

UNIVERSIDADE FEDERAL DO AMAZONAS - UFAM
INSTITUTO DE CIÊNCIAS EXATAS - ICE
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

Misturas de Distribuições t de Student Assimétricas

JOCELY NASCIMENTO LOPES

MANAUS

2008

UNIVERSIDADE FEDERAL DO AMAZONAS - UFAM
INSTITUTO DE CIÊNCIAS EXATAS - ICE
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

JOCELY NASCIMENTO LOPES

*MISTURAS DE DISTRIBUIÇÕES t DE STUDENT
ASSIMÉTRICAS*

Dissertação apresentada ao Programa de Pós-Graduação em Matemática da Universidade Federal do Amazonas, como requisito final para obtenção do título de Mestre em Matemática, área de concentração em Estatística.

Orientador: Prof. Dr. Celso Rômulo Barbosa Cabral

MANAUS
2008

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

L864m Lopes, Jocely Nascimento
Misturas de distribuições T de student assimétricas / Jocely
Nascimento Lopes. 2008
60 f.: il. color; 31 cm.

Orientador: Celso Rômulo Barbosa Cabral
Coorientador: José Raimundo Gomes Pereira
Dissertação (Mestrado em Matemática Pura e Aplicada) -
Universidade Federal do Amazonas.

1. Modelo Normal Assimétrico. 2. Modelo t de Student. 3. Modelo
Gamma. 4. Densidade Conjunta Reparametrizada. I. Cabral, Celso
Rômulo Barbosa II. Universidade Federal do Amazonas III. Título

JOCELY NASCIMENTO LOPES

MISTURAS DE DISTRIBUIÇÕES t DE STUDENT
ASSIMÉTRICAS

Dissertação apresentada ao Programa de Pós-Graduação em Matemática da Universidade Federal do Amazonas, como requisito final para obtenção do título de Mestre em Matemática, área de concentração em Estatística.

Manaus, 30 de agosto de 2008.

BANCA EXAMINADORA

.....
Prof. Dr. Celso Rômulo Barbosa Cabral, Presidente
Universidade Federal do Amazonas

.....
Prof. Dr. José Raimundo Gomes Pereira,
Universidade Federal do Amazonas

.....
Profa. Dra. Cibele Queiroz da Silva,
Universidade de Brasília

AGRADECIMENTOS

A Deus pela ajuda inestimável e pelas bênçãos que foram graciosamente concedidas durante esses anos de estudo.

A Jesus Cristo seu Filho em quem todos os tesouros da sabedoria e do conhecimento estão ocultos.

A minha mãe e meu filho que junto a Deus são os grandes responsáveis por essa vitória.

A Celso Rômulo Barbosa Cabral pela paciente orientação, Aos professores do mestrado pela experiência transmitida durante o curso.

A SEMED pelo suporte financeiro.

Aos amigos de estudo pelo apoio e excelente convivência.

É impossível expressar adequadamente toda a minha gratidão às inúmeras pessoas de onde obtive estímulo e assistência. A excelência do poder pertence a Deus, e não a nós. Exclusivamente a Ele seja todo louvor e glória.

RESUMO

Misturas de Distribuições t de Student Assimétricas

Este trabalho trata do problema de estimar parâmetros de uma mistura finita de densidades t-assimétricas. Como ferramenta para a estimação foi usado o algoritmo EM. Foi avaliada a consistência desses estimadores e realizado um experimento de aplicação da teoria desenvolvida para uma modelagem com dados reais utilizando um conjunto analisado anteriormente na literatura, relativo ao PIB per capita. Os objetivos centrais desse trabalho são apresentar uma descrição detalhada do método de estimação, via algoritmo EM, dos parâmetros do modelo finito de mistura de densidades t-assimétricas e avaliar através de um estudo de simulação se o estimador obtido é consistente.

ABSTRACT

Finite Mixtures of Skew Student-t Distributions

In this work we consider the estimation of parameters of a finite mixture of skew Student-t distributions, via EM algorithm. The main goals of this dissertation is to show a detailed description of the EM algorithm applied to this model and to evaluate the consistency of the estimator. A data set concerning the Gross Domestic Product per capita (Human Development Report), previously studied in the related literature, is analyzed.

Sumário

| | |
|---|-----------|
| Introdução | 1 |
| 1 Preliminares | 2 |
| 1.1 Algumas Definições | 2 |
| 1.1.1 O Modelo Normal Assimétrico Padrão | 2 |
| 1.1.2 O Modelo Normal Assimétrico | 2 |
| 1.1.3 Modelo Normal Truncada | 3 |
| 1.1.4 O Modelo Gamma | 3 |
| 1.1.5 O Modelo t de Student | 3 |
| 1.2 O Modelo t-Assimétrico | 4 |
| 1.2.1 Representação Estocástica | 6 |
| 1.2.2 A Densidade Conjunta de Y , γ e ζ | 8 |
| 1.2.3 A Função de Verossimilhança dos Dados Aumentados | 8 |
| 1.3 Estimação via Algoritmo EM | 10 |
| 1.3.1 Passo E | 11 |
| 1.3.2 Passo CM | 12 |
| 1.4 Reparametrização | 13 |
| 1.4.1 A Densidade Conjunta Reparametrizada | 13 |
| 1.4.2 A Função de Verossimilhança Reparametrizada | 14 |
| 1.4.3 Reformulação do Algoritmo ECME | 15 |
| 1.4.4 As Esperanças Condicionais do Passo E | 17 |
| 2 Mistura Finita de Densidades t-Assimétricas | 26 |
| 2.1 Mistura Finita de t-Assimétricas | 26 |
| 2.2 Função log-Verossimilhança para os dados completos | 27 |
| 2.3 Estimação via algoritmo ECME | 28 |
| 2.3.1 O Algoritmo ECME para Mistura Finita de Densidades t-Assimétricas | 32 |

| | | |
|----------|---|-----------|
| 3 | Consistência das Estimativas e Uma Aplicação | 34 |
| 3.1 | Estudo de Simulação | 34 |
| 3.1.1 | Processo de Simulação | 38 |
| 3.1.2 | Resultados | 38 |
| 3.2 | Aplicação | 49 |

Introdução

Em grande parte a análise estatística desenvolvida para o estudo de várias variáveis contínuas está baseada no modelo normal. No entanto, em muitas situações reais nos deparamos com dados dotados de assimetria. Para a modelagem destes tipos de dados Azzalini (1985) apresentou o modelo normal assimétrico. Nos casos em que esses conjuntos de dados apresentam valores discrepantes esse tipo de modelagem pode ser ineficaz. Em situações deste tipo, propõe-se como alternativa mais robusta que o modelo normal assimétrico, a versão assimétrica da distribuição t de Student introduzida por Azzalini & Capitanio (2003).

Nos casos de dados heterogêneos, onde sabemos que as observações pertencem a k populações distintas, mas não sabemos como discriminá-las, a modelagem se dá por mistura de distribuições. Uma introdução para esse contexto pode ser encontrada em McLachlan & Peel (2000) e Frühwirth-Schnatter (2006), com ponto de vista frequentistas e Bayesiano, respectivamente. Na literatura, existem estudos envolvendo misturas de distribuições normais, algumas referências podem ser encontradas nos livros citados anteriormente. Uma extensão desse tipo de modelagem foi introduzida por Lin *et al.* (2007b), que trata de misturas de densidades normais assimétricas. Ainda nesta direção uma modelagem mais robusta que a por mistura de distribuições normal assimétrica é introduzida por Lin *et al.* (2007a), onde foram consideradas misturas de distribuições t-assimétricas.

Neste trabalho, Apresentaremos a mistura finita de densidades t-assimétricas, a estimação de seus parâmetros via algoritmo EM e um estudo empírico das propriedades assintóticas desses estimadores. Inicialmente, definiremos o modelo t-assimétrico e o algoritmo EM para a obtenção dos estimadores dos parâmetros do modelo. Em seguida aplicaremos essa teoria para obter os estimadores da mistura finita de densidades t-assimétricas. Avaliaremos a consistência das estimativas geradas obtido pelo algoritmo e finalmente apresentaremos uma aplicação deste estudo em uma modelagem com dados reais.

Capítulo 1

Preliminares

Apresentaremos neste capítulo um estudo de estimação dos parâmetros do modelo t-assimétrico definido em Azzalini & Capitanio (2003). Inicialmente, baseado no artigo de Lin *et al.* (2007a), apresentaremos uma representação estocástica para o modelo e o algoritmo EM como ferramenta computacional para obtenção dos estimadores de seus parâmetros. Em seguida apresentaremos uma reparametrização do modelo t-assimétrico proposta por Lachos *et al.* (2009), com o objetivo de simplificar as iterações do algoritmo.

