

Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Matemática

**Propriedades Assintóticas do Algoritmo  
MCEM para Misturas de Duas Distribuições  
Lognormal na Estimação da Distribuição do  
Comprimento da Fibra da Madeira**

por

**Thiago de Lima Macedo**

**Orientadora: Daniele S. Baratela Martins Neto**

Brasília

2011

Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Matemática

**Propriedades Assintóticas do Algoritmo MCEM para  
Misturas de Duas Distribuições Lognormal na  
Estimação da Distribuição do Comprimento da Fibra  
da Madeira**

por

**Thiago de Lima Macedo**

*Dissertação apresentada ao Departamento de Matemática da Universidade de Brasília,  
como parte dos requisitos para obtenção do grau de*

**MESTRE EM MATEMÁTICA**

Brasília, 15 de julho de 2011

Comissão Examinadora:

---

Prof<sup>ª</sup>. Daniele da S. Baratela Martins Neto - Orientadora (MAT-UnB)

---

Prof<sup>ª</sup>. Cátia Regina Gonçalves - Membro (MAT-UnB)

---

Prof<sup>ª</sup>. Cira Etheowalda Guevara Otiniano - Membro (EST-UnB)

*À minha esposa Kelly Cristina e  
meus pais Francisco e Cleide.*

# Agradecimentos

---

Primeiramente, agradeço a Deus por todas as oportunidades que me foram concedidas e tornaram possível a realização desse trabalho.

À minha querida esposa Kelly Cristina, pela paciência de estar ao meu lado durante este longo período de estudos, no qual por muitas vezes não pude dedicar meu tempo à ela. Aos meus pais Francisco e Cleide, aos meus irmãos Carlos e Mariana e a todos meus familiares que sempre tiveram fé em mim e nunca deixaram de me incentivar nos momentos mais difíceis.

À professora Daniele da Silva Baratela Martins Neto, minha orientadora, pela total dedicação, paciência e ajuda à seu orientando, no qual sem ela este trabalho não seria possível de ser concluído. Aos membros da banca, Cátia Regina e Cira Etheowalda, pelas correções e valiosas sugestões que enriqueceram este trabalho.

Aos professores Ary Medino, Lucina Ávila e Marcelo Furtado pelas disciplinas ministradas. Em especial, à professora Cátia Regina que sempre mostrou disponibilidade em ajudar e no qual procurou a professora Daniele para me orientar.

Aos amigos, Mônica, Jairo, Elis Gardel, Hudson, Aristóteles, Joabe, Thaynara, Robson, Wembesom, Bruno César, Bruno Nunes, Rafaela, Tarcísio e Wesley que caminharam comigo durante todo o curso dando apoio e ajuda sempre que precisei. Em especial, aos amigos Fábio e Grayce Kelly que estudaram comigo por diversas vezes, possibilitando um melhor aprendizado durante o curso.

A todas as pessoas que de alguma forma me ajudaram a chegar ao fim desse ciclo de estudos. Em especial ao seu Manoel, que não está mais conosco.

Enfim, agradeço à SEE/DF, pelo suporte financeiro.

# Resumo

---

Neste trabalho, estudamos um método de amostragem que leva em consideração a extração de núcleos de incremento em árvores plantadas, para estimar a distribuição das fibras da madeira. Consideramos o caso em que a distribuição do comprimento das fibras é uma mistura de duas distribuições lognormal e utilizamos o algoritmo MCEM para estimar os parâmetros do modelo. Por fim, verificamos as propriedades de consistência e normalidade assintótica do estimador obtido via algoritmo MCEM, baseados no estudo de Svensson e Luna (2010).

**Palavras-chave:** Mistura de Distribuições, Distribuição Lognormal, Estimador de Máxima Verossimilhança, Algoritmo EM, Algoritmo MCEM, Normalidade Assintótica.

# Abstract

---

In this work, we study a sampling method which takes into account the extraction of increment cores to find the fibre wood length distribution in standing trees. We consider the case where the fibre length distributions is a mixture of two lognormal distributions and use the MCEM algorithm to find maximum likelihood estimates for the parameters of the model. Finally, we investigate the properties of consistency and asymptotic normality of the estimator obtained via the MCEM algorithm, based on the study by Svensson and Luna (2010).

**Keywords:** Mixture of Distributions, Lognormal Distribution, Estimator Maximum Likelihood, EM Algorithm, Algorithm MCEM, Asymptotic Normality.

# Sumário

---

<b>Introdução</b>	<b>1</b>
<b>1 Preliminares</b>	<b>5</b>
1.1 Introdução . . . . .	5
1.2 Princípio de Máxima Verossimilhança . . . . .	5
1.3 O Algoritmo EM e a Sua Versão Estocástica MCEM . . . . .	8
1.4 Mistura de Distribuições e Identificabilidade . . . . .	14
<b>2 O Modelo</b>	<b>19</b>
2.1 Introdução . . . . .	19
2.2 Comprimento da Fibra da Madeira . . . . .	21
2.2.1 A Distribuição de $X$ . . . . .	22
2.2.2 A Distribuição de $Y$ dado $X$ . . . . .	27
2.2.3 A Distribuição de $W$ . . . . .	28
2.3 Identificabilidade . . . . .	31
2.4 O Algoritmo MCEM . . . . .	35
<b>3 Propriedades Assintóticas do Algoritmo MCEM para Mistura com Dados Censurados</b>	<b>41</b>
3.1 Introdução . . . . .	41
3.2 Condições de Regularidade e Resultados Preliminares . . . . .	43
3.3 Normalidade Assintótica do Estimador de Máxima Verossimilhança . . . . .	47
3.4 Convergência do Algoritmo EM Para o EMV . . . . .	55
3.5 Convergência do Algoritmo MCEM Para o Algoritmo EM . . . . .	59
<b>Bibliografia</b>	<b>67</b>

# Introdução

---

Com o crescente interesse em melhorar a utilização dos recursos da madeira, tornou-se imprescindível o desenvolvimento de métodos para avaliar a qualidade da madeira em árvores plantadas. Uma das propriedades importantes para avaliar a qualidade da madeira é o comprimento de suas fibras e, por esta razão, é desejado cada vez mais que os procedimentos para medir tal comprimento devam ser executados de maneira rápida, com baixo custo e não destrutivos em relação à árvore.

Um método de amostragem para estimar a distribuição do comprimento das fibras da madeira, desenvolvido por Mörling et al. [13], considerado rápido e não destrutivo, é feito por meio de núcleos de incrementos. Um núcleo de incremento é uma amostra cilíndrica de madeira de  $5mm$  de diâmetro, extraída de uma árvore, ainda plantada, com uma broca especial (ver Figura 1). Porém, encontrar a distribuição do comprimento das fibras não é uma tarefa fácil, pois em um núcleo de incremento os dados podem ser censurados contendo fibras cortadas e fibras não cortadas. Outro fato que dificulta este tipo de coleta de dados, é que a amostra não contém apenas as fibras de interesse, mas também outros tipos de células, que aqui chamamos de microfibras. Nesse estudo, os dados correspondem às medidas dos comprimentos das células (fibras e microfibras) obtidos em grandes quantidades de amostras, em um processo que não distingue automaticamente se as células são microfibras ou fibras, como também não é capaz de dizer se uma célula é ou não cortada. Devido à essas incertezas, tornou-se bastante razoável supor que os comprimentos observados vêm de uma versão censurada de uma mistura de distribuições de comprimento das microfibras e fibras da árvore.

Para descrever, de modo formal, o método para medir o comprimento das fibras da madeira sem realizar a derrubada da árvore, consideramos as seguintes variáveis aleatórias:

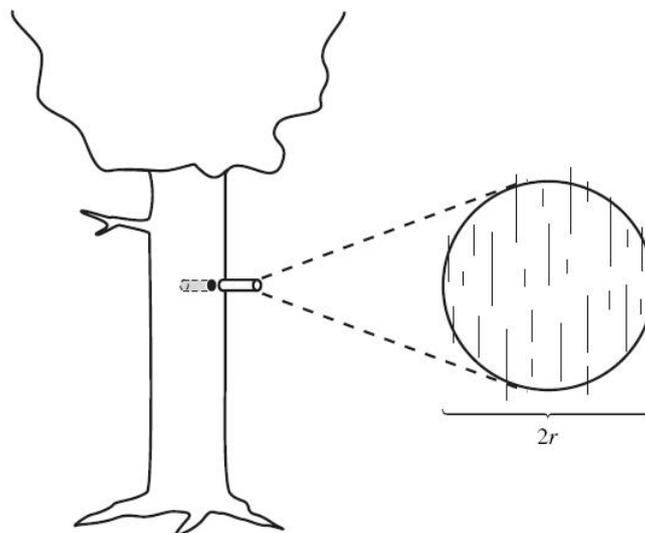


Figura 1: Um núcleo de incremento cilíndrico de diâmetro  $2r$  é extraído perpendicular à direção das fibras na árvore. A seção transversal do núcleo incremento mostra como as fibras da amostra podem ser tanto cortadas como não cortadas.

$W$  = o comprimento de uma célula (microfibra ou fibra) em uma árvore que está plantada,

$Y$  = o comprimento real de uma célula que, pelo menos aparece parcialmente no núcleo de incremento, e

$X$  = o comprimento correspondente à  $Y$ , visto no núcleo de incremento.

Considerando que a célula pode ter sido cortada, quando seu comprimento é obtido a partir do núcleo de incremento, temos que  $P(X \leq Y) = 1$ . Os comprimentos das células são obtidos a partir do núcleo de incremento e, portanto, são dados referentes à distribuição de  $X$ . Contudo, nosso interesse real é na distribuição de  $W$ , e a distribuição de  $Y$  auxilia na conexão entre às distribuições de  $X$  e de  $W$ . Aqui, supomos que as distribuições de  $X$ ,  $Y$  e  $W$  são algum tipo de mistura de distribuições de comprimento de uma microfibra e de uma fibra. Sendo assim, iniciamos este estudo supondo que  $Y$  é uma variável aleatória contínua com função de densidade  $f_Y(y, \theta)$ , onde  $\theta$  é um parâmetro desconhecido, dada por

$$f_Y(y; \theta) = \varepsilon f_{Y_1}(y; \phi_1) + (1 - \varepsilon) f_{Y_2}(y; \phi_2), \quad 0 < y < \infty \quad (1)$$

onde  $\varepsilon$  é a proporção das microfibras do núcleo de incremento e  $f_{Y_1}(y; \phi_1)$  e  $f_{Y_2}(y; \phi_2)$  são as funções densidades do comprimento real das microfibras e fibras, respectivamente,

que pelo menos aparecem parcialmente no núcleo de incremento. Desta forma temos que  $\theta = (\varepsilon, \phi_1, \phi_2)$  e é de interesse estimar tal parâmetro.

O processo de extração das medidas das células nas amostras não é capaz de registrar se cada comprimento observado vem de uma célula não cortada ou de uma célula cortada em uma ou duas partes. Pela presença deste tipo de censura e também por levar em consideração a falta de informação nos dados, é que utilizamos o algoritmo EM (Expectation - Maximization) e uma versão estocástica dele, o algoritmo MCEM (Monte Carlo EM), para estimar o parâmetro  $\theta$ .

O algoritmo EM, apresentado por Dempster et al. [7] em 1977, é um processo iterativo para estimar parâmetros via máxima verossimilhança, quando as observações são, de alguma forma, dados incompletos. O algoritmo contém dois passos onde no primeiro, passo - E, é calculado a esperança de uma função e no segundo, passo - M, é obtido o valor de  $\theta$  para o qual maximiza a função obtida no passo E. Este algoritmo apresenta alguns problemas, dentre os quais, a dificuldade do passo E. Na literatura, é apresentado em 1990 uma versão estocástica desse algoritmo, proposto por Wei e Tanner [20], chamado algoritmo MCEM, que substitui o cálculo analítico de uma esperança no passo E pelo cálculo aproximado dela. Tanto o algoritmo EM, como o algoritmo MCEM, produz uma sequência de estimadores de máxima verossimilhança de  $\theta$ , que aqui denotamos por  $\{\tilde{\theta}^{(k)}\}$  e  $\{\tilde{\theta}_S^{(k)}\}$ , respectivamente.

Este trabalho, baseado em Svensson e Luna [17] e Svensson et al. [18], tem como objetivo principal estudar o comportamento assintótico do algoritmo MCEM proposto por Svensson et al. [18], quando em (1) é assumido uma mistura de distribuições lognormal, ou seja, quando  $f_{Y_1}(y; \phi_1)$  e  $f_{Y_2}(y; \phi_2)$  são densidades pertencentes a uma família lognormal. Mais formalmente, denotamos por  $\theta^* = (\varepsilon^*, \phi_1^*, \phi_2^*)$  o verdadeiro parâmetro, onde  $\phi_i^*(\mu_i^*, \sigma_i^*)$ ,  $i = 1, 2$ , são parâmetros da distribuição lognormal e mostramos, sob certas condições de regularidade, que

$$\sqrt{n}(\tilde{\theta}_S^{(k)} - \theta^*) \xrightarrow{d} N(0, I(\theta^*)^{-1}) \quad (2)$$

com  $k, m, n \rightarrow \infty$  em algum sentido. Aqui,  $n$  é o tamanho da amostra,  $m$  é o número de gerações independentes obtidas do passo E do algoritmo MCEM e  $I(\cdot)$  é a informação de Fisher.

A normalidade assintótica (2) é mostrada em três etapas, a saber

$$(1^a) \sqrt{n}(\hat{\theta}_{MV} - \theta^*) \xrightarrow{d} N(0, I(\theta^*)^{-1}), \text{ quando } n \rightarrow \infty,$$

$$(2^a) \sqrt{n}(\tilde{\theta}^{(k)} - \hat{\theta}_{MV}) \xrightarrow{p} 0, \text{ quando } k \rightarrow \infty \text{ e } n \rightarrow \infty \text{ (em algum sentido),}$$

$$(3^a) \sqrt{n}(\tilde{\theta}_S^{(k)} - \tilde{\theta}^{(k)}) \xrightarrow{p} 0, \text{ } k \rightarrow \infty, m \rightarrow \infty \text{ e } n \rightarrow \infty \text{ (em algum sentido),}$$

onde  $\hat{\theta}_{MV}$  é o estimador de máxima verossimilhança de  $\theta^*$ . Esse desmembramento é possível, pois temos que

$$\sqrt{n}(\tilde{\theta}_S^{(k)} - \theta^*) = \sqrt{n}(\tilde{\theta}_S^{(k)} - \tilde{\theta}^{(k)}) + \sqrt{n}(\tilde{\theta}^{(k)} - \hat{\theta}_{MV}) + \sqrt{n}(\hat{\theta}_{MV} - \theta^*).$$

No Capítulo 1 apresentamos conceitos e resultados preliminares da teoria de estimação que nos auxiliam no desenvolvimento deste trabalho.

No Capítulo 2 apresentamos detalhes da modelagem matemática realizada para determinar a função de distribuição do comprimento das fibras da madeira em árvores que estão plantadas. Na Seção 2.2 apresentamos as variáveis aleatórias presentes neste estudo, bem como suas distribuições. Na Seção 2.3 tratamos da identificabilidade da classe de misturas finitas das distribuições envolvidas nesta análise. E por último, na Seção 2.4 descrevemos o algoritmo MCEM, apropriado para lidar com a falta de informação dos dados, gerada pela censura, para estimar os parâmetros do modelo.

Finalmente, no Capítulo 3, enunciamos e demonstramos o Teorema 3.1 referente a consistência e a normalidade assintótica do estimador obtido via algoritmo MCEM. Na Seção 3.2 apresentamos definições e resultados preliminares que são utilizados nas demais seções do Capítulo 3, na Seção 3.3 mostramos a convergência do EMV para o parâmetro  $\theta^*$ , na Seção 3.4 mostramos a convergência do estimador gerado via algoritmo EM para o EMV e na Seção 3.5 mostramos a convergência do estimador gerado via algoritmo MCEM para o gerado via algoritmo EM.

---

# Preliminares

---

## 1.1 Introdução

Neste capítulo, apresentamos algumas definições, conceitos e resultados que auxiliam no desenvolvimento deste trabalho tendo como referências básicas Lehmann [5], Dempster et al. [7], Wei e Tanner [20], Titteringtin et al. [19] e Atienza et al. [2]. Iniciamos na Seção 1.2 com definições necessárias para o estudo e compreensão do princípio de máxima verossimilhança e algumas de suas propriedades. Na Seção 1.3 apresentamos o algoritmo EM e o algoritmo MCEM, os quais são utilizados na estimação da distribuição do comprimento da fibra da madeira nos capítulos subsequentes. Terminamos, na Seção 1.4, com o estudo de misturas de distribuições. Nesta seção definimos a identificabilidade para uma classe de misturas finitas, apresentamos condições suficientes para garanti-la e mostramos que uma classe de misturas de distribuições lognormal é identificável.

## 1.2 Princípio de Máxima Verossimilhança

**Definição 1.1.** *Sejam  $X_1, \dots, X_n$  variáveis aleatórias independentes e identicamente distribuídas (i.i.d.), com função de distribuição comum  $F$ . Dizemos que  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  é uma amostra aleatória de tamanho  $n$  de  $F$ .*

*Se as variáveis  $X_1, X_2, \dots, X_n$  i.i.d. são contínuas com densidade comum  $f$ , podemos dizer que  $\mathbf{X}$  é uma amostra aleatória de  $f$ .*

Sejam  $X_1, X_2, \dots, X_n$  uma amostra aleatória de  $f(x; \theta)$  com respeito a uma medida  $\sigma$ -finita  $\mu$  sobre  $\mathbb{R}$ , onde  $\theta$  é um parâmetro, podendo ser um escalar ou um vetor. Aqui,

supomos  $\theta \in \Theta \subset \mathbb{R}^v$  com  $v$  um inteiro e  $v \geq 1$ . Quando  $\theta$  é desconhecido, tem-se como objetivo estimá-lo. Para isso, vejamos as seguintes definições.

**Definição 1.2.** Um estimador de  $\theta$  é uma estatística  $\hat{\theta}_n = T(X_1, X_2, \dots, X_n)$ , onde  $T$  é uma função de  $\mathbb{R}^n$  em  $\Theta$ .

**Definição 1.3.** Um estimador  $\hat{\theta}_n$  é chamado de estimador consistente de  $\theta$ , se para todo  $\varepsilon > 0$ , temos que

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1,$$

ou seja, a sequência  $\{\hat{\theta}_n\}_n$  converge em probabilidade para  $\theta$ . Notação:  $\hat{\theta}_n \xrightarrow{p} \theta$ .

**Definição 1.4.** Um estimador  $\hat{\theta}_n$  é chamado de estimador fortemente consistente de  $\theta$ , se

$$P(\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta) = 1,$$

ou seja, a sequência  $\{\hat{\theta}_n\}_n$  converge quase certamente para  $\theta$ . Notação:  $\hat{\theta}_n \xrightarrow{q.c.} \theta$ .

**Definição 1.5.** Seja  $X_1 = x_1, \dots, X_n = x_n$  uma amostra aleatória observada de  $f(x; \theta)$ . A função de verossimilhança das observações  $x_1, \dots, x_n$  é uma função de  $\theta$  definida em  $\Theta$  por

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i; \theta).$$

Um método bastante utilizado para estimar  $\theta$  é o *princípio de máxima verossimilhança*, que consiste em escolher como uma estimativa de  $\theta$  um valor  $\hat{\theta}(X_1, \dots, X_n) \in \Theta$  que maximiza a função de verossimilhança, isto é, um  $\hat{\theta}$  que satisfaz

$$L(\hat{\theta}; x_1, \dots, x_n) = \sup_{\theta \in \Theta} L(\theta; x_1, \dots, x_n).$$

Se  $\hat{\theta}$  existe, ele é chamado de *estimador de máxima verossimilhança* (EMV) de  $\theta$ . Observe, entretanto, que  $\hat{\theta}$  pode não existir, ou, no caso de existir, pode não ser único.

Note que, como a função  $\log$  é uma função monótona estritamente crescente, logo é mais conveniente trabalhar com o logaritmo da função de verossimilhança, chamada de função de log-verossimilhança dada por:

$$l(\theta) = \log L(\theta; x_1, \dots, x_n) = \sum_{i=1}^n \log f(x_i; \theta). \quad (1.1)$$

Se a função de log-verossimilhança é diferenciável em  $\Theta$ , o EMV de  $\theta = (\theta_1, \dots, \theta_v)$  pode ser encontrado resolvendo a(s) equaç(ões) de verossimilhança

$$\frac{\partial l(\theta)}{\partial \theta_i} = 0, \quad i = 1, \dots, v. \quad (1.2)$$

**Definição 1.6.** Chamamos de matriz de informação, a matriz quadrada de ordem  $v$ ,  $I(\theta) = I_{jk}(\theta)$ , no qual suas entradas são dadas por

$$I_{jk}(\theta) = \text{cov} \left[ \frac{\partial}{\partial \theta_j} \log f(X, \theta), \frac{\partial}{\partial \theta_k} \log f(X, \theta) \right], \quad \forall j, k = 1, \dots, v.$$

Nos casos em que

$$E \left( \frac{\partial}{\partial \theta_j} \log f(X; \theta) \right) = 0, \quad \forall j = 1, \dots, v.$$

e

$$E \left( \frac{\partial}{\partial \theta_j} \log f(X; \theta) \cdot \frac{\partial}{\partial \theta_k} \log f(X; \theta) \right) = -E \left( \frac{\partial^2}{\partial \theta_j \partial \theta_k} \log f(X; \theta) \right),$$

então  $I_{jk}(\theta) = -E \left( \frac{\partial^2}{\partial \theta_j \partial \theta_k} \log f(X; \theta) \right)$  e, neste caso é chamada de matriz de informação de Fisher de  $X$ .

**Teorema 1.7.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória i.i.d. de  $f(x; \theta)$  e seja  $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n) \in \Theta \subset \mathbb{R}^v$  o EMV de  $\theta$ . Sob as condições de regularidade **(A1)** - **(A7)**, dadas abaixo, tem-se que:

- (a)  $\hat{\theta}_i$  é estimador consistente de  $\theta_i$ ,  $i = 1, \dots, v$ ;
- (b)  $\sqrt{n}(\hat{\theta} - \theta^*)$  é assintoticamente normal com média zero e matriz de covariância,  $[I(\theta^*)]^{-1}$ ;
- (c)  $\hat{\theta}_i$  é assintoticamente eficiente no sentido que

$$\sqrt{n}(\hat{\theta} - \theta^*) \xrightarrow{d} N(0, [I(\theta^*)]_{ii}^{-1}),$$

onde  $\theta^*$  é o valor verdadeiro de  $\theta$ .

Condições de Regularidade:

- (A1)** Para  $\theta, \theta' \in \Theta \subset \mathbb{R}^v$ , se  $\theta \neq \theta'$ , então  $f(x; \theta) \neq f(x; \theta')$ ;
- (A2)** As densidade  $f(x; \theta)$  tem suporte comum, isto é,  $\text{supp}(f) = \{x; f(x; \theta) > 0\}$  independente de  $\theta$ ;
- (A3)** O espaço paramétrico  $\Theta$  contém um conjunto aberto  $\Theta^*$  que tem  $\theta^*$  em seu interior;

**(A4)** Para quase todo  $x$ , a densidade  $f(x; \theta)$  admite todas as terceiras derivadas  $\frac{\partial^3 f(x; \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k}$  para todo  $\theta \in \Theta^*$ ;

(A5) As primeiras e as segundas derivadas de  $\log$  de  $f$  satisfazem as seguintes equações

$$E \left[ \frac{\partial \log f(x; \theta)}{\partial \theta_i} \right] = 0, \quad i = 1, \dots, v$$

e

$$I_{ij}(\theta) = E \left[ \frac{\partial \log f(x; \theta)}{\partial \theta_i} \frac{\partial \log f(x; \theta)}{\partial \theta_j} \right] = E \left[ -\frac{\partial^2 \log f(x; \theta)}{\partial \theta_i \partial \theta_j} \right].$$

(A6) A matriz de informação  $[I(\theta)]_{d \times d}$  tem entradas  $I_{ij}(\theta)$  finitas e é uma matriz positiva definida para todo  $\theta \in \Theta^*$ ;

(A7)  $\forall \theta \in \Theta^*$ , existem funções  $M_{ijk}(x) \geq 0$ , tais que

$$\left| \frac{\partial^3 \log f(x; \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| \leq M_{ijk}(x),$$

onde

$$E(M_{ijk}(x)) < \infty, \quad \forall i, j, k.$$

A demonstração do Teorema 1.7, assim como mais detalhes deste resultado, pode ser encontrada em Lehmann [5].

## 1.3 O Algoritmo EM e a Sua Versão Estocástica MCEM

O algoritmo EM (Expectation Maximization) apresentado por Dempster et. al. [7] em 1977, é um processo iterativo para estimar parâmetros via máxima verossimilhança, quando as observações são, de alguma forma, dados incompletos. Isto é, seja  $Y$  uma variável aleatória com função de densidade  $f_Y(y; \theta)$ . Queremos determinar o EMV para  $\theta$  com base na observação  $Y = y$ . Todavia, quando o cálculo direto do EMV torna-se complexo, buscamos outra forma de resolver esse problema com a introdução de um dado latente  $z$ , de tal forma que  $c = (y, z)$  seja visto como um dado completo e a maximização da função de verossimilhança  $f_{Y,Z}(y, z; \theta) = f_C(c; \theta)$  seja mais simples. Note que  $f_C = (c; \theta)$ , produz um estimador que depende do valor introduzido  $z$ , o que não faz sentido. Assim, a proposta do algoritmo EM é obter o estimador de  $\theta$  baseado em um valor esperado de  $f_C = (c; \theta)$ . Formalizemos o algoritmo EM.

