP.

4.2.1 Mistura de Distribuições Skew-normal

Seja $Y_i|\mu_i, V_i, \lambda_i \stackrel{ind}{\sim} SN(\mu_i, V_i, \lambda_i)$, $i = 1, \dots, n$. Assuma que, *a priori*, o parâmetro λ_i é independente do vetor de parâmetros (μ_i, V_i) e que

$$\begin{cases} \lambda_i \sim N(\varepsilon, z^2), \\ V_i^{-1} \sim Gama(s/2, S/2), \\ \mu_i|V_i \sim N(m, \tau V_i), \end{cases} \quad (4.12)$$

onde $\tau > 0$ é um fator de escala, s , S e z são números reais positivos e $m, \varepsilon \in \mathbb{R}$.

A seguinte proposição apresenta as condicionais completas dos parâmetros da distribuição *skew-normal*.

Proposição 1: Considere $\theta_i = (\mu_i, V_i, \lambda_i)$, $i = 1, \dots, n$. Seja $\boldsymbol{\theta}^* = (\theta_1^*, \dots, \theta_k^*)$, $k \leq n$, o vetor formado pelos θ_i 's distintos. Se $Y_i|\theta_i \stackrel{ind}{\sim} SN(\theta_i)$ e a distribuição *a priori* de θ_i é definida como em (4.12), então

(i) a distribuição de μ_j^* , dados V_j^* , λ_j^* e $\mathbf{y}_{(j)}$, é:

$$P(\mu_j^*|V_j^*, \lambda_j^*, \mathbf{y}_{(j)}) \propto \phi(\mu_j^*; M^*, V^*) \Phi_{n_j}(\mu_j^* \boldsymbol{\lambda}^* + \mathbf{y}^*), \quad j = 1, \dots, k, \quad (4.13)$$

(ii) a distribuição de V_j^* , dados μ_j^* , λ_j^* e $\mathbf{y}_{(j)}$, é:

$$\begin{aligned} P(V_j^*|\mu_j^*, \lambda_j^*, \mathbf{y}_{(j)}) &\propto GI\left(\frac{n_j + s + 1}{2}; \frac{1}{2} \left[S + a(\mu_j^* - m)^2 + \sum_{i=1}^{n_j} (y_i - \bar{y}_{(j)})^2 + n_j(\mu_j^* - \bar{y}_{(j)})^2 \right]\right) \\ &\times \Phi_{n_j}(\lambda_j^* \mathbf{Z}), \quad j = 1, \dots, k, \end{aligned} \quad (4.14)$$

(iii) a distribuição de λ_j^* , dados μ_j^* , V_j^* e $\mathbf{y}_{(j)}$, é:

$$P(\lambda_j^*|\mu_j^*, V_j^*, \mathbf{y}_{(j)}) \propto \phi(\lambda_j^*; \varepsilon, z^2) \Phi_{n_j}(\lambda_j^* \mathbf{Z}), \quad j = 1, \dots, k, \quad (4.15)$$

onde n_j é o número de observações pertencentes ao j -ésimo *cluster*, $\mathbf{y}_{(j)} = (y_1, \dots, y_{n_j})$ é o vetor de observações pertencentes ao j -ésimo *cluster*, $\bar{y}_{(j)}$ é a média de $\mathbf{y}_{(j)}$, $M^* = (am + n_j \bar{y}_{(j)})/(a + n_j)$, $V^* = (V_j^*)/(a + n_j)$, $\boldsymbol{\lambda}^* = -(\lambda_j^* \mathbf{1}_{n_j})/(\mathbf{V}_j^*)^{\frac{1}{2}}$, $\mathbf{y}^* = (\lambda_j^* \mathbf{y}_{(j)})/(V_j^*)^{\frac{1}{2}}$ e $\mathbf{Z} = (\mathbf{y}_{(j)} - \mu_j^* \mathbf{1}_{n_j})/(\mathbf{V}_j^*)^{\frac{1}{2}}$. As demonstrações dos resultados em (4.13), (4.14) e (4.15) podem ser obtidas Apêndice A.

As condicionais completas de μ_j^* e λ_j^* , representadas em (4.13) e (4.15), pertencem a família de distribuições SUN (Unified *skew-normal*) definida na equação (10) de Arellano-Valle e Azzalini (2006).

Foram obtidas também as representações estocásticas das variáveis aleatórias $\mu_j^*|V_j^*, \lambda_j^*, \mathbf{y}_{(j)}$ e $\lambda_j^*|\mu_j^*, V_j^*, \mathbf{y}_{(j)}$. As representações estocásticas são importantes pois fornecem um modo prático de se obter os momentos de uma variável aleatória e porque, em geral, é computacionalmente mais fácil gerar amostras a partir delas. Essas representações estocásticas são apresentadas nas Proposições 2 e 3. As demonstrações das mesmas podem ser encontradas no Apêndice B.

Proposição 2: Seja $W \in \mathbb{R}$ e $U^0 \in \mathbb{R}$ duas variáveis aleatórias tais que, dados $\mathbf{y}_{(j)}$, V_j^* e λ_j^* , são independentes. Considere que $W \sim N(M^*, V^*/(1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*))$ e $U^0 \sim LTN_{n_j}(-M^* \boldsymbol{\lambda}^* - \mathbf{y}^*; \mathbf{0}, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T})$, onde $LTN(\mathbf{L}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denota uma distribuição normal truncada multivariada truncada para valores superiores a \mathbf{L} , com parametro de locação igual a $\boldsymbol{\mu}$ e parâmetro de escala igual a $\boldsymbol{\Sigma}$. Então, a variável aleatória $\mu_j^*|V_j^*, \lambda_j^*, \mathbf{y}_{(j)} = \mu_y$, cuja distribuição está representada em (4.13), é tal que

$$\mu_y \stackrel{d}{=} W + \frac{V^* \boldsymbol{\lambda}^{*T} U^0}{1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*}. \quad (4.16)$$

Proposição 3: Seja $W \in \mathbb{R}$ e $U^0 \in \mathbb{R}$ duas variáveis aleatórias tais que, dados $\mathbf{y}_{(j)}$, μ_j^* e V_j^* , são independentes. Considere que $W \sim N(\varepsilon, z^2/(1 + z^2 \mathbf{Z}^{*T} \mathbf{Z}^*))$ e $U^0 \sim LTN_{n_j}(-\varepsilon \mathbf{Z}^*; \mathbf{0}, \mathbf{I}_{n_j} + z^2 \mathbf{Z}^* \mathbf{Z}^{*T})$. Então, a variável aleatória $\lambda_j^*|\mu_j^*, V_j^*, \mathbf{y}_{(j)} = \lambda_y$, cuja distribuição está representada em (4.15), é tal que

$$\lambda_y \stackrel{d}{=} W + \frac{z^2 \mathbf{Z}^T U^0}{1 + z^2 \mathbf{Z}^T \mathbf{Z}}. \quad (4.17)$$

Os resultados apresentados nas Proposições 1, 2 e 3 estendem alguns resultados obtidos por Castro *et al.* (2009). Em seu trabalho, Castro *et al.* considera uma distribuição *a priori* imprópria para o vetor de parâmetros (μ_i, V_i) . O mesmo não pode ser aplicado neste trabalho, visto que há a necessidade de se amostrar valores desse vetor de parâmetros.

4.2.2 Estimação de Densidades Utilizando Mistura Skew-normal via Processos de Dirichlet

Neste trabalho, consideramos uma extensão do modelo MNDP, proposto por Escobar e West (1995), para estimação de densidades. Nossa proposta consiste em estimar densidades utilizando misturas de densidades *skew-normal* definidas em (2.1), cuja medida misturadora é modelada segundo um Processo de Dirichlet. Dessa forma, seja $\theta_i = (\mu_i, V_i, \lambda_i)$, $i = 1, \dots, n$, estamos considerando um modelo de MDP com a seguinte estrutura hierárquica

$$\begin{aligned}
Y_i|\theta_i &\stackrel{ind}{\sim} SN(\theta_i), i = 1, \dots, n, \\
\theta_i|G &\stackrel{iid}{\sim} G, \\
G|\alpha, G_0 &\sim DP(\alpha, G_0).
\end{aligned}
\tag{4.18}$$

Para a construção da distribuição basal do DP, supomos, *a priori*, que o parâmetro λ_i é independente do vetor de parâmetros (μ_i, V_i) . Também consideramos como distribuições *a priori* para os parâmetros da distribuição *skew-normal* as distribuições definidas em (4.12). Dessa forma, estamos assumindo que a distribuição basal G_0 é o produto entre uma distribuição normal-gama-invertida e uma distribuição normal.

A família de distribuições *skew-normal* compreende a família Gaussiana, tratando-se assim de uma de uma família mais ampla. Além disso, as distribuições *half-normal* positivas e negativas são casos limites de distribuições *skew-normal*. Esses fatos tornam o modelo MSNDP mais flexível do que o modelo MNDP para estimar densidades. Assim, esperamos que o modelo MSNDP seja mais adequado que o modelo MNDP para estimar densidades assimétricas e/ou estritamente positivas ou negativas.

Como demonstrado em Arellano-Valle *et al.* (2009) e Arellano-Valle *et al.* (2011), as distribuições *a priori* e *a posteriori* dos parâmetros da distribuição *skew-normal* pertencerem a mesma classe de distribuições. Isso indica a existência de conjugação na família de distribuições *skew-normal*. Apesar disso, o cálculo das distribuições *a posteriori* é bastante intenso do ponto de vista computacional. Essa dificuldade se deve ao fato de que essas distribuições *a posteriori* envolvem, em suas expressões, distribuições acumuladas da distribuição normal e t-Student. Devido à complexidade dos cálculos e pelo fato de o algoritmo de Escobar e West (1995) ter se mostrado menos eficiente na estimação de densidades via modelos MNDP², optamos por implementar apenas o algoritmo de MacEachern e Müller (1998).

As condicionais completas *a posteriori* de μ_j^* , V_j^* e λ_j^* , $j = 1, \dots, k$, necessárias para a implementação do algoritmo de MacEachern e Müller (1998), são aquelas estabelecidas na Proposição 1. Isso é consequência de estarmos considerando o modelo em (4.18). Note que a condicional completa de V_j^* não tem forma conhecida. Já as condicionais completas de μ_j^* e λ_j^* são distribuições *skew-normal*. Também vale salientar que a expressão de todas as condicionais completas envolve a distribuição acumulada da normal, o que dificulta o processo de se gerar amostras a partir delas.

Amostras das distribuições condicionais completas em (4.13), (4.14) e (4.15) foram obtidas através da implementação do algoritmo de Metropolis-Hastings. Consideramos como distribuições propostas a distribuição normal representada em (4.13), a distribuição gama-invertida representada em (4.14) e a distribuição normal representada em (4.15), respectivamente.

²Esse resultado foi obtido nas análises realizadas no Capítulo 5, Seção 5.1.

Após gerarmos uma amostra da distribuição *a posteriori* do vetor de parâmetros $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$, uma estimativa da distribuição preditiva *a posteriori* foi obtida a partir da técnica de aproximação de Monte Carlo descrita no Procedimento 2, o qual foi apresentado na Seção 4.1.

Capítulo 5

Análise de Dados Simulados

Neste capítulo, ilustramos a aplicação da metodologia apresentada considerando bancos de dados simulados. O primeiro objetivo dessa análise é comparar a eficiência dos algoritmos de Escobar e West (1995) e de MacEachern e Müller (1998) na estimação de densidades via modelos de MNDP. Também pretende-se averiguar o quão eficiente é o modelo proposto por Escobar e West (1995) ao estimar densidades que não envolvam misturas de distribuições normais. O segundo e mais importante objetivo é avaliar se o modelo MSNDP é mais adequado que o modelo MNDP para estimar densidades, principalmente no que diz respeito à densidades assimétricas e/ou estritamente positivas.

Neste estudo, assumimos, para ambos os modelos, os seguintes valores para os hiperparâmetros do modelo de MDP ($\alpha = 1$, $\tau = 1.5$, $s = 6$ e $S = 16$). Definimos o valor do hiperparâmetro m como sendo a média aritmética dos dados gerados, ou seja, $m = \bar{y}$. Para o modelo MSNDP, também assumimos $\varepsilon = 0$ e $z^2 = 100$. Todas as análises foram realizadas através do uso do *software* R 2.9.0, que está disponível em <http://cran.r-project.org/>.

No esquema MCMC, consideramos uma única cadeia de tamanho 200.000. Após a convergência ter sido obtida, descartou-se um *burn-in* de tamanho 150.000. Apesar das cadeias geradas terem apresentado forte autocorrelação, considerou-se um *lag* igual a 1. Resultando, assim, em uma amostra da distribuição *a posteriori* de tamanho 50.000. A escolha desse *lag* justifica-se pelo fato de estarmos estimando apenas uma medida invariante. Assim, não se faz necessário que a amostra seja não-correlacionada. Essa escolha também foi motivada pelo longo tempo computacional gasto para gerar cada cadeia ($\pm 1,5$ dias) e pela limitação do *software* em armazenar cadeias muito longas. A matriz a ser armazenada é de ordem $n \times R$, onde n é o tamanho amostral e R o tamanho da cadeia gerada.

Visto que, para cada densidade a ser estimada, necessitamos avaliar a convergência de n parâmetros e dado o grande número de análises realizadas, consideramos que a utilização de métodos gráficos para a avaliação da convergência seria muito dispendiosa. Dessa forma, optamos por avaliar a convergência através da realização do teste de Geweke (1992). Em certos casos e para alguns

parâmetros, obtivemos apenas convergência em probabilidade.

A fim de avaliar a qualidade das estimativas obtidas, calculamos o erro quadrático médio integrado (EQMI). O EQMI é definido com segue

$$EQMI = L^{-1} \sum_{i=1}^L [\hat{f}(y_i) - f(y_i)]^2, \quad (5.1)$$

onde $\hat{f}(y_i) = P(y_{n+1}^{(i)}|\mathbf{y})$ é a densidade estimada e $f(y_i)$ é a densidade teórica, ambas avaliadas no ponto y_i . Note que, quanto menor o EQMI, melhor é a estimativa de densidade obtida. Consequentemente, segundo este critério, o algoritmo mais adequado é aquele que apresentar estimativa com menor EQMI.

5.1 Comparação dos Algoritmos de Escobar e West (1995) e MacEachern e Müller (1998)

Nesta seção, comparamos a qualidade das estimativas de densidades obtidas a partir da implementação dos algoritmos propostos por Escobar e West (1995) e MacEachern e Müller (1998). Ambos os algoritmos foram implementados considerando-se as especificações do modelo MNPD, proposto por Escobar e West (1995). De agora em diante denotaremos o modelo MNPD via algoritmo de Escobar e West (1995) por MNDP-EW. Já o modelo MNDP via algoritmo de MacEachern e Müller (1998) será denotado por MNDP-MM.

Foram considerados diferentes cenários, com o intuito de verificar como diferentes situações podem afetar a qualidade das estimativas de densidade. A Tabela 5.1 exhibe todos os cenários analisados. Nos cenários de 1 a 6, os dados são gerados de misturas discretas de distribuições normais. Nesses casos, espera-se que as densidades sejam bem estimadas pelo modelo MNDP. Nos cenários 7 e 8, os dados são gerados de misturas de distribuições t-Student, que possuem caudas mais pesadas que a distribuição normal. Os demais cenários consideram misturas de distribuições assimétricas, casos onde não se espera bom desempenho do modelo MNDP. No Cenário 6 assumimos uma distribuição *a priori* não informativa para o parâmetro de escala da distribuição normal, $V_i^{-1} \sim Gama(0.1, 0.1)$. Nesse cenário, pretendemos avaliar como a falta de informação *a priori* a respeito do parâmetro V_i pode afetar as estimativas de densidade. Para cada cenário foi gerada uma amostra de tamanho $n = 100$ a partir da densidade em questão. Optamos por esse tamanho amostral por se tratar de um tamanho razoável, passível de ser obtido em estudos envolvendo dados reais, e pelo menor tempo computacional demandado para se obter as estimativas de interesse.

Tabela 5.1: Cenários simulados

Cenário	Distribuição da qual os dados foram gerados
1	$Y \sim 0.4N(0, 1) + 0.6N(5, 1)$
2	$Y \sim 0.5N(0, 8) + 0.5N(40, 8)$
3	$Y \sim 0.8N(0, 1) + 0.2N(5, 1)$
4	$Y \sim 0.1N(-2, 1.2) + 0.7N(3, 1) + 0.1N(8, 0.8)$
5	$Y \sim 0.2N(-10, 1.8) + 0.2N(2, 2.2) + 0.2N(15, 1.8) + 0.2N(25, 1.2)$
6	$Y \sim 0.5N(0, 1) + 0.5N(8, 1.5)$
7	$Y \sim T(0, 1.5, 5)$
8	$Y \sim 0.5T(2, 1, 3) + 0.5T(8, 1.3, 10)$
9	$Y \sim 0.4G(2, 0.5) + 0.6G(4, 1.5)$
10	$Y \sim 0.4Exp(0.1) + 0.6Exp(0.5)$
11	$Y \sim 0.5SN(-1, 1, 5) + 0.5SN(6, 2, -20)$
12	$Y \sim SN(0, 2, -50)$

As Figuras 5.1, 5.2 e 5.3 exibem os gráficos referentes aos cenários analisados. Cada gráfico apresenta o histograma da amostra simulada e as curvas das densidades estimadas pelos algoritmos de Escobar e West (1995) e MacEachern e Müller (1998). A Tabela 5.2 apresenta, para cada um dos cenários simulados, os Erros Quadráticos Médios Integrados referentes às estimativas de densidade obtidas pela implementação dos algoritmos de Escobar e West (1995) e MacEachern e Müller (1998). Esses erros são denotados respectivamente por $EQMI_1$ e $EQMI_2$. A Tabela 5.2 apresenta ainda a diferença entre o $EQMI_1$ e o $EQMI_2$, denotada por Dif . Diferenças positivas indicam uma melhor adequação da estimativa obtida via algoritmo de MacEachern e Müller (1998), já uma diferença negativa indica que o algoritmo de Escobar e West (1995) é o mais adequado.