1.1 Algumas Definições

Nesta seção consideramos alguns modelos probabilísticos que serão utilizados no desenvolvimento deste trabalho.

1.1.1 O Modelo Normal Assimétrico Padrão

Dizemos que Z tem distribuição normal assimétrica padrão com parâmetro $\lambda \in \mathbb{R}$, denotado por $Z \sim SN(\lambda)$, se sua função densidade de probabilidade é dada por

$$g(z; \lambda) = 2\phi(z)\Phi(\lambda z), \quad -\infty < z < \infty. \quad (1.1)$$

onde $\phi(\cdot)$ e $\Phi(\cdot)$ são as funções de densidade de probabilidade e de distribuição da normal padrão, respectivamente.

1.1.2 O Modelo Normal Assimétrico

Seja Z com distribuição normal assimétrica padrão com parâmetro de assimetria λ . Dizemos que a variável aleatória Y dada pela transformação

$$Y = \xi + \sigma Z$$

tem distribuição normal assimétrica com parâmetros de locação $\xi \in \mathbb{R}$, escala $\sigma^2 > 0$ e de assimetria $\lambda \in \mathbb{R}$, a qual denotamos $Y \sim SN(\xi, \sigma^2, \lambda)$. A função de densidade de probabilidade é dada por

$$\psi(y; \xi, \sigma^2, \lambda) = \frac{2}{\sigma} \phi\left(\frac{y - \xi}{\sigma}\right) \Phi\left(\lambda \frac{y - \xi}{\sigma}\right). \quad (1.2)$$

1.1.3 Modelo Normal Truncada

Suponha que X tenha distribuição normal com média e variância dadas por μ e σ^2 , respectivamente. Seja $Y \sim X|X \in A$, com $A = \{a_1 < x < a_2\}$ – isto significa um truncamento de X no conjunto A . Então Y tem distribuição normal truncada em A , que denotamos $Y \sim NT(\mu, \sigma^2; (a_1, a_2))$. Neste caso a função densidade de probabilidade de Y é dada por

$$f(y | \mu, \sigma^2) = \{\Phi(\alpha_2) - \Phi(\alpha_1)\}^{-1} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y - \mu)^2\right\}, a_1 < y < a_2, \quad (1.3)$$

onde $\alpha_i = \frac{(a_i - \mu)}{\sigma}, i = 1, 2$.

1.1.4 O Modelo Gamma

Dizemos que a variável aleatória X tem distribuição gamma, com parâmetros α e β , que denotamos por $X \sim \Gamma(\alpha, \beta)$, se sua função de densidade de probabilidade é dada por

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, x > 0, \alpha, \beta > 0, \quad (1.4)$$

onde Γ é a função gamma.

1.1.5 O Modelo t de Student

Suponha que Z tenha distribuição normal padrão, isto é $Z \sim N(0, 1)$, que V tenha distribuição qui-quadrado com v graus de liberdade, e que Z e V sejam independentes. Então dizemos que a variável aleatória

$$T = \frac{Z}{\sqrt{\frac{V}{v}}}$$

tem distribuição t de Student com v graus de liberdade.

Neste caso, a função de densidade de probabilidade é dada por

$$f_v(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\Gamma\left(\frac{v}{2}\right)} \left[1 + \frac{t^2}{v}\right]^{-\frac{v+1}{2}}. \quad (1.5)$$

1.2 O Modelo t-Assimétrico

Para a modelagem de dados dotados de assimetria tem-se utilizado, muito freqüentemente, a distribuição normal assimétrica (SN) de Azzalini (1985). Este tipo de modelagem pode não ser adequada para dados com essas características e que apresentam valores discrepantes. Neste caso Azzalini & Capitanio (2003) apresenta um modelo assimétrico da distribuição t de Student. Uma das características do modelo t-assimétrico é o fato de ter caudas mais pesadas que as da SN, o que permite fazer inferência sobre os dados discrepantes antes não alcançados pelo modelo normal assimétrico.

Vamos apresentar o modelo t-assimétrico introduzido por Azzalini & Capitanio (2003).

Definição 1.2.1 Sejam Z e ζ variáveis aleatórias independentes tais que $Z \sim SN(\lambda)$ e $\zeta \sim \Gamma(v/2, v/2)$, ou seja, têm distribuições normal-assimétrica com parâmetro λ e gamma com parâmetro $v/2$ respectivamente. Dizemos que a variável aleatória

$$Y = \xi + \sigma \frac{Z}{\sqrt{\zeta}} \quad (1.6)$$

tem distribuição t-Assimétrica com parâmetros de locação $\xi \in \mathfrak{R}$, de escala $\sigma^2 \in (0, \infty)$, de assimetria $\lambda \in \mathfrak{R}$ e com $v > 0$ graus de liberdade, a qual denotamos $Y \sim ST(\xi, \sigma^2, \lambda, v)$.

A função de densidade neste caso é dada por

$$f(y) = \frac{2}{\sigma} t_v(\eta) T_{v+1} \left(\lambda \eta \sqrt{\frac{v+1}{\eta^2 + v}} \right), \quad (1.7)$$

onde t_v denota a função densidade de probabilidade de uma distribuição t de Student padrão com v graus de liberdade, T_{v+1} a função de distribuição de uma t de Student padrão com $v+1$ graus de liberdade e $\eta = \frac{y-\xi}{\sigma}$.

Observe que a distribuição SN é um caso particular da t-assimétrica. Para isso em (1.6) basta fazer $v \mapsto \infty$, de tal maneira que $\zeta \mapsto 1$ com probabilidade 1.

Para demonstrar (1.7) usaremos a seguinte proposição demonstrada em Azzalini & Capitanio (2003).

Proposição 1.2.1 Se $\zeta \sim \Gamma(\alpha, \beta), \forall \alpha \in \mathfrak{R}$, então

$$E \left[\Phi(a\sqrt{\zeta}) \right] = T_{2\alpha} \left(a\sqrt{\frac{\alpha}{\beta}} \right). \quad (1.8)$$

Assim, em (1.6) suponha conhecido o valor de ζ , então

$$Y|\zeta \sim SN\left(\xi, \frac{\sigma}{\sqrt{\zeta}}, \lambda\right). \quad (1.9)$$

Vamos denotar uma densidade genericamente pela letra f , sendo que o argumento especificará à qual variável aleatória a mesma está associada. Precisamos obter a densidade conjunta

$$f(y, \zeta) = f(y|\zeta)f(\zeta), \quad (1.10)$$

uma vez que

$$f(y) = \int f(y, \zeta)d\zeta.$$

Então de (1.9) e (1.10), temos

$$\begin{aligned} f(y, \zeta) &= \frac{2}{\sqrt{\frac{\sigma^2}{\zeta}}} \left\{ (2\pi)^{-1/2} \exp \left\{ -\frac{1}{2}\zeta \left(\frac{y-\xi}{\sigma} \right)^2 \right\} \Phi \left[\lambda\zeta^{1/2} \left(\frac{y-\xi}{\sigma} \right) \right] \right\} \\ &\quad \times \frac{1}{\Gamma\left(\frac{v}{2}\right)} \left(\frac{v}{2}\right)^{\left(\frac{v}{2}\right)} \zeta^{\left(\frac{v}{2}\right)-1} \exp \left\{ -\frac{v}{2}\zeta \right\} \\ &= \frac{2}{\sigma} \left\{ (2\pi)^{-1/2} \exp \left\{ -\frac{1}{2}\zeta \left(\frac{y-\xi}{\sigma} \right)^2 \right\} \Phi \left[\lambda\zeta^{1/2} \left(\frac{y-\xi}{\sigma} \right) \right] \right\} \\ &\quad \times \frac{1}{\Gamma\left(\frac{v}{2}\right)} \left(\frac{v}{2}\right)^{\left(\frac{v}{2}\right)} \zeta^{\left(\frac{v+1}{2}\right)-1} \exp \left\{ -\frac{v}{2}\zeta \right\} \\ &= \frac{2}{\sigma} (2\pi)^{-1/2} \frac{1}{\Gamma\left(\frac{v}{2}\right)} \left(\frac{v}{2}\right)^{\left(\frac{v}{2}\right)} \zeta^{\left(\frac{v+1}{2}\right)-1} \\ &\quad \times \exp \left\{ -\frac{1}{2} \left[\left(\frac{y-\xi}{\sigma} \right)^2 + v \right] \zeta \right\} \Phi \left[\lambda\zeta^{1/2} \left(\frac{y-\xi}{\sigma} \right) \right]. \end{aligned}$$