Seja  $Y$  um vetor aleatório com valores em  $E_Y \subseteq \mathbb{R}^\ell$  e densidade  $f_Y(y; \theta)$  relativa a medida  $\mu_Y$ , onde  $\theta \in \Theta \subset \mathbb{R}^v$  é um parâmetro desconhecido. O objetivo é encontrar o EMV de  $\theta$  quando  $Y = y$  é observado, isto é, obter  $\theta_{max} = \operatorname{argmax}_\theta f_Y(y; \theta)$ . Quando este cálculo é, de alguma forma, complexo, uma alternativa é considerar a observação  $y$  incompleta e completá-la pelo dado não-observado  $z$ , de modo que o cálculo

de  $\operatorname{argmax}_\theta f_{Y,Z}(y, z; \theta)$  seja mais simples. Formalmente, seja  $Z$  um vetor aleatório com valores em  $E \subseteq \mathbb{R}^p$ ,  $\mu_y$  uma medida  $\sigma$ -finita em  $(E, \mathcal{E})$  onde  $\mathcal{E}$  é uma  $\sigma$ -álgebra de  $E$  e tais que

$$f_Y(y; \theta) = \int_E f_{Y,Z}(y, z; \theta) dz.$$

Neste caso, dado  $Y = y$  a variável  $Z$  possui uma densidade condicional (preditiva) relativa a medida  $\mu_y$  dada por

$$f_{Z|Y}(z|y; \theta) = \frac{f_{Y,Z}(y, z; \theta)}{f_Y(y; \theta)}.$$

A estimativa  $\theta_{max} = \operatorname{argmax}_\theta f_{Y,Z}(y, z; \theta)$  depende do valor acrescido  $z$  e a proposta do algoritmo EM é a substituição do cálculo do  $\operatorname{argmax}_\theta f_{Y,Z}(y, z; \theta)$  pela maximização do valor esperado de  $\log f_{Y,Z}(y, z; \theta)$  relativa a distribuição preditiva,

$$Q(\theta, \theta') = \int_E [\log f_{Y,Z}(y, z; \theta)] f_{Z|Y}(z|y; \theta') \mu_y(dz) = E[\log f_{Y,Z}(Y, Z; \theta) | \theta', y]. \quad (1.3)$$

Cada iteração do algoritmo EM consiste de dois passos: cálculo de  $Q(\theta, \theta')$  (passo E) e a determinação do  $\operatorname{argmax}_\theta Q(\theta, \theta')$  (Passo M).

Assim, no **algoritmo EM** observa-se o dado  $Y = y$  e seleciona-se um valor inicial  $\theta(0) \in \Theta$ . Na iteração  $k + 1$ ,

passo E: calcula-se a esperança

$$Q(\theta, \theta(k)) = E[\log f_{Y,Z}(Y, Z; \theta) | \theta(k), y].$$

passo M: determina-se

$$\theta(k + 1) = \operatorname{argmax}_\theta Q(\theta, \theta(k)).$$

**Exemplo 1.8.** (*Modelo do Elo Genético, Rao [14]*). Assuma que temos 4 categorias de animais e que  $Y_i$  representa o número de animais da categoria  $j$  numa amostra de tamanho  $r$ . Suponha que  $Y = (Y_1, Y_2, Y_3, Y_4)$  tem distribuição multinomial com probabilidades  $\left(\frac{1}{2} + \frac{\theta}{4}, \frac{1}{4}(1 - \theta), \frac{1}{4}(1 - \theta), \frac{\theta}{4}\right)$  e  $\theta \in (0, 1)$ . Para o dado observado  $y = (y_1, y_2, y_3, y_4)$  temos a densidade

$$f_Y(y; \theta) = \binom{r}{y_1 \ y_2 \ y_3 \ y_4} \left(\frac{1}{4}\right)^r (2 + \theta)^{y_1} (1 - \theta)^{y_2 + y_3} \theta^{y_4},$$

portando, a menos de uma constante de proporcionalidade independente de  $\theta$ , segue

$$f_Y(y; \theta) \propto (2 + \theta)^{y_1} (1 - \theta)^{y_2 + y_3} \theta^{y_4}, \quad (1.4)$$

logo,

$$\log f_Y(y; \theta) \propto y_1 \log(2 + \theta) + (y_2 + y_3) \log(1 - \theta) + y_4 \log \theta. \quad (1.5)$$

Para o cálculo do  $\theta_{max}$  em (1.4) é complexo e recai na solução da equação do 2º grau, pois derivando (1.5) em relação a  $\theta$  e igualando a 0 obtemos,

$$\frac{\partial \log f_Y(y; \theta)}{\partial \theta} = r\theta^2 - [y_1 - 2(y_2 + y_3) - y_4]\theta - 2y_4 = 0. \quad (1.6)$$

Este cálculo pode ser simplificado introduzindo uma variável latente  $Z = z$  de modo que a primeira categoria seja subdividida em duas subcategorias com probabilidades  $\frac{1}{2}$  e  $\frac{\theta}{4}$ . Neste caso, o dado completo é  $(y_1 - z, z, y_2, y_3, y_4)$  e a densidade completa e dada por,

$$f_{Y,Z}(y, z; \theta) \propto \theta^{z+y_4} (1 - \theta)^{y_2+y_3},$$

cuja maximização resulta na solução de uma equação do 1º grau:  $\theta(r - y_1 + z) - (z + y_4) = 0$ , após resolver  $\frac{\partial}{\partial \theta} \log f_{Y,Z}(y, z; \theta) = 0$ . Para melhor ilustrar a notação utilizada no algoritmo EM, temos para este exemplo:  $E = \{0, 1, \dots, r\}$ ,  $\mu_y$  é a medida contadora em  $0, 1, \dots, y_1$  e

$$f_{Z|Y}(z|y; \theta) = \frac{f_{Y,Z}(y, z; \theta)}{f_Y(y; \theta)} = \binom{y_1}{z} \left(\frac{\theta}{2 + \theta}\right)^z \left(\frac{2}{2 + \theta}\right)^{y_1 - z}. \quad (1.7)$$

Pela definição:

$$\begin{aligned} Q(\theta, \theta(k)) &= \int_E [\log f_{Y,Z}(y, z; \theta)] f_{Z|Y}(z|y; \theta(k)) \mu_y(dz) \\ &= \int_E [(z + y_4) \log \theta + (y_2 + y_3) \log(1 - \theta)] f_{Z|Y}(z|y; \theta(k)) \mu_y(dz) \\ &= y_4 \log \theta + (y_2 + y_3) \log(1 - \theta) + \log \theta \left[ \sum_{z=0}^{y_1} z f_{Z|Y}(z|y; \theta(k)) \right] + cte \\ &= \left( y_4 + \sum_{z=0}^{y_1} z f_{Z|Y}(z|y; \theta(k)) \right) \log \theta + (y_2 + y_3) \log(1 - \theta) \\ &= \left( y_4 + y_1 \left( \frac{\theta(k)}{2 + \theta(k)} \right) \right) \log \theta + (y_2 + y_3) \log(1 - \theta) \end{aligned} \quad (1.8)$$

e  $\theta(k + 1) = \operatorname{argmax}_{\theta} Q(\theta, \theta(k))$ . Então ao derivar (1.8) e igualando o resultado à 0

obtemos:

$$\theta(k+1) = \frac{(y_1 + y_4)\theta(k) + 2y_4}{r\theta(k) + 2(r - y_1)}. \quad (1.9)$$

A grande popularidade do algoritmo EM consiste no fato de termos  $\theta(k) \xrightarrow{k \rightarrow \infty} \theta_{max}$ . Observe o resultado da seguinte proposição.

**Proposição 1.9.** *Se  $Q(\theta, \theta') \geq Q(\theta', \theta)$ , então  $f_Y(y; \theta) \geq f_Y(y; \theta')$ .*

**Demonstração** Temos que

$$\log \frac{f_Y(y; \theta)}{f_Y(y; \theta')} = \log \left( \frac{f_{Y,Z}(y, z; \theta)}{f_{Z|Y}(z|y; \theta)} \cdot \frac{f_{Z|Y}(z|y; \theta')}{f_{Y,Z}(y, z; \theta')} \right).$$

Como  $\int_E f_{Z|Y}(z|y; \theta') dz = 1$ . Obtemos,

$$\begin{aligned} \log \frac{f_Y(y; \theta)}{f_Y(y; \theta')} &= \int_E \left[ \log \left( \frac{f_{Y,Z}(y, z; \theta)}{f_{Z|Y}(z|y; \theta)} \cdot \frac{f_{Z|Y}(z|y; \theta')}{f_{Y,Z}(y, z; \theta')} \right) \right] f_{Z|Y}(z|y; \theta') \mu_y(dz) \\ &= \int_E \left[ \log \frac{f_{Y,Z}(y, z; \theta)}{f_{Y,Z}(y, z; \theta')} + \log \frac{f_{Z|Y}(z|y; \theta')}{f_{Z|Y}(z|y; \theta)} \right] f_{Z|Y}(z|y; \theta') \mu_y(dz) \\ &= \int_E \left[ \log f_{Y,Z}(y, z; \theta) - \log f_{Y,Z}(y, z; \theta') - \log \frac{f_{Z|Y}(z|y; \theta)}{f_{Z|Y}(z|y; \theta')} \right] f_{Z|Y}(z|y; \theta') \mu_y(dz) \\ &= Q(\theta, \theta') - Q(\theta', \theta) - \int_E \left[ \log \frac{f_{Z|Y}(z|y; \theta)}{f_{Z|Y}(z|y; \theta')} \right] f_{Z|Y}(z|y; \theta') \mu_y(dz). \end{aligned}$$

Pela desigualdade de Jensen,

$$\int_E \left[ \log \frac{f_{Z|Y}(z|y; \theta)}{f_{Z|Y}(z|y; \theta')} \right] f_{Z|Y}(z|y; \theta') \mu_y(dz) \leq \log \left[ \int_E f_{Z|Y}(z|y; \theta) \mu_y(dz) \right] = 0.$$

Assim, se

$$Q(\theta(k+1), \theta(k)) \geq Q(\theta(k), \theta(k-1)) \geq \dots \geq Q(\theta(1), \theta(0)),$$

temos que

$$\log \frac{f_Y(y; \theta(k+1))}{f_Y(y; \theta(k))} \geq 0,$$

portanto,

$$f_Y(y; \theta(k+1)) \geq f_Y(y; \theta(k)) \geq \dots \geq f_Y(y; \theta(0)).$$

■

Agora devemos verificar se  $\lim_{k \rightarrow \infty} f_Y(y; \theta(k)) = f_{max}(y) = \max_k f_Y(y; \theta(k))$ , pois queremos saber se  $f_{max}(y)$  é um máximo local e se  $\theta(k) \xrightarrow{k \rightarrow \infty} \theta_{max} = \operatorname{argmax}_{\theta} f_Y(y; \theta)$ . Seguem os seguintes resultados.

**Teorema 1.10.** (Wu [21]). *Seja  $f_Y(y; \theta)$  contínua em  $\Theta$  e diferenciável no  $\operatorname{int}\Theta$ . Seja  $\{\theta(k)\}_{k \geq 0}$  a sequência gerada pelo algoritmo EM com  $\theta(0)$  tal que  $f_Y(y; \theta(0)) > 0$  e tal que  $\{\theta : \theta \in \Theta, f_Y(y; \theta) > f_Y(y; \theta(0))\}$  seja compacto. Assuma que para todo  $k$  tal que  $\theta(k) \in \mathcal{L}^c$ , onde  $\mathcal{L}^c = \left[ \theta : \theta \in \operatorname{int}\Theta, \frac{\partial f_Y(y; \theta')}{\partial \theta'} \Big|_{\theta'=\theta} = 0 \right]$ ,  $f_Y(y; \theta(k+1)) > f_Y(y; \theta(k))$  e que se  $\theta(k) \xrightarrow{k \rightarrow \infty} \theta^*$  com  $\theta^* \in \mathcal{L}^c$ , então  $\theta^* \in M(\theta^*)$ , onde  $M(\theta') = \{\theta : \theta = \operatorname{argmax}_{\rho} Q(\rho, \theta')\}$ .*

**Corolário 1.11.** *Seja  $\{\theta(k)\}$  a sequência gerada pelo algoritmo EM e assuma que exista um único ponto crítico  $\theta_{max}$  para o qual  $\frac{\partial f_Y(y; \theta')}{\partial \theta'} \Big|_{\theta'=\theta_{max}} = 0$ . Então, para  $Q(\theta, \theta')$  contínua em  $\theta$  e em  $\theta'$ , temos  $f_Y(y; \theta(k)) \xrightarrow{k \rightarrow \infty} f_Y(y; \theta^*)$  e  $\theta^* = \theta_{max}$ .*

Com isso, notemos que no exemplo (1.8) determinamos os valores de  $\theta_{max}$  encontrando as raízes de (1.6) que pode ser reescrita na forma:

$$\theta \left[ y_1 \frac{\theta}{2 + \theta} + y_2 + y_3 + y_4 \right] = y_1 \frac{\theta}{2 + \theta} + y_4.$$

Se  $y_i > 0$ ,  $i = 1, \dots, 4$  então temos uma única solução  $\theta_{max} \in (0, 1)$ . Pois para  $g(\theta) = (y_1 \frac{\theta}{2 + \theta} + y_4) / (y_1 \frac{\theta}{2 + \theta} + y_2 + y_3 + y_4)$  temos  $g(0) > 0$ ,  $g(1) < 1$  e  $g'(\theta) > 0$ , assim  $g(\theta) = \theta$  tem uma única solução para  $0 < \theta < 1$ . Por (1.8) temos

$$Q(\theta, \theta') = \left( y_1 \frac{\theta'}{2 + \theta'} + y_4 \right) \log \theta + (y_2 + y_3) \log(1 - \theta)$$

que é contínua em  $\theta$  e em  $\theta'$ . Assim pelo corolário (1.11) a sequência  $\{\theta(k)\}$  converge para  $\theta_{max}$  para qualquer  $\theta(0) \in (0, 1)$ .

O Teorema (1.10) nos dá condições suficientes para a convergência do algoritmo EM para algum ponto crítico de  $f_Y(y; \theta)$ , não necessariamente o máximo global. Além do mais, o limite pode depender do seu valor inicial, às vezes sua convergência pode ocorrer de forma muito lenta e também o cálculo do passo M pode ser de difícil resolução.

O Algoritmo MCEM (Wei e Tanner [20]) surge especialmente pela dificuldade de resolução do passo E do algoritmo EM, e propõe a substituição deste passo por uma aproximação Monte Carlo. O que se faz é simular os dados que faltam  $z$  a partir da distribuição condicional  $f_{Z|Y}(z|y; \theta(k))\mu_y(dz)$  no passo E na iteração de  $k+1$ , e em seguida, maximizar a esperança condicional aproximada  $Q^*(\theta, \theta(k)) = \frac{1}{m} \sum_{j=1}^m \log f_{Y, Z_j}(y, z_j; \theta)$ , ao

invés do cálculo analítico da esperança  $Q(\theta, \theta(k))$ .

**O Método MCEM:** Seja  $m \geq 1$  um inteiro fixo, observa-se o dado  $Y = y$  e seleciona-se um valor inicial  $\theta(0) \in \Theta$ . Na iteração  $k + 1$ ,

passo E: gera-se o vetor  $\mathbf{Z}(k + 1) = (Z_1(k + 1), \dots, Z_m(k + 1))$  com distribuição  $f_{\mathbf{Z}|Y}(\mathbf{z}|y; \theta(k))\mu_y^m(dz) = \prod_{j=1}^m f_{Z_j|Y}(z_j|y; \theta(k))dz_j$ . Seja  $\mathbf{z}(k + 1) = (z_1(k + 1), \dots, z_m(k + 1))$  o valor gerado. Calcula-se

$$Q^*(\theta, \theta(k)) = \frac{1}{m} \sum_{j=1}^m \log f_{Y,Z}(y, z_j; \theta) \quad (1.10)$$

passo M: Determina-se

$$\theta(k + 1) = \operatorname{argmax}_{\theta} Q^*(\theta, \theta(k)).$$

Note que  $Z_1(k + 1), \dots, Z_m(k + 1)$  são independentes e com a mesma distribuição  $f_{Z|Y}(z|y; \theta(k))\mu_y(dz)$ . Assim se  $Q(\theta, \theta(k)) = \int_E [\log f_{Y,Z}(y, z; \theta)] f_{Z|Y}(z|y; \theta(k))\mu_y(dz)$  for finita, temos pela Lei dos Grandes Números,

$$Q^*(\theta, \theta(k)) \xrightarrow{\text{q.c.}} Q(\theta, \theta(k)), \text{ com } m \rightarrow \infty, \quad (1.11)$$

onde q.c indica convergência quase certa. E isto justifica a aproximação(1.10).

**Exemplo 1.12.** Fixemos  $m$  e  $\theta(0)$  e observamos  $Y = y = (y_1, y_2, y_3, y_4)$ , a implementação do MCEM para o exemplo (1.8) é bastante simples. Na iteração  $k + 1$  geramos  $m$  variáveis independentes com distribuição binomial de parâmetros  $y_1$  e  $\frac{\theta(k)}{2 + \theta(k)}$ . Sejam  $z_1(k + 1), \dots, z_m(k + 1)$  os valores gerados. Temos no passo E:

$$Q^*(\theta, \theta(k)) = \frac{1}{m} \sum_{j=1}^m [(z_j(k + 1) + y_4) \log \theta + (y_2 + y_3) \log(1 - \theta)].$$

E maximizando  $Q^*(\theta, \theta(k))$  com relação à  $\theta$  obtemos no passo M:

$$\theta(k + 1) = \frac{\bar{z}(k + 1) + y_4}{\bar{z}(k + 1) + y_2 + y_3 + y_4}$$

onde  $\bar{z}(k + 1) = \frac{1}{m} \sum_{j=1}^m z_j(k + 1)$ .

## 1.4 Mistura de Distribuições e Identificabilidade

**Definição 1.13.** *Sejam  $f_1, \dots, f_N$  funções de densidades com respeito a alguma medida  $\sigma$ -finita  $\nu$  sobre  $\mathbb{R}$ , e  $Y$  uma variável aleatória com função densidade dada por*

$$f(y) = \sum_{i=1}^N \varepsilon_i f_i(y),$$

onde  $\varepsilon_i$ ,  $i = 1, \dots, N$ , são valores não negativos tais que  $0 \leq \varepsilon_i \leq 1$ ,  $i = 1, \dots, N$  e  $\sum_{i=1}^N \varepsilon_i = 1$ . Dizemos, assim, que a distribuição de  $Y$  é um mistura finita de distribuições,  $f$  é uma mistura finita de densidades, onde  $\varepsilon_i$  é chamada a  $i$ -ésima componente de proporção,  $f_i$  a  $i$ -ésima densidade componente da mistura e  $N$  o número de componentes.

Se as densidades componentes de uma mistura,  $f_i$ ,  $i = 1, \dots, N$ , pertencem a uma mesma família paramétrica, denotamos a função mistura de densidades da seguinte forma

$$f(y; \theta) = \sum_{i=1}^N \varepsilon_i f_i(y; \theta_i),$$

onde  $0 \leq \varepsilon_i \leq 1$ ,  $i = 1, \dots, N$ ,  $\sum_{i=1}^N \varepsilon_i = 1$  e cada função  $f_i$  é uma densidade parametrizada por  $\theta_i$ . Denotamos  $\theta = (\varepsilon_1, \dots, \varepsilon_{N-1}, \theta_1, \dots, \theta_N)$  o resumo dos parâmetros da mistura.

Todas as densidades componentes da mistura não, necessariamente, precisam pertencer a mesma família de distribuições. A única exigência é que  $f_i$  seja um densidade ou função de probabilidade.

Estimar o parâmetro  $\theta$  baseado em  $n$  observações independentes de  $Y$  com densidade  $f_Y(\cdot; \theta)$ , somente é significativa se  $\theta$  é identificável. A falta de identificabilidade pode ocasionar problemas especialmente no caso de uso de técnicas de estimação via máxima verossimilhança e simulação, feitas a partir da distribuição em questão, para inferências. Uma família de densidades paramétrica é identificável para o parâmetro  $\theta$ , se os valores de distintos  $\theta$  determinam membros distintos da família. Veja a definição a seguir.

**Definição 1.14.** *Considere a família de densidades paramétricas*

$$\mathcal{F} = \{f(y; \theta); \theta \in \Theta, y \in \mathbb{R}^r\}$$

chamada classes de funções de densidades, onde  $\Theta$  é o espaço paramétrico correspondente.

Dizemos que a classe de densidades  $\mathcal{F}$  é identificável quando:

$$f(\cdot; \theta) = f(\cdot; \theta') \text{ se, e somente se, } \theta = \theta'.$$

Veja, a seguir, um caso no qual a identificabilidade para uma classe de densidades com respeito a medida contadora, não ocorre.

**Exemplo 1.15.** : Seja  $\mathcal{F} = \left\{ p(y; \theta) = \binom{2}{y} \theta^y (1 - \theta)^{2-y}; y = 0, 1, 2, \theta \in (0, 1) \right\}$ . Note que

$$p(1; \theta) = \binom{2}{1} \theta (1 - \theta),$$

e que para  $\theta \neq \frac{1}{2}$ , temos que  $\theta \neq 1 - \theta$ , mas  $p(1; \theta) = p(1; 1 - \theta)$ . Dessa forma, esta classe de distribuições binomiais não é identificável.

Identificabilidade para mistura finita de distribuições é definida de forma um pouco diferente. Para ver o porquê disto ser necessário, suponha que  $f_Y(y; \theta) = \varepsilon f_{Y_1}(y; \phi_1) + (1 - \varepsilon) f_{Y_2}(y; \phi_2)$ , onde as funções densidades componentes da mistura pertencem a mesma família paramétrica. Se  $\theta = (\varepsilon, \phi_1, \phi_2)$  e  $\theta' = (1 - \varepsilon, \phi_2, \phi_1)$ , então para  $\theta \neq \theta'$  temos  $f_Y(y; \theta) = f_Y(y; \theta')$ . Na verdade, se todas as  $N$  funções densidades componentes da mistura de distribuições pertencem a mesma família paramétrica, então  $f_Y(y; \theta)$  é invariante sob as permutações  $N!$  dos índices das componentes em  $\theta$ . Este problema referente a falta de identificabilidade de  $\theta$  é conhecido como o problema de *label switching* de densidades de misturas. Por esta razão, a seguir, definimos a identificabilidade para o caso de mistura de distribuições. Para mais detalhes veja McLachlan [12], pag. 26.

**Definição 1.16.** Seja

$$\mathcal{F} = \{f(y; \theta); \theta \in \Theta, y \in \mathbb{R}^r\}$$

uma classe de funções de densidades com respeito a medida  $\sigma$ -finita  $\nu$  sobre  $\mathbb{R}$ . Definimos uma classe de misturas de  $\mathcal{F}$ , por

$$\mathcal{H} = \{f(y; \theta) | f(y; \theta) = \sum_{i=1}^N \varepsilon_i f_i(y; \theta_i), \varepsilon_i > 0, \sum_{i=1}^N \varepsilon_i = 1, f_i(y; \theta_i) \in \mathcal{F}, i = 1, \dots, N, \forall N \in \mathbb{Z}^+, y \in \mathbb{R}^r\},$$

onde  $\theta = (\varepsilon_1, \dots, \varepsilon_{N-1}, \theta_1, \dots, \theta_N)$ . Sejam  $f(y; \theta)$  e  $f(y; \theta^*)$  dois membros quaisquer de  $\mathcal{H}$ , dadas por

$$f(y; \theta) = \sum_{i=1}^N \varepsilon_i f_i(y; \theta_i)$$

e

$$f(y; \theta^*) = \sum_{j=1}^{N^*} \varepsilon_j^* f_j(y; \theta_j^*).$$

A classe de mistura  $\mathcal{H}$  é dita identificável para  $\theta$ , se

$$f(\cdot; \theta) \equiv f(\cdot; \theta^*),$$

implica que  $N = N^*$  e  $\forall i, \exists j$  tal que  $\varepsilon_i = \varepsilon_j^*$  e  $\theta_i = \theta_j^*$  e  $\equiv$  implica a igualdade das densidades para quase todo  $y$  relativo a medida  $\sigma$ -finita em  $\mathbb{R}^r$  para  $f(y; \theta)$ . Isto equivale a seguinte situação

$$F(y; \theta) \equiv F(y; \theta^*) \Rightarrow N = N^* \text{ e } \forall i, \exists j \text{ tal que } \varepsilon_i = \varepsilon_j^* \text{ e } \theta_i = \theta_j^*,$$

onde  $F(y; \theta)$  é uma função de distribuição correspondente a densidade  $f(y; \theta)$ .

O teorema, a seguir, é um resultado de Atienza et al. [2] e fornece condições suficientes para garantir a identificabilidade para mistura finita de distribuições de uma mesma família paramétrica.

No que segue,  $A'$  denota o conjunto dos pontos de acumulação de  $A \subset \mathbb{R}^d$ .

**Teorema 1.17.** *Seja  $\mathcal{F}$  uma família de distribuições. Seja  $M$  uma aplicação linear que transforma qualquer  $F \in \mathcal{F}$  em uma função real  $\Phi_F$  com domínio  $S(F) \subset \mathbb{R}^d$ . Seja  $S_0(F) = \{t \in S(F) : \Phi_F(t) \neq 0\}$ . Suponha que existe uma ordem total  $\prec$  em  $\mathcal{F}$ , tal que para todo  $F \in \mathcal{F}$  existe  $t(F) \in S_0(F)'$  satisfazendo:*

(a) *Se  $F_1, \dots, F_m \in \mathcal{F}$  com  $F_1 \prec F_i$  para  $2 \leq i \leq m$ , então*

$$t(F_1) \in \left[ S_0(F_1) \cap \left[ \bigcap_{i=2}^m S(F_i) \right] \right]'.$$

(b) *Se  $F_1 \prec F_2$ , então  $\lim_{t \rightarrow t(F_1)} \frac{\Phi_{F_2}(t)}{\Phi_{F_1}(t)} = 0$ .*

Então, a classe  $\mathcal{H}$  de misturas finitas de distribuições de  $\mathcal{F}$  é identificável.