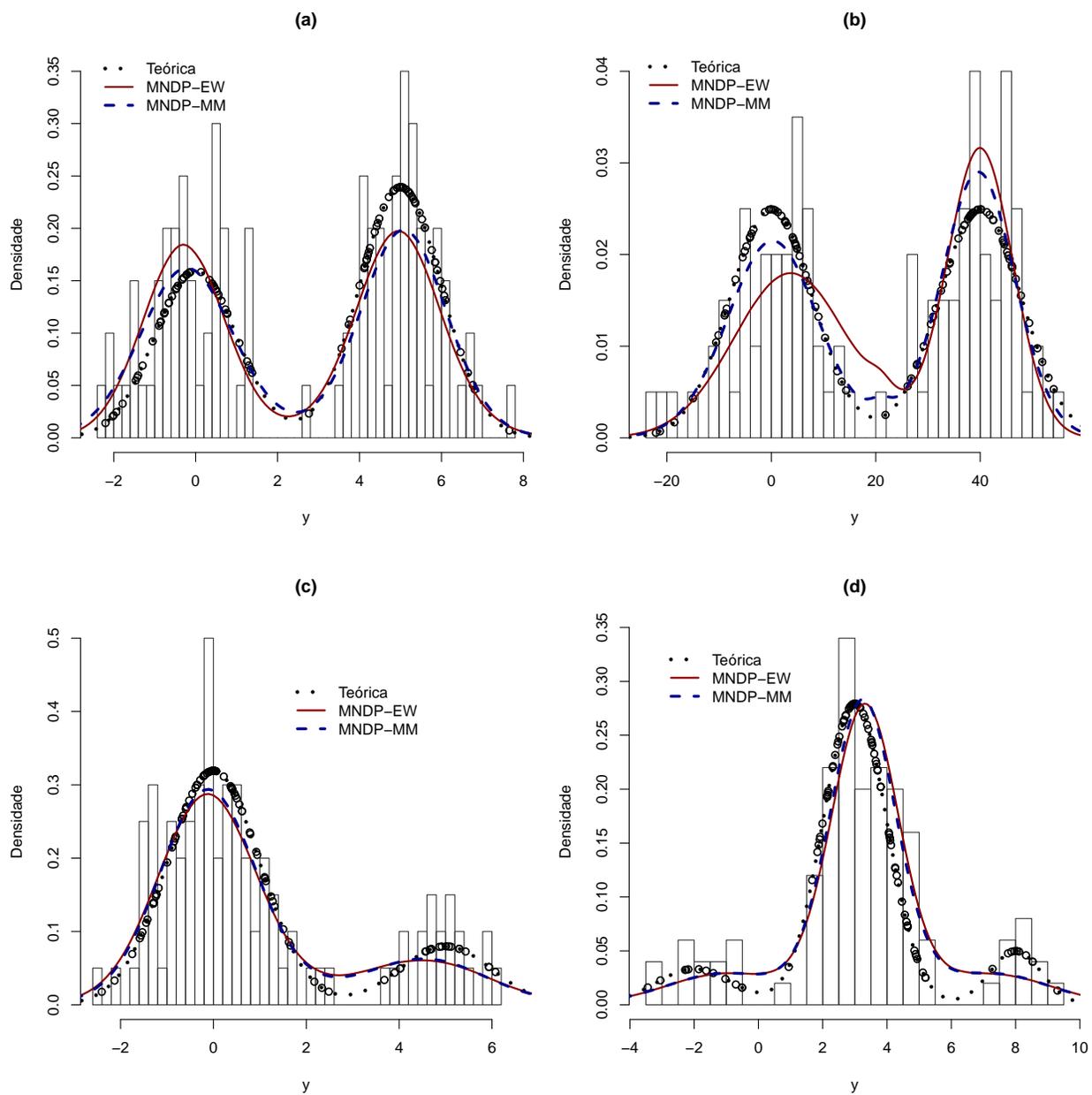


Figura 5.1: Histograma da amostra simulada e densidades estimadas pelos algoritmos de Escobar e West (1995) e MacEachern e Müller (1998). (a) **Cenário 1**; (b) **Cenário 2**; (c) **Cenário 3** e (d) **Cenário 4**.

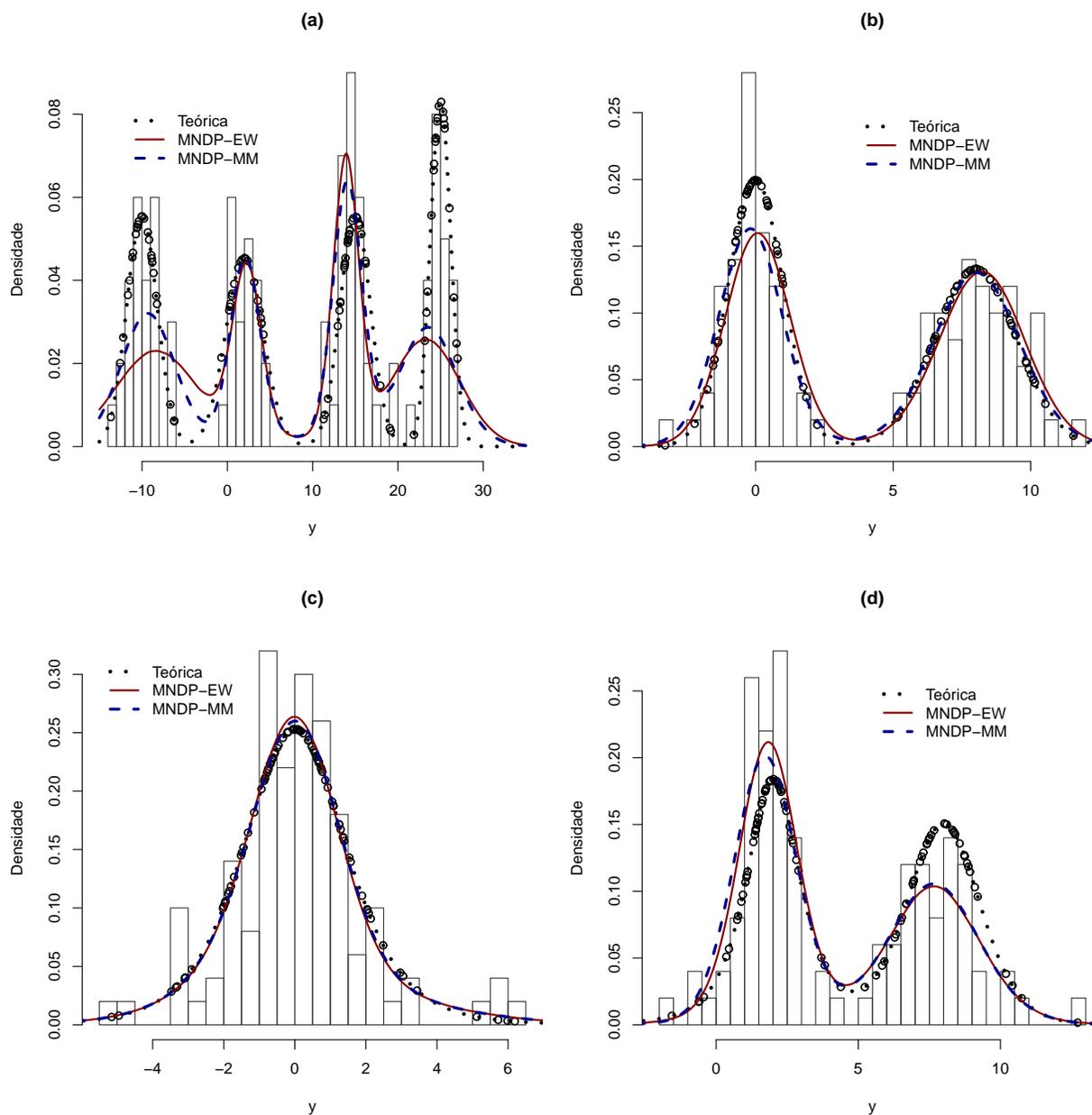


Figura 5.2: Histograma da amostra simulada e densidades estimadas pelos algoritmos de Escobar e West (1995) e MacEachern e Müller (1998). (a) **Cenário 5**; (b) **Cenário 6**; (c) **Cenário 7** e (d) **Cenário 8**.

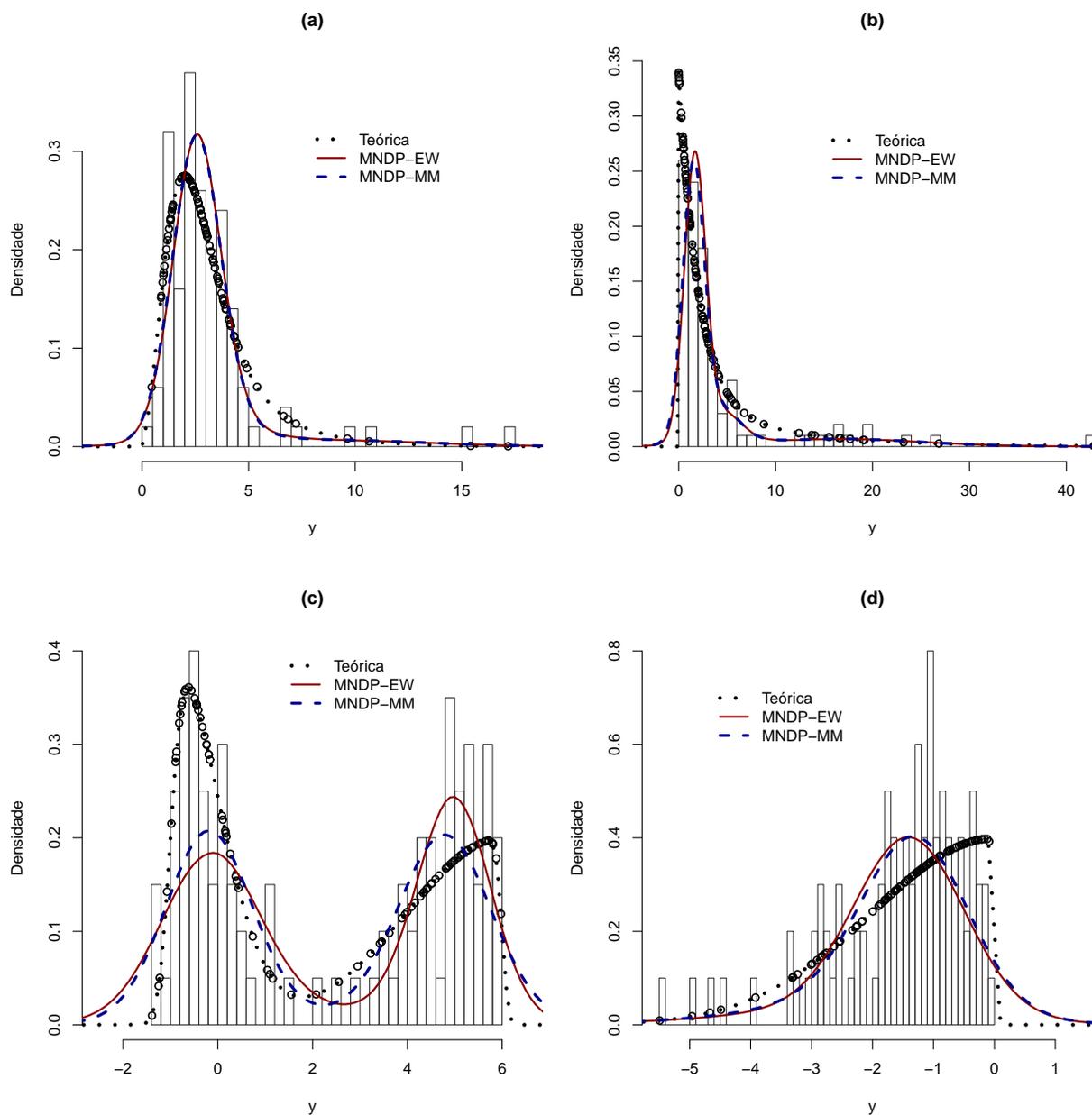


Figura 5.3: Histograma da amostra simulada e densidades estimadas pelos algoritmos de Escobar e West (1995) e MacEachern e Müller (1998). (a) **Cenário 9**; (b) **Cenário 10**; (c) **Cenário 11** e (d) **Cenário 12**.

Tabela 5.2: Erros Quadráticos Médios Integrados das estimativas de densidade obtidas.

	$EQMI_1$	$EQMI_2$	$Dif = EQMI_1 - EQMI_2$
Cenário 1	1.94×10^{-4}	1.45×10^{-4}	0.49×10^{-4}
Cenário 2	1.35×10^{-5}	0.28×10^{-5}	1.07×10^{-5}
Cenário 3	1.37×10^{-4}	1.07×10^{-4}	0.30×10^{-4}
Cenário 4	3.75×10^{-4}	2.88×10^{-4}	0.87×10^{-4}
Cenário 5	2.75×10^{-4}	2.08×10^{-4}	0.67×10^{-4}
Cenário 6	9.05×10^{-5}	8.92×10^{-5}	0.13×10^{-5}
Cenário 7	1.24×10^{-5}	0.70×10^{-5}	0.54×10^{-5}
Cenário 8	2.58×10^{-4}	2.73×10^{-4}	-0.15×10^{-4}
Cenário 9	4.63×10^{-4}	4.46×10^{-4}	0.17×10^{-4}
Cenário 10	8.81×10^{-4}	7.31×10^{-4}	1.50×10^{-4}
Cenário 11	1.21×10^{-3}	0.91×10^{-3}	0.30×10^{-3}
Cenário 12	1.22×10^{-3}	1.06×10^{-3}	0.16×10^{-3}

Analisando o gráfico (a) da Figura 5.1, podemos observar que ambos os algoritmos estimam de forma adequada a densidade teórica, captando tanto a multimodalidade quanto a variabilidade da distribuição. Tanto a análise gráfica quanto a análise da diferença entre o $EQMI_1$ e o $EQMI_2$ mostram que o algoritmo de MacEachern e Müller (1998) fornece uma melhor aproximação para as densidades teóricas. No Cenário 2, onde a dispersão da distribuição é grande, pode-se observar uma melhor qualidade da estimativa obtida pelo algoritmo de MacEachern e Müller (1998). Isso pode ser verificado através do valor positivo do Dif e da análise do gráfico (b) da Figura 5.1. Nos cenários 3 e 4, o intuito é avaliar o efeito de diferentes pesos da mistura na qualidade das estimativas. Esses cenários são representados graficamente na Figura 5.1 (c) e (d), respectivamente. Pela análise do gráfico 5.1 (c) verificamos que ambos os algoritmos produzem estimativas de densidade satisfatórias, conseguindo captar a bimodalidade da distribuição. O mesmo não ocorre no Cenário 4, uma vez que nenhuma das estimativas foi capaz de identificar as distribuições que ocorrem com menor probabilidade na mistura. Os gráficos 5.1 (c) e 5.1 (d) evidenciam uma proximidade entre as estimativas obtidas a partir dos algoritmos de Escobar e West (1995) e de MacEachern e Müller (1998). Apesar disso, a diferença entre os EQMI's indica que o algoritmo de MacEachern e Müller (1998) é o mais adequado nessas situações.

A Figura 5.2 (a) apresenta o gráfico resultante da análise do Cenário 5, onde considera-se um número maior de distribuições na mistura. Tanto o algoritmo de Escobar e West (1995) quanto o de MacEachern e Müller (1998) mostraram-se inadequados nessa situação. As estimativas obtidas conseguem captar a multimodalidade da distribuição, mas não a variabilidade da mesma. No Cenário 6, onde consideramos uma distribuição *a priori* não informativa para o parâmetro de escala,

ambos os algoritmos produziram estimativas de boa qualidade. Esse cenário está representado no gráfico (b) da Figura 5.2. Ainda nessa figura, apresentamos os gráficos referentes aos cenários 7 e 8, onde consideramos misturas de distribuições t-Student. A análise dos gráficos 5.2 (c) e 5.2 (d) revela um bom ajuste das curvas estimadas. A partir da Tabela 5.2 verificamos que o Cenário 8 apresentou um valor negativo para o Dif , indicando uma maior adequação da estimativa de Escobar e West (1995). O algoritmo de MacEachern e Müller (1998) mostrou-se mais adequado em todos os outros cenários representados na Figura 5.2.

Os cenários representados na Figura 5.3 foram analisados como o objetivo de avaliar a qualidade das estimativas de densidade produzidas pelos algoritmos de Escobar e West (1995) e de MacEachern e Müller (1998) no caso em que a densidade teórica é assimétrica. Nos Cenários 9 e 10 consideramos densidades estritamente positivas, já no Cenário 12 consideramos uma distribuição que põe massa apenas em valores não-positivos. A análise dos gráficos (a) e (b) da Figura 5.3 demonstra que apesar de ambos os algoritmos captarem a assimetria da distribuição teórica, eles não a estimam de maneira satisfatória, uma vez que as densidades geradas atribuem massa de probabilidade aos valores negativos de Y . Os gráficos (c) e (d) da Figura 5.3 evidenciam a ineficiência dos algoritmos em captar a assimetria das distribuições *skew-normal*. A partir da análise do gráfico 5.3 (d) é possível observar ainda que as estimativas dos dois algoritmos atribuem massa de probabilidade aos valores positivos de Y , o que afirma a inadequação das estimativas. A partir da Tabela 5.2 observamos que a diferença entre o $EQMI_1$ e o $EQMI_2$ de todos os cenários representados na Figura 5.3 é positiva. Isso demonstra que, nesses casos, o algoritmo de MacEachern e Müller (1998) é o mais adequado para estimar densidades.

Cabe comentar aqui que, apesar de termos obtido estimativas similares para as densidades teóricas na maioria dos cenários analisados, o algoritmo de MacEachern e Müller (1998) mostrou-se mais eficiente na estimação de densidades. Isso se deve ao fato desse algoritmo ter apresentado convergência mais rápida e ter gerado cadeias menos correlacionadas.

5.2 Comparação dos Modelos MSNDP e MNDP

Neste estudo, comparamos os modelos de mistura normal e de mistura *skew-normal* por processos de Dirichlet na estimação de densidades. Implementamos, para esse fim, o algoritmo de MacEachern e Müller (1998).

Como na Seção 5.1, foram considerados diferentes cenários. A Tabela 5.3 exibe todos os cenários considerados. Para cada cenário foi gerada uma amostra de tamanho $n = 100$ a partir da densidade em questão. Os cenários de 1 a 8 correspondem a dados gerados de misturas discretas de distribuições normais. Nesses casos, espera-se que o modelo MNDP seja mais adequado que o modelo MSNDP. Os demais cenários consideram misturas de distribuições assimétricas, casos onde se espera

melhor adequação do modelo MSNDP. Apenas no Cenário 8, assumimos uma distribuição *a priori* não informativa para o parâmetro de escala da distribuição *skew-normal*, $V_i^{-1} \sim Gama(0.1, 0.1)$, pelas mesmas razões especificadas na Seção 5.1.

Tabela 5.3: Cenários simulados

Cenário	Distribuição da qual os dados foram gerados
1	$Y \sim 0.4N(0, 1) + 0.6N(5, 1)$
2	$Y \sim 0.6N(0, 1) + 0.4N(5, 1)$
3	$Y \sim 0.2N(0, 1) + 0.8N(5, 1)$
4	$Y \sim 0.8N(0, 1) + 0.2N(5, 1)$
5	$Y \sim 0.5N(0, 8) + 0.5N(40, 10)$
6	$Y \sim 0.5N(0, 2.5) + 0.5N(10, 2.5)$
7	$Y \sim N(5, 4)$
8	$Y \sim 0.5N(0, 1) + 0.5N(8, 1.5)$, com $s = 0.2$ e $S = 0.2$
9	$Y \sim 0.6G(2, 0.5) + 0.4G(4, 1.5)$
10	$Y \sim 0.8G(2, 0.5) + 0.2G(4, 1.5)$
11	$Y \sim 0.5G(1, 0.1) + 0.5G(5, 2.5)$
12	$Y \sim G(4, 1.5)$
13	$Y \sim Exp(3)$
14	$Y \sim Exp(0.4)$
15	$Y \sim 0.5SN(0, 8, 5) + 0.5SN(40, 10, -20)$
16	$Y \sim 0.5SN(0, 2.5, 10) + 0.5SN(10, 2.5, -10)$
17	$Y \sim 0.4SN(-1, 1, 5) + 0.6SN(6, 2, -20)$
18	$Y \sim 0.6SN(-1, 1, 5) + 0.4SN(6, 2, -20)$
19	$Y \sim 0.8SN(-1, 1, 5) + 0.2SN(6, 2, -20)$
20	$Y \sim 0.2SN(-1, 1, 5) + 0.8SN(6, 2, -20)$
21	$Y \sim SN(0, 2, -100)$
22	$Y \sim SN(20, 4, 100)$
23	$Y \sim SN(0, 4, 2)$

Os gráficos referentes às análises dos cenários de 1 a 23 estão representados nas Figuras 5.4 a 5.9. A Tabela 5.4 apresenta, para cada um dos cenários simulados, os erros quadráticos médios integrados referentes às estimativas de densidade obtidas a partir dos modelos MNDP e MSNDP. Esses erros são denotados respectivamente por $EQMI_1$ e $EQMI_2$. Essa tabela apresenta ainda o valor da diferença $Dif = EQMI_1 - EQMI_2$. Diferenças positivas indicam que o modelo MNDP é o mais adequado, já uma diferença negativa indica que o modelo MSNDP é o mais adequado.