Fazendo $\eta = \frac{y-\xi}{\sigma}$, temos

$$f(y, \zeta) = \frac{2}{\sigma} \left\{ (2\pi)^{-1/2} \frac{1}{\Gamma\left(\frac{v}{2}\right)} \left(\frac{v}{2}\right)^{\frac{v}{2}} \zeta^{\frac{v+1}{2}-1} \exp \left\{ -\frac{1}{2} (\eta^2 + v) \zeta \right\} \Phi \left(\lambda\eta\sqrt{\zeta} \right) \right\} \quad (1.11)$$

Ainda em (1.11) vamos inserir o termo $\frac{\left(\frac{1}{2}(\eta^2+v)\right)^{\frac{v+1}{2}}}{\Gamma\left(\frac{v+1}{2}\right)}$ a fim de obter uma densidade da distribuição gamma com parâmetros $\left(\frac{v+1}{2}, \frac{1}{2}(\eta^2 + v)\right)$ em função de ζ , então

$$f(y) = \int f(y, \zeta)d\zeta$$

$$\begin{aligned}
&= \frac{2}{\sigma} \left\{ (2\pi)^{-1/2} \frac{1}{\Gamma(\frac{v}{2})} \left(\frac{v}{2}\right)^{\frac{v}{2}} \frac{\Gamma(\frac{v+1}{2})}{(\frac{1}{2}(\eta^2 + v))^{\frac{v+1}{2}}} \right\} \\
&\quad \times \int \Phi(\lambda\eta\sqrt{\zeta}) \frac{(\frac{1}{2}(\eta^2 + v))^{\frac{v+1}{2}}}{\Gamma(\frac{v+1}{2})} \zeta^{\frac{v+1}{2}-1} \exp\left\{-\frac{1}{2}(\eta^2 + v)\zeta\right\} d\zeta. \quad (1.12)
\end{aligned}$$

A integral é igual a

$$E\left[\Phi(\lambda\eta\sqrt{\zeta})\right].$$

Então por (1.8) a densidade (1.12) é igual a

$$f(y) = \frac{2}{\sigma} \left\{ \frac{\Gamma(\frac{v+1}{2}) \left(\frac{v}{2}\right)^{\frac{v}{2}}}{\Gamma(\frac{v}{2})\sqrt{2\pi}} \left[\frac{1}{2}(v + \eta^2)\right]^{-\frac{(v+1)}{2}} \right\} T_{v+1}\left(\sqrt{\frac{v+1}{\eta^2 + v}}\right),$$

No termo entre chaves vamos introduzir $\left(\frac{1}{v}\right)^{\frac{v+1}{2}}$, com o objetivo de obter a densidade t de Student com parâmetro v em função de η

$$\begin{aligned}
f(y) &= \frac{2}{\sigma} \left\{ \frac{\Gamma(\frac{v+1}{2}) \left(\frac{v}{2}\right)^{\frac{v}{2}} \left(\frac{1}{v}\right)^{\frac{v+1}{2}}}{\Gamma(\frac{v}{2})\sqrt{2\pi}} \left[\frac{1}{2}(v + \eta^2)\right]^{-\frac{(v+1)}{2}} \right\} T_{v+1}\left(\sqrt{\frac{v+1}{\eta^2 + v}}\right) \\
&= \frac{2}{\sigma} \left\{ \frac{\Gamma(\frac{v+1}{2}) v^{\frac{(v+1)}{2}} \left(\frac{v}{2}\right)^{\frac{v}{2}} \left(\frac{1}{2}\right)^{-\frac{(v+1)}{2}}}{\Gamma(\frac{v}{2})\sqrt{2\pi}} \left[\frac{1}{2}\left(1 + \frac{\eta^2}{v}\right)\right]^{-\frac{(v+1)}{2}} \right\} T_{v+1}\left(\sqrt{\frac{v+1}{\eta^2 + v}}\right).
\end{aligned}$$

desenvolvendo as potências, obtemos

$$f(y) = \frac{2}{\sigma} t_v(\eta) T_{v+1}\left(\lambda\eta\sqrt{\frac{v+1}{\eta^2 + v}}\right),$$

como desejado.

1.2.1 Representação Estocástica

Vamos apresentar os teoremas que garantem uma representação estocástica do modelo t-assimétrico dado na definição 1.2.1, com o objetivo de aplicar a teoria do algoritmo EM para a obtenção das estimativas de seus parâmetros.

Teorema 1.2.1 Se $Y \sim ST(\xi, \sigma^2, \lambda, v)$ então existem variáveis aleatórias γ e ζ tais que

$$\begin{aligned}
Y|\gamma, \zeta &\sim N\left(\xi + \delta_\lambda \gamma, \frac{1 - \delta_\lambda^2}{\zeta} \sigma^2\right), \\
\gamma|\zeta &\sim TN\left(0, \frac{\sigma^2}{\zeta}; (0, \infty)\right), \\
\zeta &\sim \Gamma(v/2, v/2),
\end{aligned} \quad (1.13)$$

onde $\xi, \lambda \in \Re$ e $\sigma^2, v > 0$.

Para a demonstração deste teorema precisaremos enunciar um teorema demonstrado em Henze (1986).

Teorema 1.2.2 (Teorema de Henze) Se $Z \sim SN(\lambda)$ então uma representação estocástica para z é dada por

$$Z = \delta_\lambda |U_1| + \sqrt{1 - \delta_\lambda^2} U_2, \quad (1.14)$$

onde U_1, U_2 são variáveis aleatórias independentes e têm distribuição normal padrão e $\delta_\lambda = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

Demonstração do Teorema 1.2.1: Pela definição do modelo t-assimétrico (1.2.1) temos a densidade de ζ , isto significa que a demonstração do teorema se reduz em mostrar a distribuição da $\gamma|\zeta$ e da $Y|\gamma, \zeta$.

Para usar as hipóteses do teorema de Henze, substitua (1.14) em (1.6) logo

$$\begin{aligned} Y &= \xi + \frac{\sigma}{\sqrt{\zeta}} \left(\delta_\lambda |U_1| + \sqrt{1 - \delta_\lambda^2} U_2 \right) \\ &= \xi + \delta_\lambda \frac{\sigma}{\sqrt{\zeta}} |U_1| + \sigma \frac{\sqrt{1 - \delta_\lambda^2}}{\sqrt{\zeta}} U_2. \end{aligned} \quad (1.15)$$

Seja

$$\gamma = \frac{\sigma}{\sqrt{\zeta}} |U_1|, \quad (1.16)$$

onde $|U_1| \sim TN(0, 1)$ pois do teorema de Henze vem que $U_1 \sim N(0, 1)$. Então dado que ζ seja conhecido, temos

$$\gamma|\zeta =_\zeta \sim TN \left(0, \frac{\sigma^2}{\zeta}, (0, \infty) \right),$$

o que mostra a segunda parte do teorema. Para mostrar a primeira parte, substitua (1.16) em (1.15)

$$Y = \xi + \delta_\lambda \gamma + \sigma \frac{\sqrt{1 - \delta_\lambda^2}}{\sqrt{\zeta}} U_2. \quad (1.17)$$

Como $U_2 \sim N(0, 1)$ supondo conhecidos γ e ζ , temos

$$Y|\gamma, \zeta =_\zeta \sim N \left(\xi + \delta_\lambda \gamma, \frac{\sigma^2(1 - \delta_\lambda^2)}{\zeta} \right),$$

finalizando a demonstração do teorema.

1.2.2 A Densidade Conjunta de Y , γ e ζ

Pelo teorema 1.2.1 a variável aleatória Y possui uma representação estocástica baseada nas densidades condicionais $(Y|\gamma, \zeta)$, $(\gamma|\zeta)$ e ζ . Portanto, a função de densidade conjunta de Y , γ e ζ , segundo essa representação, é dada por

$$f(\gamma, \zeta, y) = f(y|\gamma, \zeta)f(\gamma|\zeta)f(\zeta). \quad (1.18)$$