**Demonstração de 1.17 :** Suponha que

$$\sum_{i=1}^N \varepsilon_i F_i = \sum_{j=1}^{N^*} \varepsilon_j^* F_j^* \tag{1.12}$$

para algum  $\varepsilon_i, \varepsilon_j^* \in (0, 1)$  com  $\sum_{i=1}^N \varepsilon_i = 1 = \sum_{j=1}^{N^*} \varepsilon_j^*$  e  $F_1, \dots, F_N, F_1^*, \dots, F_N^* \in \mathcal{F}$ . Sem perda de generalidade, podemos assumir que  $N \leq N^*, F_i \prec F_j, F_i^* \prec F_j^*$  para  $i < j$

e  $F_1 \preceq F_1^*$ . Portanto,  $F_1 \prec F_j^*$  para todo  $j = 2, \dots, N^*$ . Seja

$$A = S_0(F_1) \cap \left[ \bigcap_{i=2}^N S(F_i) \right] \cap \left[ \bigcap_{i=2}^{N^*} S(F_i^*) \right].$$

De (a) existe  $t(F_1) \in A'$ . Para qualquer  $t \in A$ ,

$$\sum_{i=1}^N \varepsilon_i \Phi_{F_i}(t) = \sum_{j=1}^{N^*} \varepsilon_j^* \Phi_{F_j^*}(t).$$

Para  $t \in S_0(F_1)$ , podemos dividir por  $\Phi_{F_1}(t)$ . Fazendo  $t \rightarrow t(F_1)$ , obtemos

$$\sum_{i=1}^N \varepsilon_i \lim_{t \rightarrow t(F_1)} \frac{\Phi_{F_i}(t)}{\Phi_{F_1}(t)} = \sum_{i=1}^{N^*} \varepsilon_i^* \lim_{t \rightarrow t(F_1)} \frac{\Phi_{F_i^*}(t)}{\Phi_{F_1}(t)}.$$

Portanto, de (b) temos que

$$\varepsilon_1 = \varepsilon_1^* \lim_{t \rightarrow t(F_1)} \frac{\Phi_{F_1^*}(t)}{\Phi_{F_1}(t)}.$$

Agora, se temos  $F_1 \prec F_1^*$ , então  $\varepsilon_1 = 0$ , o que é uma contradição. Assim  $F_1 = F_1^*$  e  $\varepsilon_1 = \varepsilon_1^*$  e os fatores correspondentes em (1.12) são cancelados. Repetindo este processo, obtemos  $\varepsilon_i = \varepsilon_i^*$  e  $F_i = F_i^*$  para  $i = 1, 2, \dots, \min[N, N^*]$ . Finalmente  $N = N^*$ . De fato, se  $N^* > N$ , temos  $\sum_{j=N+1}^{N^*} \varepsilon_j^* F_j^*(x) = 0$ , e portanto  $\varepsilon_j^* = 0$  para  $N+1 \leq j \leq N^*$ . Logo  $\mathcal{H}$  é identificável. ■

O resultado a seguir, simplifica os pressupostos nos casos em que o ponto  $t(F) = t_0$  não depende de  $F \in \mathcal{F}$ .

**Corolário 1.18.** *Seja  $\mathcal{F}$  uma família de distribuições. Seja  $M$  uma aplicação linear que transforma qualquer  $F \in \mathcal{F}$  em uma função real  $\Phi_F$  com domínio  $S(F) \subset \mathbb{R}^d$ . Seja  $S_0(F) = \{t \in S(F) : \Phi_F(t) \neq 0\}$  e suponha que exista um ponto  $t_0$  satisfazendo*

$$t_0 \in \left[ \bigcap_{1 \leq i \leq N} S_0(F_i) \right]' \tag{1.13}$$

para qualquer coleção finita de distribuições  $F_1, \dots, F_N \in \mathcal{F}$ . Se

$$F_1 \prec F_2 \text{ se, e somente se, } \lim_{t \rightarrow t_0} \frac{\Phi_{F_2}(t)}{\Phi_{F_1}(t)} = 0, \tag{1.14}$$

é uma ordem total em  $\mathcal{F}$ , então a classe  $\mathcal{H}$  de misturas finitas de distribuições de  $\mathcal{F}$  é identificável.

**Exemplo 1.19.** Considere  $\mathcal{F}$  a classe de distribuições lognormal dada por  $\mathcal{F} = \left\{ F : F(x; \mu, \sigma) = \int_0^x \frac{\exp\left\{-\frac{(\log v - \mu)^2}{2\sigma^2}\right\}}{\sigma v \sqrt{2\pi}} dv, \mu \in \mathbb{R}, \sigma > 0, x > 0 \right\}$ . Então a classe  $\mathcal{H}$  de misturas de  $\mathcal{F}$  é identificável.

**Demonstração de 1.19 :** Seja  $M$  uma aplicação que transforma uma função de distribuição  $F \in \mathcal{F}$  em uma função geradora de momentos de  $Y = \log X$ , onde  $X$  tem distribuição lognormal. A aplicação  $M$  é linear e

$$\begin{aligned} M[F(y; \mu, \sigma)] &= E[\exp\{tY\}] = \int_{-\infty}^{\infty} \exp\{ty\} \frac{1}{\sigma \sqrt{2\pi}} \exp\left\{-\frac{(y - \mu)^2}{2\sigma^2}\right\} dy \\ &= \int_{-\infty}^{\infty} \exp\{t(z + \mu)\} \frac{1}{\sigma \sqrt{2\pi}} \exp\{-z^2/2\sigma^2\} dz \\ &= \exp\{\mu t\} \int_{-\infty}^{\infty} \frac{1}{\sigma \sqrt{2\pi}} \exp\{tz - (z^2/2\sigma^2)\} dz, \quad t \in \mathbb{R}. \end{aligned}$$

Neste caso  $S_0(F(y; \mu, \sigma)) = \mathbb{R}$ . Observe que

$$tz - \frac{z^2}{2\sigma^2} = -\frac{(z - \sigma^2 t)^2}{2\sigma^2} + \frac{\sigma^2 t^2}{2}.$$

Portanto,

$$M[F(y; \mu, \sigma)] = \exp\{\mu t\} \exp\{\sigma^2 t^2/2\} \int_{-\infty}^{\infty} \frac{1}{\sigma \sqrt{2\pi}} \exp\left\{-\frac{(z - \sigma^2 t)^2}{2\sigma^2}\right\} dy.$$

Como esta última integral representa a integral da densidade da normal  $N(\sigma^2 t, \sigma^2)$ , então

$$M[F(y; \mu, \sigma)] = \Phi_{(\mu, \sigma)}(t) = \exp\{t\mu + (t^2 \sigma^2/2)\}, \quad t \in (-\infty, +\infty) \text{ para } F \in \mathcal{F}. \text{ Note que}$$

$$\lim_{t \rightarrow \infty} \frac{\Phi_{(\mu_2, \sigma_2)}(t)}{\Phi_{(\mu_1, \sigma_1)}(t)} = \lim_{t \rightarrow \infty} \frac{\exp\left(t\mu_2 + \frac{t^2 \sigma_2^2}{2}\right)}{\exp\left(t\mu_1 + \frac{t^2 \sigma_1^2}{2}\right)} = \lim_{t \rightarrow \infty} \exp\{t(\mu_2 - \mu_1)\} \exp\left\{\frac{t^2}{2}(\sigma_2^2 - \sigma_1^2)\right\}.$$

Portanto,  $F(y; \mu_1, \sigma_1) \prec F(y; \mu_2, \sigma_2) \Leftrightarrow \lim_{t \rightarrow \infty} \frac{\Phi_{(\mu_2, \sigma_2)}(t)}{\Phi_{(\mu_1, \sigma_1)}(t)} = 0 \Leftrightarrow [\sigma_2 < \sigma_1]$  ou  $[\sigma_2 = \sigma_1 \text{ e } \mu_2 < \mu_1]$ , é uma ordem total em  $\mathcal{F}$ . Pelo corolário (1.18), a classe  $\mathcal{H}$  de mistura finita de distribuições de  $\mathcal{F}$  é identificável. ■

A identificabilidade de misturas finitas foi também mostrada para classes de misturas finitas geradas pela união de diferentes distribuições, tais como as distribuições Gamma e Weibull, em Atienza et al. [2].

---

# O Modelo

---

## 2.1 Introdução

Com o crescente interesse em melhorar a utilização dos recursos da madeira e o desempenho dos produtos derivados da madeira, tornou-se imprescindível o desenvolvimento de métodos para avaliar a qualidade da madeira de árvores plantadas, para servir a programas de melhoramento genético e avaliar os métodos silviculturais. Uma das propriedades importantes para avaliar a qualidade da madeira é o comprimento de suas fibras. Assim, é necessário que os procedimentos para medir essas propriedades da madeira em árvores sejam rápidos, de baixo custo e não destrutivos (em relação a árvore), por isso, os melhores métodos são aqueles que podem ser executados sem a derrubada da árvore. Neste trabalho apresentamos um método que foi desenvolvido por Mörling et al.[13] e Svensson et al.[18], para investigação da distribuição do comprimento das fibras.

Um método de amostragem para estimar a distribuição do comprimento das fibras da madeira, considerado rápido e não destrutivo, é feito por meio da utilização de núcleos de incrementos. Um núcleo de incremento é uma amostra cilíndrica de madeira de  $5mm$  de diâmetro, extraída de uma árvore com uma broca especial (ver Figura 2.1). Este método de extração de dados é considerado o melhor para análise de crescimento. Instrumentos ópticos para medição automática do comprimento das fibras tem sido desenvolvidos, que torna possível medir grandes quantidades de fibras num curto espaço de tempo. Contudo, encontrar a distribuição do comprimento no núcleo de incremento não é tarefa fácil. Em primeiro lugar, em uma simples amostra pode conter fibras não cortadas, como também fibras cortadas uma ou até duas vezes. Uma explicação para tal fato é que o núcleo de incremento é extraído horizontalmente na árvore, enquanto as fibras crescem

verticalmete (ver Figura 2.1). Todas as fibras com comprimento superior ao núcleo de incremento certamente serão cortadas. Esta forma de censura deve ser considerada quando as amostras de núcleos de incrementos são usadas para estimar a distribuição do comprimento das fibras em árvores que estão plantadas. Outro fato que dificulta este tipo de coleta de dados, é que a amostra não contém apenas as fibras de interesse (traqueídeos), mas também outros tipos de células, tais como parêquima radial e traqueídes radial, que aqui chamaremos de microfibras. O comprimento médio das fibras varia entre 2 e 6mm para diferentes espécies de coníferas, enquanto o comprimento médio das microfibras é cerca de 10 a 20 vezes menor. Outra questão que precisa ser levada em consideração é o problema de viés de comprimento, resultante do fato que as células maiores são mais propensas a serem coletadas em um núcleo de incremento.

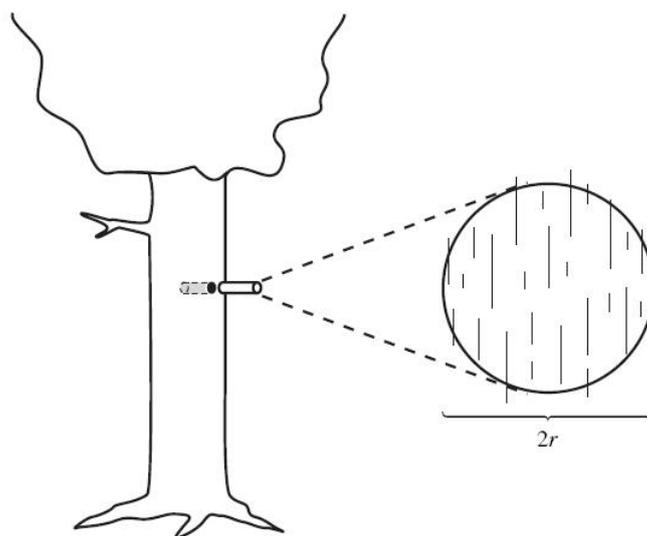


Figura 2.1: Um núcleo de incremento cilíndrico de diâmetro  $2r$  é extraído perpendicular à direção das fibras na árvore. A seção transversal do núcleo incremento mostra como as fibras da amostra podem ser tanto cortadas como não cortadas.

O procedimento para medir os comprimentos individuais das células em uma amostra de núcleos de incrementos é feito em três etapas. Na primeira, as amostras são maceradas em uma mistura de ácido acético e peróxido de hidrogênio por cerca de 48 horas a  $70^{\circ}\text{C}$ . Então, as amostras são cuidadosamente lavadas com água e as células são separadas pela agitação da solução celular. Depois disto, os comprimentos das fibras e das microfibras na subamostra resultante da interrupção, podem ser medidos num microscópio óptico ou por um analisador automático de fibras. Em um microscópio, é possível perceber a diferença entre fibras e microfibras, como também se uma célula foi ou não cortada. No

entanto, uma análise microscópica não é automatizada e portanto a quantidade de células que podem ser medidas, na prática, é limitada. Por outro lado, um analisador de fibras ópticas pode medir grandes amostras automaticamente em um curto espaço de tempo, mas não pode distinguir automaticamente microfibras de fibras, como também não é capaz de dizer se uma célula é ou não é cortada. Isto torna o analisador de fibras ópticas em um procedimento mais rápido e atraente para o estudo em questão. Agora, pela forma como as amostras são medidas, gerando incertezas quanto à classificação de uma célula em fibra ou microfibra, e também se esta é ou não cortada, é razoável supormos que os comprimentos observados vêm de uma versão censurada de uma mistura de distribuições de comprimento das microfibras e fibras na árvore. Na Seção 2.2 apresentamos as variáveis aleatórias presentes neste estudo, bem como suas distribuições. Na Seção 2.3 tratamos da identificabilidade da classe de misturas finitas das distribuições envolvidas nesta análise, e na Seção 2.4 descrevemos o algoritmo MCEM proposto por Svensson et al.[18], que é um método apropriado para lidar com a falta de informação, gerada pela censura, para estimar os parâmetros do modelo.

## 2.2 Comprimento da Fibra da Madeira

Com o objetivo de descrever o método para medir as propriedades da madeira, apresentamos inicialmente as variáveis aleatórias envolvidas no modelo, bem como as distribuições das mesmas.

Considere as seguintes variáveis aleatórias relacionadas ao modelo:

$W$  = o comprimento de uma célula (microfibra ou fibra) em uma árvore que está plantada,

$Y$  = o comprimento real de uma célula que, pelo menos aparece parcialmente no núcleo de incremento, e

$X$  = o comprimento correspondente à  $Y$ , visto no núcleo de incremento.

Considerando que a célula pode ter sido cortada quando seu comprimento é obtido a partir do núcleo de incremento, temos que  $P(X \leq Y) = 1$ . Os comprimentos das células são obtidos a partir do núcleo de incremento e, portanto, são dados referentes à distribuição de  $X$ . Contudo, nosso interesse real é na distribuição de  $W$ , e a distribuição de  $Y$  auxiliará na conexão entre às distribuições de  $X$  e de  $W$ . Como nas observações não é possível distinguir as microfibras das fibras, quando os comprimentos são medidos num analisador de fibra óptica, supomos que as distribuições de  $X$ ,  $Y$  e  $W$  são algum tipo de mistura de distribuições de comprimento de uma microfibra e de uma fibra.

Seja  $Y$  uma variável aleatória contínua com função de densidade  $f_Y(y, \theta)$  onde  $\theta$  é um parâmetro desconhecido. Para estimar o parâmetro  $\theta$  do modelo, faremos uso do algoritmo MCEM, e isto pode ser visto na Seção 2.4. Neste estudo a distribuição de  $Y$  é assumida da seguinte forma:

$$f_Y(y; \theta) = \varepsilon f_{Y_1}(y; \phi_1) + (1 - \varepsilon) f_{Y_2}(y; \phi_2), 0 < y < \infty \quad (2.1)$$

onde  $\varepsilon$  é a proporção das microfibras do núcleo de incremento e  $f_{Y_1}(y; \phi_1)$  e  $f_{Y_2}(y; \phi_2)$  são as funções densidades do comprimento real das microfibras e fibras, respectivamente, que pelo menos aparecem parcialmente no núcleo de incremento. Desta forma temos que  $\theta = (\varepsilon, \phi_1, \phi_2)$ .

### 2.2.1 A Distribuição de $X$

Seja o diâmetro do núcleo de incremento circular igual a  $2r$ . Para encontrar a distribuição de  $X$ , que está definida em  $(0, 2r]$ , notemos que para  $0 < x \leq 2r$ ,

$$\begin{aligned} P(X \leq x) &= \int_0^x f_X(x'; \theta) dx' \\ &= \int_0^x \int_0^\infty f_{X,Y}(x', y; \theta) dy dx' \\ &= \int_0^x \int_0^\infty f_{X|Y}(x'|y; \theta) f_Y(y; \theta) dy dx' \\ &= \int_0^\infty f_Y(y; \theta) \int_0^x f_{X|Y}(x'|y; \theta) dx' dy \\ &= \int_0^\infty P(X \leq x|Y = y) f_Y(y; \theta) dy. \end{aligned} \quad (2.2)$$

Observe que se  $x \leq 0$ , então  $P(X \leq x) = 0$  e que se  $x > 2r$ , então  $P(X \leq x) = 1$ . Sendo assim, devemos encontrar a distribuição condicional de  $X$  dado  $Y = y$ , que possui suporte em  $(0, y]$  se  $y \leq 2r$  e  $(0, 2r]$  se  $y > 2r$ . Note também que, a probabilidade  $P(X \leq x|Y = y) = 1$  para  $x > y$ , já que  $X$  é possivelmente a versão censurada de  $Y$ .

Primeiramente, analisamos o caso quando  $x = y$ , correspondendo a uma célula não-cortada. Uma célula pode ser não-cortada somente se o comprimento real  $y$  é menor do que o diâmetro do núcleo de incremento, isto é,  $y \leq 2r$ . Na Figura 2.2 podemos ver a representação de uma célula de comprimento real  $Y = y$ , com  $(y \leq 2r)$ , e as regiões  $T(y)$  e  $U(y)$ , auxiliares na obtenção da probabilidade  $P(X = y|Y = y)$ , onde:

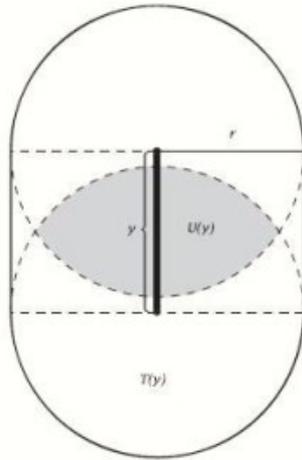


Figura 2.2:  $T(y)$  região total na qual o centro do núcleo de incremento pode estar situado de modo a ter no mínimo parte da célula, de comprimento real  $Y = y$ , no núcleo de incremento.  $U(y)$  é a região sombreada na qual, se o centro do núcleo de incremento estiver ali localizado, então a célula será não-cortada.

$T(y)$  = região total na qual o centro do núcleo de incremento pode estar situado de modo a ter no mínimo parte da célula, de comprimento real  $Y = y$ , no núcleo de incremento,

e

$U(y)$  = região sombreada na qual, se o centro do núcleo de incremento estiver ali localizado, então a célula será não-cortada.

Assim, a probabilidade da célula ser não-cortada dado que a célula aparece no núcleo de incremento é dada por:

$$p_{nc} = P(X = y|Y = y) = \begin{cases} u(y)/t(y), & \text{se } 0 \leq y \leq 2r \\ 0, & \text{se } y > 2r \end{cases} \quad (2.3)$$

onde  $t(y)$  e  $u(y)$  são áreas das regiões  $T(y)$  e  $U(y)$ , respectivamente. Essas áreas podem ser facilmente calculadas e mostradas como sendo:

$$t(y) = \pi r^2 + 2ry$$

e

$$u(y) = 2r^2 \arcsen \left( \frac{\sqrt{4r^2 - y^2}}{2r} \right) - \frac{y}{2} \sqrt{4r^2 - y^2}. \quad (2.4)$$

Da equação (9) em Mörling et al. (2003), fazendo  $C_i = (0, x]$ , para  $x < y$ , temos que:

$$P(X \leq x|Y = y) =$$

$$= \frac{1}{t(y)} \left[ t(y) - 2r^2 \arcsen \left( \frac{\sqrt{4r^2 - x^2}}{2r} \right) + \left( \frac{3x}{2} - y \right) \sqrt{4r^2 - x^2} \right], \quad 0 \leq x \leq 2r. \quad (2.5)$$

Pelo teorema fundamental do cálculo, a equação (2.5) pode ser escrita como

$$P(X \leq x|Y = y) = \begin{cases} \int_0^x f_{X|Y}(w|y)dw, & \text{se } 0 < x < 2r \text{ e } x < y \\ 1, & \text{se } x > \min(y, 2r) \end{cases}, \quad (2.6)$$

onde

$$f_{X|Y}(x|y) = \frac{8r^2 - 3x^2 + xy}{t(y)\sqrt{4r^2 - x^2}},$$

obtida derivando (2.5) com respeito a  $x$ .

Por (2.3) existe uma probabilidade positiva da célula ser não-cortada se  $y < 2r$ . Isto implica que a função de distribuição condicional de  $X$  dado  $Y = y$  tem um salto em  $y$  se  $y < 2r$ , e portanto é uma mistura de uma distribuição discreta e uma contínua, com uma probabilidade positiva de  $X = y$ , para  $y < 2r$ .

Retornando à distribuição de  $X$  e considerando que  $P(X \leq x|Y = y) = 1$  para  $0 < y < x$ , das equações (2.2), (2.4) e (2.5) podemos escrever para  $0 < x \leq 2r$

$$\begin{aligned} P(X \leq x) &= \int_0^x P(X \leq x|Y = y)f_Y(y; \theta)dy + \int_x^\infty P(X \leq x|Y = y)f_Y(y; \theta)dy \\ &= \int_0^x f_Y(y; \theta)dy + \int_x^\infty P(X \leq x|Y = y)f_Y(y; \theta)dy \\ &= P(Y \leq x) + \int_x^\infty \left[ 1 + \frac{1}{t(y)}(-u(x) + (x - y)\sqrt{4r^2 - x^2}) \right] f_Y(y; \theta)dy \\ &= 1 - \int_x^\infty \left[ u(x) - (x - y)\sqrt{4r^2 - x^2} \right] \frac{f_Y(y; \theta)}{t(y)} dy. \end{aligned} \quad (2.7)$$

A função densidade de  $X$  é dada por:

$$f_X(x; \theta) = \begin{cases} p_{nc}(x)f_Y(x; \theta) + \int_x^\infty f_{X|Y}(x|y)f_Y(y; \theta)dy, & \text{se } 0 < x < 2r \\ 0, & \text{c.c.} \end{cases}. \quad (2.8)$$

**Demonstração de 2.8 :** Para verificar (2.8), devemos mostrar que

$$\int_0^x f_X(x'; \theta) dx' = P(X \leq x), \forall x \in [0, \infty),$$

onde  $P(X \leq x)$  é dada na equação (2.7). Note que  $p_{nc}(x) = \frac{u(x)}{t(x)}$  se  $x \in [0, 2r]$ , então para  $0 < x \leq 2r$ , temos:

$$\int_0^x f_X(x'; \theta) dx' = \int_0^x \frac{u(x')}{t(x')} f_Y(x'; \theta) dx' + \int_0^x \int_{x'}^{\infty} f_{X|Y}(x'|y) f_Y(y; \theta) dy dx'. \quad (2.9)$$

Como  $f_{X|Y}(x'|y) f_Y(y; \theta)$  é mensurável e não - negativa, a ordem de integração pode ser invertida, de modo que a integral dupla em (2.9) pode ser escrita como:

$$\begin{aligned} \int_0^x \int_{x'}^{\infty} f_{X|Y}(x'|y) f_Y(y; \theta) dy dx' &= \int_0^x f_Y(y; \theta) dy \int_0^y f_{X|Y}(x'|y) dx' \\ &+ \int_x^{\infty} f_Y(y; \theta) dy \int_0^x f_{X|Y}(x'|y) dx' \\ &= \int_0^x \frac{f_Y(y; \theta)}{t(y)} \left[ t(y) - 2r^2 \arcsen \left( \frac{\sqrt{4r^2 - y^2}}{2r} \right) + \left( \frac{3y}{2} - y \right) \sqrt{4r^2 - y^2} \right] dy \\ &+ \int_x^{\infty} \frac{f_Y(y; \theta)}{t(y)} \left[ t(y) - 2r^2 \arcsen \left( \frac{\sqrt{4r^2 - x^2}}{2r} \right) + \left( \frac{3x}{2} - y \right) \sqrt{4r^2 - x^2} \right] dy. \end{aligned} \quad (2.10)$$

Sabendo que

$$\arcsen \left( \frac{\sqrt{4r^2 - y^2}}{2r} \right) = \frac{\pi}{2} - \arcsen \left( \frac{y}{2r} \right),$$

então de (2.4), temos que

$$u(y) = \pi r^2 - 2r^2 \arcsen \left( \frac{y}{2r} \right) - \frac{y}{2} \sqrt{4r^2 - y^2}. \quad (2.11)$$