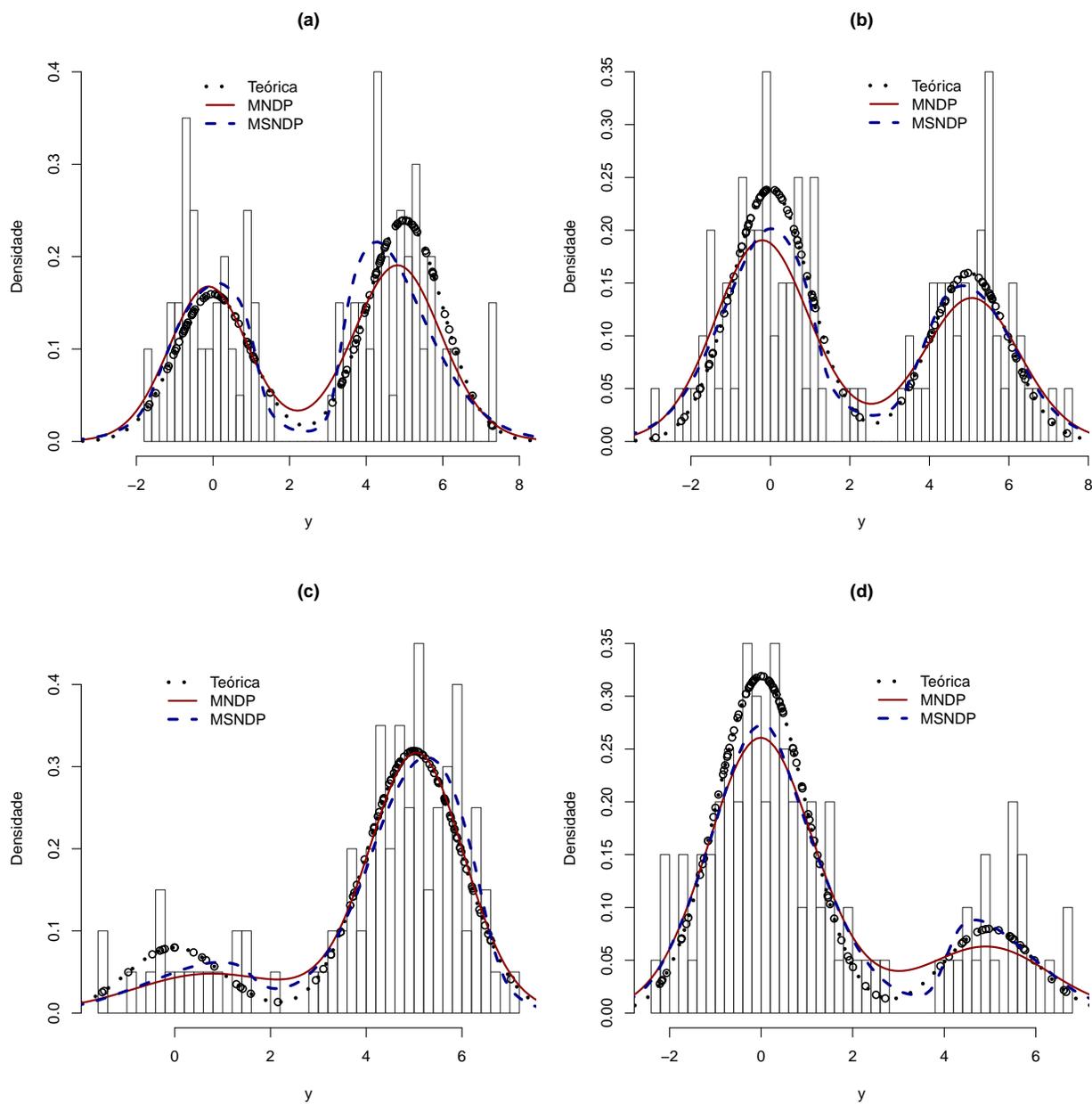


Figura 5.4: Histograma da amostra observada e densidades estimadas pelos modelos MNDP e MSNDP. (a) **Cenário 1**; (b) **Cenário 2**; (c) **Cenário 3** e (d) **Cenário 4**.

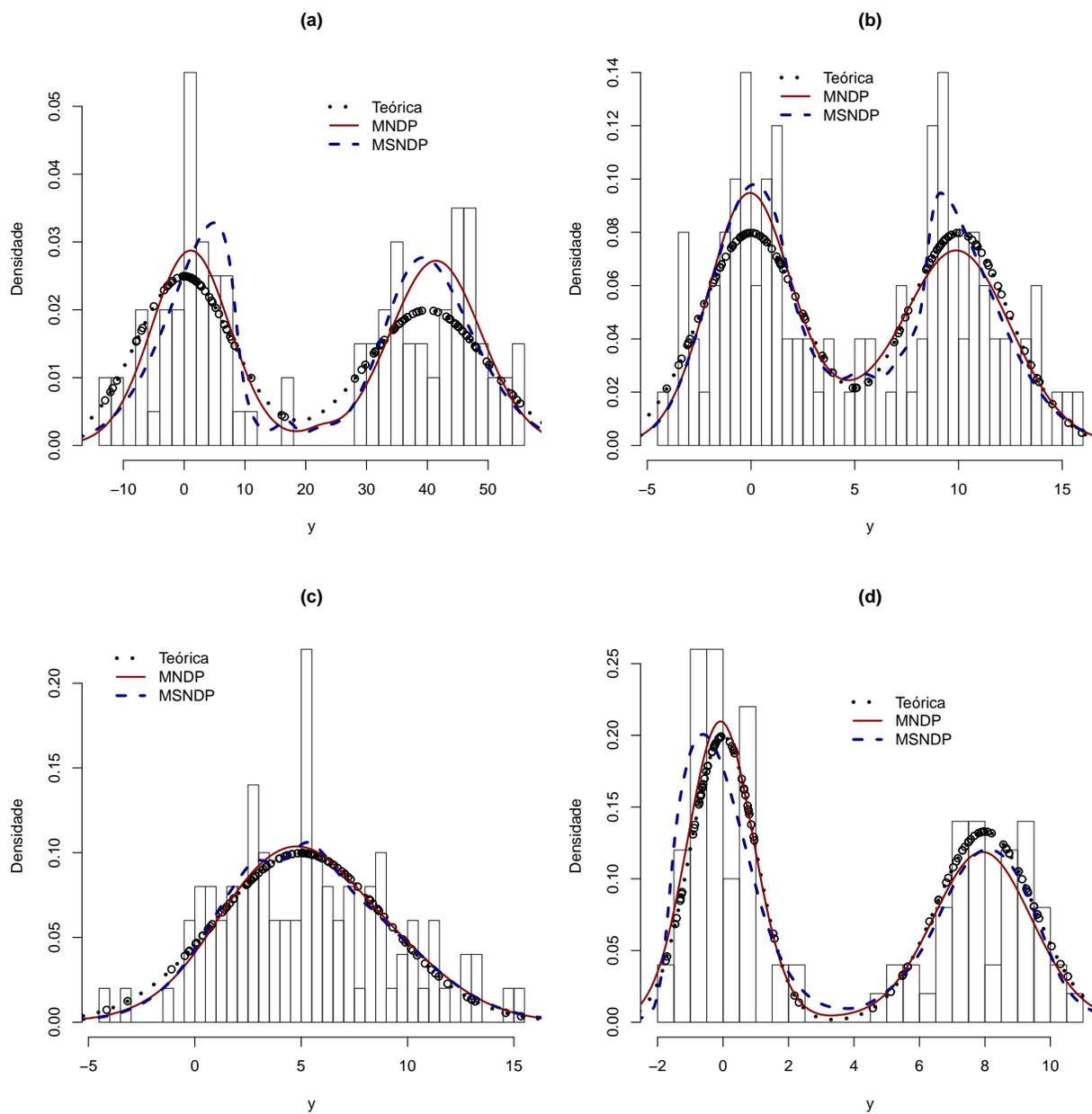


Figura 5.5: Histograma da amostra observada e densidades estimadas pelos modelos MNDP e MSNDP. (a) **Cenário 5**; (b) **Cenário 6**; (c) **Cenário 7** e (d) **Cenário 8**.

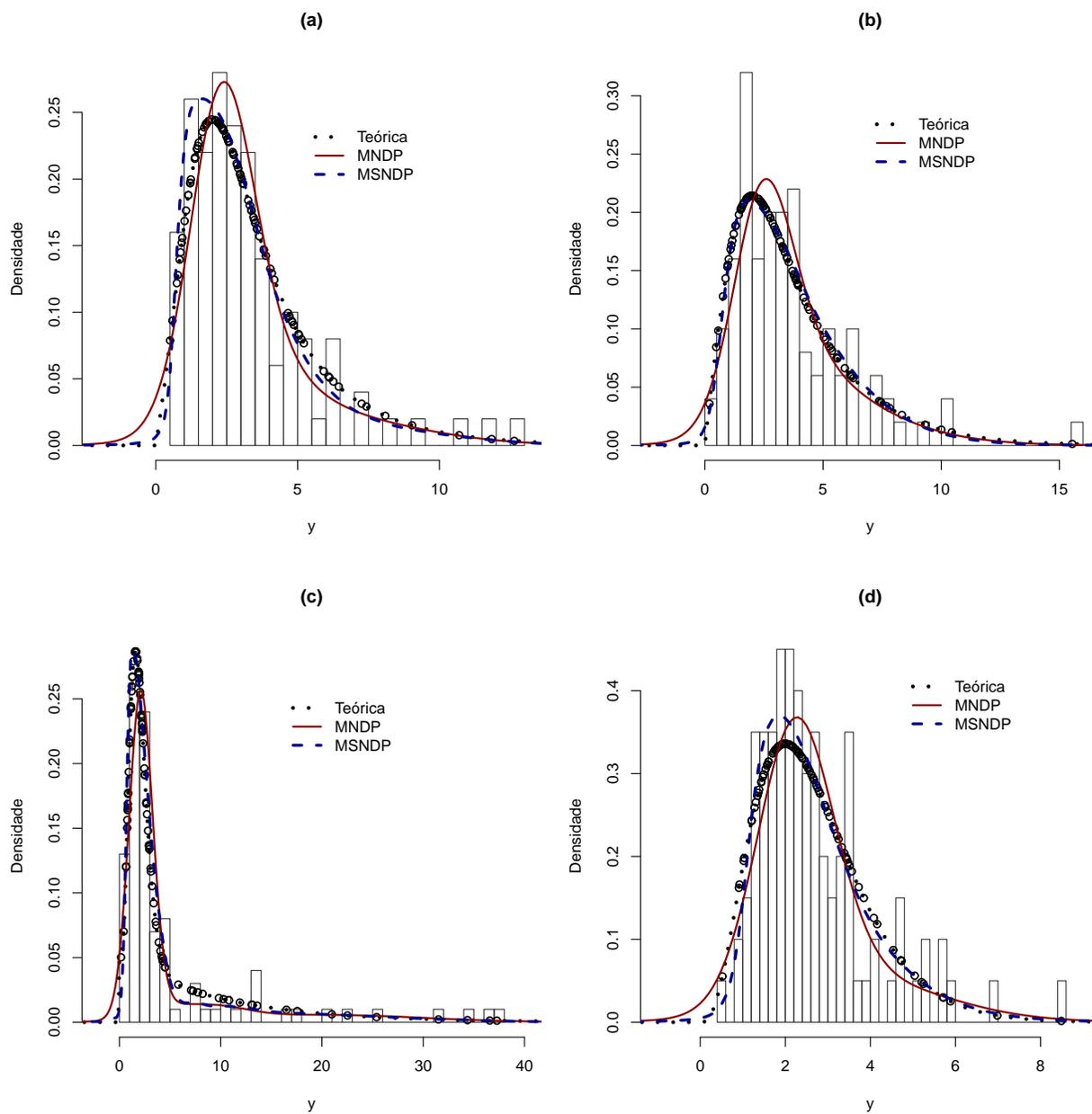


Figura 5.6: Histograma da amostra observada e densidades estimadas pelos modelos MNDP e MSNDP. (a) **Cenário 9**; (b) **Cenário 10**; (c) **Cenário 11** e (d) **Cenário 12**.

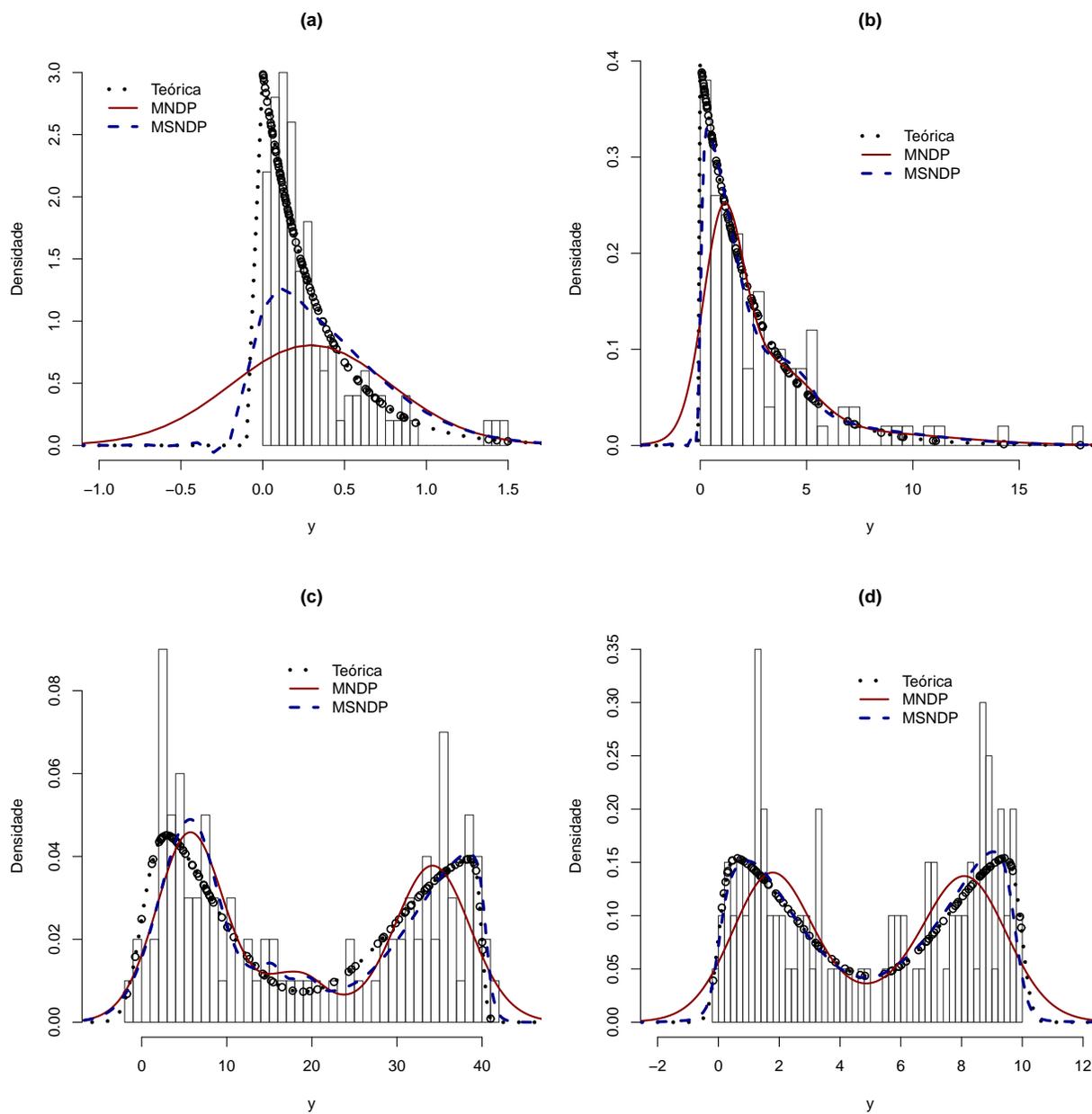


Figura 5.7: Histograma da amostra observada e densidades estimadas pelos modelos MNDP e MSNDP. (a) **Cenário 13**; (b) **Cenário 14**; (c) **Cenário 15** e (d) **Cenário 16**.

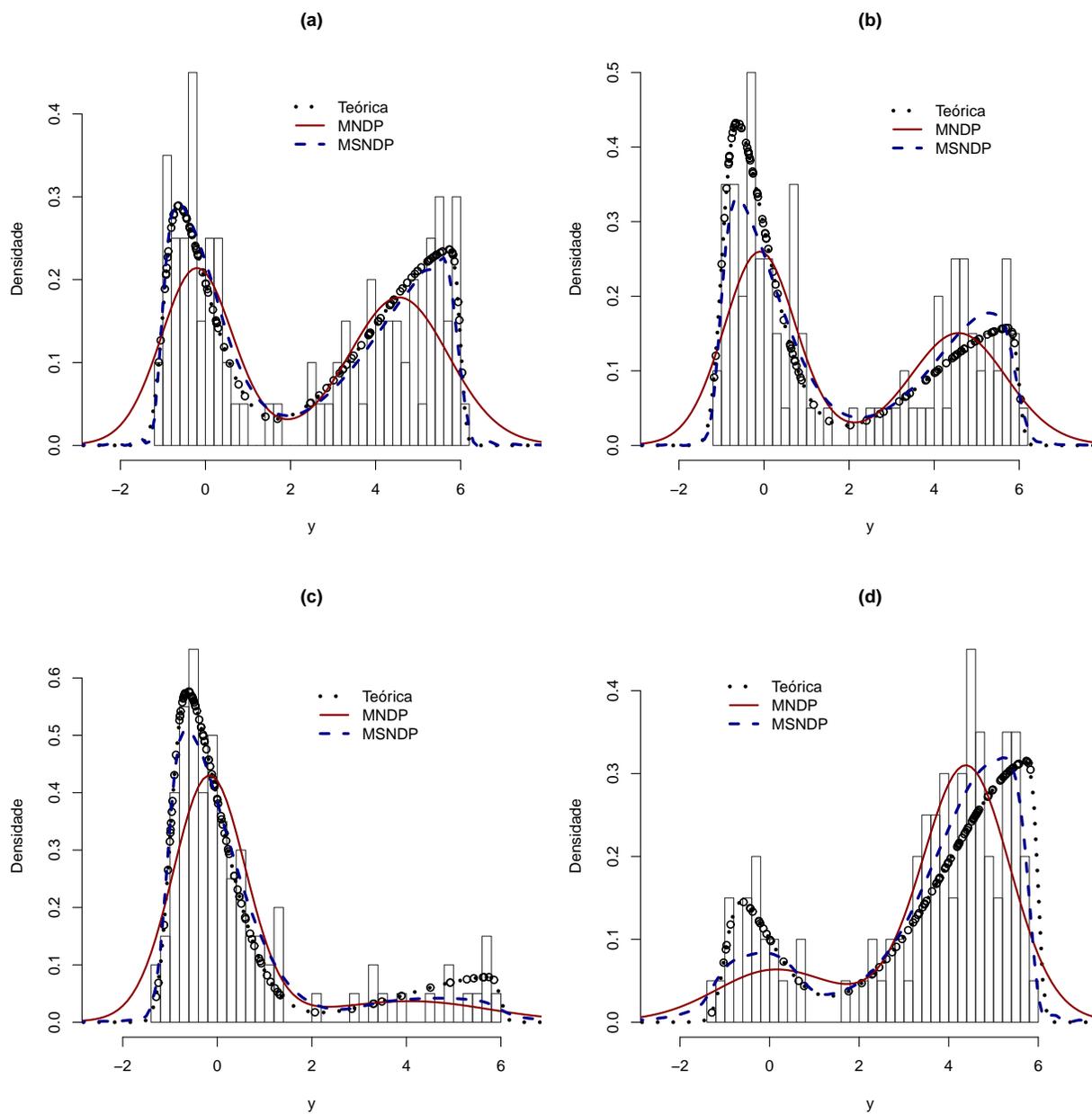


Figura 5.8: Histograma da amostra observada e densidades estimadas pelos modelos MNDP e MSNDP. (a) **Cenário 17**; (b) **Cenário 18**; (c) **Cenário 19** e (d) **Cenário 20**.