Assim,

$$\begin{aligned} f(\gamma, \zeta, y) &= \frac{1}{\sqrt{2\pi\frac{(1-\delta_\lambda^2)}{\zeta}\sigma^2}} \exp\left\{-\frac{1}{2\left(\frac{1-\delta_\lambda^2}{\zeta}\right)\sigma^2}(y - (\xi + \delta_\lambda\gamma))^2\right\} \frac{2}{\sqrt{2\pi\left(\frac{\sigma^2}{\zeta}\right)}} \exp\left\{-\frac{\gamma^2}{2\left(\frac{\sigma^2}{\zeta}\right)}\right\} \\ &\quad \times \frac{\left(\frac{v}{2}\right)^{v/2}}{\Gamma\left(\frac{v}{2}\right)} \zeta^{\left(\frac{v}{2}\right)-1} \exp\left\{-\frac{v}{2}\zeta\right\} \\ &= \frac{\zeta}{\pi\sqrt{(1-\delta_\lambda^2)\sigma^2}} \zeta^{\left(\frac{v}{2}\right)-1} \frac{\left(\frac{v}{2}\right)^{v/2}}{\Gamma\left(\frac{v}{2}\right)} \exp\left(-\frac{\zeta}{2}v\right) \\ &\quad \times \exp\left\{-\frac{\zeta}{2(1-\delta_\lambda^2)\sigma^2}(y^2 - 2y(\xi + \delta_\lambda\gamma) + (\xi + \delta_\lambda\gamma)^2)\right\} \exp\left\{-\frac{\gamma^2\zeta}{2\sigma^2}\right\} \\ &= \frac{1}{\pi\sqrt{(1-\delta_\lambda^2)\sigma^2}} \frac{\left(\frac{v}{2}\right)^{v/2}}{\Gamma\left(\frac{v}{2}\right)} \zeta^{\frac{v}{2}} \exp\left\{-\frac{\zeta}{2}v\right\} \exp\left\{-\frac{\zeta}{2(1-\delta_\lambda^2)}\left(\frac{y-\xi}{\sigma}\right)^2\right\} \\ &\quad \times \exp\left\{\frac{\gamma\zeta}{(1-\delta_\lambda^2)\sigma^2}\delta_\lambda(y-\xi)\right\} \exp\left\{-\frac{\delta_\lambda^2\gamma^2\zeta}{2(1-\delta_\lambda^2)\sigma^2} - \frac{\gamma^2\zeta}{2\sigma^2}\right\} \\ &= \frac{1}{\pi\sqrt{(1-\delta_\lambda^2)\sigma^2}} \frac{\left(\frac{v}{2}\right)^{v/2}}{\Gamma\left(\frac{v}{2}\right)} \zeta^{\frac{v}{2}} \exp\left\{-\frac{\zeta}{2}v\right\} \exp\left\{-\frac{\zeta}{2(1-\delta_\lambda^2)}\left(\frac{y-\xi}{\sigma}\right)^2\right\} \\ &\quad \times \exp\left\{\frac{\gamma\zeta}{(1-\delta_\lambda^2)\sigma^2}\delta_\lambda(y-\xi)\right\} \exp\left\{-\frac{\gamma^2\zeta}{2(1-\delta_\lambda^2)\sigma^2}\right\}. \end{aligned}$$

Fazendo $\eta = \frac{y-\xi}{\sigma}$, temos

$$\begin{aligned} f(\gamma, \zeta, y) &= \frac{1}{\pi\sqrt{1-\delta_\lambda^2}\sigma^2} \frac{(v/2)^{v/2}}{\Gamma(v/2)} \zeta^{v/2} \exp\left(-\frac{\zeta}{2(1-\delta_\lambda^2)}\eta^2\right) \exp\left(-\frac{\zeta}{2}v\right) \\ &\quad \exp\left(-\frac{\gamma^2\zeta}{2(1-\delta_\lambda^2)\sigma^2} + \frac{\gamma\zeta}{(1-\delta_\lambda^2)\sigma^2}\delta_\lambda(y-\xi)\right). \end{aligned} \quad (1.19)$$

1.2.3 A Função de Verossimilhança dos Dados Aumentados

Definiremos aqui a função de verossimilhança aumentada associada às observações provenientes da distribuição t-assimétricas dada pela representação (1.13). No que segue $\mathbf{y} = (y_1, \dots, y_n)$, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_n)$, $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_n)$ constituem amostras aleatórias das variáveis definidas no Teorema (1.2.1) e $\boldsymbol{\theta} = (\xi, \sigma^2, \lambda, v)$ é o vetor com todos os parâmetros. Denotaremos esta função por $L_c(\boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\gamma}, \boldsymbol{\zeta})$, e vamos denominá-la verossimilhança baseada nos dados aumentados, e a função de verossimilhança baseada somente nos dados observados por $L(\boldsymbol{\theta}; \mathbf{y})$.

Definição 1.2.2 Seja \mathbf{y} uma amostra aleatória da variável aleatória $Y \sim St(\xi, \sigma^2, \lambda, v)$. Então a função de verossimilhança aumentada é dada por

$$L_c(\theta; \mathbf{y}, \gamma, \zeta) = \prod_{j=1}^n f(\gamma_j, \zeta_j, y_j | \theta). \quad (1.20)$$

Para cada $j = 1, \dots, n$ a variável aleatória dada pela representação (1.13) tem a seguinte densidade

$$\begin{aligned} f(\gamma_j, \zeta_j, y_j | \theta) &= \frac{1}{\pi \sqrt{1 - \delta_\lambda^2} \sigma^2} \frac{(v/2)^{v/2}}{\Gamma(v/2)} \zeta_j^{v/2} \exp\left(-\frac{\zeta_j}{2(1 - \delta_\lambda^2) \sigma^2} \eta_j^2\right) \exp\left(-\frac{\zeta_j}{2} v\right) \\ &\quad \exp\left(-\frac{\gamma_j^2 \zeta_j}{2(1 - \delta_\lambda^2) \sigma^2} + \frac{\gamma_j \zeta_j}{2(1 - \delta_\lambda^2) \sigma^2} \delta_\lambda (y_j - \xi)\right), \end{aligned} \quad (1.21)$$

onde $\eta_j = \frac{y_j - \xi}{\sigma}$. Substituindo (1.21) em (1.20) temos

$$\begin{aligned} L_c(\theta; \mathbf{y}, \gamma, \zeta) &= \prod_{j=1}^n \left\{ \frac{1}{\pi \sqrt{1 - \delta_\lambda^2} \sigma^2} \frac{(v/2)^{v/2}}{\Gamma(v/2)} \zeta_j^{v/2} \exp\left(-\frac{\zeta_j}{2(1 - \delta_\lambda^2) \sigma^2} \eta_j^2\right) \exp\left(-\frac{\zeta_j}{2} v\right) \right. \\ &\quad \left. \exp\left(-\frac{\gamma_j^2 \zeta_j}{2(1 - \delta_\lambda^2) \sigma^2} + \frac{\gamma_j \zeta_j}{2(1 - \delta_\lambda^2) \sigma^2} \delta_\lambda (y_j - \xi)\right) \right\}. \end{aligned} \quad (1.22)$$

Analogamente obtemos a log-verossimilhança baseada nos dados aumentados, dada por

$$l_c(\theta; \mathbf{y}, \gamma, \zeta) = \log \left(\prod_{j=1}^n f_j(\gamma_j, \zeta_j, y_j | \theta) \right).$$

Aplicando a propriedade do logaritmo do produto, temos

$$\begin{aligned} l_c(\theta; \mathbf{y}, \gamma, \zeta) &= -\frac{v}{2} \sum_{j=1}^n \zeta_j - \sum_{j=1}^n \left(\frac{\eta_j^2 \zeta_j}{2(1 - \delta_\lambda^2)} \right) + \sum_{j=1}^n \left(\frac{\delta_\lambda \eta_j \gamma_j \zeta_j}{(1 - \delta_\lambda^2) \sigma} \right) \\ &\quad - \sum_{j=1}^n \left(\frac{\gamma_j^2 \zeta_j}{2(1 - \delta_\lambda^2) \sigma^2} \right) - n \log \sigma^2 - \frac{n}{2} \log(1 - \delta_\lambda^2) \\ &\quad + \frac{nv}{2} \log\left(\frac{v}{2}\right) - n \log \Gamma\left(\frac{v}{2}\right) + \frac{v}{2} \sum_{j=1}^n \log \zeta_j, \end{aligned} \quad (1.23)$$

onde $\eta_j = \frac{y_j - \xi}{\sigma}$.

1.3 Estimação via Algoritmo EM

Nesta seção mostraremos como aplicar a teoria do algoritmo EM introduzido por Dempster *et al.* (1977) para estimar os parâmetros da densidade t-assimétrica. É importante ressaltar que não temos interesse na implementação computacional desse desenvolvimento. O nosso objetivo aqui é apenas mostrar os detalhes da teoria do algoritmo EM a qual iremos aplicar na obtenção dos estimadores de uma mistura finita de densidades t-assimétricas que definiremos no próximo capítulo.

Em resumo, a teoria do algoritmo EM sugere estimação por máxima verossimilhança que se dá através de dois passos: primeiramente, o passo E onde são calculadas esperanças condicionais envolvendo a função de log-verossimilhança aumentada. Desse processo resulta uma função Q dos parâmetros. No passo M são atualizados os estimadores obtidos pela maximização desta função. Repetidas iteradas deste método fornecem uma sequência monótona de estimativas dos parâmetros da densidade t-assimétrica, processo este que é interrompido logo que se cumpra uma condição de parada pré-estabelecida.

O algoritmo é desenvolvido a seguir.