Lembrando também que  $t(y) = \pi r^2 + 2ry$ , de (2.10), temos que:

$$\begin{aligned}
& \int_0^x \int_{x'}^{\infty} f_{X|Y}(x'|y) f_Y(y; \theta) dy dx' = \\
& = \int_0^x \frac{f_Y(y; \theta)}{t(y)} \left[ t(y) - 2r^2 \left( \frac{\pi}{2} - \arcsen \left( \frac{y}{2r} \right) \right) + \frac{y}{2} \sqrt{4r^2 - y^2} \right] dy + \\
& + \int_x^{\infty} \frac{f_Y(y; \theta)}{t(y)} \left[ t(y) - 2r^2 \left( \frac{\pi}{2} - \arcsen \left( \frac{x}{2r} \right) \right) + \left( \frac{3x}{2} - y \right) \sqrt{4r^2 - x^2} \right] dy = \\
& = \int_0^x (t(y) - u(y)) \frac{f_Y(y; \theta)}{t(y)} dy + \\
& + \int_x^{\infty} (t(y) - u(x) + (x - y) \sqrt{4r^2 - x^2}) \frac{f_Y(y; \theta)}{t(y)} dy. \tag{2.12}
\end{aligned}$$

Finalmente, de (2.9) e (2.11) obtemos:

$$\int_0^x f_X(x'; \theta) dx' = 1 - \int_x^{\infty} [u(x) - (x - y) \sqrt{4r^2 - x^2}] \frac{f_Y(y; \theta)}{t(y)} dy.$$

■

Do que acabamos de mostrar, temos que  $X$  é uma variável aleatória contínua. Note que  $Y$  tendo uma distribuição de mistura conforme em (2.1), então  $X$  terá também uma distribuição de mistura onde sua densidade é dada por:

$$f_X(x; \theta) = \varepsilon f_{X_1}(x; \phi_1) + (1 - \varepsilon) f_{X_2}(x; \phi_2) \tag{2.13}$$

onde para  $t = 1, 2$

$$\begin{aligned}
f_{X_t}(x; \phi_t) &= p_{nc}(x) f_{Y_t}(x; \phi_t) + \int_x^{\infty} f_{X/Y}(x/y) f_{Y_t}(y; \phi_t) dy \\
&= p_{nc}(x) f_{Y_t}(x; \phi_t) + \int_x^{\infty} \frac{R(x, y)}{\sqrt{4r^2 - x^2}} f_{Y_t}(y; \phi_t) dy, \text{ se } 0 < x < 2r, \tag{2.14}
\end{aligned}$$

e 0 caso contrário. Aqui

$$R(x, y) = \frac{8r^2 - 3x^2 + yx}{\pi r^2 + 2ry}.$$

### 2.2.2 A Distribuição de $Y$ dado $X$

Vamos agora analisar a distribuição condicional de  $Y$  dado  $X = x$ , que tem suporte no intervalo  $[x, \infty)$  com  $x \leq 2r$ . Note que

$$P(Y \in (y, y+\Delta y) | X \in (x, x+\Delta x)) = \frac{P(X \in (x, x + \Delta x) | Y \in (y, y + \Delta y))P(Y \in (y, y + \Delta y))}{P(X \in (x, x + \Delta x))}. \quad (2.15)$$

Como  $f_{X|Y}(x|y)$  e  $f_Y(y; \theta)$  são contínuas com respeito a  $y$ , segue por (2.6) que para  $y > x$ , (2.15) pode ser aproximada por  $f_{Y|X}(y|x; \theta)\Delta y$ , onde

$$f_{Y|X}(y|x; \theta) = \frac{f_{X|Y}(x|y)f_Y(y; \theta)}{f_X(x; \theta)}.$$

Note que  $f_{Y|X}(y|x; \theta)$  também é contínua em relação a  $y$ . Assim, para  $y > x$  e  $0 < x < 2r$  podemos escrever

$$P(Y \leq y | X = x) = 1 - P(Y > y | X = x) = 1 - \int_y^\infty f_{Y|X}(y'|x; \theta) dy'. \quad (2.16)$$

Resta, portanto, descrevermos a probabilidade condicional de uma célula ser não-cortada, dado que o seu comprimento observado no núcleo de incremento é  $x$ , ou seja,  $P(Y = x | X = x)$ . Esta probabilidade pode ser obtida da seguinte forma:

$$\begin{aligned} P(Y = x | X = x) &= \lim_{\Delta x \rightarrow 0^+} P(Y \leq x + \Delta x | X = x) \\ &= 1 - \int_x^\infty f_{Y|X}(y|x; \theta) dy \\ &= 1 - \int_x^\infty \frac{f_{X|Y}(x/y)f_Y(y; \theta)}{f_X(x; \theta)} dy \\ &= 1 - \frac{1}{f_X(x; \theta)} (f_X(x; \theta) - p_{nc}(x)f_Y(x, \theta)) \\ &= \frac{p_{nc}(x)f_Y(x; \theta)}{f_X(x; \theta)}, \text{ para } 0 < x < 2r. \end{aligned} \quad (2.17)$$

Então temos uma probabilidade positiva de uma célula ser não-cortada com comprimento  $x$  observado, lembrando que  $p_{nc}(y)$  é a probabilidade de uma célula ser não-cortada no núcleo de incremento, dado que o seu comprimento real é  $Y = y$ . Das equações (2.16) e (2.17), temos que a distribuição condicional de  $Y$  dado  $X = x$  é uma mistura de uma distribuição discreta e uma distribuição contínua com probabilidade positiva de  $Y = x$ . Desta forma, temos que para  $0 < x < 2r$ ,

$$P(Y \leq y|X = x) = \begin{cases} 0, & \text{se } y < x \\ \frac{p_{nc}(x)f_Y(x; \theta)}{f_X(x; \theta)}, & \text{se } y = x \\ \frac{p_{nc}(x)f_Y(x; \theta)}{f_X(x; \theta)} + \int_x^y f_{Y|X}(y'|x; \theta)dy', & \text{se } y > x. \end{cases} \quad (2.18)$$

Observe que se a distribuição de  $Y$  é uma mistura, assim como dado em (2.1), então a distribuição condicional de  $Y$  dado  $X$  também é. Neste caso em particular, é uma mistura de duas distribuições, de  $Y_1$  e  $Y_2$  dado  $X = x$  com componente de proporção de mistura  $\varepsilon$ , e é dada da seguinte forma:

$$P(Y_t \leq y|X = x) = \begin{cases} 0, & \text{se } y < x \\ \frac{p_{nc}(x)f_{Y_t}(x; \theta_t)}{f_X(x; \theta)}, & \text{se } y = x \\ \frac{p_{nc}(x)f_{Y_t}(x; \theta_t)}{f_X(x; \theta)} + \int_x^y f_{Y_t|X}(y'|x; \theta)dy', & \text{se } y > x, \end{cases}$$

onde  $f_{Y_t|X}(y|x; \theta) = \frac{f_{X|Y}(x|y)f_{Y_t}(y; \theta)}{f_X(x; \theta)}$ , para  $t = 1, 2$ .

### 2.2.3 A Distribuição de $W$

A partir da distribuição de  $Y$  podemos encontrar a distribuição do comprimento das células  $W$  em uma árvore que está plantada. Pela definição das variáveis aleatórias  $W$  e  $Y$ , podemos observar que a probabilidade de uma célula de comprimento real  $W = y$  aparecer ao menos parcialmente no núcleo de incremento é proporcional à área  $t(y) = \pi r^2 + 2ry$  de  $T(y)$  (ver Figura 2.2). Sabendo que  $f_Y(y; \theta)$  é a função densidade de  $Y$  e  $f_W(y; \theta)$  é a função densidade de  $W$ , pela fórmula de Bayes temos que:

$$f_Y(y; \theta) \propto (\pi r^2 + 2ry)f_W(y; \theta). \quad (2.19)$$

Do fato de  $f_Y(y; \theta)$  ser uma densidade e, conseqüentemente,  $\int_0^\infty f_Y(y; \theta)dy = 1$ , é possível calcular a constante de normalização  $[\pi r^2 + 2rE[W]]^{-1}$  de (2.19). Logo, a função densidade

do comprimento da célula numa árvore plantada é dada por:

$$f_W(y; \theta) = \frac{\pi r^2 + 2rE[W]}{\pi r^2 + 2ry} f_Y(y; \theta). \quad (2.20)$$

Desta forma, se  $f_Y(y; \theta)$  é conhecida, então  $f_W(y; \theta)$  pode ser determinada, uma vez que  $E[W]$  depende apenas de  $f_Y(y; \theta)$ , isto porque

$$E[W] = \frac{1}{2} \left[ \int_0^\infty \frac{f_Y(y; \theta)}{\pi r + 2y} dy \right]^{-1} - \frac{\pi r}{2}. \quad (2.21)$$

**Demonstração de 2.21 :** Integrando ambos os lados da equação (2.20) e rearranjando os termos, temos:

$$\begin{aligned} \int_0^\infty f_W(y; \theta) dy &= \int_0^\infty \frac{\pi r^2 f_Y(y; \theta)}{\pi r^2 + 2ry} dy + \int_0^\infty \frac{2rE[w]f_Y(y; \theta)}{\pi r^2 + 2ry} dy \Leftrightarrow \\ \Leftrightarrow 1 &= \pi r^2 \int_0^\infty \frac{f_Y(y; \theta)}{\pi r^2 + 2ry} dy + 2rE[W] \int_0^\infty \frac{f_Y(y; \theta)}{\pi r^2 + 2ry} dy \Leftrightarrow \\ \Leftrightarrow 2rE[W] &= \frac{1 - \pi r^2 \int_0^\infty \frac{f_Y(y; \theta)}{\pi r^2 + 2ry} dy}{\int_0^\infty \frac{f_Y(y; \theta)}{\pi r^2 + 2ry} dy} \Leftrightarrow \\ \Leftrightarrow E[W] &= \frac{1}{2} \left[ \int_0^\infty \frac{f_Y(y; \theta)}{\pi r + 2y} dy \right]^{-1} - \frac{\pi r}{2}. \end{aligned}$$

■

A função densidade de  $W$  pode ser escrita como a seguinte mistura de distribuições:

$$f_W(y; \theta) = \tilde{\varepsilon} f_{W_1}(y; \phi_1) + (1 - \tilde{\varepsilon}) f_{W_2}(y; \phi_2), \quad (2.22)$$

onde  $\tilde{\varepsilon} = \tilde{\varepsilon}(\theta)$  é a proporção de microfibras na árvore que está plantada. Na equação acima,  $f_{W_1}(y; \phi_1)$  e  $f_{W_2}(y; \phi_2)$  são as funções densidades dos comprimentos das microfibras e das fibras, respectivamente, na árvore plantada. Ao utilizar a equação (2.20) para as funções densidades  $f_{W_1}(y; \phi_1)$  e  $f_{W_2}(y; \phi_2)$ , temos que:

$$f_{W_t}(y; \phi_t) = \frac{\pi r^2 + 2rE_t[W]}{\pi r^2 + 2ry} f_{Y_t}(y; \phi_t), \text{ para } t = 1, 2. \quad (2.23)$$

O fato da proporção  $\tilde{\varepsilon}$  ser uma função do parâmetro  $\theta$  e não ser igual a proporção de microfibras no núcleo de incremento,  $\varepsilon$ , é explicado pela fórmula (2.24) abaixo. Isto é uma consequência do comprimento ser tendencioso. Entretanto, se  $\varepsilon$  é conhecido, então

$\tilde{\varepsilon}$  pode ser determinado por

$$\tilde{\varepsilon} = \frac{\varepsilon[\pi r + 2E[W]]}{\pi r + 2E_1[W]}. \quad (2.24)$$

**Demonstração de 2.24 :** Substituindo as funções  $f_{W_1}(y; \phi_1)$  e  $f_{W_2}(y; \phi_2)$  de (2.23) em (2.22), temos a função densidade de  $W$ , ou seja,

$$f_W(y; \theta) = \tilde{\varepsilon} \left[ \frac{\pi r^2 + 2rE_1[W]}{\pi r^2 + 2ry} \right] f_{Y_1}(y; \phi_1) + (1 - \tilde{\varepsilon}) \left[ \frac{\pi r^2 + 2rE_2[W]}{\pi r^2 + 2ry} \right] f_{Y_2}(y; \phi_2), \quad (2.25)$$

onde  $E_t(W) = \int_0^\infty w f_{W_t}(w; \phi_t) dw$ . Daí, pelo fato de  $(1 - \tilde{\varepsilon})E_2[W] = E[W] - \tilde{\varepsilon}E_1[W]$ , podemos escrever a equação (2.25) como:

$$\begin{aligned} f_W(y; \theta) &= \tilde{\varepsilon} \left[ \frac{\pi r^2 + 2rE_1[W]}{\pi r^2 + 2ry} \right] f_{Y_1}(y; \phi_1) + \\ &+ \frac{(1 - \tilde{\varepsilon})\pi r^2}{\pi r^2 + 2ry} f_{Y_2}(y; \phi_2) + \frac{2r(E[W] - \tilde{\varepsilon}E_1[W])}{\pi r^2 + 2ry} f_{Y_2}(y; \phi_2). \end{aligned} \quad (2.26)$$

Por outro lado, substituindo (2.1) na equação (2.20) obtemos outra expressão para a função densidade de  $W$ , da seguinte maneira:

$$f_W(y; \theta) = \frac{\pi r^2 + 2rE[W]}{\pi r^2 + 2ry} (\varepsilon f_{Y_1}(y; \phi_1) + (1 - \varepsilon) f_{Y_2}(y; \phi_2)). \quad (2.27)$$

Portanto das equações (2.26) e (2.27), temos que:

$$\begin{aligned} &\tilde{\varepsilon} \left[ \frac{\pi r^2 + 2rE_1[W]}{\pi r^2 + 2ry} \right] f_{Y_1}(y; \phi_1) + \frac{(1 - \tilde{\varepsilon})\pi r^2}{\pi r^2 + 2ry} f_{Y_2}(y; \phi_2) + \frac{2rE[W]}{\pi r^2 + 2ry} f_{Y_2}(y; \phi_2) - \\ &- \frac{2r\tilde{\varepsilon}E_1[W]}{\pi r^2 + 2ry} f_{Y_2}(y; \phi_2) - \frac{\pi r^2 \varepsilon f_{Y_1}(y; \phi_1)}{\pi r^2 + 2ry} - \frac{\pi r^2 (1 - \varepsilon) f_{Y_2}(y; \phi_2)}{\pi r^2 + 2ry} - \\ &- \frac{2r\varepsilon E[W] f_{Y_1}(y; \phi_1)}{\pi r^2 + 2ry} - \frac{2r(1 - \varepsilon)E[W] f_{Y_2}(y; \phi_2)}{\pi r^2 + 2ry} = 0 \Leftrightarrow \end{aligned}$$

$$\begin{aligned} \Leftrightarrow &\pi r^2 f_{Y_2}(y; \phi_2) (1 - \tilde{\varepsilon} - 1 + \varepsilon) + \pi r^2 f_{Y_1}(y; \phi_1) (\tilde{\varepsilon} - \varepsilon) + 2r\tilde{\varepsilon}E_1[W] f_{Y_1}(y; \phi_1) + \\ &+ 2rE[W] f_{Y_2}(y; \phi_2) - 2r\tilde{\varepsilon}E_1[W] f_{Y_2}(y; \phi_2) - 2r\varepsilon E[W] f_{Y_1}(y; \phi_1) - \\ &- 2rE[W] f_{Y_2}(y; \phi_2) + 2r\varepsilon E[W] f_{Y_2}(y; \phi_2) = 0 \Leftrightarrow \end{aligned}$$

$$\Leftrightarrow \pi r^2(\tilde{\varepsilon} - \varepsilon)[f_{Y_1}(y; \phi_1) - f_{Y_2}(y; \phi_2)] + 2r\tilde{\varepsilon}E_1[W][f_{Y_1}(y; \phi_1) - f_{Y_2}(y; \phi_2)] - 2r\varepsilon E[W][f_{Y_1}(y; \phi_1) - f_{Y_2}(y; \phi_2)] = 0 \Leftrightarrow$$

$$\Leftrightarrow \pi r^2\tilde{\varepsilon} - \pi r^2\varepsilon + 2r\tilde{\varepsilon}E_1[W] - 2r\varepsilon E[W] = 0 \Leftrightarrow$$

$$\Leftrightarrow \tilde{\varepsilon}[\pi r^2 + 2rE_1[W]] = \varepsilon[\pi r^2 + 2rE[W]] \Leftrightarrow$$

$$\Leftrightarrow \tilde{\varepsilon} = \frac{\varepsilon[\pi r + 2E[W]]}{\pi r + 2E_1[W]}.$$

■

## 2.3 Identificabilidade

Na Seção 1.4 definimos e explicamos identificabilidade de uma classe de misturas de distribuições paramétricas. Nesta seção, por intermédio do Teorema 2.1 apresentado a seguir, damos condições suficientes para garantir a identificabilidade da classe de misturas finitas de  $f_{X_t}(x; \phi_t)$  dada por (2.14).

**Teorema 2.1.** *A classe de misturas finitas de  $f_{X_t}(x; \phi_t)$  é identificável se a correspondente classe de misturas de  $f_{Y_t}(y; \phi_t)$  é identificável e  $f_{Y_t}(y; \phi_t)$  é analítica em  $y$ .*

**Demonstração:**

Note que pela equação (2.7), a função de distribuição de  $X_t$  é dada por

$$F_{X_t}(x; \phi_t) = 1 - \int_x^\infty A(x, y) f_{Y_t}(y; \phi_t) dy, \quad x \in (0, 2r),$$

onde

$$A(x, y) = \frac{2r^2 \arcsen\left(\frac{1}{2r} \sqrt{4r^2 - x^2}\right) + \left(y - \frac{3x}{2}\right) \sqrt{4r^2 - x^2}}{\pi r^2 + 2ry}.$$

Para  $x \in (0, 2r)$  temos que

$$\sum_{t=1}^N \varepsilon_t (1 - F_{X_t}(x; \theta_t)) = \sum_{s=1}^{N^*} \varepsilon_s^* (1 - F_{X_s}(x; \theta_s^*)) \Leftrightarrow$$

$$\begin{aligned} \Leftrightarrow \sum_{t=1}^N \varepsilon_t \left( \int_x^\infty A(x, y) f_{Y_t}(y; \phi_t) dy \right) - \sum_{s=1}^{N^*} \varepsilon_s^* \left( \int_x^\infty A(x, y) f_{Y_s}(y; \phi_s^*) dy \right) = 0 &\Leftrightarrow \\ \Leftrightarrow \int_x^\infty A(x, y) g(y) dy = 0, &\quad (2.28) \end{aligned}$$

onde

$$g(y) = \sum_{t=1}^N \varepsilon_t f_{Y_t}(y; \phi_t) - \sum_{s=1}^{N^*} \varepsilon_s^* f_{Y_s}(y; \phi_s^*). \quad (2.29)$$

Como, por hipótese, as misturas finitas da classe de  $f_{Y_t}(\cdot; \phi_t)$  são identificáveis, teremos a identificabilidade para misturas finitas da classe de  $f_{X_t}(\cdot; \phi_t)$  se, para  $x \in (0, 2r)$  a igualdade (2.28) implicar que  $g(y) = 0$ , para todo  $y \in (0, \infty)$ . Sabemos que

$$\int_0^\infty g(y) dy = 0,$$

pois as funções que  $f_{Y_t}(\cdot; \phi_t)$  são densidades e  $\sum_{t=1}^N \varepsilon_t = \sum_{s=1}^{N^*} \varepsilon_s^* = 1$ . Seja

$$G(y) = \frac{g(y)}{\pi r^2 + 2ry}. \quad (2.30)$$

Então a equação (2.28) é equivalente à

$$b(x) \int_x^\infty G(y) dy + \sqrt{4r^2 - x^2} \int_x^\infty y G(y) dy = 0, \quad (2.31)$$

para todo  $x \in (0, 2r)$ , onde

$$b(x) = 2r^2 \arcsen \left( \frac{1}{2r} \sqrt{4r^2 - x^2} \right) - \frac{3x}{2} \sqrt{4r^2 - x^2}.$$

Assim, a equação (2.31) é equivalente à

$$\begin{aligned} \int_x^\infty y G(y) dy &= \frac{-b(x)}{\sqrt{4r^2 - x^2}} \int_x^\infty G(y) dy \\ &= B(x) \int_x^\infty G(y) dy, \end{aligned} \quad (2.32)$$

onde  $B(x) = \frac{-b(x)}{\sqrt{4r^2 - x^2}}$ . Tomando a derivada de (2.32) com respeito a  $x$ , temos que

$$B'(x) \int_x^\infty G(y)dy - B(x)G(x) = -xG(x),$$

e portanto

$$-G(x) = c(x) \int_x^\infty G(y)dy, \quad (2.33)$$

onde  $c(x) = \frac{B'(x)}{(x - B(x))}$  e  $B'(x)$  é a derivada de  $B(x)$ . Seja

$$H(x) = \int_x^\infty G(y)dy, \quad (2.34)$$

então, da equação (2.33), temos que

$$H'(x) - c(x)H(x) = 0, \quad x \in (0, 2r). \quad (2.35)$$

Esta equação diferencial tem a seguinte solução

$$H(x) = K \exp(C(x)), \quad \text{onde } C(x) = \int_0^x c(y)dy, \quad (2.36)$$

e  $K$  é uma constante. Das equações de (2.33) à (2.36), temos que

$$-G(x) = H'(x) = Kc(x) \exp(C(x)), \quad (2.37)$$

e que

$$-G(0) = Kc(0). \quad (2.38)$$

Note que  $c(0) > 0$ , pois

$$\begin{aligned} c(0) &= \frac{B'(0)}{-B(0)} = \frac{-b'(0)\frac{1}{2r} + xb(0)(4r^2 - x^2)^{-3/2}}{4r^2 - x^2} \left( \frac{1}{-B(0)} \right) \\ &= \frac{-b'(0)}{8r^3(B(0))} = \frac{-b'(0)2r}{8r^3b(0)} = \frac{-b'(0)}{4r^2b(0)}, \end{aligned}$$

como  $b'(0) = -3r$ , então

$$c(0) = \frac{3r}{4r^2b(0)} = \frac{3}{4\pi r^3} > 0.$$

Se  $G(0) = 0$ , temos por (2.38) que  $K = 0$  e assim por (2.36) que  $H(x) = 0$  para todo

$x \in (0, 2r)$ . Assim, de (2.34) temos que

$$\int_x^\infty G(y)dy = 0, \quad \forall x \in (0, 2r),$$

e, portanto

$$G(y) = 0, \quad \forall y \in (0, 2r).$$

Esta última afirmação pode ser verificada por continuidade e usando o fato de que  $G$  é função analítica. Sendo assim, de (2.30) temos que

$$g(y) = \sum_{t=1}^N \varepsilon_t f_{Y_t}(y; \phi_t) - \sum_{s=1}^{N^*} \varepsilon_s^* f_{Y_s}(y; \phi_s^*) = 0, \quad y \in (0, 2r). \quad (2.39)$$

Como  $f_{Y_t}(y; \phi_t)$  é analítica em  $y$ , então por (2.29)  $g$  também é analítica. Agora, de (2.39) e do Teorema de Identificabilidade para Funções Analíticas (veja Lima [9]), segue que

$$g(y) = 0, \quad \forall y \in (0, \infty).$$

Daí, e pela hipótese de que a classe de misturas de  $f_{Y_t}(\cdot; \phi_t)$  é identificável, temos que

$$N = N^*$$

e

$$\forall t, \exists s \text{ tal que } \varepsilon_t = \varepsilon_s^*,$$

mostrando assim, que a classe de misturas finitas de  $f_{X_t}(\cdot; \phi_t)$  é identificável. Resta, portanto, mostrar que  $G(0) = 0$ .

Suponha, por contradição, que  $G(0) \neq 0$ . Então  $K = \frac{-G(0)}{c(0)}$  e por (2.37) segue que

$$G(x) = \frac{G(0)c(x)}{c(0)} \exp(C(x)),$$

e, por (2.30)

$$g(y) = (\pi r^2 + 2ry) \frac{G(0)}{c(0)} c(y) \exp(C(y)), \quad y \in (0, 2r). \quad (2.40)$$

Sabemos que o lado esquerdo de (2.40) é analítica em  $y$ , pois  $f_{Y_t}(y; \phi_t)$  é analítica em  $y$ , para  $y \in (0, \infty)$ . No entanto, o lado direito não pode ser analítica, pois  $c(y)$  e  $c(y) \exp(C(y))$  tem singularidade em  $y = 2r$ . Portanto  $G(0)$  deve ser 0. Assim, a prova do Teorema 2.1 está completa. ■

Logo, se sabemos que a classe de misturas finitas de  $f_{Y_t}(y; \phi_t)$  é identificável e  $f_{Y_t}(y; \phi_t)$  é analítica em  $y$ , então a classe de misturas finitas de  $f_{X_t}(x; \phi_t)$  é identificável. A identificabilidade da classe de misturas finitas de  $f_{X_t}(x; \phi_t)$ , quando a correspondente  $f_{X_t}(x; \phi_t)$  é uma densidade lognormal, segue diretamente do Teorema 2.1, já que no caso lognormal,  $f_{Y_t}(y; \phi_t)$  é analítica em  $y$  e as classes de misturas finitas de distribuições lognormais são identificáveis, como visto no exemplo (1.19) na Seção 1.4.

## 2.4 O Algoritmo MCEM

O analisador de fibra ótica não é capaz de registrar se o comprimento  $x$  observado no núcleo de incremento, vem de uma célula não-cortada ( $x = y$ ) ou de uma célula que foi cortada uma ou duas vezes ( $x < y$ ). Aqui mostramos como o algoritmo MCEM pode ser usado para estimar  $\theta$ , que leva em consideração a falta de informação nos dados.

Conforme foi visto na Seção 2.2, temos as seguintes variáveis aleatórias:

$Y$  = o comprimento real de uma célula que, pelo menos aparece parcialmente no núcleo de incremento, e

$X$  = o comprimento correspondente à  $Y$ , visto no núcleo de incremento.