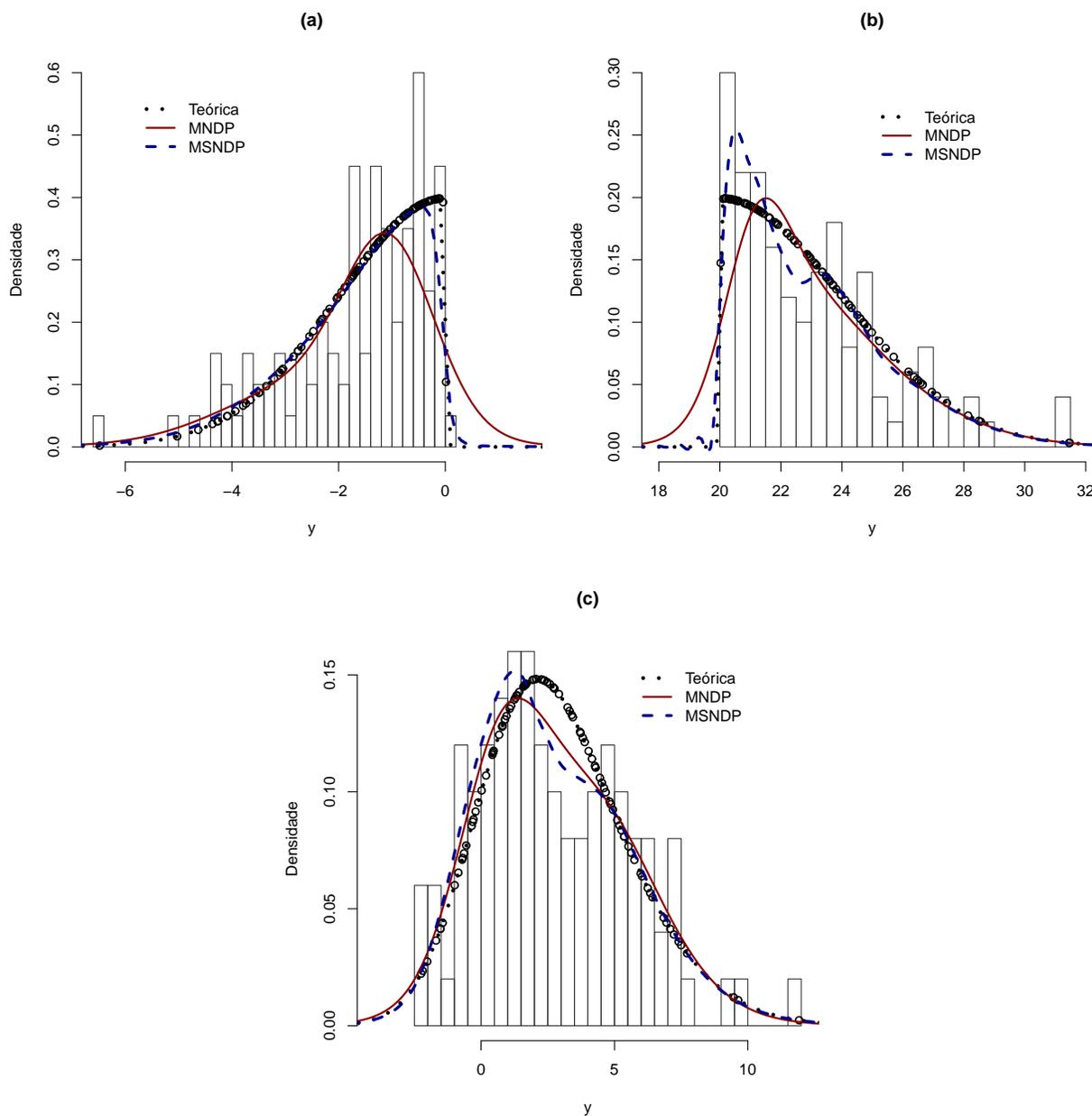


Figura 5.9: Histograma da amostra observada e densidades estimadas pelos modelos MNDP e MSNDP. (a) **Cenário 21**; (b) **Cenário 22** e (c) **Cenário 23**.

Tabela 5.4: Erros Quadráticos Médios Integrados das estimativas de densidade obtidas.

	$EQMI_1$	$EQMI_2$	$Dif = EQMI_1 - EQMI_2$
Cenário 1	1.69×10^{-4}	3.93×10^{-4}	-2.24×10^{-4}
Cenário 2	1.97×10^{-4}	0.99×10^{-4}	0.98×10^{-4}
Cenário 3	0.99×10^{-4}	1.44×10^{-4}	-0.45×10^{-4}
Cenário 4	1.99×10^{-4}	1.36×10^{-4}	0.63×10^{-4}
Cenário 5	0.91×10^{-5}	1.81×10^{-5}	-0.90×10^{-5}
Cenário 6	1.91×10^{-5}	4.07×10^{-5}	-2.16×10^{-5}
Cenário 7	5.49×10^{-6}	8.98×10^{-6}	-3.49×10^{-6}
Cenário 8	0.34×10^{-4}	2.07×10^{-4}	-1.73×10^{-4}
Cenário 9	1.20×10^{-4}	0.74×10^{-4}	0.46×10^{-4}
Cenário 10	1.20×10^{-4}	0.13×10^{-4}	1.07×10^{-4}
Cenário 11	1.77×10^{-4}	0.61×10^{-4}	1.16×10^{-4}
Cenário 12	2.42×10^{-4}	1.46×10^{-4}	0.96×10^{-4}
Cenário 13	9.50×10^{-2}	4.63×10^{-2}	4.87×10^{-2}
Cenário 14	1.30×10^{-3}	0.43×10^{-3}	0.87×10^{-3}
Cenário 15	2.69×10^{-5}	1.96×10^{-5}	0.73×10^{-5}
Cenário 16	3.16×10^{-4}	0.45×10^{-4}	2.71×10^{-4}
Cenário 17	6.96×10^{-4}	0.75×10^{-4}	6.21×10^{-4}
Cenário 18	1.20×10^{-3}	0.29×10^{-3}	0.91×10^{-3}
Cenário 19	1.28×10^{-3}	0.22×10^{-3}	1.06×10^{-3}
Cenário 20	1.19×10^{-3}	0.51×10^{-3}	0.68×10^{-3}
Cenário 21	1.19×10^{-3}	0.18×10^{-3}	1.01×10^{-3}
Cenário 22	1.74×10^{-4}	0.78×10^{-4}	0.96×10^{-4}
Cenário 23	2.84×10^{-5}	5.91×10^{-5}	-3.07×10^{-5}

Para os cenários 1, 13, 18 e 21, avaliamos ainda a diferença (em valor absoluto) entre a densidade estimada e a densidade teórica ao longo do domínio de $f_Y(y)$. Essa diferença será denotada por $D(y_i)$ e é definida por

$$D(y_i) = |\hat{f}(y_i) - f(y_i)|, y_i \in \mathcal{Y}.$$

Essa análise foi realizada para as estimativas obtidas a partir dos modelos MNDP e MSNDP. As diferenças $D(y_i)$, referentes a ambos os modelos, são representadas graficamente na Figura 5.10.

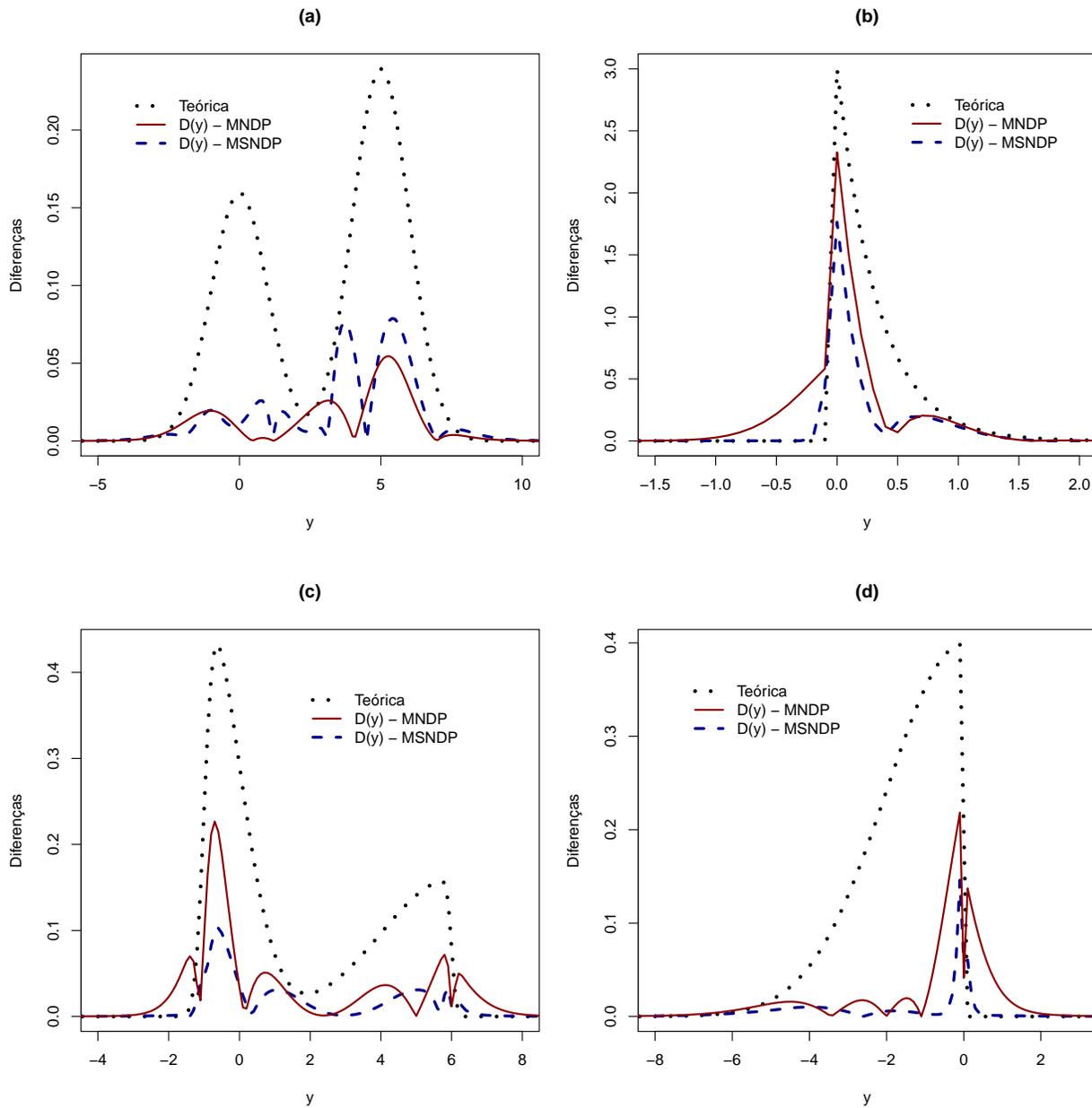


Figura 5.10: Curvas das diferenças $D(y_i)$ para os modelos MNDP e MSNDP. (a) **Cenário 1**; (b) **Cenário 13**; (c) **Cenário 18** e (d) **Cenário 21**.

A Figura 5.4 apresenta os gráficos referentes aos cenários de 1 a 4. Nesses cenários os dados foram gerados de uma mistura de duas distribuições normais. A diferença entre as misturas consideradas está nos pesos atribuídos a cada uma das distribuições pertencentes a mistura. Pela análise dos gráficos (a), (b) e (d) da Figura 5.4, é possível observar que tanto o modelo MNDP quanto o modelo MSNDP produzem estimativas de densidade satisfatórias, captando a bimodalidade e a variabilidade da distribuição teórica. No Cenário 3, representado pelo gráfico 5.4 (c), verifica-se que ambos os modelos não identificam corretamente a distribuição com menos peso na mistura. A

análise do Dif dos cenários 1 e 3, apresentado na Tabela 5.4, revela uma melhor adequação do modelo MNDP. Já nos cenários 2 e 4, o modelo MSNDP foi o que se mostrou mais adequado.

Os cenários de 5 a 8 estão graficamente representados na Figura 5.5. Analisando essa figura, observamos que as estimativas obtidas pelo modelo MNDP apresentam qualidade superior àquelas obtidas pelo modelo MSNDP em todos os cenários representados. Verifica-se ainda que as estimativas obtidas pelo modelo MSNDP são mais afetadas pela distribuição dos dados simulados. No Cenário 8, representado no gráfico 5.5 (c), consideramos uma distribuição *a priori* não informativa para o parâmetro de escala da distribuição *skew-normal*. Nesse cenário, ambos os modelos de estimação produziram estimativas adequadas. O valor negativo da diferença Dif obtido nesses quatro cenários indica uma maior conformidade das estimativas obtidas via modelo MNDP.

Os gráficos apresentados na Figura 5.6 mostram diferentes misturas de distribuições gama. Esses gráficos demonstram que os modelos MNDP e MSNDP conseguem captar a assimetria das distribuições teóricas, porém apenas o modelo MSNDP estima de forma satisfatória a cauda à esquerda dessas distribuições. Note que, em todos os casos, o modelo MSNDP põe pouca ou nenhuma massa de probabilidade nos valores negativos. Isso evidencia o ganho que se tem ao se utilizar o modelo MSNDP na estimação de densidades estritamente positivas. Salientamos a ótima qualidade das estimativas obtidas pelo modelo MSNDP nos cenários 10 e 11. A melhor adequação do modelo MSNDP também pode ser verificada analisando-se o valor da diferença entre os EQMI's referente aos cenários 9 a 12, que é apresentado na Tabela 5.4.

Os gráficos (a) e (b) da Figura 5.7 resultam da análise dos cenários 13 e 14, respectivamente. Em ambos os cenários, os dados foram gerados a partir de distribuições exponenciais. Note que no Cenário 13, onde consideramos uma exponencial com decaimento rápido, ambos os modelos fornecem estimativas de baixa qualidade. Já no Cenário 14, onde o decaimento da distribuição é mais lento, verifica-se uma considerável melhora na qualidade das estimativas. Nos cenários 15 e 16 as densidades teóricas são misturas de distribuições *skew-normal*. A densidade avaliada no Cenário 15 apresenta uma alta variabilidade e a avaliada no Cenário 16 é uma mistura simétrica. Pela análise dos gráficos (c) e (d) da Figura 5.7 verificamos que o modelo MNDP falha ao estimar a densidade teórica de ambos os cenários, uma vez que a estimativa obtida não capta a assimetria da mesma. O gráfico 5.7 (c) demonstra ainda que a curva estimada pelo modelo MSNDP não estima corretamente a distribuição à direita da mistura. Já o gráfico 5.7 (d) demonstra a ótima qualidade da estimativa obtida pelo modelo MSNDP. Tanto a análise gráfica quanto a análise do Dif obtido para os cenários 13 a 16 evidenciam uma melhor adequação do modelo MSNDP para estimar as densidades consideradas.

A Figura 5.8 apresenta os gráficos referentes aos cenários de 17 a 20. Nesses cenários os dados foram gerados de uma mistura de distribuições *skew-normal*. A diferença entre as misturas

consideradas está nos pesos atribuídos às distribuições pertencentes a mistura. Os gráficos (a), (b) e (c) da Figura 5.8 evidenciam o ótimo ajuste das curvas estimadas pelo modelo MSNDP. No Cenário 20, representado pelo gráfico 5.8 (d), observamos que apenas a distribuição com maior peso na mistura foi adequadamente estimada pelo modelo MSNDP. A Figura 5.8 evidencia ainda a incapacidade do modelo MNDP em estimar distribuições assimétricas com mais de uma moda. A partir da Tabela 5.4 verificamos que o modelo mais adequado para estimar densidades nos cenários de 17 a 20 é o modelo MSNDP.

Os gráficos (a) e (b) da Figura 5.9 apresentam distribuições *skew-normal* com forte assimetria negativa e positiva, respectivamente. Ao examinar esses gráficos observamos que ambos os modelos captam a assimetria das densidades teóricas. Apesar disso, observamos que o modelo MNDP põe massa em valores que possuem probabilidade de ocorrência zero, tornando esse modelo inadequado para estimar densidades com forte assimetria. Em contrapartida, a utilização do modelo MSNDP para estimar densidades com essa característica é bastante apropriada. A análise do *Dif* dos cenários 21 e 22 confirma a superioridade do modelo MSNDP nessas situações. O gráfico (c) da Figura 5.9 apresenta uma distribuição *skew-normal* com leve assimetria positiva. A análise desse gráfico juntamente com a análise do valor do *Dif* referente ao Cenário 23 revelam um melhor ajuste da curva estimada pelo modelo MNDP. Examinando os gráficos 5.9 (b) e 5.9 (c) observamos ainda que as estimativas obtidas pelo modelo MSNDP são mais afetadas pela distribuição dos dados simulados.

A Figura 5.10 demonstra que, no caso em que a densidade teórica é uma mistura de distribuições normais (gráfico 5.10 (a)), as diferenças $D(y_i)$ referentes ao modelo MSNDP são superiores às diferenças referentes ao modelo MNDP na maior parte do domínio de $f_Y(y)$. Já nos casos onde a densidade teórica é uma mistura de distribuições assimétricas (gráficos 5.10 (b), (c) e (d)), observa-se a superioridade do modelo MSNDP ao estimar os pontos de mais alta densidade e as caudas da densidade teórica.

Vale salientar que, em todas as análises realizadas nesse capítulo, consideramos uma amostra de tamanho $n = 100$. Certamente um aumento no tamanho amostral irá acarretar em uma melhoria na qualidade de todas as estimativas obtidas, visto que o histograma dos dados estará mais próximo da densidade teórica.

5.2.1 Análise da Distribuição a Posteriori do Parâmetro de Assimetria

Nesta seção, avaliamos o comportamento do parâmetro de assimetria da distribuição *skew-normal* em alguns cenários específicos. Para tal, estimamos a média da distribuição *a posteriori* de λ_i , $i = 1, \dots, n$, a partir da amostra gerada no estudo de dados simulados. Construímos ainda o intervalo HDP de 95% correspondente a cada estimativa obtida. Essas análises foram realizadas

para os cenários 1, 13, 18 e 21 e os resultados obtidos são apresentados na Figura 5.11.

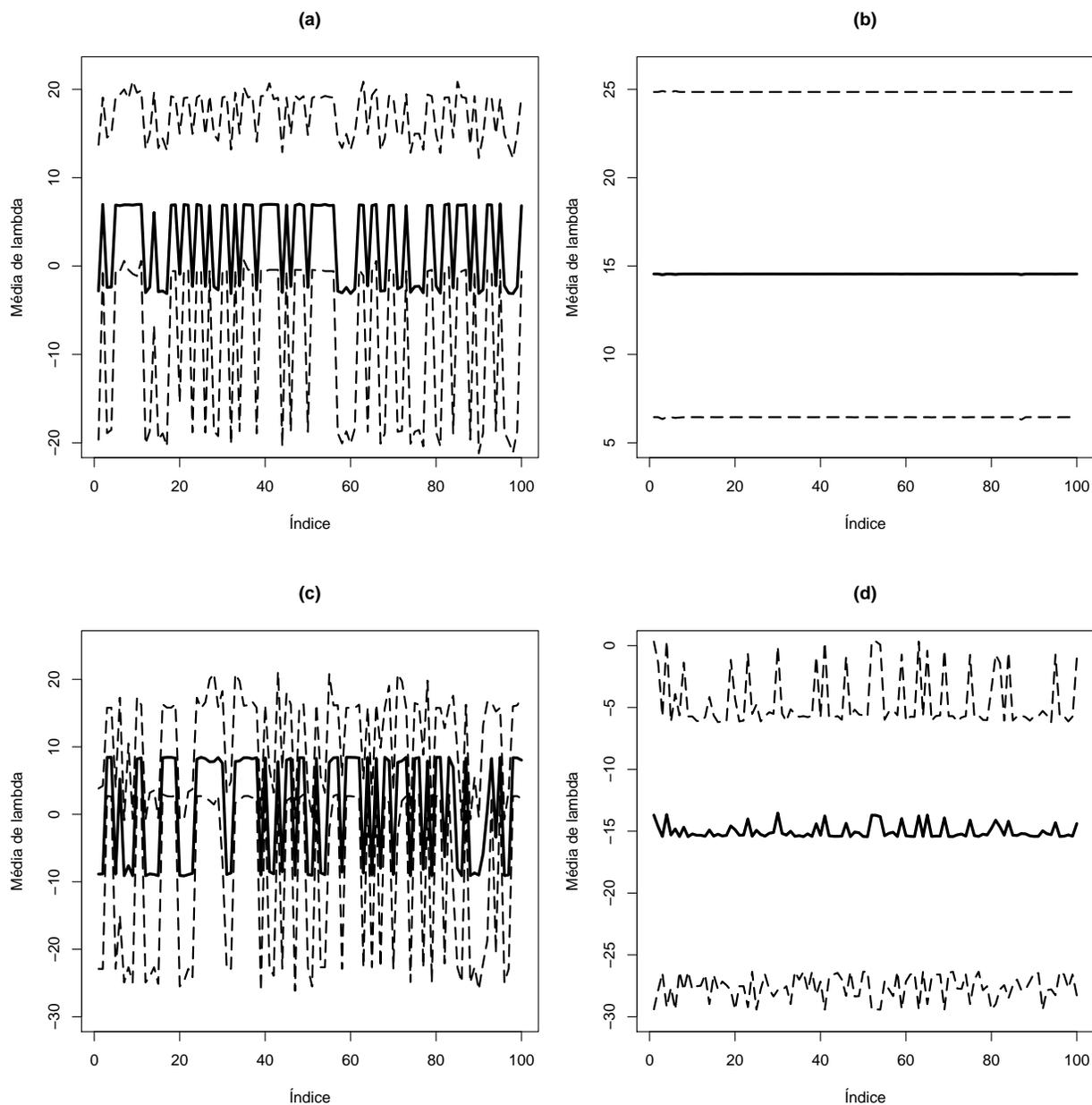


Figura 5.11: Estimativa da média da distribuição *a posteriori* do parâmetro de assimetria λ_i e intervalo HPD de 95%. (a) Cenário 1; (b) Cenário 13; (c) Cenário 18 e (d) Cenário 21.