O algoritmo EM é definido de tal maneira que, sob condições gerais, encontra-se uma sequência $\{\theta^{(k)}\}$ que converge para um ponto estacionário de $L(\theta)$. Dado um valor $\theta^{(k)}$ gerado pelo algoritmo, obteremos uma atualização $\theta^{(k+1)}$ por:

- i) **Passo E:** Calcular as Esperanças Condicionais;
- ii) **Passo M:** Escolher $\theta^{(k+1)} \in \operatorname{argmax}_{\theta \in \Omega} Q(\theta|\theta^{(k)})$.

Observações:

A notação $\operatorname{argmax}_{\theta \in \Omega} Q(\theta|\theta^{(k)})$ significa o conjunto dos valores de $\theta \in \Omega$ que maximizam $Q(\theta|\theta^{(k)})$ sobre Ω , onde Ω é o espaço paramétrico.

A aproximação inicial é dada a partir de um “chute” $\theta^{(0)}$. Para este trabalho, com o objetivo de estudarmos a consistência das estimativas geradas pelo algoritmo, escolhemos como valores iniciais os verdadeiros valores dos parâmetros, e a partir destes será gerada uma sequência de aproximações $\theta^{(k)} = \{\theta^{(k)} : k = 0, 1, \dots\}$ de modo que:

$$L(\theta^{(k)}) \leq L(\theta^{(k+1)})$$

- iii) **Condição de Parada** As iterações do algoritmo EM são repetidas continuamente até que o valor da norma

$$\|L(\theta^{(k+1)})/L(\theta^{(k)}) - 1\|$$

seja pequeno. É comum estabelecer um limite, um $p > 0$ muito pequeno, de modo que as iterações são repetidas até que a norma acima seja menor que este valor. Para este trabalho vamos tomar $p = 10^{-6}$, ou seja, o algoritmo será encerrado assim que tivermos

$$\|L(\theta^{(k+1)})/L(\theta^{(k)}) - 1\| \leq 10^{-6}$$

para algum k inteiro positivo.

Apresentaremos a seguir como explorar a extensão ECM Meng & Rubin (1993) do algoritmo EM, no qual ele subdivide o passo M em uma sucessão de passos CM dependentes entre si.

1.3.1 Passo E

O Passo E (esperança) do algoritmo consiste em determinar a Q-função

$$Q(\theta|\hat{\theta}^{(k)}) = E \left[l_c(\theta|y, \gamma, \zeta) | \mathbf{y}, \hat{\theta}^{(k)} \right],$$

onde $\hat{\theta}^{(k)} = (\hat{\xi}^{(k)}, \hat{\sigma}^{2(k)}, \hat{\lambda}^{(k)}, \hat{v}^{(k)})$ representa uma dada aproximação (k -ésima iteração) de θ . Ou seja, para obter a quantidade acima devemos calcular a esperança condicional da função log-verossimilhança (1.23), dado que se tenha uma aproximação inicial de θ . Então a Q-função fica definida por

$$\begin{aligned} Q(\theta|\hat{\theta}^{(k)}) &= -\frac{v}{2} \sum_{j=1}^n E \left[\zeta_j | y_j; \hat{\theta}^{(k)} \right] - \sum_{j=1}^n \left(\frac{\eta_j^2 E \left[\zeta_j | y_j; \hat{\theta}^{(k)} \right]}{2(1 - \delta_\lambda^2)} \right) + \sum_{j=1}^n \left(\frac{\delta_\lambda \eta_j E \left[\gamma_j \zeta_j | y_j; \hat{\theta}^{(k)} \right]}{(1 - \delta_\lambda^2) \sigma} \right) \\ &\quad - \sum_{j=1}^n \left(\frac{E \left[\gamma_j^2 \zeta_j | y_j; \hat{\theta}^{(k)} \right]}{2(1 - \delta_\lambda^2) \sigma^2} \right) - n \log \sigma^2 - \frac{n}{2} \log(1 - \delta_\lambda^2) + \frac{nv}{2} \log \left(\frac{v}{2} \right) - n \log \Gamma \left(\frac{v}{2} \right) \\ &\quad + \frac{v}{2} \sum_{j=1}^n E \left[\log \zeta_j | y_j; \hat{\theta}^{(k)} \right]. \end{aligned}$$

Tomando os termos

$$\begin{aligned} \hat{S}_{1j}^{(k)} &= E[\zeta_j | y_j; \hat{\theta}^{(k)}], \\ \hat{S}_{2j}^{(k)} &= E[\gamma_j \zeta_j | y_j; \hat{\theta}^{(k)}], \\ \hat{S}_{3j}^{(k)} &= E[\gamma_j^2 \zeta_j | y_j; \hat{\theta}^{(k)}], \\ \hat{S}_{4j}^{(k)} &= E[\log \zeta_j | y_j; \hat{\theta}^{(k)}] \end{aligned} \tag{1.24}$$

Então,

$$\begin{aligned} Q(\theta|\hat{\theta}^{(k)}) &= -\frac{v}{2} \sum_{j=1}^n \hat{S}_{1j}^{(k)} - \sum_{j=1}^n \left(\frac{\eta_j^2 \hat{S}_{1j}^{(k)}}{2(1 - \delta_\lambda^2)} \right) + \sum_{j=1}^n \left(\frac{\delta_\lambda \eta_j \hat{S}_{2j}^{(k)}}{(1 - \delta_\lambda^2) \sigma} \right) - \sum_{j=1}^n \left(\frac{\hat{S}_{3j}^{(k)}}{2(1 - \delta_\lambda^2) \sigma^2} \right) \\ &\quad - n \log \sigma^2 - \frac{n}{2} \log(1 - \delta_\lambda^2) + \frac{nv}{2} \log \left(\frac{v}{2} \right) - n \log \Gamma \left(\frac{v}{2} \right) + \frac{v}{2} \sum_{j=1}^n \hat{S}_{4j}^{(k)}. \end{aligned} \tag{1.25}$$

Logo o passo E fica resumido a

Dado $\theta = \hat{\theta}^{(k)}$, um valor inicial para os parâmetros da densidade, calcular

$$\hat{S}_{1j}^{(k)}, \hat{S}_{2j}^{(k)}, \hat{S}_{3j}^{(k)} \text{ e } \hat{S}_{4j}^{(k)}$$

das equações dadas em (1.24) para $j = 1, \dots, n$.

1.3.2 Passo CM

Vamos maximizar a Q-função definida no passo anterior a partir do valor inicial atribuído ao vetor de parâmetros θ , definindo os seguintes passos:

Passo CM 1: Obtenha a atualização da estimativa $\hat{\xi}^{(k)}$ maximizando (1.25) sobre ξ , então

$$\hat{\xi}^{(k+1)} = \frac{\sum_{j=1}^n \hat{S}_{1j}^{(k)} y_j - \hat{\delta}_\lambda^{(k)} \sum_{j=1}^n \hat{S}_{2j}^{(k)}}{\sum_{j=1}^n \hat{S}_{1j}^{(k)}}. \quad (1.26)$$

Passo CM 2: Fixando $\xi = \hat{\xi}^{(k+1)}$, obtenha uma atualização da estimativa $\hat{\sigma}^{2(k)}$ maximizando (1.25) sobre σ^2 , então

$$\hat{\sigma}^{2(k+1)} = \frac{\sum_{j=1}^n \left(\hat{S}_{1j}^{(k)} (y_j - \hat{\xi}^{(k+1)})^2 - 2\hat{\delta}_\lambda^{(k)} \hat{S}_{2j}^{(k)} (y_j - \hat{\xi}^{(k+1)}) + \hat{S}_{3j}^{(k)} \right)}{2n(1 - \hat{\delta}_\lambda^{(k)})}. \quad (1.27)$$

Passo CM 3: Fixando $\xi = \hat{\xi}^{k+1}$ e $\sigma^2 = \hat{\sigma}^{2(k+1)}$, obtenha uma atualização da estimativa $\hat{\lambda}^{(k)}$ solucionado a equação

$$\begin{aligned} n\delta_\lambda(1 - \delta_\lambda^2) - \delta_\lambda \left(\sum_{j=1}^n \frac{\hat{S}_{1j}^{(k)} (y_j - \hat{\xi}^{(k+1)})^2}{\hat{\sigma}^{2(k+1)}} + \sum_{j=1}^n \frac{\hat{S}_{3j}^{(k)}}{\hat{\sigma}^{2(k+1)}} \right) \\ + (1 - \delta_\lambda^2) \sum_{j=1}^n \frac{\hat{S}_{2j}^{(k)} (y_j - \hat{\xi}^{(k+1)})}{\hat{\sigma}^{2(k+1)}} = 0. \end{aligned} \quad (1.28)$$

Passo CM 4: Fixando $\xi = \hat{\xi}^{k+1}$, $\sigma^2 = \hat{\sigma}^{2(k+1)}$ e $\lambda = \hat{\lambda}^{(k+1)}$, obtenha uma atualização da estimativa $\hat{v}^{(k)}$ solucionado a equação

$$\log\left(\frac{v}{2}\right) + 1 - DG\left(\frac{v}{2}\right) + \frac{1}{n} \sum_{j=1}^n \left(\hat{S}_{4j}^{(k)} - \hat{S}_{1j}^{(k)} \right) = 0, \quad (1.29)$$

onde $DG(x) = \Gamma'(x)/\Gamma(x)$.