Seja  $X = x$  o valor observado de  $X$  e  $Y = y$  o valor observado de  $Y$ . Pela equação (2.1) temos que  $f_Y(y; \theta) = \varepsilon f_{Y_1}(y; \phi_1) + (1 - \varepsilon) f_{Y_2}(y; \phi_2)$ . Todavia, é complexo executar o cálculo direto do EMV para encontrar  $\theta$  por meio da função de verossimilhança (dado incompleto)

$$\log f_Y(y; \theta) = \log[\varepsilon f_{Y_1}(y; \phi_1) + (1 - \varepsilon) f_{Y_2}(y; \phi_2)], \quad (2.41)$$

devido ao fato desta função ser o logaritmo de uma soma de termos. Para facilitar esse cálculo, temos a alternativa de considerar  $y$  como um dado-incompleto e completá-lo com um dado não-observado  $z$ , de modo que a maximização da verossimilhança dado-completo  $f_{Y,Z}(y, z; \theta)$  seja mais simples. Neste caso, consideramos a variável aleatória  $Z$  definida por

$$Z = \begin{cases} 1, & \text{se a observação é uma microfibrã} \\ 0, & \text{se a observação é uma fibrã} \end{cases},$$

com função de probabilidade

$$f_Z(z; \theta) = \begin{cases} \varepsilon, & \text{se } z = 1 \\ 1 - \varepsilon, & \text{se } z = 0 \end{cases}. \quad (2.42)$$

Desta forma,  $C = (Y, Z)$  é um vetor dado-completo e a função densidade condicional de  $Y$  dado  $Z$  é dada por

$$f_{Y|Z}(y|z; \theta) = \begin{cases} f_{Y_1}(y; \phi_1), & \text{se } z = 1 \\ f_{Y_2}(y; \phi_2), & \text{se } z = 0 \end{cases}, \quad y > 0.$$

Denotando o dado-completo por  $c = (y, z)$ , temos que a função log-verossimilhança dado-completo é dada por

$$\begin{aligned} \log f_C(c; \theta) &= \log f_{Y,Z}(y, z; \theta) = \log[f_{Y|Z}(y|z; \theta)f_Z(z; \theta)] = \\ &= z \log[\varepsilon f_{Y_1}(y; \phi_1)] + (1 - z) \log[(1 - \varepsilon)f_{Y_2}(y; \phi_2)]. \end{aligned} \quad (2.43)$$

Se o dado completo  $c$  fosse conhecido, o EMV de  $\theta$  poderia ser obtido diretamente de (2.43), de forma mais simples que em (2.41), visto que agora temos soma de logaritmos. No entanto, somente o comprimento da célula é observado no núcleo de incremento, isto é, conhecemos apenas a observação  $X = x$ . Sendo assim, podemos obter o EMV de  $\theta$  via algoritmo EM, substituindo o cálculo do  $\arg \max_{\theta} \log f_C(c; \theta)$  pela maximização do valor esperado do  $\log f_C(c; \theta)$  relativo à distribuição condicionada à  $X = x$ ,

$$\begin{aligned} Q(\theta, \theta') &= E[\log f_C(c; \theta) | x; \theta'] = E[Z(\log \varepsilon + \log f_{Y_1}(Y; \phi_1)) | x; \theta'] + \\ &+ E[(1 - Z)(\log(1 - \varepsilon) + \log f_{Y_2}(Y; \phi_2)) | x; \theta']. \end{aligned} \quad (2.44)$$

Note, pelo teorema de Bayes, que

$$\begin{aligned} P(Z = z | X = x, \theta') &= \frac{f_{X|Z}(x|z; \theta')f_Z(z; \theta')}{f_X(x; \theta')} \\ &= \begin{cases} \frac{\varepsilon f_{X_1}(x; \phi'_1)}{f_X(x; \theta')}, & \text{se } z = 1 \\ \frac{(1 - \varepsilon)f_{X_2}(x; \phi'_2)}{f_X(x; \theta')}, & \text{se } z = 0 \end{cases}. \end{aligned} \quad (2.45)$$

Denote  $\tau(x; \theta') = E(Z|x; \theta')$ , assim

$$\tau(x; \theta') = P(Z = 1 | x; \theta') = \frac{\varepsilon' f_{X_1}(x; \phi'_1)}{f_X(x; \theta')}.$$

Agora, usando o condicionamento em  $Z$ , a equação (2.44) pode ser reescrita como

$$\begin{aligned} Q(\theta, \theta') &= \tau(x; \theta') \log \varepsilon + (1 - \tau(x; \theta')) \log(1 - \varepsilon) + \\ &+ \tau(x; \theta') E[\log f_{Y_1}(Y; \phi_1) \mid x, Z = 1; \theta'] + \\ &+ (1 - \tau(x; \theta')) E[\log f_{Y_2}(Y; \phi_2) \mid x, Z = 0; \theta']. \end{aligned} \quad (2.46)$$

**Algoritmo EM.** Observa-se  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  valores independentes da variável  $X$ . Selecciona-se  $\theta^{(0)} \in \Theta$ . Na iteração  $k + 1$ ,

passo E: Calcula-se a esperança

$$\begin{aligned} Q(\theta, \theta^{(k)}) &= E \left[ \log \prod_{i=1}^n f_{Y_i, Z_i}(Y_i, Z_i; \theta) \mid \mathbf{x}; \theta^{(k)} \right] \\ &= \sum_{i=1}^n E [\log f_{Y_i, Z_i}(Y_i, Z_i; \theta) \mid \mathbf{x}; \theta^{(k)}], \end{aligned} \quad (2.47)$$

onde  $Y_1, \dots, Y_n$  são variáveis aleatórias independentes e identicamente distribuídas, onde cada  $Y_i$  tem função de densidade dada por (2.1), e  $Z_1, \dots, Z_n$  são variáveis aleatórias independentes e identicamente distribuídas, onde cada  $Z_i$  tem função de probabilidade dada por (2.42). Os vetores  $(Y_1, Z_1), \dots, (Y_n, Z_n)$  são também independentes e identicamente distribuídos. De (2.44) e (2.46), temos que (2.47) pode ser calculado da seguinte forma

$$\begin{aligned} Q(\theta, \theta^{(k)}) &= \sum_{i=1}^n (\tau(x_i; \theta') \log \varepsilon + (1 - \tau(x_i; \theta')) \log(1 - \varepsilon)) + \\ &+ \sum_{i=1}^n \tau(x_i; \theta') E_{\theta'} [\log f_{Y_1}(Y_i; \phi_1) \mid x_i, Z_i = 1] + \\ &+ \sum_{i=1}^n (1 - \tau(x_i; \theta')) E_{\theta'} [\log f_{Y_2}(Y_i; \phi_2) \mid x_i, Z_i = 0]. \end{aligned} \quad (2.48)$$

passo M: Determina-se

$$\theta^{(k+1)} = \arg \max_{\theta} Q(\theta, \theta^{(k)}).$$

Na iteração  $(k + 1)$ , o cálculo do EMV de  $\theta$  requer resolver a equação

$$\frac{\partial Q(\theta, \theta^{(k)})}{\partial \theta} = 0. \quad (2.49)$$

Podemos maximizar a expressão (2.48) em termos de  $\varepsilon$ ,  $\phi_1$  e  $\phi_2$  independentemente. Fazendo a derivada parcial com respeito à  $\varepsilon$  em (2.49) igual à zero, temos que:

$$\begin{aligned} \frac{\partial Q(\theta, \theta^{(k)})}{\partial \varepsilon} &= \sum_{i=1}^n \left( \tau(x_i; \theta^{(k)}) \frac{1}{\varepsilon} - \frac{1}{(1-\varepsilon)} + \tau(x_i; \theta^{(k)}) \frac{1}{(1-\varepsilon)} \right) = \\ &= \sum_{i=1}^n \left( \tau(x_i; \theta^{(k)}) \frac{1}{\varepsilon(1-\varepsilon)} - \frac{1}{(1-\varepsilon)} \right) = 0. \end{aligned}$$

Daí, segue que o EMV atualizado da proporção das microfibras é

$$\hat{\varepsilon}^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \tau(x_i; \theta^{(k)}). \quad (2.50)$$

Como visto na equação (2.45), então as probabilidades  $\tau(x_i; \theta^{(k)})$ ,  $i = 1, \dots, n$ , são calculadas por

$$\tau(x_i; \theta^{(k)}) = \frac{\varepsilon^{(k)} f_{X_1}(x_i; \phi_1^{(k)})}{f_X(x_i; \theta^{(k)})}. \quad (2.51)$$

**Exemplo 2.2.** *Se as distribuições de  $Y_1$  e  $Y_2$  pertencem a uma família de distribuições lognormais, as soluções de (2.49) para os parâmetros existem de forma fechada.*

De fato, considere

$$f_{Y_t}(y, \phi_t) = \frac{1}{\sigma_t \sqrt{2\pi} y} \exp \left\{ -\frac{(\log y - \mu_t)^2}{2\sigma_t^2} \right\}, \quad y > 0, \quad t = 1, 2.$$

Calculando, primeiramente, a derivada parcial em (2.49) com respeito à  $\mu_1$ , temos que

$$\begin{aligned} &\frac{\partial Q(\theta, \theta^{(k)})}{\partial \mu_1} = \\ &= \frac{\partial}{\partial \mu_1} \left\{ \sum_{i=1}^n \tau(x_i; \theta^{(k)}) E \left[ \log \left( \frac{1}{\sigma_1 \sqrt{2\pi} Y_i} \exp \left\{ -\frac{(\log Y_i - \mu_1)^2}{2\sigma_1^2} \right\} \right) \mid x_i, Z_i = 1; \theta^{(k)} \right] \right\} \\ &= \frac{\partial}{\partial \mu_1} \left\{ \sum_{i=1}^n \tau(x_i; \theta^{(k)}) E \left[ \log(2\pi)^{-1/2} - \log \sigma_1 - \log Y_i - \frac{(\log Y_i - \mu_1)^2}{2\sigma_1^2} \mid x_i, Z_i = 1; \theta^{(k)} \right] \right\} \\ &= \sum_{i=1}^n \tau(x_i; \theta^{(k)}) \left[ E \frac{\partial}{\partial \mu_1} \left( \frac{2\mu_1 \log Y_i - \mu_1^2}{2\sigma_1^2} \mid x_i, Z_i = 1; \theta^{(k)} \right) \right] \\ &= \sum_{i=1}^n \tau(x_i; \theta^{(k)}) \left[ E \left( \frac{\log Y_i}{\sigma_1^2} \mid x_i, Z_i = 1; \theta^{(k)} \right) - \frac{\mu_1}{\sigma_1^2} \right]. \end{aligned}$$

Agora, resolvendo a equação  $\frac{\partial Q(\theta, \theta^{(k)})}{\partial \mu_1} = 0$ , obtemos

$$\hat{\mu}_1^{(k+1)} = \frac{\sum_{i=1}^n \tau(x_i; \theta^{(k)}) E(\log Y_i | x_i, Z_i = 1; \theta^{(k)})}{\sum_{i=1}^n \tau(x_i; \theta^{(k)})}. \quad (2.52)$$

Calculando agora a derivada parcial em (2.49) com respeito à  $\sigma_1$ , temos que

$$\begin{aligned} \frac{\partial Q(\theta, \theta^{(k)})}{\partial \sigma_1} &= \\ &= \frac{\partial}{\partial \sigma_1} \left\{ \sum_{i=1}^n \tau(x_i; \theta^{(k)}) E \left[ \log \left( \frac{1}{\sigma_1 \sqrt{2\pi}} \frac{1}{Y_i} \exp \left\{ -\frac{(\log Y_i - \mu_1)^2}{2\sigma_1^2} \right\} \right) | x_i, Z_i = 1; \theta^{(k)} \right] \right\} \\ &= \frac{\partial}{\partial \sigma_1} \left\{ \sum_{i=1}^n \tau(x_i; \theta^{(k)}) E \left[ \log(2\pi)^{-1/2} - \log \sigma_1 - \log Y_i - \frac{(\log Y_i - \mu_1)^2}{2\sigma_1^2} | x_i, Z_i = 1; \theta^{(k)} \right] \right\} \\ &= \sum_{i=1}^n \tau(x_i; \theta^{(k)}) E \left[ \frac{(\log Y_i)^2 - 2\mu_1 \log Y_i + \mu_1^2}{\sigma_1^3} - \frac{1}{\sigma_1} | x_i, Z_i = 1; \theta^{(k)} \right]. \end{aligned}$$

Resolvendo  $\frac{\partial Q(\theta, \theta^{(k)})}{\partial \sigma_1} = 0$  de (2.52), encontramos

$$\sum_{i=1}^n \tau(x_i; \theta^{(k)}) \sigma_1^2 = \sum_{i=1}^n \tau(x_i; \theta^{(k)}) E[(\log Y_i)^2 | x_i, Z_i = 1] - \mu_1 \sum_{i=1}^n \tau(x_i; \theta^{(k)}),$$

e portanto

$$(\hat{\sigma}_1^2)^{(k+1)} = \frac{\sum_{i=1}^n \tau(x_i; \theta^{(k)}) E[(\log Y_i)^2 | x_i, Z_i = 1; \theta^{(k)}]}{\sum_{i=1}^n \tau(x_i; \theta^{(k)})} - (\hat{\mu}_1^{(k+1)})^2. \quad (2.53)$$

De modo similar para  $\hat{\mu}_2^{(k+1)}$  e  $(\hat{\sigma}_2^2)^{(k+1)}$ , basta reescrever (2.52) e (2.53) com  $Z_i = 0$  no lugar de  $Z_i = 1$  e  $1 - \tau(x_i; \theta^{(k)})$  no lugar de  $\tau(x_i; \theta^{(k)})$ .

Um problema para o cálculo dos EMV's dos parâmetros envolvidos neste estudo, consiste na obtenção das esperanças condicionais que aparecem nas estimativas, e que nem sempre existem de forma fechada. A alternativa aqui, porém,

é usar a versão estocástica do algoritmo EM, chamada de **algoritmo MCEM**, de forma semelhante ao apresentado na Seção 1.3. Neste caso as esperanças  $E(\log Y_i \mid x_i, Z_i = 1; \theta^{(k)})$  e  $E[(\log Y_i)^2 \mid x_i, Z_i = 1; \theta^{(k)}]$  são aproximadas, simulando  $m$  valores de  $y'_1, \dots, y'_m$  da distribuição condicional de  $Y$  dado  $X$  para microfibras com os parâmetros  $\phi_1^{(k)}$  e usando

$$\frac{1}{m} \sum_{j=1}^m \log y'_j \quad e \quad \frac{1}{m} \sum_{j=1}^m (\log y'_j)^2 \quad (2.54)$$

para as respectivas esperanças condicionais. As esperanças condicionais para as fibras são aproximadas de forma semelhante. No capítulo seguinte, provaremos a consistência e a normalidade assintótica deste algoritmo MCEM, quando  $f_Y(y; \theta)$  é uma mistura de duas distribuições lognormal.

---

# Propriedades Assintóticas do Algoritmo MCEM para Mistura com Dados Censurados

---

## 3.1 Introdução

O objetivo deste capítulo é tratar a consistência e a normalidade assintótica do estimador obtido via algoritmo MCEM. A normalidade assintótica é mostrada em três etapas e, antes de detalhar o modo como isso é feito, apresentamos algumas notações e definições preliminares.

Seja  $\Theta$  o espaço paramétrico, que para um modelo de misturas de duas lognormal, corresponde ao espaço paramétrico  $\Theta = ((0, 1) \times \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+)$  onde  $\mathbb{R}^+ = (0, \infty)$ . Seja ainda:

$\theta^* = (\varepsilon^*, \phi_1^*, \phi_2^*)$  = o verdadeiro valor do parâmetro que gerou os dados observados no núcleo de incremento;

$\hat{\theta}_{MV}$  = o estimador de máxima verossimilhança de  $\theta^*$ ;

$\tilde{\theta}^{(k)}$  = o estimador de  $\theta^*$  no passo  $k$  da sequência EM;

$\tilde{\theta}_S^{(k)}$  = o estimador de  $\theta^*$  no passo  $k$  da sequência MCEM proposto.

Seja também  $\Theta'$  um conjunto compacto de  $\Theta$ , tal que  $\theta^*$  é um ponto interior e em quase todo  $x$ , com  $\theta \in \Theta'$ , temos que  $f_X(x; \theta) = f_X(x; \theta^*)$  se, e somente se,  $\theta = \theta^*$ .

Note que  $\tilde{\theta}_S^{(k)}$  depende também de  $m$ , número de valores simulados para a integração Monte Carlo, como visto no final do Capítulo 2. Neste capítulo, mostramos a consistência

e a normalidade assintótica de  $\tilde{\theta}_S^{(k)}$ , onde  $k$ ,  $m$  e  $n$  tendem simultaneamente para infinito de forma ainda a ser estabelecida, quando  $f_Y(y; \theta)$  é uma mistura de duas distribuições lognormal.

Um problema para distribuições de mistura é que a função de logverossimilhança  $l(\theta) = \sum_{i=1}^n \log f_X(x_i; \theta)$ , pode atingir o seu maior máximo local em diferentes escolhas de  $\theta \in \Theta$ . Por exemplo, isto pode ocorrer no chamado problema *label switching*, como mencionado na Seção 1.4. Contudo, no nosso estudo, este problema é facilmente evitado pelo fato de termos uma ordem natural das funções densidades de  $f_{X_1}(x; \phi_1)$  e  $f_{X_2}(x; \phi_2)$ , uma vez que o valor esperado do comprimento das microfibras é muito menor do que o comprimento das fibras.

Na prova da consistência e a normalidade assintótica, assumimos que a matriz de informação de Fisher  $I(\theta) = E[S_{\mathbf{X}}(\theta; \mathbf{X})S_{\mathbf{X}}^T(\theta; \mathbf{X}); \theta]$  é bem definida e definida positiva em  $\theta^*$ . Aqui  $S_x(\theta; X) = \frac{\partial \log f_{\mathbf{X}}(\mathbf{x}; \theta)}{\partial \theta}$ . A demonstração do teorema formulado a seguir, é o objetivo principal deste trabalho.

**Teorema 3.1.** *Assuma que  $f_Y(y; \theta)$  é uma mistura de duas distribuições lognormal. Então*

$$\sqrt{n}(\tilde{\theta}_S^{(k)} - \theta^*) \xrightarrow{d} N(0, I(\theta^*)^{-1}) \text{ com } k, m, n \rightarrow \infty \text{ cooperativamente,}$$

onde  $\tilde{\theta}_S^{(k)} = \arg \max_{\theta \in \Theta'} Q(\theta; \tilde{\theta}_S^{(k-1)})$  e  $\xrightarrow{d}$  denota a convergência em distribuição.

O termo *cooperativamente*, que especifica de qual forma  $k, m$  e  $n$  devem simultaneamente tender ao infinito, e está definido na Seção 3.5, dependerá de  $M$  em (3.32) e da sequência  $k(n)$  no Lema 3.8. Contudo, a demonstração do Teorema 3.1 é feito em três etapas, isto é, em seções distintas, observando que

$$\sqrt{n}(\tilde{\theta}_S^{(k)} - \theta^*) = \sqrt{n}(\tilde{\theta}_S^{(k)} - \tilde{\theta}^{(k)}) + \sqrt{n}(\tilde{\theta}^{(k)} - \hat{\theta}_{MV}) + \sqrt{n}(\hat{\theta}_{MV} - \theta^*).$$

Sendo assim, na Seção 3.2 apresentamos definições e resultados que serão utilizados nas demais seções deste capítulo, na Seção 3.3 mostramos a convergência do EMV para o parâmetro  $\theta^*$ , na Seção 3.4 mostramos a convergência do estimador gerado via algoritmo EM para o EMV e na Seção 3.5 mostramos a convergência do estimador gerado via algoritmo MCEM para o gerado via algoritmo EM.

## 3.2 Condições de Regularidade e Resultados Preliminares

Apresentamos nesta seção algumas condições e alguns resultados que servem de base para as demonstrações das propriedades assintóticas do estimador obtido via algoritmo MCEM. Esses resultados preliminares são encontrados em Redner e Walker [15].

Nesta seção, consideramos  $\theta = (\theta_1, \theta_2, \dots, \theta_v)$  um vetor  $v$ -dimensional pertencente ao espaço paramétrico  $\Theta$ . Algumas condições de regularidades utilizadas na análise assintótica do algoritmo MCEM, são relacionadas a seguir.

**Condição RW 1.** *Existe uma vizinhança  $\tilde{\Theta} \subset \Theta$  de  $\theta^*$  tal que para todo  $\theta \in \tilde{\Theta}$ , para quase todo  $x \in (0, 2r)$  e para  $i, j = 1, \dots, v$ , as derivadas parciais  $\frac{\partial f_X}{\partial \theta_i}$  e  $\frac{\partial^2 f_X}{\partial \theta_i \partial \theta_j}$  existem e satisfazem*

$$\left| \frac{\partial f_X(x; \theta)}{\partial \theta_i} \right| \leq g_i(x), \quad \left| \frac{\partial^2 f_X(x; \theta)}{\partial \theta_i \partial \theta_j} \right| \leq g_{ij}(x),$$

onde  $g_i(x)$  e  $g_{ij}(x)$  são integráveis em  $(0, 2r)$  com respeito à medida de Lebesgue.

**Condição RW 2.** *Para todo  $\theta \in \tilde{\Theta}$ , para quase todo  $x \in (0, 2r)$  e para  $i, j, k = 1, \dots, v$ , as derivadas parciais  $\frac{\partial^3 f_X}{\partial \theta_i \partial \theta_j \partial \theta_k}$  existem e satisfazem*

$$\left| \frac{\partial^3 \log f_X(x; \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| \leq g_{ijk}(x),$$

onde  $g_{ijk}(x)$  satisfaz

$$\int_0^{2r} g_{ijk}(x) f_X(x; \theta^*) dx < \infty.$$

Para a condição RW3, é necessário introduzirmos a seguinte notação. Para  $\theta \in \Theta$  e  $\rho > 0$ , seja  $N_\rho[\theta]$  uma bola fechada de raio  $\rho$  e centro  $\theta$  em  $\Theta$ , e defina  $f_X(x; \theta, \rho) = \sup_{\theta' \in N_\rho[\theta]} f_X(x; \theta')$  e  $f_X^*(x; \theta, \rho) = \max\{1, f_X(x; \theta, \rho)\}$ .

**Condição RW 3.** *Para cada  $\theta \in \Theta$  e  $\rho > 0$  suficientemente pequeno,*

$$\int_0^{2r} \log f_X^*(x; \theta, \rho) f_X(x; \theta^*) dx < \infty.$$

**Condição RW 4.**  $\int_0^{2r} \log f_X(x; \theta^*) f_X(x; \theta^*) dx < \infty.$

**Condição RW 5.** *Com probabilidade 1, a função  $Q(\theta, \theta')$  é contínua em  $\theta$  e  $\theta'$  em  $\Theta'$ , onde  $Q(\theta, \theta')$  é a esperança calculada no passo  $E$  do algoritmo EM (veja (2.47)) e,  $\Theta'$  é definido no início do capítulo.*

**Condição RW 6.** *As funções  $Q(\theta, \theta')$ , dada em (2.47), e  $l(\theta) = \sum_{i=1}^n \log f_X(x_i; \theta)$  são ambas diferenciáveis em  $\theta$ , para  $\theta$  e  $\theta'$  em  $\Theta'$ , quando  $n$  for suficientemente grande.*

A seguir, relacionamos três proposições de Redner Walker [15], que são utilizadas nas demonstrações do Lema 3.7 e do Lema 3.8.

**Proposição 3.2.** *Se a matriz de informação de Fisher é bem definida e positiva definida em  $\theta^*$ , se as condições RW(1) e RW(2) são satisfeitas, e se qualquer vizinhança suficientemente pequena de  $\theta^*$  em  $\Theta$  é dada, então com probabilidade 1, existe, para todo  $n$  suficientemente grande, uma única solução  $\hat{\theta}_{MV}$  das equações de verossimilhança  $\partial \log f_X(x; \theta) / \partial \theta = 0$  nesta vizinhança, e esta solução maximiza localmente a função logverossimilhança. Além disso,  $\sqrt{n}(\hat{\theta}_{MV} - \theta^*)$  é assintoticamente normal distribuído com média zero e matriz de covariância  $I(\theta^*)^{-1}$ .*

Note que a Proposição 3.2 garante que numa vizinhança suficientemente pequena de  $\theta^*$  em  $\Theta$ , o EMV é fortemente consistente e assintoticamente normal distribuído.

**Proposição 3.3.** *Seja  $\Theta'$  qualquer subconjunto compacto de  $\Theta$  que contém  $\theta^*$  em seu interior, e o conjunto*

$$C = \{\theta \in \Theta' : f_X(x; \theta) = f_X(x; \theta^*) \text{ em quase todo } x\}.$$

*Se as condições RW(3) e RW(4) são satisfeitas e  $D$  é um subconjunto fechado qualquer de  $\Theta'$  que não intercepta  $C$ , então com probabilidade 1,*

$$\lim_{n \rightarrow \infty} \sup_{\theta \in D} \frac{\prod_{i=1}^n f_X(x_i; \theta)}{\prod_{i=1}^n f_X(x_i; \theta^*)} = 0.$$

Esta proposição assegura que  $\hat{\theta}_{MV}$ , para  $n$  suficientemente grande, atinge o maior máximo local num subconjunto compacto qualquer de  $\Theta$  contendo  $\theta^*$  em seu interior (com exceção do problema *label switching* mencionado na Seção 1.4).

**Proposição 3.4.** *Sob todas as condições de regularidade, RW1 a RW6, temos com probabilidade 1, para  $n$  suficientemente grande, que o único estimador de máxima verossimilhança fortemente consistente,  $\hat{\theta}_{MV}$ , é bem definido em  $\Theta'$  e  $\lim_{k \rightarrow \infty} \tilde{\theta}^{(k)} = \hat{\theta}_{MV}$ , quando  $\theta^{(0)}$  é suficientemente próximo à  $\hat{\theta}_{MV}$ .*

Agora, o próximo resultado pode ser encontrado em Casella e Berger [3] e será utilizado nas demonstrações dos Lemas 3.6, 3.7 e 3.8.