Pela análise do gráfico 5.11 (a) podemos observar que no Cenário 1, onde consideramos uma mistura de distribuições simétricas, o valor médio de lambda oscila em torno de zero. Isso já era de se esperar, visto que para $\lambda = 0$ a densidade da *skew-normal* se reduz à densidade de uma normal. No caso em que a densidade teórica é estritamente positiva (negativa) verificamos, também como esperado, que o parâmetro de assimetria assume em média valores positivos (negativos), mais

especificamente valores em torno de 15 (-15). Essa avaliação foi realizada com base na análise dos gráficos 5.11 (b) e (d), respectivamente. Já no Cénario 18, representado graficamente pela Figura 5.11 (c), verifica-se novamente que o valor médio de λ oscila em torno de zero, porém a variabilidade dos valores médios observados é bem mais ampla que a variabilidade apresentada no Cénario 1. Em todos os casos avaliados observa-se ainda uma baixa precisão dos intervalos HPD contruídos.

Capítulo 6

Estudo de Caso - Duração de Erupções do Géiser Old Faithful

Nesta seção, realizamos um estudo empregando dados reais. O conjunto de dados utilizado refere-se à duração (em minutos) de 107 erupções do géiser Old Faithful, que está localizado no Parque Nacional de Yellowstone, EUA. Esse conjunto de dados encontra-se disponível em Silverman (1986). A densidade que originou os dados foi estimada a partir dos modelos MNDP-EW, MNDP-MM e MSNDP. O intuito é comparar, através da análise gráfica, as estimativas obtidas a partir da implementação desses modelos. O conjunto de dados analisado é apresentado na Tabela 6.1.

Tabela 6.1: Duração de 107 erupções do géiser Old Faithful.

Duração da Erupção (em minutos)									
4.37	3.95	4.13	1.67	2.00	3.75	1.80	4.08	4.42	3.58
4.70	4.83	4.53	4.63	3.73	4.00	3.70	3.92	3.80	3.77
1.68	3.87	4.10	1.83	4.60	4.33	2.50	4.73	4.60	1.77
1.75	1.73	4.12	2.03	2.93	1.82	2.27	3.72	2.25	4.13
4.35	3.92	4.00	2.72	4.65	1.67	2.93	4.50	1.83	3.67
1.77	3.20	4.93	4.03	4.18	3.50	4.63	4.40	3.33	3.43
4.25	2.33	3.68	1.73	4.58	4.20	4.00	4.58	4.00	4.50
4.10	4.57	1.85	3.10	3.50	4.43	1.97	3.50	4.25	1.95
4.05	3.58	3.83	4.62	4.62	1.90	3.93	1.80	1.83	1.90
1.90	3.70	1.85	1.88	4.03	4.08	4.07	4.28	3.73	4.33
4.00	4.25	3.80	3.52	1.97	3.43	4.50			

Nesse estudo, assumimos os seguintes valores para os hiperparâmetros dos modelos de MDP ($m = \bar{y}$, $\alpha = 1$, $\tau = 1.5$, $s = 6$, $S = 16$, $\varepsilon = 0$ e $z^2 = 100$). No esquema MCMC, consideramos uma única cadeia de tamanho 500.000. Após a convergência ter sido obtida, descartou-se um *burn-in* de tamanho 490.000 e considerou-se um *lag* igual a 1. Resultando, assim, em uma amostra da distribuição *a posteriori* de tamanho 10.000. Os algoritmos utilizados

neste estudo foram implementados na linguagem de programação R, através do uso do *software* R 2.9.0.

A seguir, apresentamos o gráfico contendo as curvas das densidades estimadas pelos modelos MNDP-EW, MNDP-MM e MSNDP e o histograma da amostra observada. O histograma apresentado na Figura 6.1 indica que a distribuição da duração das erupções do géiser Old Faithful apresenta assimetria e bimodalidade.

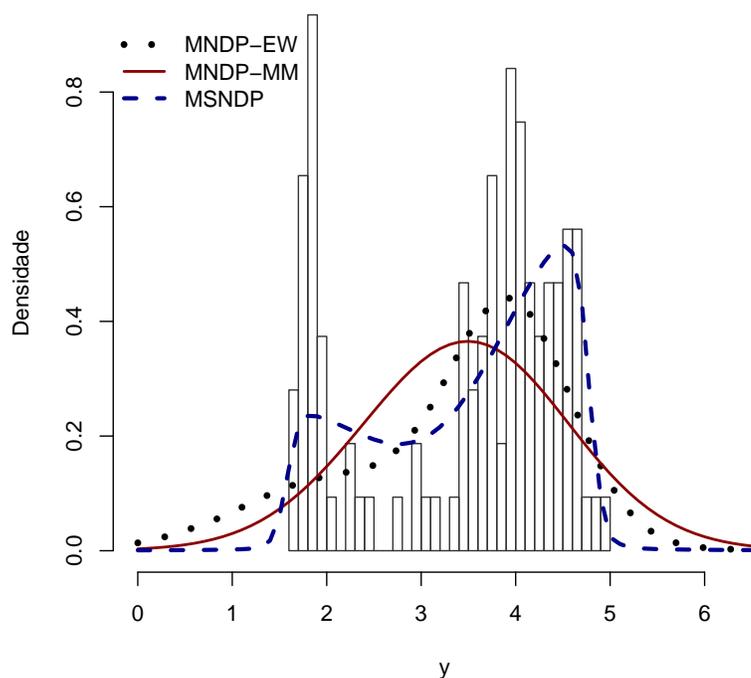


Figura 6.1: Histograma da amostra observada e densidades estimadas pelos modelos MNDP-EW, MNDP-MM e MSNDP.

Analisando a Figura 6.1, pode-se verificar que o modelo MSNDP foi o único que forneceu uma estimativa de densidade que indica tanto a assimetria quanto a bimodalidade que parece haver na distribuição dos dados. Também é possível observar que a curva estimada pelo modelo MSNDP possui caudas mais leves que as curvas estimadas pelos outros dois modelos, estimando de forma mais adequada a distribuição empírica dos dados. O modelo MNDP-EW indicou apenas a assimetria da distribuição. Já a curva estimada pelo modelo MNDP-MM é unimodal e aproximadamente simétrica, o que está muito distante da evidência empírica fornecida pelos dados. Em resumo, o modelo proposto capturou melhor as características apontadas no histograma dos dados.

Capítulo 7

Conclusões e Discussão

Neste trabalho, utilizamos o modelo hierárquico de misturas por processos de Dirichlet com a finalidade de estimarmos densidades. O principal objetivo do trabalho foi estender o modelo MNDP para estimação de densidades, proposto por Escobar e West (1995), considerando uma mistura de distribuições *skew-normal* no primeiro estágio do modelo. Comparou-se a qualidade das estimativas de densidade obtidas a partir do modelo proposto (MSNDP) com àquelas obtidas pelo modelo MNDP. Essa comparação foi realizada implementando-se o algoritmo de MacEachern e Müller (1998). Tivemos como objetivo secundário, comparar os algoritmos de Escobar e West (1995) e de MacEachern e Müller (1998) na estimação de densidades via modelo MNDP. A fim de atingir nossas metas, realizamos dois estudos envolvendo dados simulados e um estudo envolvendo dados reais. Os resultados desses estudos são apresentados nos capítulos 5 e 6.

Os resultados obtidos na análise de dados simulados apresentada na Seção 5.1 mostraram que, de uma forma geral, tanto o algoritmo proposto por Escobar e West (1995) quanto o proposto por MacEachern e Müller (1998) produzem aproximações razoáveis para a densidade teórica. No entanto, constatamos que o algoritmo de MacEachern e Müller (1998) converge mais rapidamente e gera cadeias menos correlacionadas. Na maioria os cenários analisados, observamos uma proximidade entre as curvas estimadas pelos dois algoritmos. Porém, a análise gráfica e a análise da diferença *Dif* revelaram uma melhor adequação das curvas estimadas pelo algoritmo de MacEachern e Müller (1998) em quase todos os cenários considerados. Observamos também que a distribuição da qual os dados foram gerados influi significativamente na qualidade das estimativas. Constatamos, como esperado, que o modelo de MNDP fornece melhores estimativas para densidades simétricas ou que são misturas de distribuições simétricas. No caso de distribuições assimétricas, o modelo tende a estimar de forma inadequada as caudas da distribuição, atribuindo massa não desprezível a valores da variável aleatória que originalmente possuem probabilidade de ocorrência zero. Por exemplo, em situações onde a densidade teórica é estritamente positiva, o modelo MNDP tende a fornecer estimativas que atribuem massa de probabilidade a valores negativos da variável aleatória.

Com base nos resultados obtidos na análise de dados simulados apresentada na Seção 5.2, concluímos que o modelo MNDP é mais adequado para estimar densidades com leve assimetria ou que sejam provenientes de misturas de distribuições simétricas. Já o modelo MSNDP mostrou-se mais adequado na estimação de densidades assimétricas e/ou que põem massa apenas em valores não-negativos ou não-positivos. Nos casos em que a densidade estimada apresenta pelo menos uma dessas características, a qualidade das estimativas obtidas pelo modelo proposto neste trabalho revelou-se muito superior àquelas obtidas pelo modelo MNDP. Esses resultados ratificam as suposições iniciais que motivaram a realização deste trabalho.

Como pôde ser observado, os modelos MNDP-EW, MNDP-MM e MSNDP produziram estimativas de densidade satisfatórias na maioria das análises realizadas. Apesar disso, todos eles apresentaram problemas na convergência de alguns parâmetros e longas cadeias tiveram que ser geradas para atingi-la. O único estudo que não apresentou tal problema foi o estudo que envolve dados reais. A análise de convergência foi realizada através do teste de Geweke (1992), disponível no pacote *coda* do *software* R. Esse problema de convergência se deu com mais intensidade em alguns modelos do que em outros. O modelo que apresentou melhor comportamento com respeito à convergência foi o modelo MNDP-MM. O cenário considerado também influenciou na qualidade da convergência da cadeia para a distribuição *a posteriori* dos parâmetros. Cenários que envolvem densidades com grande número de modas, que possuem alta variabilidade ou uma assimetria muito acentuada (*e.g.* exponencial, alguns casos de distribuições gama, etc.), em geral, apresentaram convergência mais lenta. Comportamento similar foi observado nos cenários onde consideramos uma distribuição *a priori* não informativa para o parâmetro de escala. Tentamos gerar cadeias com tamanhos superiores a 200.000, porém o *software* R acusou erros na capacidade de armazenagem.

Outro problema, encontrado ao longo do desenvolvimento deste trabalho, deu-se na implementação do modelo MSNDP¹. Mais especificamente, no processo de obtenção de amostras das distribuições condicionais completas *a posteriori* apresentadas nas equações (4.13) e (4.15). Visto que a tarefa de amostrar diretamente dessas distribuições é bastante complicada, optamos por gerar amostras a partir das representações estocásticas apresentadas nas Proposições 2 e 3. No entanto, o fato de não termos conseguido gerar valores da distribuição normal truncada multivariada impossibilitou a obtenção de amostras a partir dessas representações. Dessa forma, amostras das condicionais completas de μ_j^* e λ_j^* foram obtidas através da implementação do algoritmo de Metropolis-Hastings.

Como proposta para trabalhos futuros, estamos considerando uma outra extensão do modelo de Escobar e West (1995), na qual assumimos distribuições mais flexíveis para a distribuição basal do processo de Dirichlet.

¹Para maiores detalhes a respeito do modelo MSNDP veja a Seção 4.2.2.

Apêndice A

Demonstrações

Neste apêndice, apresentamos as demonstrações dos resultados apresentados nas equações (4.5), (4.8), (4.10), (4.13), (4.14) e (4.15).

- Cálculo da distribuição condicional completa *a posteriori* dada pela equação (4.5).

$$\begin{aligned}
P(\theta_i | \boldsymbol{\theta}_{(-i)}, \mathbf{y}) &= \frac{P(\theta_i, \boldsymbol{\theta}_{(-i)}, \mathbf{y})}{P(\boldsymbol{\theta}_{(-i)}, \mathbf{y})} = \frac{P(\mathbf{y} | \theta_i, \boldsymbol{\theta}_{(-i)}) P(\theta_i | \boldsymbol{\theta}_{(-i)})}{\int P(\mathbf{y} | \theta_i, \boldsymbol{\theta}_{(-i)}) P(\theta_i | \boldsymbol{\theta}_{(-i)}) d\theta_i} \\
&= \frac{\prod_{j=1}^n N(\mu_j, V_j) \left[\frac{\alpha}{\alpha+n-1} G_0(\theta_i) + \frac{1}{\alpha+n-1} \sum_{j=1, j \neq i}^n \delta_{\theta_j}(\theta_i) \right]}{\int \prod_{j=1}^n N(\mu_j, V_j) \left[\frac{\alpha}{\alpha+n-1} G_0(\theta_i) + \frac{1}{\alpha+n-1} \sum_{j=1, j \neq i}^n \delta_{\theta_j}(\theta_i) \right] d\theta_i} \\
&= \frac{\frac{\prod_{j=1, j \neq i}^n N(\mu_j, V_j)}{\alpha+n-1} \left[\alpha N(\mu_i, V_i) G_0(\theta_i) + N(\mu_i, V_i) \sum_{j=1, j \neq i}^n \delta_{\theta_j}(\theta_i) \right]}{\int \frac{\prod_{j=1, j \neq i}^n N(\mu_j, V_j)}{\alpha+n-1} \left[\alpha N(\mu_i, V_i) G_0(\theta_i) + N(\mu_i, V_i) \sum_{j=1, j \neq i}^n \delta_{\theta_j}(\theta_i) \right] d\theta_i} \\
&= \frac{\alpha N(\mu_i, V_i) G_0(\theta_i) + \sum_{j=1, j \neq i}^n N(\mu_j, V_j) \delta_{\theta_j}(\theta_i)}{\int \alpha N(\mu_i, V_i) G_0(\theta_i) + \sum_{j=1, j \neq i}^n N(\mu_j, V_j) \delta_{\theta_j}(\theta_i) d\theta_i}. \tag{A.1}
\end{aligned}$$

Segue de (A.1) que

$$P(\theta_i | \boldsymbol{\theta}_{(-i)}, \mathbf{y}) \propto \alpha N(\mu_i, V_i) G_0(\theta_i) + \sum_{j=1, j \neq i}^n N(\mu_j, V_j) \delta_{\theta_j}(\theta_i).$$

Temos, agora, que mostrar que $\alpha N(\mu_i, V_i) G_0(\theta_i) = q_0 G_i(\theta_i)$. Tem-se que

$$\begin{aligned}
 & \alpha N(\mu_i, V_i) G_0(\theta_i) = \alpha N(\mu_i, V_i) NGI(m, \tau, s/2, S/2) = \alpha N(\mu_i, V_i) N(m, \tau V_i) GI(s/2, S/2) = \\
 & = \alpha \frac{1}{(2\pi V_i)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2V_i}(y_i - \mu_i)^2\right\} \frac{1}{(2\pi\tau V_i)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2\tau V_i}(\mu_i - m)^2\right\} GI(s/2, S/2) \\
 & = \underbrace{\frac{\alpha GI(s/2, S/2)}{2\pi V_i \tau^{\frac{1}{2}}}}_K \exp\left\{-\frac{1}{2V_i}\left[y_i^2 - 2\mu_i y_i + \mu_i^2 + \frac{\mu_i^2}{\tau} - \frac{2\mu_i m}{\tau} + \frac{m^2}{\tau}\right]\right\} \\
 & = K \exp\left\{-\frac{1}{2V_i}\left(y_i^2 + \frac{m^2}{\tau}\right)\right\} \exp\left\{-\frac{1}{2}\left(\frac{\tau+1}{\tau V_i}\right)\left[\mu_i^2 - 2\mu_i\left(\frac{\tau y_i + m}{\tau}\right)\left(\frac{\tau}{\tau+1}\right)\right]\right\} \\
 & = K \exp\left\{-\frac{1}{2V_i}\left(y_i^2 + \frac{m^2}{\tau}\right)\right\} \exp\left\{-\frac{1}{2}\left(\frac{\tau+1}{\tau V_i}\right)\left[\mu_i^2 - 2\mu_i\left(\frac{\tau y_i + m}{\tau+1}\right) \pm \left(\frac{\tau y_i + m}{\tau+1}\right)^2\right]\right\} \\
 & = K \exp\left\{-\frac{1}{2V_i}\left(y_i^2 + \frac{m^2}{\tau}\right)\right\} \exp\left\{-\frac{1}{2}\left(\frac{\tau+1}{\tau V_i}\right)\left[\mu_i^2 - 2\mu_i\left(\frac{\tau y_i + m}{\tau+1}\right) \pm \left(\frac{\tau y_i + m}{\tau+1}\right)^2\right]\right\} \\
 & = K \exp\left\{-\frac{1}{2V_i}\left(y_i^2 + \frac{m^2}{\tau}\right) - \frac{1}{2}\left(\frac{\tau+1}{\tau V_i}\right)\left[-\frac{(\tau y_i + m)^2}{(\tau+1)^2}\right]\right\} \left(\frac{2\pi\tau V_i}{\tau+1}\right)^{\frac{1}{2}} \left[\frac{1}{2\pi}\left(\frac{\tau+1}{\tau V_i}\right)\right]^{\frac{1}{2}} \times \\
 & \times \exp\left\{-\frac{1}{2}\left(\frac{\tau+1}{\tau V_i}\right)\left[\mu_i - \left(\frac{\tau y_i + m}{\tau+1}\right)\right]^2\right\} \\
 & = \frac{\alpha GI(s/2, S/2)}{[2\pi V_i(\tau+1)]^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left[\frac{y_i^2}{\tau V_i + V_i} - \frac{2y_i m}{\tau V_i + V_i} + \frac{m^2}{\tau V_i + V_i}\right]\right\} N\left[\left(\frac{\tau y_i + m}{\tau+1}\right), \left(\frac{\tau V_i}{\tau+1}\right)\right] \\
 & = \frac{\alpha}{[2\pi V_i(\tau+1)]^{\frac{1}{2}}} \left(\frac{S}{2}\right)^{\frac{s}{2}} \left(\frac{1}{V_i}\right)^{\frac{s}{2}+1} \exp\left\{-\frac{S}{2V_i}\right\} \exp\left\{-\frac{(y_i - m)^2}{2(\tau V_i + V_i)}\right\} N(x_i, X V_i) \\
 & = \frac{\alpha}{[2\pi(\tau+1)]^{\frac{1}{2}}} \frac{\left(\frac{S}{2}\right)^{\frac{s}{2}}}{\Gamma\left(\frac{s}{2}\right)} \frac{\Gamma\left(\frac{s+1}{2}\right)}{\left(\frac{S}{2} + \frac{(y_i - m)^2}{2(\tau+1)}\right)^{\frac{s+1}{2}}} \frac{\left(\frac{S}{2} + \frac{(y_i - m)^2}{2(\tau+1)}\right)^{\frac{s+1}{2}}}{\Gamma\left(\frac{s+1}{2}\right)} \left(\frac{1}{V_i}\right)^{\frac{s+1}{2}+1} \times \\
 & \times \exp\left\{-\frac{1}{V_i}\left(\frac{S}{2} + \frac{(y_i - m)^2}{2(\tau+1)}\right)\right\} N(x_i, X V_i) \\
 & = \frac{\alpha}{[2\pi(\tau+1)]^{\frac{1}{2}}} \frac{\left(\frac{S}{2}\right)^{\frac{s}{2}}}{\Gamma\left(\frac{s}{2}\right)} \frac{\Gamma\left(\frac{s+1}{2}\right)}{\left(\frac{S}{2} + \frac{(y_i - m)^2}{2(\tau+1)}\right)^{\frac{s+1}{2}}} \left(\frac{S}{2}\right)^{\frac{1}{2}} \times GI\left[\frac{s+1}{2}, \frac{1}{2}\left(S + \frac{(y_i - m)^2}{\tau+1}\right)\right] N(x_i, X V_i) \\
 & = \alpha \frac{\Gamma\left(\frac{s+1}{2}\right)}{\sqrt{\pi} \Gamma\left(\frac{s}{2}\right)} [(\tau+1)S]^{-\frac{1}{2}} \left[\frac{1}{1 + \frac{(y_i - m)^2}{S(\tau+1)}}\right]^{\frac{s+1}{2}} GI\left[\frac{s+1}{2}, \frac{S_i}{2}\right] N(x_i, X V_i)
 \end{aligned}$$

$$\begin{aligned}
 \alpha N(\mu_i, V_i) G_0(\theta_i) &= \\
 &= \alpha \frac{\Gamma\left(\frac{s+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{s}{2}\right)} \left[s(\tau+1) \frac{S}{s} \right]^{-\frac{1}{2}} \left[1 + \frac{(y_i - m)^2}{s(\tau+1) \frac{S}{s}} \right]^{-\frac{s+1}{2}} GI\left[\frac{s+1}{2}, \frac{S_i}{2}\right] N(x_i, XV_i) \\
 &= \alpha \frac{\Gamma\left(\frac{s+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{s}{2}\right)} [sM]^{-\frac{1}{2}} \left[1 + \frac{(y_i - m)^2}{sM} \right]^{-\frac{s+1}{2}} GI\left[\frac{s+1}{2}, \frac{S_i}{2}\right] N(x_i, XV_i) \\
 &= \alpha T_s(m, M) GI\left[\frac{s+1}{2}, \frac{S_i}{2}\right] N(x_i, XV_i) = q_0 G_i(\theta_i).
 \end{aligned}$$