Note que nos passos CM 3 e CM 4 requerem que seja determinadas as raízes das equações (1.28) e (1.29) para a obtenção dos parâmetros λ e v , respectivamente. De acordo com Liu e Rubin (1994) em algumas situações uni-dimensionais a execução destes passos pode ser bastante lenta. Visando uma melhora no processo das iterações usaremos o algoritmo ECME proposto por Liu & Rubin (1994) que reduz os passos CM 3 e CM 4 em um único passo que consiste em maximizar a função log-verossimilhança restrita dos dados observados pela distribuição t-assimétrica, o chamado passo CML.

$$(\hat{\lambda}^{(k+1)}, \hat{v}^{(k+1)}) = \underset{\lambda, v}{\operatorname{argmax}} \sum_{j=1}^n \log \left\{ t_v(\eta_j^{(k+1)}) T_{v+1} \left(\lambda \eta_j^{(k+1)} \sqrt{\frac{v+1}{\eta_j^{2(k+1)} + v}} \right) \right\}. \quad (1.30)$$

1.4 Reparametrização

A obtenção do passo CML, definido na seção anterior, tem um certo custo computacional, pois a cada interação serão fixados as estimativas atuais de ξ e σ^2 restando a maximização da função de log-verossimilhança da t-assimétrica, usando a densidade (1.7), em função do λ e v simultaneamente. Visando melhorar esse processo vamos apresentar nessa seção uma reparametrização proposta por Lachos *et al.* (2009) e a partir dessa nova reparametrização redefiniremos os passos do algoritmo ECME.

A reparametrização do modelo no teorema 1.2.1 será definida pela seguinte transformação

$$(\sigma^2, \lambda) \mapsto (\alpha^2, \beta). \quad (1.31)$$

onde $\beta = \delta_\lambda \sigma$ e $\alpha^2 = (1 - \delta_\lambda^2) \sigma^2$. Logo, temos

$$Y|\gamma, \zeta \sim N\left(\xi + \beta\gamma, \frac{\alpha^2}{\zeta}\right);$$

$$\gamma|\zeta \sim TN\left(0, \frac{1}{\zeta}; (0, \infty)\right);$$

$$\zeta \sim \Gamma(v/2, v/2).$$

1.4.1 A Densidade Conjunta Reparametrizada

A partir da reparametrização, vamos desenvolver a nova densidade conjunta.

Usando os mesmos argumentos da densidade original, temos

$$\begin{aligned} f(\gamma, \zeta, y) &= \frac{1}{\sqrt{2\pi\frac{\alpha^2}{\zeta}}} \exp\left\{-\frac{1}{2\left(\frac{\alpha^2}{\zeta}\right)}(y - (\xi + \beta\gamma))^2\right\} \frac{1}{\sqrt{2\pi\left(\frac{1}{\zeta}\right)}} \exp\left\{-\frac{\gamma^2}{2\left(\frac{1}{\zeta}\right)}\right\} \\ &\quad \times \frac{\left(\frac{v}{2}\right)^{v/2}}{\Gamma\left(\frac{v}{2}\right)} \zeta^{\left(\frac{v}{2}\right)-1} \exp\left\{-\frac{v}{2}\zeta\right\} \\ &= \frac{\zeta}{2\pi\alpha} \exp\left(-\frac{\zeta}{2\alpha^2}(y - (\xi + \beta\gamma))^2\right) \exp\left\{-\frac{\gamma^2\zeta}{2}\right\} \frac{\left(\frac{v}{2}\right)^{v/2}}{\Gamma\left(\frac{v}{2}\right)} \zeta^{\left(\frac{v}{2}\right)} \exp\left\{-\frac{v}{2}\zeta\right\} \\ &= \frac{1}{2\pi\alpha} \frac{\left(\frac{v}{2}\right)^{v/2}}{\Gamma\left(\frac{v}{2}\right)} \zeta^{\left(\frac{v}{2}\right)} \exp\left\{-\frac{\zeta}{2\alpha^2}(y^2 - 2y(\xi + \beta\gamma) + (\xi + \beta\gamma)^2)\right\} \\ &\quad \times \exp\left\{-\frac{\gamma^2\zeta}{2}\right\} \exp\left\{-\frac{v}{2}\zeta\right\} \\ &= \frac{1}{2\pi\alpha} \frac{\left(\frac{v}{2}\right)^{v/2}}{\Gamma\left(\frac{v}{2}\right)} \zeta^{\left(\frac{v}{2}\right)} \exp\left\{-\frac{\zeta}{2}\left(\frac{y - \xi}{\alpha}\right)^2 + \frac{\beta\gamma\zeta}{\alpha^2}(y - \xi) - \frac{\beta^2\gamma^2\zeta}{2\alpha^2}\right\} \\ &\quad \times \exp\left\{-\frac{\gamma^2\zeta}{2}\right\} \exp\left\{-\frac{v}{2}\zeta\right\} \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2\pi\alpha} \frac{\left(\frac{v}{2}\right)^{v/2}}{\Gamma\left(\frac{v}{2}\right)} \zeta^{\left(\frac{v}{2}\right)} \exp\left\{-\frac{v}{2}\zeta\right\} \exp\left\{-\frac{\zeta}{2}\left(\frac{y-\xi}{\alpha}\right)^2\right\} \\
&\times \exp\left\{\frac{\beta\gamma\zeta}{\alpha}\left(\frac{y-\xi}{\alpha}\right)\right\} \exp\left\{-\frac{\gamma^2\zeta}{2}\left(1+\frac{\beta^2}{\alpha^2}\right)\right\}.
\end{aligned}$$

Fazendo $\rho = \frac{y-\xi}{\alpha}$, obtemos

$$\begin{aligned}
f(\gamma, \zeta, y) &= \frac{1}{2\pi\alpha} \frac{\left(\frac{v}{2}\right)^{v/2}}{\Gamma\left(\frac{v}{2}\right)} \zeta^{\left(\frac{v}{2}\right)} \exp\left\{-\frac{v}{2}\zeta\right\} \exp\left\{-\frac{\zeta}{2}\rho^2\right\} \\
&\times \exp\left\{\frac{\rho\beta\gamma\zeta}{\alpha} - \frac{\gamma^2\zeta}{2}\left(1+\frac{\beta^2}{\alpha^2}\right)\right\}.
\end{aligned}$$

1.4.2 A Função de Verossimilhança Reparametrizada

Pela nova representação a função de verossimilhança aumentada de θ também será reparametrizada, usaremos $\Theta = (\xi, \alpha^2, \beta, v)$ para indicar essa reparametrização. Logo a função de verossimilhança aumentada é dada por

$$L_c(\Theta|\mathbf{y}, \gamma, \zeta) = \prod_{j=1}^n f(\gamma_j, \zeta_j, y_j; \Theta), \quad (1.32)$$

onde para cada $j = 1, \dots, n$ o modelo t-assimétrico tem a seguinte representação

$$\begin{aligned}
Y_j|\gamma_j, \zeta_j &\sim N\left(\xi + \beta\gamma_j, \frac{\alpha^2}{\zeta_j}\right), \\
\gamma_j|\zeta_j &\sim TN\left(0, \frac{1}{\zeta_j}; (0, \infty)\right), \\
\zeta_j &\sim \Gamma(v/2, v/2).
\end{aligned}$$

Com densidade

$$\begin{aligned}
f(\gamma_j, \zeta_j, y_j) &= \frac{1}{2\pi\alpha} \frac{\left(\frac{v}{2}\right)^{v/2}}{\Gamma\left(\frac{v}{2}\right)} \zeta_j^{\left(\frac{v}{2}\right)} \exp\left\{-\frac{v}{2}\zeta_j\right\} \exp\left\{-\frac{\zeta_j}{2}\rho_j^2\right\} \\
&\times \exp\left\{\frac{\beta\gamma_j\zeta_j\rho_j}{\alpha} - \frac{\gamma_j^2\zeta_j}{2}\left(1+\frac{\beta^2}{\alpha^2}\right)\right\}, \quad (1.33)
\end{aligned}$$

onde $\rho_j = \frac{y_j-\xi}{\alpha}$, substituindo (1.33) em (1.32) temos

$$\begin{aligned}
L_c(\Theta|\mathbf{y}, \gamma, \zeta) &= \prod_{j=1}^n \left\{ \frac{1}{2\pi\alpha} \frac{\left(\frac{v}{2}\right)^{v/2}}{\Gamma\left(\frac{v}{2}\right)} \zeta_j^{\left(\frac{v}{2}\right)} \exp\left\{-\frac{v}{2}\zeta_j\right\} \exp\left\{-\frac{\zeta_j}{2}\rho_j^2\right\} \right. \\
&\times \left. \exp\left\{\frac{\beta\gamma_j\zeta_j\rho_j}{\alpha} - \frac{\gamma_j^2\zeta_j}{2}\left(1+\frac{\beta^2}{\alpha^2}\right)\right\} \right\}.
\end{aligned}$$