**Proposição 3.5.** *Se  $f_{Y_t}(\cdot; \phi_t)$ , com  $t = 1, 2$ , é diferenciável em  $\phi_t$  e existir uma função  $g(\cdot; \phi_t)$  e uma constante  $\delta_0 > 0$  tal que*

$$\left| \frac{\partial f_{Y_t}(y; \phi_t)}{\partial \phi_t} \right|_{\phi_t = \phi'_t} \leq g(y; \phi_t), \quad \forall \phi_t \text{ tal que } |\phi'_t - \phi_t| \leq \delta_0,$$

e

$$\int_{-\infty}^{\infty} g(y; \phi_t) dy < \infty,$$

então, temos que

$$\frac{\partial}{\partial \phi_t} \int_{-\infty}^{\infty} f_{Y_t}(y; \phi_t) dy = \int_{-\infty}^{\infty} \frac{\partial}{\partial \phi_t} f_{Y_t}(y; \phi_t) dy.$$

Na sequência, apresentamos um lema que especifica condições para  $f_Y(y; \theta)$ , sob as quais, a condição RW1 é satisfeita para a função densidade  $f_X(x; \theta)$ , quando  $f_Y(y; \theta)$  é uma mistura de duas distribuições. Neste caso  $\theta_1 = \varepsilon$  e  $(\theta_2, \dots, \theta_v) = (\phi_1, \phi_2)$ .

**Lema 3.6.** *A condição RW(1) é satisfeita para  $f_X(x; \theta)$  se existe uma vizinhança  $\tilde{\Theta} \subset \Theta$  de  $\theta^*$  tal que para todo  $\theta \in \tilde{\Theta}$  e quase todo  $y \in (0, \infty)$ ,*

$$|f_{Y_t}(y; \phi_t)| < h_t(y) \tag{3.1}$$

e as derivadas parciais  $\partial f_{Y_t}/\partial \theta_i$  e  $\partial^2 f_{Y_t}/\partial \theta_i \partial \theta_j$ , existem para  $i, j = 2, \dots, v$ , e satisfazem

$$\left| \frac{\partial f_{Y_t}(y; \phi_t)}{\partial \theta_i} \right| \leq h_{ti}(y) \quad \text{e} \quad \left| \frac{\partial^2 f_{Y_t}(y; \phi_t)}{\partial \theta_i \partial \theta_j} \right| \leq h_{tij}(y), \tag{3.2}$$

$t = 1, 2$ , onde  $h_t(y)$ ,  $h_{ti}(y)$  e  $h_{tij}(y)$  são integráveis em  $(0, \infty)$  com respeito a medida de Lebesgue.

Seja  $\nabla_d$  uma notação conjunta para qualquer derivada parcial de ordem  $d$ , para uma função, com respeito aos parâmetros  $\theta_i$ ,  $i = 1, \dots, v$ .

**Demonstração do Lema 3.6 :** Conforme foi visto anteriormente na equação (2.13), temos que

$$f_X(x; \theta) = \varepsilon f_{X_1}(x; \phi_1) + (1 - \varepsilon) f_{X_2}(x; \phi_2).$$

Daí,

$$\left| \frac{\partial f_X(x; \theta)}{\partial \varepsilon} \right| = |f_{X_1}(x; \phi_1) - f_{X_2}(x; \phi_2)| \leq f_{X_1}(x; \phi_1) + f_{X_2}(x; \phi_2). \tag{3.3}$$

Pela equação (2.14), sabemos que

$$f_{X_t}(x; \phi_t) = p_{nc}(x) f_{Y_t}(x; \phi_t) + \int_x^{\infty} \frac{R(x, y)}{\sqrt{4r^2 - x^2}} f_{Y_t}(y; \phi_t) dy, \quad \text{se } 0 < x < 2r,$$

onde

$$R(x, y) = \frac{8r^2 - 3x^2 + yx}{\pi r^2 + 2ry}.$$

Para um valor fixo de  $x \in (0, 2r)$ , temos que  $\frac{\partial R(x, y)}{\partial y} > 0$  ou  $< 0$ , dependendo apenas de  $x$ . Logo,  $R(x, y)$  é crescente ou decrescente como função de  $y$  em  $[x, \infty)$ . Temos ainda que

$$R(x, x) = \frac{8r^2 - 2x^2}{\pi r^2 + 2rx} < \frac{8r^2}{\pi r^2 + 2rx} < \frac{8r^2}{\pi r^2} = \frac{8}{\pi}$$

e

$$\lim_{y \rightarrow \infty} R(x, y) = \frac{x}{2r} < 1.$$

Assim,

$$|R(x, y)| \leq \frac{8}{\pi} \text{ para } x \in (0, 2r), y \in [x, \infty). \quad (3.4)$$

E portando, da equação (3.1), temos que

$$\begin{aligned} \left| \frac{\partial f_X(x; \theta)}{\partial \varepsilon} \right| &\leq f_{X_1}(x; \phi_1) + f_{X_2}(x; \phi_2) \\ &= \sum_{t=1}^2 \left[ p_{nc}(x) f_{Y_t}(x; \phi_t) + \int_x^\infty \frac{R(x, y)}{\sqrt{4r^2 - x^2}} f_{Y_t}(y; \phi_t) dy \right] \\ &< \sum_{t=1}^2 \left[ p_{nc}(x) h_t(x) + \int_x^\infty \frac{R(x, y)}{\sqrt{4r^2 - x^2}} h_t(y) dy \right] \\ &\stackrel{Por(3.4)}{\leq} \sum_{t=1}^2 \left[ p_{nc}(x) h_t(x) + \int_0^\infty \frac{8}{\pi \sqrt{4r^2 - x^2}} h_t(y) dy \right] \\ &\leq \sum_{t=1}^2 \left[ h_t(x) + \frac{8}{\pi \sqrt{4r^2 - x^2}} \int_0^\infty h_t(y) dy \right] =: g_1(x). \end{aligned}$$

Segue, da hipótese que  $h_t(x)$  é integrável em  $(0, \infty)$ , e que

$$\int_0^{2r} \frac{1}{\sqrt{4r^2 - x^2}} dx = \arcsen \frac{x}{2r} \Big|_0^{2r} = \frac{\pi}{2},$$

que  $g_1(x)$  é integrável em  $(0, 2r)$ . Uma vez que  $f_X(x; \theta)$  é linear em  $\varepsilon$ , a derivada de segunda ordem de  $f_X(x; \theta)$  com respeito a  $\varepsilon$  é zero. Também, observamos que

$$\frac{\partial^2 f_X(x; \theta)}{\partial \phi_i \partial \phi_j} = 0, \text{ com } i, j = 1, 2 \text{ e } i \neq j.$$

Então, para provarmos que a condição RW1 é satisfeita, precisamos encontrar funções

integráveis, dominando as derivadas parciais de primeira e segunda ordem de  $f_{X_t}(x; \phi_t)$ , com respeito a  $\phi_t$ ,  $t = 1, 2$ . Contudo, de (2.14) e (3.4) temos que

$$\begin{aligned} \nabla_d f_{X_t}(x; \phi_t) &= p_{nc}(x) \nabla_d f_{Y_t}(x; \phi_t) + \nabla_d \int_x^\infty \frac{R(x, y)}{\sqrt{4r^2 - x^2}} f_{Y_t}(y; \phi_t) dy \\ &\leq p_{nc}(x) \nabla_d f_{Y_t}(x; \phi_t) + \frac{8}{\pi \sqrt{4r^2 - x^2}} \nabla_d \int_x^\infty f_{Y_t}(y; \phi_t) dy, \end{aligned} \quad (3.5)$$

se  $0 < x < 2r$ , onde  $d = 1, 2$ . Note que podemos derivar sobre o sinal da integral no segundo termo de (3.5) utilizando a Proposição 3.5, pois por hipótese, existe uma vizinhança  $\tilde{\Theta} \subset \Theta$  de  $\theta^*$  tal que para todo  $\theta \in \tilde{\Theta}$  e quase todo  $y \in (0, \infty)$ ,

$$|f_{Y_t}(y; \phi_t)| < h_t(y)$$

e as derivadas parciais  $\partial f_{Y_t}/\partial \theta_i$  e  $\partial^2 f_{Y_t}/\partial \theta_i \partial \theta_j$ , existem para  $i, j = 2, \dots, v$ , e satisfazem

$$\left| \frac{\partial f_{Y_t}(y; \phi_t)}{\partial \theta_i} \right| \leq h_{ti}(y) \quad e \quad \left| \frac{\partial^2 f_{Y_t}(y; \phi_t)}{\partial \theta_i \partial \theta_j} \right| \leq h_{tij}(y),$$

$t = 1, 2$ , onde  $h_t(y)$ ,  $h_{ti}(y)$  e  $h_{tij}(y)$  são integráveis em  $(0, \infty)$  com respeito a medida de Lebesgue. Portanto, de (3.5) segue para  $\theta_i$ ,  $i = 2, \dots, v$ ,  $d = 1, 2$ , que

$$\nabla_d f_{X_t}(x; \phi_t) \leq p_{nc}(x) \nabla_d f_{Y_t}(x; \phi_t) + \frac{8}{\pi \sqrt{4r^2 - x^2}} \int_x^\infty \nabla_d f_{Y_t}(y; \phi_t) dy, \quad \text{se } 0 < x < 2r,$$

e assim

$$|\nabla_d f_{X_t}(x, \phi_t)| \leq |\nabla_d f_{Y_t}(x, \phi_t)| + \left| \frac{8}{\pi \sqrt{4r^2 - x^2}} \right| \int_x^\infty |\nabla_d f_{Y_t}(y, \phi_t)| dy.$$

Então,

$$|\nabla_d f_{X_t}(x, \phi_t)| \leq h_{t \cdot}(x) + \frac{8}{\pi \sqrt{4r^2 - x^2}} \int_0^\infty h_{t \cdot}(y) dy < \infty, \quad (3.6)$$

onde  $h_{t \cdot}(y)$  é uma notação conjunta para  $h_{ti}(y)$  e  $h_{tij}(y)$  em (3.2). Assim, o lado direito de (3.6) é integrável. Isto completa a demonstração do Lema 3.6.  $\blacksquare$

### 3.3 Normalidade Assintótica do Estimador de Máxima Verossimilhança

Nesta seção, verificamos a normalidade assintótica do EMV de  $\theta$ . Lembre-se que quando falamos sobre o EMV de  $\theta$ , queremos denotar qualquer escolha de  $\theta \in \Theta$  onde

$l(\theta) = \sum_{i=1}^n f_X(x_i; \theta)$  atinge seu maior máximo local.

**Lema 3.7.** *Assuma que  $f_Y(y; \theta)$  é uma mistura de duas distribuições lognormal. Então*

$$\sqrt{n}(\hat{\theta}_{MV} - \theta^*) \xrightarrow{d} N(0, I(\theta^*)^{-1}) \text{ com } n \rightarrow \infty.$$

**Demonstração do Lema 3.7 :** Essa demonstração é feita, considerando  $f_Y(y; \theta)$  uma densidade de mistura de duas lognormais, como dado em (2.1), e verificando as condições RW1 à RW4. O resultado segue diretamente das Proposições 3.2 e 3.3.

*Condição RW1:* Iniciamos a verificação da condição RW1, observando que pelo Lema 3.6, é suficiente mostrar que  $|f_{Y_t}(y; \phi_t)|$  e  $|\nabla_d f_{Y_t}(y; \phi_t)|$ , para  $t = 1, 2$  e  $d = 1, 2$ , são dominados por funções integráveis (independente de  $\theta$ ), para modelos lognormal. Afim de verificar esse fato, note que para modelos lognormal com  $\theta_1 = \varepsilon$ ,  $\theta_2 = \mu_1$ ,  $\theta_3 = \sigma_1^2$ ,  $\theta_4 = \mu_2$  e  $\theta_5 = \sigma_2^2$ , onde  $\phi_t = (\mu_t, \sigma_t)$ , temos que  $|f_{Y_t}(y; \phi_t)|$  e  $|\nabla_d f_{Y_t}(y; \phi_t)|$  para  $d = 1, 2, 3$ , são somas de no máximo quatro termos da forma

$$|T_L^K(y, \mu_t, \sigma_t)| = C \frac{|\log y - \mu_t|^K}{\sigma_t^L} f_{Y_t}(y; \phi_t), \quad t = 1, 2, \quad (3.7)$$

onde  $K, L$  e  $C$  são inteiros não negativos. Se para cada  $K$  e  $L$  podemos encontrar uma função  $h_L^K(y, t)$  integrável em  $[0, \infty)$ , tal que

$$|T_L^K(y, \mu_t, \sigma_t)| \leq h_L^K(y, t),$$

então temos que a condição RW1 é satisfeita.

Seja  $\Psi_t$  uma vizinhança de  $\phi_t^*$  definida por  $\mu_t \in [\mu_t^* - m_t, \mu_t^* + m_t]$  e  $\sigma_t \in \left[ \frac{\sigma_t^*}{s_t}, s_t \sigma_t^* \right]$ , para algum número finito de constantes  $m_t$  e  $s_t$ , onde  $m_t > 0$  e  $s_t > 1$  com  $t = 1, 2$ . Seja também,  $m_t^* = \max(|\mu_t^* - m_t|, |\mu_t^* + m_t|)$ , então

$$\begin{aligned} (\log y - \mu_t)^2 &\geq (\log y)^2 - 2|\mu_t| |\log y| + \mu_t^2 \\ &\geq (\log y)^2 - 2m_t^* |\log y| = (|\log y| - m_t^*)^2 - (m_t^*)^2, \end{aligned}$$

e para todo  $\phi_t \in \Psi_t$ , temos que

$$\begin{aligned}
|T_L^K(y, \mu_t, \sigma_t)| &= C \frac{|\log y - \mu_t|^K}{\sigma_t^L} \frac{1}{\sigma_t \sqrt{2\pi}} \frac{1}{y} \exp \left\{ -\frac{(\log y - \mu_t)^2}{2\sigma_t^2} \right\} \\
&\leq C \frac{(|\log y| + m_t^*)^K}{\left(\frac{\sigma_t^*}{s_t}\right)^{L+1} y \sqrt{2\pi}} \exp \left\{ -\frac{(|\log y| - m_t^*)^2 + (m_t^*)^2}{2(s_t \sigma_t^*)^2} \right\} \\
&=: h_L^K(y, t).
\end{aligned} \tag{3.8}$$

Daí, temos que para  $t = 1, 2$

$$\int_0^\infty h_L^K(y, t) dy = \int_0^\infty C \frac{(|\log y| + m_t^*)^K}{\left(\frac{\sigma_t^*}{s_t}\right)^{L+1} y \sqrt{2\pi}} \exp \left\{ -\frac{(|\log y| - m_t^*)^2 + (m_t^*)^2}{2(s_t \sigma_t^*)^2} \right\} dy.$$

Aplicando Binômio de Newton e fazendo  $\log y = z$ , segue que

$$\int_0^\infty h_L^K(y, t) dy \propto \sum_{i=0}^K (m_t^*)^{K-i} \binom{K}{i} \int_{-\infty}^{+\infty} \frac{|z|^i}{\sigma_t^* s_t \sqrt{2\pi}} \exp \left\{ -\frac{(|z| - m_t^*)^2}{2(s_t \sigma_t^*)^2} \right\} dz.$$

Conseqüentemente, denotando  $\varphi(\cdot)$  a função de densidade da normal padrão, e ainda sabendo que todos os momentos da distribuição normal padrão existem e que a mesma é simétrica em torno de zero, segue que

$$\int_0^\infty h_L^K(y, t) dy \propto 2 \sum_{i=0}^K (m_t^*)^{K-i} \binom{K}{i} \int_0^{+\infty} z^i \varphi \left( \frac{z - m_t^*}{s_t \sigma_t^*} \right) dz < \infty. \tag{3.9}$$

Agora, se o conjunto

$$\tilde{\Theta} = (\mathcal{E}, \Psi_1, \Psi_2), \tag{3.10}$$

onde  $\mathcal{E}$  é uma vizinhança fechada qualquer de  $\varepsilon^*$ , tal que  $\mathcal{E} \subset (0, 1)$  e  $\Psi_t$ ,  $t = 1, 2$  é definido anteriormente, temos então, a partir da equação (3.9), que a condição RW1 é satisfeita para modelos lognormal, para todo  $\theta \in \tilde{\Theta}$ .

Condição RW2: Temos que

$$\begin{aligned} \frac{\partial^3 \log f_X(x; \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} &= \frac{\frac{\partial^3 f_X(x; \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k}}{f_X(x; \theta)} - \frac{\frac{\partial f_X(x; \theta)}{\partial \theta_i} \frac{\partial^2 f_X(x; \theta)}{\partial \theta_j \partial \theta_k}}{f_X^2(x; \theta)} - \frac{\frac{\partial f_X(x; \theta)}{\partial \theta_j} \frac{\partial^2 f_X(x; \theta)}{\partial \theta_i \partial \theta_k}}{f_X^2(x; \theta)} \\ &- \frac{\frac{\partial f_X(x; \theta)}{\partial \theta_k} \frac{\partial^2 f_X(x; \theta)}{\partial \theta_i \partial \theta_j}}{f_X^2(x; \theta)} + 2 \frac{\frac{\partial f_X(x; \theta)}{\partial \theta_i} \frac{\partial f_X(x; \theta)}{\partial \theta_j} \frac{\partial f_X(x; \theta)}{\partial \theta_k}}{f_X^3(x; \theta)}. \end{aligned} \quad (3.11)$$

Note que

$$\frac{\partial^3 f_X(x; \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} = 0, \quad \text{se} \begin{cases} \theta_i = \theta_j = \varepsilon; \\ \theta_i = \theta_k = \varepsilon; \\ \theta_j = \theta_k = \varepsilon; \\ \theta_i = \phi_1 \text{ e } \theta_j = \phi_2; \text{ ou vice-versa;} \\ \theta_i = \phi_1 \text{ e } \theta_k = \phi_2; \text{ ou vice-versa;} \\ \theta_j = \phi_1 \text{ e } \theta_k = \phi_2; \text{ ou vice-versa.} \end{cases}$$

Assim, temos que

$$\frac{\partial^3 \log f_X(x; \theta)}{\partial \varepsilon^3} = 2 \left( \frac{\frac{\partial f_X(x; \theta)}{\partial \varepsilon}}{f_X(x; \theta)} \right)^3,$$

e como  $f_X(x; \theta) = \varepsilon f_{X_1}(x; \phi_1) + (1 - \varepsilon) f_{X_2}(x; \phi_2)$  para  $0 < x < 2r$ , visto na equação (2.13), então a partir da equação (3.3), segue que

$$\left| \frac{\frac{\partial f_X(x; \theta)}{\partial \varepsilon}}{f_X(x; \theta)} \right| \leq \frac{1}{\varepsilon} + \frac{1}{1 - \varepsilon} \quad (3.12)$$

para  $0 < x < 2r$ . Daí, para  $\varepsilon \in \mathcal{E}$  a equação (3.12) é limitada superiormente por uma constante, logo a condição RW2 é satisfeita para  $\theta_i = \theta_j = \theta_k = \varepsilon$ .

Vamos verificar agora a condição RW2 para  $\theta_i$ , com  $i = 2, \dots, 5$ . Da equação (2.14), temos que

$$f_{X_i}(x; \phi_t) = p_{nc}(x) f_{Y_i}(x; \phi_t) + \int_x^\infty \frac{R(x, y)}{\sqrt{4r^2 - x^2}} f_{Y_i}(y; \phi_t) dy, \quad \text{se } 0 < x < 2r.$$

Para o modelo lognormal, vimos que

$$\begin{aligned} |T_L^K(y; \mu_t, \sigma_t)| &\stackrel{3.7}{=} C \frac{|\log y - \mu_t|^K}{\sigma_t^L} f_Y(y, \phi_t) \\ &\stackrel{3.8}{\leq} C \frac{(|\log y| + m_t^*)^K}{\frac{\sigma_t^{*L+1}}{s_t} y \sqrt{2\pi}} \exp \left\{ -\frac{(|\log y - m_t^*|)^2 + (m_t^*)^2}{2(s_t \sigma_t^*)^2} \right\} =: h_L^K(y, t), \end{aligned}$$

e por (3.9), temos que

$$\int_0^\infty h_L^K(y, t) \propto 2 \sum_{i=0}^K (m_t^*)^{K-i} \binom{K}{i} \int_0^\infty z^i \varphi \left( \frac{z - m_t^*}{s_t \sigma_t^*} dz \right) < \infty,$$

portanto, pela Proposição 3.5, é possível derivar sobre o sinal da integral para  $d = 3$ . Logo, para  $d = 1, 2, 3$  e  $\theta_i$ , com  $i = 2, 3$ , segue que

$$|\nabla_d f_X(x, \theta)| \leq \varepsilon \left[ p_{nc}(x) |\nabla_d f_{Y_1}(x; \phi_1)| + \int_x^\infty \frac{|R(x, y) \nabla_d f_{Y_1}(y; \phi_1)|}{\sqrt{4r^2 - x^2}} dy \right], \quad (3.13)$$

e para  $d = 1, 2, 3$  e  $\theta_i$ , com  $i = 4, 5$ , segue que

$$|\nabla_d f_X(x, \theta)| \leq (1 - \varepsilon) \left[ p_{nc}(x) |\nabla_d f_{Y_2}(x; \phi_2)| + \int_x^\infty \frac{|R(x, y) \nabla_d f_{Y_2}(y; \phi_2)|}{\sqrt{4r^2 - x^2}} dy \right]. \quad (3.14)$$

Logo, obtemos de (2.14), (3.13) e (3.14), para  $d = 1, 2, 3$

$$\frac{|\nabla_d f_X(x; \theta)|}{f_X(x; \theta)} \leq A_d(x; \phi_t) + B_d(x; \phi_t),$$

$$\text{onde } A_d(x; \phi_t) = \frac{|\nabla_d f_{Y_t}(x; \phi_t)|}{f_{Y_t}(x; \phi_t)} \text{ e } B_d(x; \phi_t) = \frac{\int_x^\infty |R(x, y) \nabla_d f_{Y_t}(y; \phi_t)| dy}{\int_x^\infty R(x, y) f_{Y_t}(y; \phi_t) dy},$$

com  $x \in (0, 2r)$ . Observe que por (3.6), temos que

$$\int_0^\infty |R(x, y) \nabla_d f_{Y_t}(y; \phi_t)| dy \leq \int_0^\infty \frac{8}{\pi} h_t \cdot(y) dy < \infty,$$

portanto o numerador de  $B_d(x; \phi_t)$  é limitado superiormente para  $d = 1, 2$ . Note que, por (3.7) até (3.9), o numerador de  $B_d(x; \phi_t)$  também é limitado superiormente para  $d = 3$ . Agora, observe que um limite inferior pode ser obtido para o denominador de  $B_d(x; \phi_t)$ , pois  $R(x, y) > R(2r, y) > 0$  para  $x \in (0, 2r)$  e  $y \in (2r, 6r)$ , onde  $R(2r, y)$  é crescente como

uma função de  $y$  e  $f_{Y_t}(y; \theta) > 0$  é contínua, assim temos

$$\begin{aligned} \int_x^\infty R(x, y) f_{Y_t}(y; \phi_t) dy &\geq \int_{3r}^{4r} R(2r, y) f_{Y_t}(y; \phi_t) dy \geq R(2r, 3r) \int_{3r}^{4r} f_{Y_t}(y; \phi_t) dy \\ &\geq \frac{2r}{(\pi + 6)} \min_{y \in (3r, 4r)} f_{Y_t}(y; \phi_t) = C_{dt} > 0, \quad x \in (0, 2r). \end{aligned} \quad (3.15)$$

Portanto, podemos concluir que  $B_d(x; \phi_t)$  é limitado superiormente.

Seja  $\Lambda(x; \phi_t)$  o produto de um, dois ou três fatores de  $A_d(x; \phi_t)$ , com  $d = 1, 2, 3$  e  $t = 1, 2$ . Então, precisamos apenas mostrar que  $\Lambda(x; \phi_t) f_X(x; \theta^*)$  é integrável em  $(0, 2r)$ , uma vez que

$$\left| \frac{\partial^3 \log f_X(x; \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| \leq \begin{cases} \text{Que somas dos produtos de } A_d(x; \phi_t) \text{ com } A_d(x; \phi_t), \\ A_d(x; \phi_t) \text{ com } B_d(x; \phi_t) \text{ e } B_d(x; \phi_t) \text{ com } B_d(x; \phi_t), \end{cases}$$

e  $B_d(x; \phi_t)$  é limitado superiormente. Da equação (3.7), temos que  $A_d(x; \phi_t)$  é uma soma de no máximo quatro termos da forma

$$C \frac{|\log x - \mu_t|^K}{\sigma_t^L}, \quad (3.16)$$

com  $K, L$  e  $C$  sendo inteiros finitos e não-negativos. Agora, para todo  $\phi_t \in \Psi_t$ ,

$$\frac{|\log x - \mu_t|^K}{\sigma_t^L} \leq \frac{(|\log x| + m_t^*)^K}{\left(\frac{\sigma_t^*}{s_t}\right)^L} = \frac{1}{\left(\frac{\sigma_t^*}{s_t}\right)^L} \sum_{i=0}^K |\log x|^i (m_t^*)^{K-i} \binom{K}{i}.$$

Assim, a menos dos fatores constantes, precisamos apenas mostrar que para todo  $j$  positivo, temos que

$$\int_0^{2r} |\log x|^j f_X(x; \theta^*) dx < \infty. \quad (3.17)$$

Portanto, para qualquer  $c \in (0, 2r)$ , o lado esquerdo de (3.17) pode ser escrito como

$$\int_0^{2r-c} |\log x|^j f_X(x; \theta^*) dx + \int_{2r-c}^{2r} |\log x|^j f_X(x; \theta^*) dx. \quad (3.18)$$

De (2.14) e (3.4), para  $x \in (0, 2r - c]$  e  $c > 0$ , segue que

$$f_{X_t}(x; \phi_t^*) \leq f_{Y_t}(x; \phi_t^*) + \frac{8}{\pi \sqrt{2r + x} \sqrt{2r - x}} \leq m_y(\phi_t^*) + \frac{8}{\pi \sqrt{2r} \sqrt{c}}, \quad (3.19)$$

onde  $m_y(\phi_t^*) \equiv \max_{x \in (0, 2r)} f_{Y_t}(x; \phi_t^*)$ ,  $t = 1, 2$ . Portanto  $f_X(x; \theta^*)$  é limitado em  $(0, 2r - c]$ . Considerando também o fato de que  $\int_0^1 |\log x|^j dx = \Gamma(j + 1) = j!$ , vemos que o primeiro termo de (3.18),  $\int_0^{2r-c} |\log x|^j f_X(x; \theta^*) dx$ , é finito. O segundo termo de (3.18),  $\int_{2r-c}^{2r} |\log x|^j f_X(x; \theta^*) dx$ , também é finito, devido ao fato de  $|\log x|^j$  ser limitado superiormente para  $x \in (1, 2r)$  e  $f_X(x; \theta^*)$  ser uma função densidade. Logo, a condição RW2 é satisfeita para o modelo lognormal em  $\theta_i$ , com  $i = 2, \dots, 5$ .