Portanto, tem-se que:

$$P(\theta_i | \boldsymbol{\theta}_{(-i)}, \mathbf{y}) \propto q_0 G_i(\theta_i) + \sum_{j=1, j \neq i}^n q_j \delta_{\theta_j}(\theta_i).$$

- Demonstração do resultado apresentado na equação (4.8):

$$\begin{aligned}
 P(Y_{n+1} | \boldsymbol{\theta}) &= \int P(Y_{n+1}, \theta_{n+1} | \boldsymbol{\theta}) d\theta_{n+1} \\
 &= \int P(Y_{n+1} | \boldsymbol{\theta}, \theta_{n+1}) P(\theta_{n+1} | \boldsymbol{\theta}) d\theta_{n+1} \\
 &= \int P(Y_{n+1} | \theta_{n+1}) P(\theta_{n+1} | \boldsymbol{\theta}) d\theta_{n+1} \\
 &= \int N(\mu_{n+1}, V_{n+1}) P(\theta_{n+1} | \boldsymbol{\theta}) d\theta_{n+1} \\
 &= \frac{\alpha}{\alpha + n} \int N(\mu_{n+1}, V_{n+1}) G_0(\theta_{n+1}) d\theta_{n+1} + \frac{1}{\alpha + n} N(\mu_{n+1}, V_{n+1}) \sum_{i=1}^n \delta_{\theta_i}(\theta_{n+1}) \\
 &= \frac{\alpha}{\alpha + n} \int N(\mu_{n+1}, V_{n+1}) G_0(\theta_{n+1}) d\theta_{n+1} + \frac{1}{\alpha + n} \sum_{i=1}^n N(\mu_i, V_i).
 \end{aligned}$$

Como G_0 é uma distribuição Normal-Gama-Invertida dada em (4.2), segue que

$$\begin{aligned}
 &\int N(\mu_i, V_i) G_0(\theta_i) d\theta_i = \\
 &= \iint N(\mu_i, V_i) NGI(m, \tau, s/2, S/2) d\mu_i dV_i \\
 &= \iint N(\mu_i, V_i) N(m, \tau V_i) GI(s/2, S/2) d\mu_i dV_i
 \end{aligned}$$

$$\begin{aligned}
 & \int N(\mu_i, V_i) G_0(\theta_i) d\theta_i = \\
 &= \iint \frac{1}{(2\pi V_i)^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2V_i} (y_i - \mu_i)^2 \right\} \frac{1}{(2\pi\tau V_i)^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2\tau V_i} (\mu_i - m)^2 \right\} GI(s/2, S/2) d\mu_i dV_i \\
 &= \iint \underbrace{\frac{GI(s/2, S/2)}{2\pi V_i \tau^{\frac{1}{2}}}}_K \exp \left\{ -\frac{1}{2V_i} \left[y_i^2 - 2\mu_i y_i + \mu_i^2 + \frac{\mu_i^2}{\tau} - \frac{2\mu_i m}{\tau} + \frac{m^2}{\tau} \right] \right\} d\mu_i dV_i \\
 &= \iint \underbrace{K \exp \left\{ -\frac{1}{2V_i} \left(y_i^2 + \frac{m^2}{\tau} \right) \right\}}_Q \exp \left\{ -\frac{1}{2} \left(\frac{\tau+1}{\tau V_i} \right) \left[\mu_i^2 - 2\mu_i \left(\frac{\tau y_i + m}{\tau} \right) \left(\frac{\tau}{\tau+1} \right) \right] \right\} d\mu_i dV_i \\
 &= \iint K \cdot Q \exp \left\{ -\frac{1}{2} \left(\frac{\tau+1}{\tau V_i} \right) \left[\mu_i^2 - 2\mu_i \left(\frac{\tau y_i + m}{\tau+1} \right) \pm \left(\frac{\tau y_i + m}{\tau+1} \right)^2 \right] \right\} d\mu_i dV_i \\
 &= \int K \cdot Q \exp \left\{ -\frac{1}{2} \left(\frac{\tau+1}{\tau V_i} \right) \left[-\left(\frac{\tau y_i + m}{\tau+1} \right)^2 \right] \right\} \left(\frac{2\pi\tau V_i}{\tau+1} \right)^{\frac{1}{2}} \int \left[\frac{1}{2\pi} \left(\frac{\tau+1}{\tau V_i} \right) \right]^{\frac{1}{2}} \times \\
 &\times \exp \left\{ -\frac{1}{2} \left(\frac{\tau+1}{\tau V_i} \right) \left[\mu_i - \left(\frac{\tau y_i + m}{\tau+1} \right) \right]^2 \right\} d\mu_i dV_i \\
 &= \int K \left(\frac{2\pi\tau V_i}{\tau+1} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2V_i} \left(y_i^2 + \frac{m^2}{\tau} \right) - \frac{1}{2} \left(\frac{\tau+1}{\tau V_i} \right) \left[-\frac{(\tau y_i + m)^2}{(\tau+1)^2} \right] \right\} dV_i \\
 &= \int \frac{GI(s/2, S/2)}{[2\pi V_i (\tau+1)]^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \left[\frac{y_i^2}{V_i} + \frac{m^2}{\tau V_i} - \frac{(\tau y_i + m)^2}{\tau V_i (\tau+1)} \right] \right\} dV_i \\
 &= \int \frac{GI(s/2, S/2)}{[2\pi V_i (\tau+1)]^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \left[\frac{y_i^2}{V_i} - \frac{\tau y_i^2}{\tau V_i + V_i} - \frac{2y_i m}{\tau V_i + V_i} + \frac{m^2}{\tau V_i} - \frac{m^2}{\tau(\tau V_i + V_i)} \right] \right\} dV_i \\
 &= \int \frac{GI(s/2, S/2)}{[2\pi V_i (\tau+1)]^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \left[\frac{y_i^2}{\tau V_i + V_i} - \frac{2y_i m}{\tau V_i + V_i} + \frac{m^2}{\tau V_i + V_i} \right] \right\} dV_i \\
 &= \int \frac{1}{[2\pi V_i (\tau+1)]^{\frac{1}{2}}} \frac{\left(\frac{S}{2}\right)^{\frac{s}{2}}}{\Gamma\left(\frac{s}{2}\right)} \left(\frac{1}{V_i}\right)^{\frac{s}{2}+1} \exp \left\{ -\frac{S}{2V_i} \right\} \exp \left\{ -\frac{(y_i - m)^2}{2(\tau V_i + V_i)} \right\} dV_i \\
 &= \frac{1}{[2\pi(\tau+1)]^{\frac{1}{2}}} \frac{\left(\frac{S}{2}\right)^{\frac{s}{2}}}{\Gamma\left(\frac{s}{2}\right)} \frac{\Gamma\left(\frac{s+1}{2}\right)}{\left(\frac{S}{2} + \frac{(y_i - m)^2}{2(\tau+1)}\right)^{\frac{s+1}{2}}} \int \frac{\left(\frac{S}{2} + \frac{(y_i - m)^2}{2(\tau+1)}\right)^{\frac{s+1}{2}}}{\Gamma\left(\frac{s+1}{2}\right)} \left(\frac{1}{V_i}\right)^{\frac{s+1}{2}+1} \times \\
 &\times \exp \left\{ -\frac{1}{V_i} \left(\frac{S}{2} + \frac{(y_i - m)^2}{2(\tau+1)} \right) \right\} dV_i \times \frac{\left(\frac{S}{2}\right)^{\frac{1}{2}}}{\left(\frac{S}{2}\right)^{\frac{1}{2}}}
 \end{aligned}$$

$$\begin{aligned}
 & \int N(\mu_i, V_i) G_0(\theta_i) d\theta_i = \\
 &= \frac{\Gamma\left(\frac{s+1}{2}\right)}{\sqrt{\pi} \Gamma\left(\frac{s}{2}\right)} [(\tau+1)S]^{-\frac{1}{2}} \left[\frac{1}{1 + \frac{(y_i-m)^2}{S(\tau+1)}} \right]^{\frac{s+1}{2}} \\
 &= \frac{\Gamma\left(\frac{s+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{s}{2}\right)} \left[s(\tau+1) \frac{S}{s} \right]^{-\frac{1}{2}} \left[1 + \frac{(y_i-m)^2}{s(\tau+1) \frac{S}{s}} \right]^{-\frac{s+1}{2}} \\
 &= \frac{\Gamma\left(\frac{s+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{s}{2}\right)} [sM]^{-\frac{1}{2}} \left[1 + \frac{(y_i-m)^2}{sM} \right]^{-\frac{s+1}{2}},
 \end{aligned}$$

que é a densidade da distribuição t-Student com parâmetros s , m e $M^{1/2}$.

Portanto, tem-se que:

$$P(Y_{n+1}|\boldsymbol{\theta}) = \frac{\alpha}{\alpha+n} T_s(m, M) + \frac{1}{\alpha+n} \sum_{i=1}^n N(\mu_i, V_i);$$

concluindo a prova.

- Cálculo da distribuição condicional *a posteriori* de $\theta_j^* = (\mu_j^*, V_j^*)$, $j = 1, \dots, k$ dada pela equação (4.10).

$$\begin{aligned}
 P(\mu_j^*, V_j^* | \mathbf{y}_{(j)}) &\propto \prod_{i:c_i=j} P(y_i | \mu_j^*, V_j^*) P(\mu_j^* | V_j^*) P(V_j^*) \\
 &\propto \left(\frac{1}{V_j^*} \right)^{\frac{n_j}{2}} \exp \left\{ -\frac{1}{2V_j^*} \sum_{i=1}^{n_j} (y_i - \mu_j^*)^2 \right\} \left(\frac{1}{V_j^*} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2\tau V_j^*} (\mu_j^* - m)^2 \right\} \left(\frac{1}{V_j^*} \right)^{\frac{s}{2}+1} \exp \left\{ -\frac{S}{2V_j^*} \right\} \\
 &\propto \left(\frac{1}{V_j^*} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2V_j^*} \left[\sum_{i=1}^{n_j} (y_i - \bar{y}_{(j)})^2 + n_j (\mu_j^* - \bar{y}_{(j)})^2 \right] \right\} \exp \left\{ -\frac{1}{2\tau V_j^*} (\mu_j^* - m)^2 \right\} \times \\
 &\times \left(\frac{1}{V_j^*} \right)^{\frac{n_j+s}{2}+1} \exp \left\{ -\frac{S}{2V_j^*} \right\} \\
 &\propto \left(\frac{1}{V_j^*} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2V_j^*} \left[n_j (\mu_j^* - \bar{y}_{(j)})^2 + a (\mu_j^* - m)^2 \right] \right\} \left(\frac{1}{V_j^*} \right)^{\frac{n_j+s}{2}+1} \exp \left\{ -\frac{1}{2V_j^*} \left[\sum_{i=1}^{n_j} (y_i - \bar{y}_{(j)})^2 + S \right] \right\}
 \end{aligned}$$

$$\begin{aligned}
 & P(\mu_j^*, V_j^* | \mathbf{y}_{(j)}) \propto \\
 & \propto \left(\frac{1}{V_j^*} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2V_j^*} \left[n_j(\mu_j^{*2} - 2\bar{y}_{(j)}\mu_j^*) + a(\mu_j^{*2} - 2m\mu_j^*) \right] \right\} \exp \left\{ -\frac{1}{2V_j^*} \left[n_j\bar{y}_{(j)}^2 + am^2 \right] \right\} \times \\
 & \times \left(\frac{1}{V_j^*} \right)^{\frac{n_j+s}{2}+1} \exp \left\{ -\frac{1}{2V_j^*} \left[\sum_{i=1}^{n_j} (y_i - \bar{y}_{(j)})^2 + S \right] \right\} \\
 & \propto \left(\frac{1}{V_j^*} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2V_j^*} \left[(a + n_j)\mu_j^{*2} - 2\mu_j^*(am + n_j\bar{y}_{(j)}) \right] \right\} \exp \left\{ -\frac{1}{2V_j^*} \left[n_j\bar{y}_{(j)}^2 + am^2 \right] \right\} \times \\
 & \times \left(\frac{1}{V_j^*} \right)^{\frac{n_j+s}{2}+1} \exp \left\{ -\frac{1}{2V_j^*} \left[\sum_{i=1}^{n_j} (y_i - \bar{y}_{(j)})^2 + S \right] \right\} \\
 & \propto \left(\frac{1}{V_j^*} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left(\frac{a + n_j}{V_j^*} \right) \left[\mu_j^{*2} - 2\mu_j^* \left(\frac{am + n_j\bar{y}_{(j)}}{a + n_j} \right) \pm \left(\frac{am + n_j\bar{y}_{(j)}}{a + n_j} \right)^2 \right] \right\} \left(\frac{1}{V_j^*} \right)^{\frac{n_j+s}{2}+1} \times \\
 & \times \exp \left\{ -\frac{1}{2V_j^*} \left[n_j\bar{y}_{(j)}^2 + am^2 \right] \right\} \exp \left\{ -\frac{1}{2V_j^*} \left[\sum_{i=1}^{n_j} (y_i - \bar{y}_{(j)})^2 + S \right] \right\} \\
 & \propto \left(\frac{1}{V_j^*} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left(\frac{a + n_j}{V_j^*} \right) \left[\mu_j^* - \left(\frac{am + n_j\bar{y}_{(j)}}{a + n_j} \right) \right]^2 \right\} \left(\frac{1}{V_j^*} \right)^{\frac{n_j+s}{2}+1} \times \\
 & \times \exp \left\{ -\frac{1}{2V_j^*} \left[n_j\bar{y}_{(j)}^2 + am^2 - \frac{(am + n_j\bar{y}_{(j)})^2}{a + n_j} \right] \right\} \exp \left\{ -\frac{1}{2V_j^*} \left[\sum_{i=1}^{n_j} (y_i - \bar{y}_{(j)})^2 + S \right] \right\} \\
 & \propto \left(\frac{a + n_j}{V_j^*} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left(\frac{a + n_j}{V_j^*} \right) \left[\mu_j^* - \left(\frac{am + n_j\bar{y}_{(j)}}{a + n_j} \right) \right]^2 \right\} \left(\frac{1}{V_j^*} \right)^{\frac{n_j+s}{2}+1} \times \\
 & \times \exp \left\{ -\frac{1}{2V_j^*} \left[\frac{an_j}{a + n_j} (\bar{y}_{(j)} - m)^2 \right] \right\} \exp \left\{ -\frac{1}{2V_j^*} \left[\sum_{i=1}^{n_j} (y_i - \bar{y}_{(j)})^2 + S \right] \right\} \\
 & \propto N \left[\frac{am + n_j\bar{y}_{(j)}}{a + n_j}; \frac{V_j^*}{a + n_j} \right] \left(\frac{1}{V_j^*} \right)^{\frac{n_j+s}{2}+1} \exp \left\{ -\frac{1}{2V_j^*} \left[\frac{an_j}{a + n_j} (\bar{y}_{(j)} - m)^2 + \sum_{i=1}^{n_j} (y_i - \bar{y}_{(j)})^2 + S \right] \right\} \\
 & \propto N \left[\frac{am + n_j\bar{y}_{(j)}}{a + n_j}; \frac{V_j^*}{a + n_j} \right] GI \left(\frac{n_j + s}{2}; \frac{1}{2} \left[S + \sum_{i=1}^{n_j} (y_i - \bar{y}_{(j)})^2 + \frac{an_j}{2(a + n_j)} (\bar{y}_{(j)} - m)^2 \right] \right).
 \end{aligned}$$

Portanto, tem-se que:

$$P(\mu_j^*, V_j^* | \mathbf{y}_{(j)}) \propto P(\mu_j^* | V_j^*, \mathbf{y}_{(j)}) P(V_j^* | \mathbf{y}_{(j)}),$$

onde

$$\begin{cases} V_j^{*-1} | \mathbf{y}_{(j)} \sim G \left(\frac{n_j + s}{2}; \frac{1}{2} \left[S + \sum_{i=1}^{n_j} (y_i - \bar{y}_{(j)})^2 + \frac{an_j}{2(a+n_j)} (\bar{y}_{(j)} - m)^2 \right] \right); \\ (\mu_j^* | V_j^*, \mathbf{y}_{(j)}) \sim N \left(\frac{am + n_j \bar{y}_{(j)}}{a + n_j}; \frac{V_j^*}{a + n_j} \right). \end{cases}$$

- Cálculo da distribuição condicional completa *a posteriori* de μ_j^* , $j = 1, \dots, k$, dada pela equação (4.13).