Assim, a nova função log-verossimilhança aumentada é dada por

$$\begin{aligned}
l_c(\Theta|\mathbf{y}, \gamma, \zeta) &= -\frac{v}{2} \sum_{j=1}^n \zeta_j - \sum_{j=1}^n \left(\frac{\zeta_j}{2} \rho_j^2 \right) + \sum_{j=1}^n \left(\frac{\beta \gamma_j \zeta_j \rho_j}{\alpha} \right) - \sum_{j=1}^n \left[\frac{\gamma_j^2 \zeta_j}{2} \left(1 + \frac{\beta^2}{\alpha^2} \right) \right] - n \log 2\pi \\
&\quad - \frac{n}{2} \log \alpha^2 + \frac{nv}{2} \log \left(\frac{v}{2} \right) - n \log \Gamma \left(\frac{v}{2} \right) + \frac{v}{2} \sum_{j=1}^n \log \zeta_j,
\end{aligned}$$

onde $\rho_j = \frac{y_j - \xi}{\alpha}$.

1.4.3 Reformulação do Algoritmo ECME

Neste seção vamos redefinir os passos do algoritmo EM usando a reparametrização dada na seção anterior.

Passo E: A Q-função segundo a reparametrização fica assim definida

$$\begin{aligned}
Q(\Theta|\hat{\Theta}^{(k)}) &= -\frac{v}{2} \sum_{j=1}^n E \left[\zeta_j | y_j; \hat{\Theta}^{(k)} \right] - \sum_{j=1}^n \left(-\frac{\rho_j^2}{2} E \left[\zeta_j | y_j; \hat{\Theta}^{(k)} \right] \right) + \sum_{j=1}^n \left(\frac{\rho_j \beta}{\alpha} E \left[\gamma_j \zeta_j | y_j; \hat{\Theta}^{(k)} \right] \right) \\
&\quad - \sum_{j=1}^n \left[\frac{1}{2} \left(1 + \frac{\beta^2}{\alpha^2} \right) E \left[\gamma_j^2 \zeta_j | y_j; \hat{\Theta}^{(k)} \right] \right] - n \log 2\pi \\
&\quad - \frac{n}{2} \log \alpha^2 + \frac{nv}{2} \log \left(\frac{v}{2} \right) - \log \Gamma \left(\frac{v}{2} \right) + \frac{v}{2} \sum_{j=1}^n E \left[\log \zeta_j | y_j; \hat{\Theta}^{(k)} \right]. \tag{1.34}
\end{aligned}$$

Onde $\hat{\Theta}^{(k)} = (\hat{\xi}^{(k)}, \hat{\alpha}^{2(k)}, \hat{\beta}^{(k)}, \hat{v}^{(k)})$ representa uma dada aproximação de Θ .

Tomemos os termos

$$\begin{aligned}
\hat{s}_{1j}^{(k)} &= E[\zeta_j | y_j; \hat{\Theta}^{(k)}], \\
\hat{s}_{2j}^{(k)} &= E[\gamma_j \zeta_j | y_j; \hat{\Theta}^{(k)}], \\
\hat{s}_{3j}^{(k)} &= E[\gamma_j^2 \zeta_j | y_j; \hat{\Theta}^{(k)}], \\
\hat{s}_{4j}^{(k)} &= E[\log \zeta_j | y_j; \hat{\Theta}^{(k)}]. \tag{1.35}
\end{aligned}$$

Substituindo em (1.34), temos

$$\begin{aligned}
Q(\Theta|\hat{\Theta}^{(k)}) &= -\frac{v}{2} \sum_{j=1}^n \hat{s}_{1j}^{(k)} - \sum_{j=1}^n \left(\frac{\rho_j^2}{2} \hat{s}_{1j}^{(k)} \right) + \sum_{j=1}^n \left(\frac{\rho_j \beta}{\alpha} \hat{s}_{2j}^{(k)} \right) - \sum_{j=1}^n \left[\frac{1}{2} \left(1 + \frac{\beta^2}{\alpha^2} \right) \hat{s}_{3j}^{(k)} \right] \\
&\quad - n \log 2\pi - \frac{n}{2} \log \alpha^2 + \frac{nv}{2} \log \left(\frac{v}{2} \right) - n \log \Gamma \left(\frac{v}{2} \right) + \frac{v}{2} \sum_{j=1}^n \hat{s}_{4j}^{(k)}.
\end{aligned}$$

Logo o Passo E fica definido como segue:

Dado $\Theta = \hat{\Theta}^{(k)}$, um valor inicial para os parâmetros da densidade, calcular

$$\hat{s}_{1j}^{(k)}, \hat{s}_{2j}^{(k)} \text{ e } \hat{s}_{3j}^{(k)}$$

das equações dadas em (1.35) para $j = 1, \dots, n$.

Passo CM 1: Obtenha uma atualização da estimativa $\hat{\xi}^{(k)}$ maximizando (1.34) sobre ξ , então

$$\hat{\xi}^{(k+1)} = \frac{\sum_{j=1}^n \hat{s}_{1j}^{(k)} y_j - \sum_{j=1}^n \hat{s}_{2j}^{(k)}}{\sum_{j=1}^n \hat{s}_{1j}^{(k)}}. \quad (1.36)$$

Passo CM 2: Fixando $\xi = \hat{\xi}^{k+1}$, obtenha uma atualização da estimativa $\hat{\alpha}^{2(k)}$ maximizando (1.34) sobre α^2 , então

$$\hat{\alpha}^{2(k+1)} = \frac{\sum_{j=1}^n \left((y_j - \hat{\xi}^{(k+1)})^2 \hat{s}_{1j}^{(k)} - 2(y_j - \hat{\xi}^{(k+1)}) \beta^{(k)} \hat{s}_{2j}^{(k)} + \beta^{(k)2} \hat{s}_{3j}^{(k)} \right)}{n}. \quad (1.37)$$

Passo CM 3: Fixando $\xi = \hat{\xi}^{k+1}$ e $\alpha^2 = \hat{\alpha}^{2(k+1)}$, obtenha uma atualização da estimativa $\hat{\beta}^{(k+1)}$ maximizando (1.34) sobre β , então

$$\hat{\beta}^{(k+1)} = \frac{\sum_{j=1}^n \left((y_j - \hat{\xi}^{(k+1)}) \hat{s}_{2j}^{(k)} \right)}{\sum_{j=1}^n \hat{s}_{3j}^{(k)}}. \quad (1.38)$$

Passo CM 4: Fixando $\xi = \hat{\xi}^{k+1}$ e $\alpha^2 = \hat{\alpha}^{2(k+1)}$ e $\beta = \hat{\beta}^{(k+1)}$, obtenha uma atualização da estimativa $\hat{v}^{(k+1)}$ solucionando a equação

$$\log\left(\frac{v}{2}\right) + 1 - DG\left(\frac{v}{2}\right) + \frac{1}{n} \sum_{j=1}^n \left(\hat{s}_{4j}^{(k)} - \hat{s}_{1j}^{(k)} \right) = 0. \quad (1.39)$$

Com a nova representação, apenas o passo **CM 4** não tem solução fechada. Então vamos substituir o passo **CM 4** pelo passo **CML** usando a extensão ECME do algoritmo EM. Esta substituição justifica sua ausência do o termo $\hat{s}_{4j}^{(k)}$ no passo E.

Passo CML: Fixados $\xi = \hat{\xi}^{k+1}$, $\alpha^2 = \hat{\alpha}^{2(k+1)}$ e $\hat{\beta}^{(k+1)}$, calcular

$$\hat{v}^{(k+1)} = \underset{x}{\operatorname{argmax}} \sum_{j=1}^n \log \left\{ t_v(\eta_{rj}^{(k+1)}) T_{v+1} \left(\frac{\beta}{\alpha} \eta_{rj}^{(k+1)} \sqrt{\frac{v+1}{\eta_{rj}^{2(k+1)} + v}} \right) \right\}. \quad (1.40)$$

A inclusão dos termos $\frac{\beta}{\alpha}$ e η_r são justificados pela transformação (1.31).