Note que também acabamos de verificar, implicitamente, a condição RW2 para  $\theta_i = \varepsilon$ ,  $\theta_j = \mu_t$  ou  $\sigma_t$  e  $\theta_k = \mu_t$  ou  $\sigma_t$ , com suas respectivas permutações. Por exemplo

$$\left| \frac{\partial^2 f_{X_t}(x; \phi_t)}{\partial \mu_t \partial \sigma_t} \right| \leq \left| \frac{\partial^3 f_X(x; \theta)}{\partial \varepsilon \partial \mu_t \partial \sigma_t} \right| \leq A_d(x; \phi_t) + B_d(x; \phi_t),$$

onde todas permutações citadas acima seguem de forma analoga. Logo, a condição RW2 é satisfeita para o modelo lognormal em todo  $\theta \in \tilde{\Theta}$ .

*Condição RW4:* Verifiquemos agora a condição RW4. Para todo  $c \in (0, 2r)$ , temos que:

$$\begin{aligned} \int_0^{2r} \log f_X(x; \theta^*) f_X(x; \theta^*) dx &= \int_0^{2r-c} \log f_X(x; \theta^*) f_X(x; \theta^*) dx + \\ &+ \int_{2r-c}^{2r} \log f_X(x; \theta^*) f_X(x; \theta^*) dx. \end{aligned} \quad (3.20)$$

Sendo  $|x \log x| \leq e^{-1}$ , com  $x \in (0, 1]$ , conseqüentemente, uma vez que a função densidade  $f_X(x; \theta^*)$  seja limitada em  $x \in (0, 2r]$ , então  $|\log f_X(x; \theta^*) f_X(x; \theta^*)|$  também é limitada em  $x \in (0, 2r - c]$ . Portanto, o primeiro termo do lado direito de (3.20) é finito. De (2.13) e (3.19) temos que

$$\begin{aligned} f_X(x; \theta^*) &= \varepsilon f_{X_1}(x; \phi_1^*) + (1 - \varepsilon) f_{X_2}(x; \phi_2^*) \\ &\leq f_{X_1}(x; \phi_1^*) + f_{X_2}(x; \phi_2^*) \\ &\leq 2m_y(\phi^*) + \frac{16}{\pi \sqrt{2r} \sqrt{2r - x}}, \end{aligned} \quad (3.21)$$

onde  $m_y(\phi^*) = \max(m_y(\phi_1^*), m_y(\phi_2^*))$ . Seja  $c$ , tal que para  $x \in (2r - c, 2r)$ ,

$$m_y(\phi^*) < \frac{8}{\pi \sqrt{2r} \sqrt{2r - x}}. \quad (3.22)$$

Logo de (3.21), temos que

$$f_X(x; \theta^*) \leq \frac{D}{\sqrt{2r-x}}, \quad (3.23)$$

onde  $D = \frac{32}{\pi\sqrt{2r}}$ . Então, para  $x \in (2r-c, 2r)$ , temos que

$$\begin{aligned} f_X(x; \theta^*) \log f_X(x; \theta^*) &\leq D(2r-x)^{-\frac{1}{2}} \log(D(2r-x)^{-\frac{1}{2}}) = \\ &= D(2r-x)^{-\frac{1}{2}} \log(D) - \frac{1}{2} D(2r-x)^{-\frac{1}{2}} \log(2r-x). \end{aligned}$$

Observamos que

$$\int_{2r-c}^{2r} (2r-x)^{-\frac{1}{2}} dx = 2\sqrt{c}$$

e

$$\int_{2r-c}^{2r} (2r-x)^{-\frac{1}{2}} \log(2r-x) dx = \int_0^c (z)^{-\frac{1}{2}} \log z dz < \infty.$$

Portando, a condição RW4 é satisfeita para o modelo lognormal.

*Condição RW3:* Finalmente, basta verificar se RW3 é satisfeita quando  $f_Y(y; \theta)$  é uma mistura de duas distribuições lognormal. Isto é,

$$\int_0^{2r} \log f_X^*(x; \theta, \rho) f_X(x; \theta^*) dx < \infty.$$

Observamos que para  $\theta \in \Theta$ , podemos encontrar  $\rho > 0$  tal que  $N_\rho[\theta] \subset \Theta$ , uma vez que  $\Theta$  é um espaço aberto. Como  $f_X(x; \theta, \rho) = \sup_{\theta' \in N_\rho[\theta]} f_X(x; \theta')$  é contínua em  $x$  e  $\theta$ , então, por raciocínio análogo feito em (3.21) à (3.23), para cada  $\theta \in \Theta$  podemos encontrar um  $c_\rho(\theta) > 0$  tal que para  $x \in (2r - c_\rho(\theta), 2r)$ ,

$$f_X(x; \theta, \rho) \leq \frac{d_\rho(\theta)}{\sqrt{2r-x}}.$$

Por (3.23) sabemos que  $f_X(x; \theta^*) \leq \frac{D}{\sqrt{2r-x}}$  e lembrando que  $f_X^*(x; \theta, \rho) = \max\{1, f_X(x; \theta, \rho)\}$ , então pelo argumento feito na Condição RW4, temos que para um dado  $\theta \in \Theta$

$$\int_{2-c_\rho(\theta)}^{2r} \log f_X^*(x; \theta, \rho) f_X(x; \theta^*) dx < \infty.$$

Notemos, uma vez que  $f_X(x; \theta^*)$  e  $\log f_X^*(x; \theta, \rho)$  são limitadas para  $x \in (0, 2r - c_\rho(\theta))$ , então

$$\int_0^{2r-c_\rho(\theta)} \log f_X^*(x; \theta, \rho) f_X(x; \theta^*) dx < \infty,$$

e assim RW3 é satisfeita para o modelo lognormal.

Consequentemente, das condições RW1 e RW2, segue pela Proposição 3.2, que  $\hat{\theta}_{MV}$  é fortemente consistente. Das condições RW3 e RW4, segue pela Proposição 3.3, que  $\hat{\theta}_{MV}$ , para  $n$  suficientemente grande, atinge o maior máximo local num subconjunto compacto qualquer de  $\Theta$  contendo  $\theta^*$  em seu interior. Finalizando assim, a demonstração do Lema 3.7. ■

### 3.4 Convergência do Algoritmo EM Para o EMV

Na Seção 3.2, verificamos as condições de Redner e Walker [15], sob as quais uma sequência de estimativas gerada via algoritmo EM, converge quase certamente para o EMV,  $\hat{\theta}_{MV}$ , quando o valor inicial  $\theta^{(0)}$  é suficientemente próximo de  $\hat{\theta}_{MV}$ . Caso a exigência de que  $\theta^{(0)}$  esteja próximo do  $\hat{\theta}_{MV}$  não ocorra, pode ocasionar um problema quando a função logverossimilhança, para o caso de mistura de distribuições, tenha vários pontos de máximo local. Na prática, este problema pode ser contornado e, em casos como esse, o que geralmente é feito é executar o algoritmo EM, iniciando em diferentes valores, e escolher como  $\hat{\theta}_{MV}$  aquele que resultar no maior máximo local.

Nesta seção, mostramos a convergência da sequência de estimadores obtidos via algoritmo EM para  $\hat{\theta}_{MV}$  em um subconjunto compacto  $\Theta'$  de  $\Theta$  que contém  $\theta^*$  em seu interior. A ideia aqui é verificar as hipóteses da Proposição 3.4. Lembremos que o subconjunto  $\Theta'$  é tal que para  $\theta \in \Theta'$ , para quase todo  $x$ ,  $f_X(x, \theta) = f_X(x, \theta^*)$  se, e somente se,  $\theta = \theta^*$ .

Seja  $\Theta'$  tal que  $\Theta' \subset \tilde{\Theta} = (\mathcal{E}, \Psi_1, \Psi_2)$ , como em (3.10). Para um  $\theta^{(0)} \in \Theta'$ , denotamos  $\{\tilde{\theta}^{(k)}\}_{k=0}^{\infty}$  uma sequência gerada pelo algoritmo EM em  $\Theta'$ , ou seja

$$\tilde{\theta}^{(k+1)} = \arg \max_{\theta \in \Theta'} Q(\theta, \tilde{\theta}^{(k)}), \quad k = 0, 1, 2, \dots,$$

onde  $Q(\theta, \tilde{\theta})$  foi definido em (2.48).

**Lema 3.8.** *Assuma que  $f_Y(y; \theta)$  é uma mistura de duas distribuições lognormal. Então existe uma sequência  $\{k(n)\}_{n=1}^{\infty}$  com  $k(n) \rightarrow \infty$  conforme  $n \rightarrow \infty$  tal que*

$$\sqrt{n}(\tilde{\theta}^{(k(n))} - \hat{\theta}_{MV}) \xrightarrow{P} 0, \quad \text{com } n \rightarrow \infty.$$

**Demonstração do Lema 3.8 :** Essa demonstração é feita, verificando as condições RW1 à RW6, onde consideramos  $f_Y(y; \theta)$  uma densidade de mistura de duas lognormais, conforme dado em (2.1). O resultado segue diretamente da Proposição 3.4. Note que as condições RW1 à RW4 são estabelecidas na Seção 3.3, uma vez que  $\Theta' \subseteq \tilde{\Theta}$ . Portanto, basta verificarmos as condições RW5 e RW6.

*Condições RW5 e RW6:* De acordo com o que vimos na Seção 2.3 de identificabilidade, podemos encontrar um conjunto compacto  $\Theta'$  que contém  $\theta^*$  em seu interior. A partir da condição RW1, temos que todas as derivadas parciais  $\partial f_X/\partial\theta_i$  existem para  $(\theta_1, \theta_2, \dots, \theta_5) \in \Theta'$  e assim temos que  $l(\theta) = \sum_{i=1}^n \log f_X(x_i; \theta)$  é diferenciável em  $\theta \in \Theta'$ . Resta agora mostrar que  $Q(\theta, \theta')$  é contínua e diferenciável em  $\theta$ . Contudo, a demonstração será feita separadamente para cada termo de  $Q(\theta, \theta')$ . Da equação (2.48), temos que

$$Q(\theta, \theta') = \sum_{i=1}^n (\tau_1(x_i; \theta') \log \varepsilon + \tau_2(x_i; \theta') \log(1-\varepsilon)) + \sum_{t=1}^2 \sum_{i=1}^n \tau_t(x_i; \theta') E[\log f_{Y_t}(Y_{ti}; \phi_t) | x_i; \phi'_t],$$

onde

$$\tau_1(x_i; \theta') = \tau(x_i; \theta'), \quad \tau_2(x_i; \theta') = 1 - \tau(x_i; \theta') \quad \text{e} \quad \tau(x_i; \theta') = \frac{\varepsilon' f_{X_1}(x_i; \phi'_1)}{f_X(x_i; \theta')}.$$

A partir deste ponto e até o fim desta demonstração, consideramos  $t = 1, 2$ .

Primeiro Termo: Uma vez que para modelos lognormal  $f_Y(y; \theta')$  é contínua em  $\theta' \in \Theta'$ , então, por (2.14),  $f_{X_t}(x_i; \theta')$  também é contínua em  $\theta' \in \Theta'$  e, conseqüentemente,  $\tau_t(x_i; \theta')$  também. Pela Condição RW1, temos que existem as derivadas parciais

$$\frac{\partial f_X}{\partial \theta_i} \quad \text{e} \quad \frac{\partial^2 f_X}{\partial \theta_i \partial \theta_j},$$

e isto assegura a diferenciabilidade de  $\tau_t(x_i; \theta')$  para  $\theta \in \Theta'$ , quando  $f_X(x_i; \theta') \neq 0$ . Contudo, se  $f_X(x_i; \theta') = 0$ , então  $f_{X_1}(x_i; \phi'_1) = 0$  e assim  $\tau_t(x_i; \theta') = 0$ . Portanto,  $\sum_{i=1}^n (\tau_1(x_i; \theta') \log \varepsilon + \tau_2(x_i; \theta') \log(1-\varepsilon))$  é contínua em  $\theta$  e  $\theta'$  e diferenciável em  $\theta$ , para  $\theta$  e  $\theta'$  em  $\Theta'$ .

Segundo Termo: Observe que no segundo termo de  $Q(\theta, \theta')$ , resta mostrarmos a continuidade e diferenciabilidade de  $E[\log f_{Y_t}(Y_{ti}; \phi_t) | x_i; \phi'_t]$ . Pela Seção 2.2.2, vimos que a função de distribuição condicional de  $Y_t$  dado  $X_t = x$  é uma mistura de uma distribuição discreta e uma distribuição contínua, dada por

$$P(Y_t \leq y | X_t = x) = \begin{cases} 0, & \text{se } y < x \\ \frac{p_{nc}(x) f_{Y_t}(x; \theta_t)}{f_{X_t}(x; \theta)}, & \text{se } y = x \\ \frac{p_{nc}(x) f_{Y_t}(x; \theta_t)}{f_{X_t}(x; \theta)} + \int_x^y f_{Y_t|X_t}(y' | x; \theta) dy', & \text{se } y > x, \end{cases}$$

onde  $f_{Y_t|X_t}(y|x; \theta) = \frac{f_{X|Y}(x|y)f_{Y_t}(y; \theta)}{f_{X_t}(x; \theta)}$ , para  $t = 1, 2$ . Conforme visto na Seção 2.2.2, para  $0 < x < 2r$  e  $x < y < \infty$ , temos que  $f_{Y|X}(y|x)$  não é uma densidade da parte da distribuição contínua de  $Y$  dado  $X$ . Contudo, sabemos que

$$f_{Y|X}(y|x; \theta) = \frac{f_{X|Y}(x|y)f_Y(y; \theta)}{f_X(x; \theta)}$$

e por (2.14) temos que

$$\begin{aligned} \frac{1}{1 - \pi(x, \phi_t)} f_{Y|X}(y|x; \theta) &= \frac{f_{X|Y}(x|y)f_Y(y; \theta)}{(1 - \pi(x, \phi_t))f_X(x; \theta)} \\ &= \frac{R(x, y)f_Y(y; \theta)}{(1 - \pi(x, \phi_t))\sqrt{4r^2 - x^2}f_X(x; \theta)}, \end{aligned}$$

onde  $\pi(x, \phi_t)$  é o parâmetro de mistura dado por

$$\pi(x, \phi_t) = P(Y_t = x|X_t = x) = \frac{p_{nc}(x)f_{Y_t}(x; \phi_t)}{f_{X_t}(x; \phi_t)}, \quad (3.24)$$

lembrando que  $p_{nc}(x) = \frac{u(y)}{t(y)}$ , se  $0 \leq x \leq 2r$  e  $p_{nc}(x) = 0$ , se  $y > 2r$ , conforme foi definido em (2.3). Daí e por (2.17), temos que

$$\begin{aligned} \int_x^\infty \frac{1}{1 - \pi(x, \phi_t)} f_{Y|X}(y|x; \theta) &= \int_x^\infty \frac{R(x, y)f_Y(y; \theta)}{(1 - \pi(x, \phi_t))\sqrt{4r^2 - x^2}f_X(x; \theta)} \\ &= \frac{1 - \pi(x, \phi_t)}{1 - \pi(x, \phi_t)} = 1. \end{aligned}$$

Portanto, a distribuição contínua tem densidade

$$\frac{R(x, y)f_{Y_t}(y; \phi_t)}{(1 - \pi(x, \phi_t))\sqrt{4r^2 - x^2}f_{X_t}(x; \phi_t)}, \quad x < y < \infty. \quad (3.25)$$

Das equações (3.25) e (3.24), temos que

$$\begin{aligned} E[\log f_{Y_t}(Y_{ti}; \phi_t)|x_i; \phi'_t] &= \pi(x_i; \phi'_t) \log f_{Y_t}(x_i; \phi_t) + \\ &+ \frac{(f_{X_t}(x_i; \phi'_t))^{-1}}{\sqrt{4r^2 - x_i^2}} \int_{x_i}^\infty \log f_{Y_t}(y'; \phi_t) R(x_i, y') f_{Y_t}(y'; \phi_t) dy'. \end{aligned} \quad (3.26)$$

Uma vez que  $f_{Y_t}(x; \phi_t)$  é contínua em  $\phi_t$ , então  $f_{X_t}(x; \phi_t)$  também é contínua em  $\phi_t$  e

consequentemente (3.26) é contínua em  $\phi$  e  $\phi' \in \Theta'$ . Se for possível tomar a derivada com respeito à  $\phi_t$  sobre o sinal da integral no segundo termo da equação (3.26), então a esperança condicional  $Q(\theta, \theta')$  será diferenciável em  $\theta$ , para  $\theta$  e  $\theta'$  em  $\Theta'$ . Para isso, usamos a Proposição 3.5. De (3.4), sabemos que

$$|R(x, y)| \leq \frac{8}{\pi},$$

para  $x \in (0, 2r)$ ,  $y \in [x, \infty)$ . Daí, segue que

$$|\log f_{Y_t}(y; \phi_t) R(x_i, y) f_{Y_t}(y; \phi'_t)| \leq \left| \frac{8}{\pi} \log f_{Y_t}(y; \phi_t) f_{Y_t}(y; \phi'_t) \right|. \quad (3.27)$$

Conforme foi visto anteriormente em (3.16), para modelos de mistura composta por distribuições lognormal, temos que  $\left| \frac{\partial \log f_{Y_t}(y; \phi_t)}{\partial \theta_i} \right|$ , com  $i = 2, \dots, 5$ , é uma soma de no máximo dois termos da forma

$$\frac{|\log y - \mu_t|^k}{\sigma_t^l},$$

com  $k$  e  $l$  sendo inteiros não negativos. Daí e por (3.27), temos que

$$\begin{aligned} \left| \frac{\partial}{\partial \theta_i} [\log f_{Y_t}(y; \phi_t) R(x_i, y) f_{Y_t}(y; \phi'_t)] \right| &\leq \left| \frac{8}{\pi} \frac{\partial}{\partial \theta_i} [\log f_{Y_t}(y; \phi_t) f_{Y_t}(y; \phi'_t)] \right| \\ &= \left| \frac{8}{\pi} \frac{|\log y - \mu_t|^k}{\sigma_t^l} f_{Y_t}(y; \phi'_t) \right| \end{aligned} \quad (3.28)$$

Portanto, se existe uma função  $w(y)$  integrável, tal que

$$\frac{|\log y - \mu_t|^k}{\sigma_t^l} f_{Y_t}(y; \phi'_t) \leq w(y),$$

então, pela Proposição 3.5, é permitido tomar as derivadas sob o sinal da integral em (3.26). Observe que para  $\phi$  e  $\phi' \in \Theta' \subseteq \tilde{\Theta}$ , temos por (3.8), visto na demonstração do Lema 3.7, que

$$\frac{|\log y - \mu_t|^k}{\sigma_t^l} f_{Y_t}(y; \phi'_t) \leq \frac{(|\log y| + m_t^*)^k}{\left(\frac{\sigma_t^*}{s_t}\right)^{l+1} y \sqrt{2\pi}} \exp \left\{ -\frac{(|\log y| - m_t^*)^2 + (m_t^*)^2}{2(s_t \sigma_t^*)^2} \right\} := w(y),$$

no qual é integrável por (3.9). Assim,

$$\begin{aligned} \nabla_1 E [\log f_{Y_t}(Y_{ti}; \phi_t) | x_i; \phi'_t] &= \nabla_1 [\pi(x_i; \phi'_t) \log f_{Y_t}(x_i; \phi_t)] + \\ &+ \nabla_1 \frac{(f_{X_t}(x_i; \phi'_t))^{-1}}{\sqrt{4r^2 - x_i^2}} \int_{x_i}^{\infty} \nabla_1 [\log f_{Y_t}(y'; \phi_t) R(x_i, y') f_{Y_t}(y'; \phi'_t)] dy'. \end{aligned}$$

Portanto,  $\sum_{t=1}^2 \sum_{i=1}^n \tau_t(x_i; \theta') E [\log f_{Y_t}(Y_{ti}; \phi_t) | x_i; \phi'_t]$  é contínua em  $\theta$  e  $\theta' \in \Theta'$  e diferenciável em  $\theta$ , para  $\theta$  e  $\theta' \in \Theta'$ . Então, RW5 e RW6 são satisfeitas quando  $\Theta' \subseteq \tilde{\Theta}$ . Pela Proposição 3.4, temos que para todo  $n$  suficientemente grande,  $\tilde{\theta}^{(k)} - \hat{\theta}_{MV} \xrightarrow{q.c.} 0$ , quando  $k \rightarrow \infty$ , sempre que  $\theta^{(0)}$  é suficientemente próximo à  $\hat{\theta}_{MV}$ . Portanto,  $\tilde{\theta}^{(k)} - \hat{\theta}_{MV} \xrightarrow{p.} 0$  com  $k \rightarrow \infty$  e assim existe uma sequência  $\{k(n)\}_{n=1}^{\infty}$  com  $k(n) \rightarrow \infty$  com  $n \rightarrow \infty$  tal que  $\sqrt{n}(\tilde{\theta}^{(k(n))} - \hat{\theta}_{MV}) \xrightarrow{p.} 0$ . ■

### 3.5 Convergência do Algoritmo MCEM Para o Algoritmo EM

Nesta seção, verificamos a convergência do algoritmo MCEM proposto para o algoritmo EM. Fazemos isso, assumindo que ambos algoritmos possuem o mesmo valor inicial e que a maximização do parâmetro é feita sobre o espaço compacto de parâmetros  $\Theta'$ . Observe que no Lema 3.10 apresentado logo abaixo,  $m$  denota o número de valores gerados em cada iteração do algoritmo MCEM.

A convergência mostrada no Lema 3.10 é obtida quando  $k$ ,  $m$  e  $n$  tendem ao infinito cooperativamente. O termo *cooperativamente* é definido a seguir.

**Definição 3.9.** Dizemos que  $m$ ,  $k$  e  $n$  tendem para o infinito cooperativamente se  $m$ ,  $k$  e  $n$  tendem ao  $\infty$ , se  $k = k(n)$  é um sequência satisfazendo o Lema 3.8, e

- (i) Se  $M < \frac{1}{5}$  então  $\frac{1}{m} = o\left(\frac{1}{n^2}\right)$ ;
- (ii) Se  $M = \frac{1}{5}$  então  $\frac{1}{m} = o\left(\frac{1}{n^2 k^2(n)}\right)$ ;
- (iii) Se  $M > \frac{1}{5}$  então  $\frac{1}{m} = o\left(\frac{1}{n^2 M^{2k(n)}}\right)$ .