$$\begin{aligned} P(\mu_j^* | V_j^*, \lambda_j^*, \mathbf{y}_{(j)}) &\propto \prod_{i:c_i=j} P(y_i | \mu_j^*, V_j^*, \lambda_j^*) P(\mu_j^* | V_j^*) \\ &\propto \phi_{n_j} \left(\frac{\mathbf{y}_{(j)} - \mu_j^* \mathbf{1}_{n_j}}{V_j^{*\frac{1}{2}}} \right) \Phi_{n_j} \left(\lambda_j^* \left(\frac{\mathbf{y}_{(j)} - \mu_j^* \mathbf{1}_{n_j}}{V_j^{*\frac{1}{2}}} \right) \right) \phi(\mu_j^*; m, \tau V_j^*) \\ &\propto \exp \left\{ -\frac{1}{2V_j^*} \|\mathbf{y}_{(j)} - \mu_j^* \mathbf{1}_{n_j}\|^2 \right\} \exp \left\{ -\frac{1}{2\tau V_j^*} (\mu_j^* - m)^2 \right\} \Phi_{n_j} \left(\lambda_j^* \left(\frac{\mathbf{y}_{(j)} - \mu_j^* \mathbf{1}_{n_j}}{V_j^{*\frac{1}{2}}} \right) \right) \\ &\propto \exp \left\{ -\frac{1}{2V_j^*} \sum_{i=1}^{n_j} (y_i - \mu_j^*)^2 \right\} \exp \left\{ -\frac{1}{2\tau V_j^*} (\mu_j^* - m)^2 \right\} \Phi_{n_j} \left(\lambda_j^* \left(\frac{\mathbf{y}_{(j)} - \mu_j^* \mathbf{1}_{n_j}}{V_j^{*\frac{1}{2}}} \right) \right) \\ &\propto \exp \left\{ -\frac{1}{2V_j^*} \left[\sum_{i=1}^{n_j} (y_i - \bar{y}_{(j)})^2 + n_j (\mu_j^* - \bar{y}_{(j)})^2 \right] \right\} \exp \left\{ -\frac{1}{2\tau V_j^*} (\mu_j^* - m)^2 \right\} \Phi_{n_j} \left(\lambda_j^* \left(\frac{\mathbf{y}_{(j)} - \mu_j^* \mathbf{1}_{n_j}}{V_j^{*\frac{1}{2}}} \right) \right) \\ &\propto \exp \left\{ -\frac{1}{2V_j^*} \left[n_j (\mu_j^{*2} - 2\bar{y}_{(j)} \mu_j^*) + a (\mu_j^{*2} - 2m \mu_j^*) \right] \right\} \Phi_{n_j} \left(\lambda_j^* \left(\frac{\mathbf{y}_{(j)} - \mu_j^* \mathbf{1}_{n_j}}{V_j^{*\frac{1}{2}}} \right) \right) \\ &\propto \exp \left\{ -\frac{1}{2V_j^*} \left[(a + n_j) \mu_j^{*2} - 2\mu_j^* (am + n_j \bar{y}_{(j)}) \right] \right\} \Phi_{n_j} \left(\lambda_j^* \left(\frac{\mathbf{y}_{(j)} - \mu_j^* \mathbf{1}_{n_j}}{V_j^{*\frac{1}{2}}} \right) \right) \\ &\propto \exp \left\{ -\frac{1}{2} \left(\frac{a + n_j}{V_j^*} \right) \left[\mu_j^{*2} - 2\mu_j^* \left(\frac{am + n_j \bar{y}_{(j)}}{a + n_j} \right) \pm \left(\frac{am + n_j \bar{y}_{(j)}}{a + n_j} \right)^2 \right] \right\} \Phi_{n_j} \left(\lambda_j^* \left(\frac{\mathbf{y}_{(j)} - \mu_j^* \mathbf{1}_{n_j}}{V_j^{*\frac{1}{2}}} \right) \right) \\ &\propto \exp \left\{ -\frac{1}{2} \left(\frac{a + n_j}{V_j^*} \right) \left[\mu_j^* - \left(\frac{am + n_j \bar{y}_{(j)}}{a + n_j} \right) \right]^2 \right\} \Phi_{n_j} \left(\lambda_j^* \left(\frac{\mathbf{y}_{(j)} - \mu_j^* \mathbf{1}_{n_j}}{V_j^{*\frac{1}{2}}} \right) \right) \end{aligned}$$

$$\begin{aligned}
 & P(\mu_j^* | V_j^*, \lambda_j^*, \mathbf{y}_{(j)}) \propto \\
 & \propto \phi \left(\mu_j^*; \underbrace{\frac{am + n_j \bar{y}_{(j)}}{a + n_j}}_{M^*}, \underbrace{\frac{V_j^*}{a + n_j}}_{V^*} \right) \Phi_{n_j} \left(\mu_j^* \left(\underbrace{-\frac{\lambda_j^* \mathbf{1}_{n_j}}{V_j^{*\frac{1}{2}}}}_{\boldsymbol{\lambda}^*} + \underbrace{\frac{\lambda_j^* \mathbf{y}_{(j)}}{V_j^{*\frac{1}{2}}}}_{\mathbf{y}^*} \right) \right).
 \end{aligned}$$

Portanto, tem-se que:

$$P(\mu_j^* | V_j^*, \lambda_j^*, \mathbf{y}_{(j)}) \propto \phi(\mu_j^*; M^*, V^*) \Phi_{n_j}(\mu_j^* \boldsymbol{\lambda}^* + \mathbf{y}^*). \quad (\text{A.2})$$

- Cálculo da distribuição condicional completa *a posteriori* de V_j^* , $j = 1, \dots, k$, dada pela equação (4.14).

$$\begin{aligned}
 & P(V_j^* | \mu_j^*, \lambda_j^*, \mathbf{y}_{(j)}) \propto \prod_{i:c_i=j} P(y_i | \mu_j^*, V_j^*, \lambda_j^*) P(\mu_j^* | V_j^*) P(V_j^*) \\
 & \propto \left(\frac{1}{V_j^*} \right)^{\frac{n_j}{2}} \phi_{n_j} \left(\frac{\mathbf{y}_{(j)} - \mu_j^* \mathbf{1}_{n_j}}{V_j^{*\frac{1}{2}}} \right) \Phi_{n_j} \left(\lambda_j^* \left(\frac{\mathbf{y}_{(j)} - \mu_j^* \mathbf{1}_{n_j}}{V_j^{*\frac{1}{2}}} \right) \right) \left(\frac{1}{V_j^*} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2\tau V_j^*} (\mu_j^* - m)^2 \right\} \times \\
 & \times \left(\frac{1}{V_j^*} \right)^{\frac{s}{2}+1} \exp \left\{ -\frac{S}{2V_j^*} \right\} \\
 & \propto \left(\frac{1}{V_j^*} \right)^{\frac{n_j+s+1}{2}+1} \exp \left\{ -\frac{1}{2V_j^*} \sum_{i=1}^{n_j} (y_i - \mu_j^*)^2 \right\} \exp \left\{ -\frac{1}{2\tau V_j^*} (\mu_j^* - m)^2 \right\} \exp \left\{ -\frac{S}{2V_j^*} \right\} \times \\
 & \times \Phi_{n_j} \left(\lambda_j^* \left(\frac{\mathbf{y}_{(j)} - \mu_j^* \mathbf{1}_{n_j}}{V_j^{*\frac{1}{2}}} \right) \right) \\
 & \propto \left(\frac{1}{V_j^*} \right)^{\frac{n_j+s+1}{2}+1} \exp \left\{ -\frac{1}{2V_j^*} \left[\sum_{i=1}^{n_j} (y_i - \bar{y}_{(j)})^2 + n_j (\mu_j^* - \bar{y}_{(j)})^2 \right] \right\} \exp \left\{ -\frac{1}{2\tau V_j^*} (\mu_j^* - m)^2 \right\} \times \\
 & \times \exp \left\{ -\frac{S}{2V_j^*} \right\} \Phi_{n_j} \left(\lambda_j^* \left(\frac{\mathbf{y}_{(j)} - \mu_j^* \mathbf{1}_{n_j}}{V_j^{*\frac{1}{2}}} \right) \right)
 \end{aligned}$$

$$\begin{aligned}
 & P(V_j^* | \mu_j^*, \lambda_j^*, \mathbf{y}_{(j)}) \propto \\
 & \propto \left(\frac{1}{V_j^*} \right)^{\frac{n_j + s + 1}{2} + 1} \exp \left\{ -\frac{1}{2V_j^*} \left[S + a(\mu_j^* - m)^2 + \sum_{i=1}^{n_j} (y_i - \bar{y}_{(j)})^2 + n_j(\mu_j^* - \bar{y}_{(j)})^2 \right] \right\} \times \\
 & \times \Phi_{n_j} \left(\lambda_j^* \underbrace{\left(\frac{\mathbf{y}_{(j)} - \mu_j^* \mathbf{1}_{n_j}}{V_j^{*\frac{1}{2}}} \right)}_{\mathbf{Z}} \right).
 \end{aligned}$$

Portanto, tem-se que:

$$P(V_j^* | \mu_j^*, \lambda_j^*, \mathbf{y}_{(j)}) \propto GI \left(\frac{n_j + s + 1}{2}; \frac{1}{2} \left[S + a(\mu_j^* - m)^2 + \sum_{i=1}^{n_j} (y_i - \bar{y}_{(j)})^2 + n_j(\mu_j^* - \bar{y}_{(j)})^2 \right] \right) \Phi_{n_j}(\lambda_j^* \mathbf{Z}).$$

- Cálculo da distribuição condicional completa *a posteriori* de λ_j^* , $j = 1, \dots, k$, dada pela equação (4.15).

$$\begin{aligned}
 P(\lambda_j^* | \mu_j^*, V_j^*, \mathbf{y}_{(j)}) & \propto \prod_{i:c_i=j} P(y_i | \mu_j^*, V_j^*, \lambda_j^*) P(\lambda_j^*) \\
 & \propto \Phi_{n_j} \left(\lambda_j^* \left(\frac{\mathbf{y}_{(j)} - \mu_j^* \mathbf{1}_{n_j}}{V_j^{*\frac{1}{2}}} \right) \right) \phi(\lambda_j^*; \varepsilon, z^2).
 \end{aligned}$$

Portanto, tem-se que:

$$P(\lambda_j^* | \mu_j^*, V_j^*, \mathbf{y}_{(j)}) \propto \phi(\lambda_j^*; \varepsilon, z^2) \Phi_{n_j}(\lambda_j^* \mathbf{Z}). \tag{A.3}$$

Apêndice B

Determinação das Representações Estocásticas

Neste apêndice, apresentamos as demonstrações das proposições 2 e 3, que descrevem as representações estocásticas de $\mu_j^*|V_j^*, \lambda_j^*, \mathbf{y}_{(j)}$ e de $\lambda_j^*|\mu_j^*, V_j^*, \mathbf{y}_{(j)}$, respectivamente. Mas antes, apresentamos alguns resultados importantes para a determinação dessas representações.

Resultado 1: Seja $X \sim N(\mu, \sigma^2)$. Segue de alguns resultados de Probabilidade que a mgf de X é dada por

$$M_X(t) = \exp \left\{ \mu t + \frac{t^2 \sigma^2}{2} \right\}. \quad (\text{B.1})$$

Resultado 2: Seja $U^L \sim LTN_n(\mathbf{L}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$. Tem-se que a mgf de U^L é dada por

$$M_{U^L}(t) = \exp \left\{ \mathbf{t}^T \boldsymbol{\mu} + \frac{1}{2} \mathbf{t}^T \boldsymbol{\Sigma} \mathbf{t} \right\} \frac{\Phi_n(\boldsymbol{\mu} + \boldsymbol{\Sigma} \mathbf{t}; \mathbf{L}, \boldsymbol{\Sigma})}{\Phi_n(\boldsymbol{\mu}; \mathbf{L}, \boldsymbol{\Sigma})}. \quad (\text{B.2})$$

Resultado 3: (Lemma) Se \mathbf{X} é um vetor não singular e $\mathbf{X} \sim N_n(\mathbf{d}, \mathbf{D})$, então para qualquer vetor fixado \mathbf{a} ($m \times 1$) e qualquer matriz \mathbf{A} ($m \times n$), temos que

$$\begin{aligned} E[\Phi_m(\mathbf{A}\mathbf{X} + \mathbf{a}; \mathbf{b}, \mathbf{B})] &= \int \Phi_m(\mathbf{A}\mathbf{X} + \mathbf{a}; \mathbf{b}, \mathbf{B}) \frac{1}{\mathbf{D}^{\frac{1}{2}}} \phi_n(\mathbf{X}; \mathbf{d}, \mathbf{D}) \mathbf{X} \\ &= \Phi_m(\mathbf{A}\mathbf{d} + \mathbf{a}; \mathbf{b}, \mathbf{B} + \mathbf{A}\mathbf{D}\mathbf{A}^T). \end{aligned} \quad (\text{B.3})$$

Resultado 4: Seja

$$\begin{cases} Y_i | \mu_j^*, V_j^*, \lambda_j^* \stackrel{iid}{\sim} SN(\mu_j^*, V_j^*, \lambda_j^*), j = 1, \dots, k, i = 1, \dots, n_j; \\ \mu_j^* | V_j^* \sim N(m, \tau V_j^*); \\ V_j^{*-1} \sim Gama(s/2, S/2); \\ \lambda_j^* \sim N(\varepsilon, z^2), \end{cases}$$

onde n_j é o número de observações pertencentes ao j -ésimo *cluster*. Considere, *a priori*, que μ_j^* e V_j^* são independentes de λ_j^* .

Dessa forma, a distribuição condicional completa *a posteriori* de μ_j^* é dada por

$$P(\mu_j^* | V_j^*, \lambda_j^*, \mathbf{y}_{(j)}) = \frac{\left(\frac{1}{V^*}\right)^{\frac{1}{2}} \phi(\mu_j^*; M^*, V^*) \Phi_{n_j}(\mu_j^* \boldsymbol{\lambda}^* + \mathbf{y}^*)}{\Phi_{n_j}(M^* \boldsymbol{\lambda}^* + \mathbf{y}^*; \mathbf{0}, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T})}, \quad (\text{B.4})$$

onde $\mathbf{y}_{(j)} = (y_1, \dots, y_{n_j})$, $M^* = (am + n_j \bar{y}_{(j)}) / (a + n_j)$, $V^* = (V_j^*) / (a + n_j)$, $\boldsymbol{\lambda}^* = -(\lambda_j^* \mathbf{1}_{n_j}) / (\mathbf{V}_j^*)^{\frac{1}{2}}$ e $\mathbf{y}^* = (\lambda_j^* \mathbf{y}_{(j)}) / (V_j^*)^{\frac{1}{2}}$.

Demonstração:

$$\begin{aligned} P(\mu_j^* | V_j^*, \lambda_j^*, \mathbf{y}_{(j)}) &= \frac{\prod_{i:c_i=j} P(y_i | \mu_j^*, V_j^*, \lambda_j^*) P(\mu_j^* | V_j^*)}{\int \prod_{i:c_i=j} P(y_i | \mu_j^*, V_j^*, \lambda_j^*) P(\mu_j^* | V_j^*) d\mu_j^*} \\ &= \frac{\phi(\mu_j^*; M^*, V^*) \Phi_{n_j}(\mu_j^* \boldsymbol{\lambda}^* + \mathbf{y}^*)}{\int \phi(\mu_j^*; M^*, V^*) \Phi_{n_j}(\mu_j^* \boldsymbol{\lambda}^* + \mathbf{y}^*) d\mu_j^*} \times \frac{\left(\frac{1}{V^*}\right)^{\frac{1}{2}}}{\left(\frac{1}{V^*}\right)^{\frac{1}{2}}} \quad (\text{aplicando o resultado obtido em A.2}) \\ &= \frac{\left(\frac{1}{V^*}\right)^{\frac{1}{2}} \phi(\mu_j^*; M^*, V^*) \Phi_{n_j}(\mu_j^* \boldsymbol{\lambda}^* + \mathbf{y}^*)}{\Phi_{n_j}(\mu_j^* \boldsymbol{\lambda}^* + \mathbf{y}^*; \mathbf{0}, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T})} \quad (\text{aplicando o Resultado 3}). \end{aligned}$$

Já a distribuição condicional completa *a posteriori* de λ_j^* é dada por

$$P(\lambda_j^* | \mu_j^*, V_j^*, \mathbf{y}_{(j)}) = \frac{\frac{1}{z} \phi(\lambda_j^*; \varepsilon, z^2) \Phi_{n_j}(\lambda_j^* \mathbf{Z})}{\Phi_{n_j}(\varepsilon \mathbf{Z}; \mathbf{0}, \mathbf{I}_{n_j} + z^2 \mathbf{Z}^T \mathbf{Z})}, \quad (\text{B.5})$$

onde $\mathbf{Z} = (\mathbf{y}_{(j)} - \mu_j^* \mathbf{1}_{n_j}) / (\mathbf{V}_j^*)^{\frac{1}{2}}$.

Demonstração:

$$\begin{aligned} P(\lambda_j^* | \mu_j^*, V_j^*, \mathbf{y}_{(j)}) &= \frac{\prod_{i:c_i=j} P(y_i | \mu_j^*, V_j^*, \lambda_j^*) P(\lambda_j^*)}{\int \prod_{i:c_i=j} P(y_i | \mu_j^*, V_j^*, \lambda_j^*) P(\lambda_j^*) d\lambda_j^*} \\ &= \frac{\phi(\lambda_j^*; \varepsilon, z^2) \Phi_{n_j}(\lambda_j^* \mathbf{Z})}{\int \phi(\lambda_j^*; \varepsilon, z^2) \Phi_{n_j}(\lambda_j^* \mathbf{Z}) d\lambda_j^*} \times \frac{\frac{1}{z}}{\frac{1}{z}} \quad (\text{aplicando o resultado obtido em A.3}) \\ &= \frac{\frac{1}{z} \phi(\lambda_j^*; \varepsilon, z^2) \Phi_{n_j}(\lambda_j^* \mathbf{Z})}{\Phi_{n_j}(\varepsilon \mathbf{Z}; \mathbf{0}, \mathbf{I}_{n_j} + z^2 \mathbf{Z}^T \mathbf{Z})} \quad (\text{aplicando o Resultado 3}). \end{aligned}$$

- Representação estocástica da distribuição condicional completa *a posteriori* de μ_j^* , $j = 1, \dots, k$.

A fim de provar que a representação estocástica de $\mu_j^* | V_j^*, \lambda_j^*, \mathbf{y}_{(j)} = \mu_y$ é dada pela equação (), iremos mostrar que

$$M_{\mu_y}(t) = M_W(t) \times M_{U^0} \left(\frac{V^* \boldsymbol{\lambda}^* t}{1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*} \right).$$

Cálculo da mgf de μ_y .

$$\begin{aligned} M_{\mu_y}(t) &= \int_{-\infty}^{\infty} \exp\{\mu_j^* t\} P(\mu_j^* | V_j^*, \lambda_j^*, \mathbf{y}_{(j)}) d\mu_j^* \\ &= \int_{-\infty}^{\infty} \exp\{\mu_j^* t\} \frac{\left(\frac{1}{V^*}\right)^{\frac{1}{2}} \phi(\mu_j^*; M^*, V^*) \Phi_{n_j}(\mu_j^* \boldsymbol{\lambda}^* + \mathbf{y}^*)}{\Phi_{n_j}(M^* \boldsymbol{\lambda}^* + \mathbf{y}^*; \mathbf{0}, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T})} d\mu_j^* \quad (\text{aplicando a equação B.5}) \\ &= \int_{-\infty}^{\infty} \left(\frac{1}{V^*}\right)^{\frac{1}{2}} \exp\left\{\mu_j^* t - \frac{1}{2V^*}(\mu_j^* - M^*)^2\right\} \frac{\Phi_{n_j}(\mu_j^* \boldsymbol{\lambda}^* + \mathbf{y}^*)}{\Phi_{n_j}(M^* \boldsymbol{\lambda}^* + \mathbf{y}^*; \mathbf{0}, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T})} d\mu_j^* \\ &= \int_{-\infty}^{\infty} \left(\frac{1}{V^*}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2V^*} \left[\mu_j^{*2} - 2\mu_j^*(V^*t + M^*) + M^{*2}\right]\right\} \frac{\Phi_{n_j}(\mu_j^* \boldsymbol{\lambda}^* + \mathbf{y}^*)}{\Phi_{n_j}(M^* \boldsymbol{\lambda}^* + \mathbf{y}^*; \mathbf{0}, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T})} d\mu_j^* \\ &= \int_{-\infty}^{\infty} \left(\frac{1}{V^*}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2V^*} [\mu_j^* - (V^*t + M^*)]^2\right\} \exp\left\{-\frac{1}{2V^*} [M^{*2} - (V^*t + M^*)^2]\right\} \times \\ &\quad \times \frac{\Phi_{n_j}(\mu_j^* \boldsymbol{\lambda}^* + \mathbf{y}^*)}{\Phi_{n_j}(M^* \boldsymbol{\lambda}^* + \mathbf{y}^*; \mathbf{0}, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T})} d\mu_j^* \\ &= \int_{-\infty}^{\infty} \left(\frac{1}{V^*}\right)^{\frac{1}{2}} \exp\left\{tM^* + \frac{t^2V^*}{2}\right\} \frac{\phi(\mu_j^*; M^* + V^*t, V^*) \Phi_{n_j}(\mu_j^* \boldsymbol{\lambda}^* + \mathbf{y}^*)}{\Phi_{n_j}(M^* \boldsymbol{\lambda}^* + \mathbf{y}^*; \mathbf{0}, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T})} d\mu_j^* \\ &= \exp\left\{tM^* + \frac{t^2V^*}{2}\right\} \frac{\Phi_{n_j}(M^* \boldsymbol{\lambda}^* + \mathbf{y}^* + V^* \boldsymbol{\lambda}^* t; \mathbf{0}, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T})}{\Phi_{n_j}(M^* \boldsymbol{\lambda}^* + \mathbf{y}^*; \mathbf{0}, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T})} \quad (\text{aplicando o Resultado 3}). \end{aligned}$$

Uma vez que $W \sim N(M^*, V^*/(1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*))$, tem-se, pelo Resultado 1, que a mgf de W é dada por

$$M_W(t) = \exp\left\{tM^* + \frac{t^2V^*}{2(1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*)}\right\}.$$

Já a mgf de U^0 é calculada da seguinte maneira

$$\begin{aligned}
 M_{U^0} \left(\frac{V^* \boldsymbol{\lambda}^* t}{1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*} \right) &= \exp \left\{ \frac{1}{2} \left(\frac{V^* \boldsymbol{\lambda}^* t}{1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*} \right)^T (\mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T}) \left(\frac{V^* \boldsymbol{\lambda}^* t}{1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*} \right) \right\} \times \\
 &\times \frac{\Phi_{n_j} \left((\mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T}) \left(\frac{V^* \boldsymbol{\lambda}^* t}{1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*} \right); -M^* \boldsymbol{\lambda}^* - \mathbf{y}^*, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T} \right)}{\Phi_{n_j}(\mathbf{0}; -M^* \boldsymbol{\lambda}^* - \mathbf{y}^*, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T})} \\
 &= \exp \left\{ \frac{1}{2} \left[\left(\frac{V^* t}{1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*} \right)^2 \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^* + \left(\frac{V^* t}{1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*} \right)^2 V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^* \right] \right\} \times \\
 &\times \frac{\Phi_{n_j} \left(\frac{V^* \boldsymbol{\lambda}^* t}{1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*} + \frac{V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T} V^* \boldsymbol{\lambda}^* t}{1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*}; -M^* \boldsymbol{\lambda}^* - \mathbf{y}^*, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T} \right)}{\Phi_{n_j}(\mathbf{0}; -M^* \boldsymbol{\lambda}^* - \mathbf{y}^*, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T})} \\
 &= \exp \left\{ \frac{1}{2} \left[\frac{(V^* t)^2}{(1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*)^2} \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^* (1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*) \right] \right\} \times \\
 &\times \frac{\Phi_{n_j} \left(\frac{V^* \boldsymbol{\lambda}^* t}{1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*} (1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*); -M^* \boldsymbol{\lambda}^* - \mathbf{y}^*, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T} \right)}{\Phi_{n_j}(\mathbf{0}; -M^* \boldsymbol{\lambda}^* - \mathbf{y}^*, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T})} \\
 &= \exp \left\{ \frac{1}{2} \left(\frac{V^{*2} \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^* t^2}{1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*} \right) \right\} \frac{\Phi_{n_j} \left(M^* \boldsymbol{\lambda}^* + \mathbf{y}^* + V^* \boldsymbol{\lambda}^* t; \mathbf{0}, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T} \right)}{\Phi_{n_j}(M^* \boldsymbol{\lambda}^* + \mathbf{y}^*; \mathbf{0}, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T})}.
 \end{aligned}$$

Assim, tem-se que

$$\begin{aligned}
 M_W(t) \times M_{U^0} \left(\frac{V^* \boldsymbol{\lambda}^* t}{1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*} \right) &= \\
 &= \exp \left\{ tM^* + \frac{1}{2} \left[\frac{t^2 V^* + V^{*2} \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^* t^2}{1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*} \right] \right\} \frac{\Phi_{n_j} \left(M^* \boldsymbol{\lambda}^* + \mathbf{y}^* + V^* \boldsymbol{\lambda}^* t; \mathbf{0}, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T} \right)}{\Phi_{n_j}(M^* \boldsymbol{\lambda}^* + \mathbf{y}^*; \mathbf{0}, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T})} \\
 &= \exp \left\{ tM^* + \frac{1}{2} \left[\frac{t^2 V^* (1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*)}{1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*} \right] \right\} \frac{\Phi_{n_j} \left(M^* \boldsymbol{\lambda}^* + \mathbf{y}^* + V^* \boldsymbol{\lambda}^* t; \mathbf{0}, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T} \right)}{\Phi_{n_j}(M^* \boldsymbol{\lambda}^* + \mathbf{y}^*; \mathbf{0}, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T})} \\
 &= \exp \left\{ tM^* + \frac{t^2 V^*}{2} \right\} \frac{\Phi_{n_j} \left(M^* \boldsymbol{\lambda}^* + \mathbf{y}^* + V^* \boldsymbol{\lambda}^* t; \mathbf{0}, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T} \right)}{\Phi_{n_j}(M^* \boldsymbol{\lambda}^* + \mathbf{y}^*; \mathbf{0}, \mathbf{I}_{n_j} + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}^{*T})} \\
 &= M_{\mu_y}(t).
 \end{aligned}$$

Portanto,

$$\mu_y \stackrel{d}{=} W + \frac{V^* \boldsymbol{\lambda}^{*T} U^0}{1 + V^* \boldsymbol{\lambda}^{*T} \boldsymbol{\lambda}^*}.$$

- Representação estocástica da distribuição condicional completa *a posteriori* de λ_j^* , $j = 1, \dots, k$.

A fim de provar que a representação estocástica de $\lambda_j^* | \mu_j^*, V_j^*, \mathbf{y}_{(j)} = \lambda_y$ é dada pela equação (), iremos mostrar que

$$M_{\lambda_y}(t) = M_W(t) \times M_{U^0} \left(\frac{z^2 \mathbf{Z} t}{1 + z^2 \mathbf{Z}^T \mathbf{Z}} \right).$$

Cálculo da mgf de λ_y .

$$\begin{aligned} M_{\lambda_y}(t) &= \int_{-\infty}^{\infty} \exp\{\lambda_j^* t\} P(\lambda_j^* | \mu_j^*, V_j^*, \mathbf{y}_{(j)}) d\lambda_j^* \\ &= \int_{-\infty}^{\infty} \exp\{\lambda_j^* t\} \frac{\frac{1}{z} \phi(\lambda_j^*; \varepsilon, z^2) \Phi_{n_j}(\lambda_j^* \mathbf{Z})}{\Phi_{n_j}(\varepsilon \mathbf{Z}; \mathbf{0}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T)} d\lambda_j^* \quad (\text{aplicando a equação B.5}) \\ &= \int_{-\infty}^{\infty} \frac{1}{z} \exp\{\lambda_j^* t - \frac{1}{2z^2} (\lambda_j^* - \varepsilon)^2\} \frac{\Phi_{n_j}(\lambda_j^* \mathbf{Z})}{\Phi_{n_j}(\varepsilon \mathbf{Z}; \mathbf{0}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T)} d\lambda_j^* \\ &= \int_{-\infty}^{\infty} \frac{1}{z} \exp\left\{-\frac{1}{2z^2} [\lambda_j^{*2} - 2\lambda_j^*(z^2 t + \varepsilon) + \varepsilon^2]\right\} \frac{\Phi_{n_j}(\lambda_j^* \mathbf{Z})}{\Phi_{n_j}(\varepsilon \mathbf{Z}; \mathbf{0}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T)} d\lambda_j^* \\ &= \int_{-\infty}^{\infty} \frac{1}{z} \exp\left\{-\frac{1}{2z^2} [\lambda_j^* - (z^2 t + \varepsilon)]^2\right\} \exp\left\{-\frac{1}{2z^2} [\varepsilon^2 - (z^2 t + \varepsilon)^2]\right\} \frac{\Phi_{n_j}(\lambda_j^* \mathbf{Z})}{\Phi_{n_j}(\varepsilon \mathbf{Z}; \mathbf{0}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T)} d\lambda_j^* \\ &= \int_{-\infty}^{\infty} \frac{1}{z} \exp\left\{t\varepsilon + \frac{t^2 z^2}{2}\right\} \frac{\phi(\lambda_j^*; \varepsilon + z^2 t, z^2) \Phi_{n_j}(\lambda_j^* \mathbf{Z})}{\Phi_{n_j}(\varepsilon \mathbf{Z}; \mathbf{0}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T)} d\lambda_j^* \\ &= \exp\left\{t\varepsilon + \frac{t^2 z^2}{2}\right\} \frac{\Phi_{n_j}(\varepsilon \mathbf{Z} + z^2 t \mathbf{Z}; \mathbf{0}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T)}{\Phi_{n_j}(\varepsilon \mathbf{Z}; \mathbf{0}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T)} \quad (\text{aplicando o Resultado 3}). \end{aligned}$$

Uma vez que $W \sim N(\varepsilon, z^2/(1 + z^2 \mathbf{Z}^T \mathbf{Z}))$, tem-se, pelo Resultado 1, que a mgf de W é dada por

$$M_W(t) = \exp\left\{t\varepsilon + \frac{t^2 z^2}{2(1 + z^2 \mathbf{Z}^T \mathbf{Z})}\right\}.$$

Já a mgf de U^0 é calculada da seguinte maneira

$$\begin{aligned} M_{U^0} \left(\frac{z^2 \mathbf{Z} t}{1 + z^2 \mathbf{Z}^T \mathbf{Z}} \right) &= \exp\left\{\frac{1}{2} \left(\frac{z^2 \mathbf{Z} t}{1 + z^2 \mathbf{Z}^T \mathbf{Z}} \right)^T (\mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T) \left(\frac{z^2 \mathbf{Z} t}{1 + z^2 \mathbf{Z}^T \mathbf{Z}} \right)\right\} \times \\ &\times \frac{\Phi_{n_j} \left((\mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T) \left(\frac{z^2 \mathbf{Z} t}{1 + z^2 \mathbf{Z}^T \mathbf{Z}} \right); -\varepsilon \mathbf{Z}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T \right)}{\Phi_{n_j}(\mathbf{0}; -\varepsilon \mathbf{Z}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T)} \end{aligned}$$

$$\begin{aligned}
 M_{U^0} \left(\frac{z^2 \mathbf{Z} t}{1 + z^2 \mathbf{Z}^T \mathbf{Z}} \right) &= \exp \left\{ \frac{1}{2} \left[\left(\frac{z^2 t}{1 + z^2 \mathbf{Z}^T \mathbf{Z}} \right)^2 \mathbf{Z}^T \mathbf{Z} + \left(\frac{z^2 t}{1 + z^2 \mathbf{Z}^T \mathbf{Z}} \right)^2 z^2 \mathbf{Z}^T \mathbf{Z} \mathbf{Z}^T \mathbf{Z} \right] \right\} \times \\
 &\times \frac{\Phi_{n_j} \left(\frac{z^2 \mathbf{Z} t}{1 + z^2 \mathbf{Z}^T \mathbf{Z}} + \frac{z^2 \mathbf{Z} \mathbf{Z}^T z^2 \mathbf{Z} t}{1 + z^2 \mathbf{Z}^T \mathbf{Z}}; -\varepsilon \mathbf{Z}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T \right)}{\Phi_{n_j}(\mathbf{0}; -\varepsilon \mathbf{Z}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T)} \\
 &= \exp \left\{ \frac{1}{2} \left[\frac{z^4 t^2}{(1 + z^2 \mathbf{Z}^T \mathbf{Z})^2} \mathbf{Z}^T \mathbf{Z} (1 + z^2 \mathbf{Z}^T \mathbf{Z}) \right] \right\} \frac{\Phi_{n_j} \left(\frac{z^2 \mathbf{Z} t}{1 + z^2 \mathbf{Z}^T \mathbf{Z}} (1 + z^2 \mathbf{Z}^T \mathbf{Z}); -\varepsilon \mathbf{Z}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T \right)}{\Phi_{n_j}(\mathbf{0}; -\varepsilon \mathbf{Z}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T)} \\
 &= \exp \left\{ \frac{1}{2} \left(\frac{z^4 \mathbf{Z}^T \mathbf{Z} t^2}{1 + z^2 \mathbf{Z}^T \mathbf{Z}} \right) \right\} \frac{\Phi_{n_j}(\varepsilon \mathbf{Z} + z^2 \mathbf{Z} t; \mathbf{0}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T)}{\Phi_{n_j}(\varepsilon \mathbf{Z}; \mathbf{0}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T)}.
 \end{aligned}$$

Assim, tem-se que

$$\begin{aligned}
 M_W(t) \times M_{U^0} \left(\frac{z^2 \mathbf{Z} t}{1 + z^2 \mathbf{Z}^T \mathbf{Z}} \right) &= \\
 &= \exp \left\{ t\varepsilon + \frac{1}{2} \left[\frac{t^2 z^2 + z^4 \mathbf{Z}^T \mathbf{Z} t^2}{1 + z^2 \mathbf{Z}^T \mathbf{Z}} \right] \right\} \frac{\Phi_{n_j}(\varepsilon \mathbf{Z} + z^2 \mathbf{Z} t; \mathbf{0}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T)}{\Phi_{n_j}(\varepsilon \mathbf{Z}; \mathbf{0}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T)} \\
 &= \exp \left\{ t\varepsilon + \frac{1}{2} \left[\frac{t^2 z^2 (1 + z^2 \mathbf{Z}^T \mathbf{Z})}{1 + z^2 \mathbf{Z}^T \mathbf{Z}} \right] \right\} \frac{\Phi_{n_j}(\varepsilon \mathbf{Z} + z^2 \mathbf{Z} t; \mathbf{0}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T)}{\Phi_{n_j}(\varepsilon \mathbf{Z}; \mathbf{0}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T)} \\
 &= \exp \left\{ t\varepsilon + \frac{t^2 z^2}{2} \right\} \frac{\Phi_{n_j}(\varepsilon \mathbf{Z} + z^2 \mathbf{Z} t; \mathbf{0}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T)}{\Phi_{n_j}(\varepsilon \mathbf{Z}; \mathbf{0}, \mathbf{I}_{n_j} + z^2 \mathbf{Z} \mathbf{Z}^T)} \\
 &= M_{\lambda_y}(t).
 \end{aligned}$$

Portanto,

$$\lambda_y \stackrel{d}{=} W + \frac{z^2 \mathbf{Z}^T U^0}{1 + z^2 \mathbf{Z}^T \mathbf{Z}}.$$

Referências Bibliográficas

- [1] ASSUNÇÃO, R. M. (2009). Notas de aula do curso Dirichlet Process and Spectral Graph Theory *Departamento de Estatística, ICEX, UFMG*.
- [2] ARELLANO-VALLE, R. B.; AZZALINI, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, Vol. 33, pp. 561-574.
- [3] ARELLANO-VALLE, R. B.; CASTRO, L. M.; LOSCHI, R. H. (2011). Change point detection in the skew-normal model parameters. *Communications in Statistics - Theory and Methods*, no prelo 2011.
- [4] ARELLANO-VALLE, R. B.; CASTRO, L. M.; GENTON, M. G.; GÓMEZ, H. W. (2008). Bayesian inference for shape mixtures of skewed distributions, with application to regression analysis. *International Society for Bayesian Analysis*, Vol. 3, No. 3, pp. 513-540.
- [5] ARELLANO-VALLE, R. B.; GENTON, M. G.; LOSCHI, R. H. (2009). Shape mixtures of multivariate skew-normal distributions. *Journal of Multivariate Analysis*, Vol. 100, pp. 91-101.
- [6] AZZALINI, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, Vol. 12, pp. 171-178.
- [7] BLACKWELL, D.; MACQUEEN, J. B. (1973). Ferguson distributions via Pólya urn schemes. *Annals of Statistics*, Vol. 1, pp. 353-355.
- [8] CASTRO, L. M.; LOSCHI, R. H.; ARELLANO-VALLE, R. B. (2009). Bayesian inference for the skew normal distribution: An application to change point problem. *Unpublished*.
- [9] DIAS, R. (2002). A review of non-parametric curve estimation methods with application to Econometrics. *Economia*, Vol. 3, No. 1, pp. 31-75.
- [10] ESCOBAR, M. D. (1994). Estimating Normal Means with a Dirichlet Process Prior. *Journal of the American Statistical Association*, Vol. 89, No. 425, pp. 268-277.
- [11] ESCOBAR, M. D.; WEST, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, Vol. 90, No. 430, pp. 577-588.
- [12] FERREIRA, J. T. A. S.; STEEL, M. F. J. (2006). A constructive representation of univariate skewed distributions, *Journal of the American Statistical Association*, Vol. 101, pp. 823-829.
- [13] FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, Vol. 1, No. 2, pp. 209-230.
- [14] GELFAND, A. E; SMITH A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, Vol. 85, No. 410, pp. 398-409.

- [15] LISEO, B.; LOPERFIDO, N. (2006). Default bayesian analysis of the skew-normal distribution. *Journal of Statistics Planning and Inference*, Vol. 136, pp. 373-389.
- [16] LIU T. I.; LEE J. C.; YEN S. Y. (2007). Finite mixture modelling using the skew normal distribution. *Statistica Sinica*, Vol. 17, pp. 909-927.
- [17] McEACHERN, S. N. (1994). Estimating normal means with a conjugate style Dirichlet process prior. *Communications in Statistics: Simulation and Computation*, Vol. 23, pp. 727-741.
- [18] McEACHERN, S. N. and MULLER, P. (1998). Estimating mixture of Dirichlet Process models. *Journal of Computational and Graphical Statistics*, Vol. 7, No. 2, pp. 223-238.
- [19] NEAL, R. (2000). Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics*, Vol. 9, pp. 249-265.
- [20] O'HAGAN, A.; LEONHARD, T. (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, Vol. 63, pp. 201-202.
- [21] PETERS, C. A.; VALAFAR, F. (2003). Comparison of three nonparametric density estimation techniques using Bayes classifiers applied to microarray data analysis. *Proceedings of the International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences*, 119-125.
- [22] SETHURAMAN, J. (1994) A constructive definition of Dirichlet priors. *Statistica Sinica*, Vol. 4, pp. 639-650.
- [23] SILVERMAN, B. W. (1986). Density Estimation for Statistics and Data Analysis. Londres(ed.). *Chapman and Hall*.
- [24] TEH, Y. W. (2010) Dirichlet Process. *University College London*. Disponível em: <<http://www.gatsbyucl.ac.uk/ywteh/research/npbayes/dp.pdf>>.
- [25] TEH, Y. W. (2007). Dirichlet Processes: Tutorial and Practical Course. *Gatsby Computational Neuroscience Unit, University College London*. Disponível em: <<http://www.gatsbyucl.ac.uk/ywteh/teaching/npbayes/mlss2007.pdf>>.
- [26] VALLEJOS, A. J. (2008). Bayesian semiparametric methods for the analysis of complex data. *Tese (Doutorado em Ciências) - Faculdade de Ciências, Universidade Católica de Leuven*
- [27] WIKIPEDIA. Non-parametric statistics. 2010. Disponível em: <http://www.en.wikipedia.org/wiki/Non-parametric_statistics>.
- [28] ZAPATA, Y. E. O. (2007). Analisis Bayesiano no Paramétrico Utilizando Procesos Skew Dirichlet. *Tese (Doutorado em Estatística) - Departamento de Estatística, Facultad de Matemática, Pontificia Universidad Católica do Chile*.