**Lema 3.10.** *Assuma que  $f_Y(y; \theta)$  é uma mistura de duas distribuições lognormal. Então se  $k$ ,  $m$  e  $n$  tendem cooperativamente para o infinito, temos que*

$$\sqrt{n}(\tilde{\theta}_S^{(k)} - \tilde{\theta}^{(k)}) \xrightarrow{p} 0. \quad (3.29)$$

**Demonstração do Lema 3.10 :** Essa demonstração é feita, provando que

$$\lim_{k,m,n \rightarrow \infty} E[\sqrt{n}|\tilde{\theta}_{iS}^{(k)} - \tilde{\theta}_i^{(k)}|; \theta^*] = 0, \quad i = 1, \dots, 5,$$

pois, pela desigualdade de Chebyshev, temos que

$$P(\sqrt{n}|(\tilde{\theta}_S^{(k)} - \tilde{\theta}^{(k)})| \geq \epsilon) \leq \frac{1}{\epsilon} E[\sqrt{n}|\tilde{\theta}_{iS}^{(k)} - \tilde{\theta}_i^{(k)}|; \theta^*].$$

Assim, se  $\lim_{k,m,n \rightarrow \infty} E[\sqrt{n}|\tilde{\theta}_{iS}^{(k)} - \tilde{\theta}_i^{(k)}|; \theta^*] = 0$ ,  $i = 1, \dots, 5$ , então  $\sqrt{n}(\tilde{\theta}_S^{(k)} - \tilde{\theta}^{(k)}) \xrightarrow{p} 0$ . Como queremos mostrar a convergência dos estimadores gerados via algoritmo MCEM para os estimadores gerados via algoritmo EM, então vamos relembrá-los. Das equações (2.50) à (2.53), estudadas no final da Seção 2.4, temos que os estimadores EM são dados por

$$\hat{\epsilon}^{(k+1)} = \hat{\epsilon}(\theta^{(k)}) = \frac{1}{n} \sum_{i=1}^n \tau(x_i; \theta^{(k)}), \quad \tau(x_i; \theta^{(k)}) = \frac{\epsilon^{(k)} f_{X_1}(x_i; \phi_1^{(k)})}{f_X(x_i; \theta^{(k)})},$$

$$\hat{\mu}_1^{(k+1)} = \hat{\mu}_1(\theta^{(k)}) = \frac{\sum_{i=1}^n \tau(x_i; \theta^{(k)}) E(\log Y_i | x_i, Z_i = 1; \theta^{(k)})}{\sum_{i=1}^n \tau(x_i; \theta^{(k)})},$$

e

$$(\hat{\sigma}_1^2)^{(k+1)} = \hat{\sigma}_1^2(\theta^{(k)}) = \frac{\sum_{i=1}^n \tau(x_i; \theta^{(k)}) E[(\log Y_i)^2 | x_i, Z_i = 1; \theta^{(k)}]}{\sum_{i=1}^n \tau(x_i; \theta^{(k)})} - (\hat{\mu}_1^{(k+1)})^2.$$

Para  $\hat{\mu}_2^{(k+1)}$  e  $(\hat{\sigma}_2^2)^{(k+1)}$ , basta reescrever  $\hat{\mu}_1^{(k+1)}$  e  $(\hat{\sigma}_1^2)^{(k+1)}$  com  $Z_i = 0$  no lugar de  $Z_i = 1$  e  $1 - \tau(x_i; \theta^{(k)})$  no lugar de  $\tau(x_i; \theta^{(k)})$ . Agora, para o algoritmo MCEM, por (2.54)

temos que os estimadores são dados por

$$\hat{\mu}_{1S}^{(k+1)} = \hat{\mu}_{1S}(\theta^{(k)}) = \frac{\sum_{i=1}^n \tau(x_i; \theta^{(k)}) \frac{1}{m} \sum_{j=1}^m \log y'_j}{\sum_{i=1}^n \tau(x_i; \theta^{(k)})},$$

e

$$(\hat{\sigma}_{1S}^2)^{(k+1)} = \hat{\sigma}_{1S}^2(\theta^{(k)}) = \frac{\sum_{i=1}^n \tau(x_i; \theta^{(k)}) \frac{1}{m} \sum_{j=1}^m (\log y'_j)^2}{\sum_{i=1}^n \tau(x_i; \theta^{(k)})} - (\hat{\mu}_{1S}^{(k+1)})^2.$$

Para  $\hat{\mu}_{2S}^{(k+1)}$  e  $(\hat{\sigma}_{2S}^2)^{(k+1)}$ , basta reescrever  $\hat{\mu}_{1S}^{(k+1)}$  e  $(\hat{\sigma}_{1S}^2)^{(k+1)}$  com  $1 - \tau(x_i; \theta^{(k)})$  no lugar de  $\tau(x_i; \theta^{(k)})$ , onde  $E(\log Y_i | x_i, Z_i = 0; \theta^{(k)})$  e  $E[(\log Y_i)^2 | x_i, Z_i = 0; \theta^{(k)}]$  são aproximadas, gerando  $m$  valores  $y'_1, \dots, y'_m$  da distribuição condicional de  $Y$  dado  $X$  para fibras com parâmetros  $\phi_2$ , usando

$$\frac{1}{m} \sum_{j=1}^m \log y'_j \quad \text{e} \quad \frac{1}{m} \sum_{j=1}^m (\log y'_j)^2.$$

Seja  $\tilde{\theta}(\theta')$  a notação para qualquer uma das funções  $\hat{\varepsilon}^{(k+1)}$ ,  $\hat{\mu}_t^{(k+1)}$  e  $(\hat{\sigma}_t^2)^{(k+1)}$ , com  $t = 1, 2$ . Note que  $\tilde{\theta}(\theta')$  são funções contínuas e diferenciáveis no conjunto compacto  $\Theta'$ , pois são dadas em função de  $f_X(x; \theta')$ , e também por (3.25) e (3.26), temos que

$$\begin{aligned} E[(\log Y_{ti})^{2a} | x_i; \phi'_t] &= p_{nc}(x) (\log x_i)^{2a} \frac{f_{Y_t}(x; \phi'_t)}{f_{X_t}(x; \phi'_t)} + \\ &+ \frac{(f_{X_t}(x_i; \phi'_t))^{-1}}{\sqrt{4r^2 - x_i^2}} \int_{x_i}^{\infty} R(x_i, y) (\log y)^{2a} f_{Y_t}(y; \phi'_t) dy, \quad a = \frac{1}{2}, 1, 2, \end{aligned} \quad (3.30)$$

para  $t = 1, 2$ . Agora, observe que a esperança condicional  $E[(\log Y_{ti})^{2a} | x_i; \phi'_t]$  é contínua e diferenciável com respeito à  $\phi'_t$ , com  $t = 1, 2$ , uma vez que por demonstração análoga à feita no Lema 3.7 em (3.7) até (3.9) e fazendo  $z = \log y$ , temos que

$$\begin{aligned} E[(\log Y_{ti})^{2a} | x_i, \phi'_t] &\leq (\log x_i)^{2a} + \frac{8(f_{X_t}(x_i; \phi'_t))^{-1}}{\pi \sqrt{4r^2 - x_i^2}} \int_0^{\infty} (\log y)^{2a} f_{Y_t}(y; \phi'_t) dy \\ &\leq (\log x_i)^{2a} + \frac{8}{\pi} \left( \int_{x_i}^{\infty} R(x_i, y) f_{Y_t}(y; \phi'_t) dy \right)^{-1} \int_{-\infty}^{\infty} z^{2a} \varphi \left( \frac{z - \mu_t}{\sigma_t} \right) dz \\ &\leq (\log x_i)^{2a} + C_a, \end{aligned} \quad (3.31)$$

para  $x_i \in (0, 2r)$ , onde  $C_a > 0$  é finito, pois aqui usamos (3.15) e o fato de todos os momentos da distribuição normal existirem. Então, uma vez que as funções  $\tilde{\theta}(\theta')$  são contínuas e diferenciáveis sobre o conjunto compacto  $\Theta'$  e juntamente com a equação (3.30), para algum  $M$  finito e positivo, temos que

$$|\tilde{\theta}_i(\theta) - \tilde{\theta}_i(\theta')| \leq M \sum_{j=1}^5 |\theta_j - \theta'_j|, \quad i = 1, \dots, 5 \quad (3.32)$$

é satisfeita para  $\theta, \theta' \in \Theta'$ . Aplicando (3.32) recursivamente e pelo fato de que o algoritmo MCEM proposto tem o mesmo valor inicial do algoritmo EM, temos que para  $i = 1, \dots, 5$

$$\begin{aligned} |\tilde{\theta}_{iS}^{(k)} - \tilde{\theta}_i^{(k)}| &= |\tilde{\theta}_{iS}(\tilde{\theta}_S^{(k-1)}) - \tilde{\theta}_i(\tilde{\theta}_S^{(k-1)})| \\ &\leq |\tilde{\theta}_{iS}(\tilde{\theta}_S^{(k-1)}) - \tilde{\theta}_i(\tilde{\theta}_S^{(k-1)})| + |\tilde{\theta}_i(\tilde{\theta}_S^{(k-1)}) - \tilde{\theta}_i(\tilde{\theta}^{(k-1)})| \\ &\leq |\tilde{\theta}_{iS}(\tilde{\theta}_S^{(k-1)}) - \tilde{\theta}_i(\tilde{\theta}_S^{(k-1)})| + M \sum_{j=1}^5 |\tilde{\theta}_{jS}^{(k-1)} - \tilde{\theta}_j^{(k-1)}| \\ &\leq |\tilde{\theta}_{iS}(\tilde{\theta}_S^{(k-1)}) - \tilde{\theta}_i(\tilde{\theta}_S^{(k-1)})| + \\ &+ M \sum_{j=1}^5 \left\{ |\tilde{\theta}_{jS}(\tilde{\theta}_S^{(k-2)}) - \tilde{\theta}_j(\tilde{\theta}_S^{(k-2)})| + M \sum_{h=1}^5 |\tilde{\theta}_{hS}^{(k-2)} - \tilde{\theta}_h^{(k-2)}| \right\} \\ &\leq \dots \leq |\tilde{\theta}_{iS}(\tilde{\theta}_S^{(k-1)}) - \tilde{\theta}_i(\tilde{\theta}_S^{(k-1)})| + \\ &+ \frac{1}{5} \sum_{j=2}^5 \sum_{h=1}^{k-1} |\tilde{\theta}_{jS}(\tilde{\theta}_S^{(h-1)}) - \tilde{\theta}_j(\tilde{\theta}_S^{(h-1)})| (5M)^{k-h}. \end{aligned} \quad (3.33)$$

Observe que  $\hat{\varepsilon}(\theta^{(k)})$  é atualizado pelo  $x_i$  observado,  $i = 1, \dots, n$  e depende exclusivamente de  $\theta^{(k)}$ , e portanto  $\hat{\varepsilon}_S(\theta^{(k)}) \equiv \hat{\varepsilon}(\theta^{(k)})$ , isto é,  $\hat{\varepsilon}_S(\tilde{\theta}_S^{(h-1)}) - \hat{\varepsilon}(\tilde{\theta}_S^{(h-1)}) = 0$ ,  $h = 1, \dots, k-1$ , lembrando que ambos algoritmos possuem o mesmo valor inicial. Agora, para qualquer  $\theta' \in \Theta'$ , temos que para  $t = 1, 2$ ,

$$E[|\tilde{\mu}_{tS}(\theta') - \tilde{\mu}_t(\theta')|^2; \theta^*] = E \left[ E' \left[ \frac{|\sum_{i=1}^n (\tau_t(x_i; \theta') \tilde{U}'_{ti})|^2}{(\sum_{i=1}^n \tau_t(x_i; \theta'))^2} \middle| x, \theta' \right]; \theta^* \right], \quad (3.34)$$

onde  $\tilde{U}'_{ti} = \frac{1}{m} \sum_{j=1}^m U'_{tij}$  e  $U'_{tij} = \log Y'_{tij} - E'[\log Y'_{tij} | x_i, \phi'_t]$ .

Aqui,  $\{Y'_{tij}\}_{j=1}^m$  são independentes e identicamente distribuídas, valores gerados a partir da distribuição condicional de  $Y_t$  dado  $x_i$ , com parâmetros  $\phi'_t$ ,  $t = 1, 2$ ,  $i = 1, \dots, n$  e a esperança  $E'$  com respeito à distribuição gerada. Note que  $\{\tilde{U}'_{ti}\}_{i=1}^n$  são variáveis

aleatórias com

$$E'[\tilde{U}'_{tij}|x_i, \phi'_t] = 0, \quad \text{e} \quad E'[|\tilde{U}'_{ti}|^2|x_i, \phi'_t] \leq \frac{1}{m} E'[(\log Y'_{tij})^2|x_i, \phi'_t].$$

Observamos, que de forma análoga à (3.31) e utilizando (2.14), para qualquer  $a$  inteiro finito, temos que

$$E'[(\log Y'_{tij})^{2a}|x_i, \phi'_t] \leq (\log x_i)^{2a} + C_a, \quad (3.35)$$

para  $x_i \in (0, 2r)$ . Assim, por (3.34) e (3.35), temos que

$$\begin{aligned} E'[|\tilde{\mu}_{tS}(\theta') - \tilde{\mu}_t(\theta')|^2|x, \theta'] &\leq \frac{2 \sum_{i=1}^n \tau_t^2(x_i; \theta') E'[|\tilde{U}'_{ti}|^2|x_i, \phi'_t]}{(\sum_{i=1}^n \tau_t(x_i; \theta'))^2} \\ &\leq \frac{2}{m} \sum_{i=1}^n (\log x_i)^2 + \frac{2n}{m} C_1, \end{aligned} \quad (3.36)$$

onde  $\frac{\sum_{i=1}^n \tau_t^2(x_i; \theta')}{(\sum_{i=1}^n \tau_t(x_i; \theta'))^2} \leq 1$ . Pela desigualdade de Jensen, para qualquer variável aleatória  $X$  com esperança finita, temos que

$$E[|X|] \leq (E[X^2])^{1/2}. \quad (3.37)$$

Portanto, temos que

$$\begin{aligned} E[|\tilde{\mu}_{tS}(\theta') - \tilde{\mu}_t(\theta')|; \theta^*] &\leq (E[E'[|\tilde{\mu}_{tS}(\theta') - \tilde{\mu}_t(\theta')|^2|x, \theta']; \theta^*])^{1/2} \\ &\leq \left( \frac{2n}{m} \int_0^{2r} (\log x)^2 f_{X_t}(x; \phi_t^*) dx + \frac{2nC_1}{m} \right)^{1/2} \\ &= O(\sqrt{n/m}), \end{aligned} \quad (3.38)$$

uma vez que pela equação (3.17), sabemos que  $\int_0^{2r} |\log x|^j f_X(x; \theta^*) dx < \infty \forall j > 0$ . Agora, seja  $\tilde{P}'_{ti} = (1/m) \sum_{j=1}^m P'_{tij}$  e  $P'_{tij} = (\log Y'_{tij})^2 - E'[(\log Y'_{tij})^2|x_i, \phi'_t]$ , então pela desigualdade triangular, temos que

$$|\tilde{\sigma}_{tS}^2(\theta') - \tilde{\sigma}_t^2(\theta')| \leq \frac{|\sum_{i=1}^n (\tau_t(x_i; \theta') \tilde{P}'_{ti})|}{\sum_{i=1}^n \tau_t(x_i; \theta')} + |\tilde{\mu}_{tS}^2(\theta') - \tilde{\mu}_t^2(\theta')|, \quad (3.39)$$

$t = 1, 2$ . Observe que  $E'[P'_{tij}|x_i, \phi'_t] = 0$  e pela equação (3.31), temos que

$$E'[|\tilde{P}'_{ti}|^2|x_i, \phi'_t] \leq \frac{1}{m} ((\log x_i)^4 + C_2),$$

e conseqüentemente, para o primeiro termo de (3.39), temos que

$$\begin{aligned} \frac{E'[\sum_{i=1}^n \tau_t(x_i; \theta') \tilde{P}'_{ti} | x, \phi'_t]}{(\sum_{i=1}^n \tau_t(x_i; \theta'))^2} &\leq \frac{2 \sum_{i=1}^n \tau_t^2(x_i; \theta') E'[\tilde{P}'_{ti} | x_i, \phi'_t]}{(\sum_{i=1}^n \tau_t(x_i; \theta'))^2} \\ &\leq \frac{2}{m} \sum_{i=1}^n (\log x_i)^4 + \frac{2n}{m} C_2. \end{aligned}$$

Conforme raciocínio similar ao feito em (3.35), temos que

$$E \left[ \frac{|\sum_{i=1}^n \tau_t(x_i; \theta') \tilde{P}'_{ti}|^2}{(\sum_{i=1}^n \tau_t(x_i; \theta'))^2}; \theta^* \right] = O(n/m). \quad (3.40)$$

Para o segundo termo da equação (3.39), por (3.37) e (3.38) e pelo fato de  $\tilde{\mu}_t(\theta') \in \Theta'$  ser limitado, temos que

$$\begin{aligned} E[|\tilde{\mu}_{tS}^2(\theta') - \tilde{\mu}_t^2(\theta')|; \theta^*] &\leq E[|\tilde{\mu}_{tS}(\theta') - \tilde{\mu}_t(\theta')|^2; \theta^*] + 2E[|\tilde{\mu}_t(\theta')| |\tilde{\mu}_{tS}(\theta') - \tilde{\mu}_t(\theta')|; \theta^*] \\ &\leq O(n/m) + O(\sqrt{n/m}). \end{aligned} \quad (3.41)$$

Assim, de (3.37), (3.39) e (3.41), obtemos

$$E[|\tilde{\sigma}_{tS}(\theta') - \tilde{\sigma}_t(\theta')|; \theta^*] = O(\sqrt{n/m}). \quad (3.42)$$

De (3.33), (3.38) e (3.42), temos portando para  $i = 1, \dots, 5$ ,

$$\begin{aligned} E[\sqrt{n}|\tilde{\theta}_{iS}^{(k)} - \tilde{\theta}_i^{(k)}|; \theta^*] &\leq \sqrt{n}E[|\tilde{\theta}_{iS}(\tilde{\theta}_S^{(k-1)}) - \tilde{\theta}_i(\tilde{\theta}_S^{(k-1)})|; \theta^*] + \\ &+ \frac{1}{5} \sum_{h=1}^{k-1} (5M)^{k-h} \sum_{j=2}^5 \sqrt{n}E[|\tilde{\theta}_{jS}(\tilde{\theta}_S^{(h-1)}) - \tilde{\theta}_j(\tilde{\theta}_S^{(h-1)})|; \theta^*] \\ &= O(n/\sqrt{m}) + O(n/\sqrt{m}) \sum_{h=1}^{k-1} (5M)^{k-h} \\ &\leq O(n/\sqrt{m}) \sum_{h=1}^k (5M)^{k-h}. \end{aligned} \quad (3.43)$$

Analizando o limite em (3.43) quando  $k$ ,  $m$  e  $n$  tendem ao infinito cooperativamente, temos que:

(a) *Caso*  $M = \frac{1}{5}$  : Do fato que  $\frac{n^2 k^2(n)}{m} \xrightarrow{k, m, n \rightarrow \infty} 0$ , segue também que  $\frac{nk(n)}{\sqrt{m}} \xrightarrow{k, m, n \rightarrow \infty}$

0. Além do mais, para  $M = \frac{1}{5}$ , temos que  $\sum_{h=1}^k (5M)^{k-h} = k$ . Daí, para  $k$ ,  $m$  e  $n$

suficientemente grandes, temos que

$$O\left(\frac{n}{\sqrt{m}}\right) \sum_{h=1}^k (5M)^{k-h} = O\left(\frac{n}{\sqrt{m}}\right) k \leq cte \frac{n}{\sqrt{m}} k, \text{ para alguma } cte > 0.$$

Portanto, para  $M = \frac{1}{5}$ , temos que  $E[\sqrt{n}|\tilde{\theta}_{iS}^{(k)} - \tilde{\theta}_i^{(k)}|; \theta^*] \xrightarrow{k, m, n \rightarrow \infty} 0$ .

(b) *Caso*  $M < \frac{1}{5}$  : Do fato que  $\frac{n^2}{m} \xrightarrow{k, m, n \rightarrow \infty} 0$ , então  $\frac{n}{\sqrt{m}} \xrightarrow{k, m, n \rightarrow \infty} 0$ . Além do mais, para  $M < \frac{1}{5}$ , temos que

$$\sum_{h=1}^k (5M)^{k-h} = (5M)^k \left[ \frac{1}{5M-1} - \frac{1}{(5M-1)(5M)^k} \right] = \frac{(5M)^k - 1}{5M-1}. \quad (3.44)$$

Daí, para  $k$ ,  $m$  e  $n$  suficientemente grandes, temos que

$$O\left(\frac{n}{\sqrt{m}}\right) \sum_{h=1}^k (5M)^{k-h} = O\left(\frac{n}{\sqrt{m}}\right) \frac{(5M)^k - 1}{5M-1} \leq cte \frac{n}{\sqrt{m}} \frac{(5M)^k - 1}{5M-1}, \text{ para alguma } cte > 0.$$

Portanto, para  $M < \frac{1}{5}$ , temos que  $E[\sqrt{n}|\tilde{\theta}_{iS}^{(k)} - \tilde{\theta}_i^{(k)}|; \theta^*] \xrightarrow{k, m, n \rightarrow \infty} 0$ .

(c) *Caso*  $M > \frac{1}{5}$  : Sabemos que  $\frac{n^2 M^{2k(n)}}{m} \xrightarrow{k, m, n \rightarrow \infty} 0$ , e assim,  $\frac{nM^{k(n)}}{\sqrt{m}} \xrightarrow{k, m, n \rightarrow \infty} 0$ . Daí, para  $k$ ,  $m$  e  $n$  suficientemente grandes, temos que

$$O\left(\frac{n}{\sqrt{m}}\right) \sum_{h=1}^k (5M)^{k-h} = O\left(\frac{n}{\sqrt{m}}\right) \frac{(5M)^k - 1}{5M-1} \leq cte \frac{n}{\sqrt{m}} \frac{(5M)^k}{5M-1}, \text{ para alguma } cte > 0.$$

Portanto, para  $M > \frac{1}{5}$ , temos também que  $5M > 1 > \frac{1}{5}$  e consequentemente,  $E[\sqrt{n}|\tilde{\theta}_{iS}^{(k)} - \tilde{\theta}_i^{(k)}|; \theta^*] \xrightarrow{k, m, n \rightarrow \infty} 0$ .

Então, para  $M \neq \frac{1}{5}$  e  $M = \frac{1}{5}$ , concluímos que  $E[\sqrt{n}|\tilde{\theta}_{iS}^{(k)} - \tilde{\theta}_i^{(k)}|; \theta^*] \xrightarrow{k, m, n \rightarrow \infty} 0$ , quando  $m$ ,  $k$  e  $n$  tendem ao infinito cooperativamente. E isto completa a prova do Lema 3.10. ■

Portanto, através dos Lemas 3.7, 3.8 e 3.10, mostramos a consistência e a normalidade assintótica do algoritmo MCEM proposto por Svensson et al. [18], para uma mistura de duas distribuições lognormal quando os dados da amostra são censurados.

Note que todas as provas neste capítulo foram feitas para uma mistura de duas componentes, mas valem para qualquer mistura com número finito de componentes. Nas demonstrações, usamos principalmente o fato de que para as funções de densidades lognormal independentes, as suas diversas derivadas parciais são limitadas e integráveis.

E a outra parte principal das demonstrações, consistiu em mostrar a continuidade e a diferenciabilidade dos momentos dos logaritmos das densidades lognormal independentes. Daí as provas são extensíveis a qualquer mistura finita de distribuições lognormais.

Acreditamos também ser possível generalizar as provas aqui realizadas, para o caso em que a distribuição do comprimento das células, que pelo menos aparecem parcialmente no núcleo de incremento, pertencem a uma mistura finita de distribuições de uma família exponencial. Em Svensson et al. [18] podemos encontrar um estudo de simulação e uma aplicação do modelo estudado neste trabalho e também apresentado pelos autores.

# Referências Bibliográficas

---

- [1] Ash, R. B. e Doléans-Dade, C. A., *Probability and Measure Theory*, 2nd Ed, Academic Press, New York, 2000.
- [2] Atienza, N., Garcia-Heras, J. and Munoz-Pichardo, J. M., *A New Condition for Identifiability of Finite Mixture Distributions*, *Metrika*, vol.63, 2003, 215-221.
- [3] Casella, G. e Berger, R. L., *Statistical Inference*, 2nd, Duxbury Press, Califórnia, 2002.
- [4] Casella, G. e Lehmann, E. L., *Theory of Point Estimation*, 2nd, Springer, New York, 1998.
- [5] Lehmann, E. L., *Theory of Point Estimation*, 2nd, Wiley, New York, 1983.
- [6] Chung, K. L., *A Course of Probability Theory*, 3rd Ed, Academic Press, New York, 2001.
- [7] Dempster, A. P., Laird N. M. e Rubin, D. B., *Maximum Likelihood from Incomplete Data via the EM Algorithm*, *Journal of the Royal Statistical Society, Series B*, vol. 39, n°.1, 1977, 1-38.
- [8] Feller, W., *An Introduction to Probability Theory and Its Applications*, 2nd Ed, John Wiley & Sons, Inc., New York, 1971.
- [9] Lima , E. L., *Curso de Análise*, vol.1, IMPA, Rio de Janeiro, 2010, 403.
- [10] Magalhães, M.N., *Probabilidade e Variáveis Aleatórias*, Edusp, São Paulo, 2006.
- [11] Martins Neto, D. S. B., *Ergodicidade Uniforme das Cadeias de Markov Associadas aos Algoritmos MCEM e StEM*, Universidade de Brasília, Brasília, 2003.

- 
- [12] McLachlan, G. e Peel, D., *Finite Mixture Models*, John Wiley & Sons, Inc., New York, 2000.
- [13] Mörling, T., Sjöstedt-deLuna, S., Svensson, I., Fries, A. and Ericsson, T., *A Method to Estimate Fibre Length Distribution in Conifers Based on Wood Samples from Increment Cores*, *Holzforschung* **74**,2003, 248-254.
- [14] Rao, C. R., *Linear Statistical Inference and Its Applications*, 2.ed., New York: John Wiley & Sons, 1973, 552p.
- [15] Redner, R. A. e Walker, H. F., *Mixture Densities, Maximum Likelihood and The EM Algorithm*, *SIAM Review*, vol. 26, n<sup>o</sup>.2, 1984, 195-239.
- [16] Spanos, A. *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data*, Cambridge University Press, New York, 1999.
- [17] Svensson, I., Sjöstedt-deLuna, S. *Asymptotic Properties of a Stochastic EM Algorithm for Mixtures with Censored Data*, *J. Statisc.* **140**, 2010, 111-127.
- [18] Svensson, I., Sjöstedt-deLuna, S., Bondesson, L., *Estimation of Fibre Length Distributions from Censored Data through an EM Algorithm*, *J. Statisc.* **33**, 2006, 503-522.
- [19] Titterington, D. M., Smith, A. F. M., Makov, U. E., *Statistical Analysis of Finite Mixture Distributions*, John Wiley & Sons, Inc., New York, 1985.
- [20] Wei, G.C.G. e Tanner, M.A. *A Monte Carlo Implementation of the EM Algorithm and Poor Man's Data Augmentation Algorithms*, *Journal of American Statistical Association* **85** (411), 1990, 699-704.
- [21] Wu, C.F.J. *On the Convergence Properties of the EM Algorithm*, *The Annals of Statistics* **11** (1), 1983, 95-103.