1.4.4 As Esperanças Condicionais do Passo E

Vamos apresentar o desenvolvimento das esperanças condicionais dadas no Passo E do algoritmo ECME. Para isso usaremos a representação estocástica inicial (teorema 1.2.1). Em seguida faremos a reparametrização dos resultados obtidos.

Para melhor entendimento deste desenvolvimento vamos ignorar os índices, calcular as densidade condicionais $f(\zeta|y)$, $f(\gamma|\zeta, y)$ em seguida usá-la para determinar as esperanças desejadas. Para tanto vamos obter inicialmente a densidade conjunta $f(\gamma, y)$.

A densidade conjunta $f(\zeta, y)$ é obtida integrando (1.19) em relação a γ

$$f(\zeta, y) = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \zeta^{\frac{v-1}{2}} \frac{\zeta^{\frac{v}{2}}}{\Gamma^{\frac{v}{2}}} \exp\left(-\frac{\zeta}{2}(\eta^2 + v)\right) \Phi(\lambda\eta\sqrt{\zeta}). \quad (1.41)$$

Logo a função de densidade condicional ζ dado Y é dada pela razão de (1.41) e (1.19), então

$$\begin{aligned} f(\zeta|y) &= \frac{f(\zeta, y)}{f(y)} \\ &= \frac{\sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \zeta^{\frac{v-1}{2}} \frac{\zeta^{\frac{v}{2}}}{\Gamma^{\frac{v}{2}}} \exp\left(-\frac{\zeta}{2}(\eta^2 + v)\right) \Phi(\lambda\eta\sqrt{\zeta})}{\frac{2}{\sigma} \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right) \sqrt{v\pi}} \left[\frac{v+\eta^2}{v}\right]^{-\left(\frac{v+1}{2}\right)} T_{v+1}\left(\lambda\eta\sqrt{\frac{v+1}{\eta^2+v}}\right)} \\ &= \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \zeta^{\frac{v-1}{2}} \frac{\zeta^{\frac{v}{2}}}{\Gamma^{\frac{v}{2}}} \times \frac{\sigma \Gamma\left(\frac{v}{2}\right) \sqrt{v\pi}}{2 \Gamma\left(\frac{v+1}{2}\right)} \left[\frac{v+\eta^2}{v}\right]^{\left(\frac{v+1}{2}\right)} \\ &\quad \times \exp\left(-\frac{\zeta}{2}(\eta^2 + v)\right) \Phi(\lambda\eta\sqrt{\zeta}) \left\{T_{v+1}\left(\lambda\eta\sqrt{\frac{v+1}{\eta^2+v}}\right)\right\}^{-1} \\ &= 2^{-1/2} \frac{1}{2} \left(\frac{v}{2}\right) v^{\left(\frac{v}{2}\right)} v^{1/2} v^{-\frac{v+1}{2}} \left[\frac{v+\eta^2}{v}\right]^{\left(\frac{v+1}{2}\right)} \left\{\Gamma\left(\frac{v+1}{2}\right) T_{v+1}\left(\lambda\eta\sqrt{\frac{v+1}{\eta^2+v}}\right)\right\}^{-1} \\ &\quad \times \zeta^{\frac{v-1}{2}} \exp\left(-\frac{\zeta}{2}(\eta^2 + v)\right) \Phi(\lambda\eta\sqrt{\zeta}) \\ &= \left[\frac{v+\eta^2}{2}\right]^{\left(\frac{v+1}{2}\right)} \left\{\Gamma\left(\frac{v+1}{2}\right) T_{v+1}\left(\lambda\eta\sqrt{\frac{v+1}{\eta^2+v}}\right)\right\}^{-1} \zeta^{\frac{v-1}{2}} \exp\left(-\frac{\zeta}{2}(\eta^2 + v)\right) \Phi(\lambda\eta\sqrt{\zeta}) \\ &= b \zeta^{\frac{v-1}{2}} \exp\left(-\frac{\zeta}{2}(\eta^2 + v)\right) \Phi(\lambda\eta\sqrt{\zeta}). \end{aligned}$$

$$\therefore f(\zeta|Y) = b \zeta^{\frac{v-1}{2}} \exp\left(-\frac{\zeta}{2}(\eta^2 + v)\right) \Phi(\lambda\eta\sqrt{\zeta}), \quad (1.42)$$

onde

$$b = \left[\frac{v+\eta^2}{2}\right]^{\left(\frac{v+1}{2}\right)} \left\{\Gamma\left(\frac{v+1}{2}\right) T_{v+1}\left(\lambda\eta\sqrt{\frac{v+1}{\eta^2+v}}\right)\right\}^{-1}. \quad (1.43)$$

Para determinar a função de densidade condicional γ dado ζ basta obter a razão de (1.19) e (1.42), então

$$\begin{aligned}
f(\gamma|\zeta, y) &= \frac{f(\gamma, \zeta, y)}{f(\zeta, y)} \\
&= \frac{\frac{1}{\pi\sqrt{(1-\delta_\lambda^2)\sigma^2}} \frac{(\frac{v}{2})^{v/2}}{\Gamma(\frac{v}{2})} \zeta^{\frac{v}{2}} \exp\left\{-\frac{\zeta}{2(1-\delta_\lambda^2)}\eta^2\right\} \exp\left\{-\frac{\zeta}{2}v - \frac{\gamma^2\zeta}{2(1-\delta_\lambda^2)\sigma^2} + \frac{\gamma\zeta}{(1-\delta_\lambda^2)\sigma^2}\delta_\lambda(y-\xi)\right\}}{\sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \zeta^{\frac{v-1}{2}} \frac{v^{\frac{v}{2}}}{\Gamma(\frac{v}{2})} \exp\left\{-\frac{\zeta}{2}(\eta^2+v)\right\} \Phi(\lambda\eta\sqrt{\zeta})} \\
&= \frac{1}{\pi\sqrt{(1-\delta_\lambda^2)\sigma^2}} \zeta^{\frac{v}{2}} \left(\sqrt{\frac{\pi}{2}} \sigma \zeta^{-\frac{v-1}{2}}\right) \\
&\times \exp\left\{-\frac{\zeta}{2(1-\delta_\lambda^2)}\eta^2 - \frac{\zeta}{2}v - \frac{\gamma^2\zeta}{2(1-\delta_\lambda^2)\sigma^2} + \frac{\gamma\zeta}{(1-\delta_\lambda^2)\sigma^2}\delta_\lambda(y-\xi) + \frac{\zeta}{2}(\eta^2+v)\right\} \\
&\times (\Phi(\lambda\eta\sqrt{\zeta}))^{-1} \\
&= \frac{1}{\sqrt{2\pi}} \frac{\sqrt{\zeta}}{\sigma\sqrt{1-\delta_\lambda^2}} \exp\left\{\frac{-\zeta(\gamma - (y-\xi)\delta_\lambda)^2}{2(1-\delta_\lambda^2)\sigma^2}\right\} \Phi(\lambda\eta\sqrt{\zeta})^{-1},
\end{aligned}$$

logo

$$\gamma|\zeta, Y \sim TN\left(\delta_\lambda(y-\xi), \frac{(1-\delta_\lambda^2)\sigma^2}{\zeta}; (0, \infty)\right). \quad (1.44)$$

Agora vamos determinar as esperanças condicionais necessárias no Passo E.

i) Vamos mostrar que a esperança condicional de ζ dado $Y = y$ é dada por

$$E[\zeta|y] = \left(\frac{v+1}{\eta^2+v}\right) \frac{T_{v+3}\left(M\sqrt{\frac{v+3}{v+1}}\right)}{T_{v+1}(M)}, \quad (1.45)$$

onde

$$M = \lambda\eta\sqrt{\frac{v+1}{\eta^2+v}}. \quad (1.46)$$

De fato, por definição

$$E[\zeta|Y] = \int_0^\infty \zeta f(\zeta|y) d\zeta. \quad (1.47)$$

Substituindo (1.42) em (1.47), temos

$$E[\zeta|Y] = \int_0^\infty b\zeta^{\frac{v+1}{2}} \exp\left(-\frac{\zeta}{2}(\eta^2+v)\right) \Phi(\lambda\eta\sqrt{\zeta}) d\zeta.$$

Introduziremos os termos $\frac{\Gamma(\frac{v+3}{2})}{\left(\frac{\eta^2+v}{2}\right)^{\frac{v+3}{2}}}$ com objetivo de obter a densidade de

uma gamma com parâmetros $\left(\frac{v+3}{2}, \frac{\eta^2+v}{2}\right)$ em função de ζ . Então

$$E[\zeta|Y] = b \frac{\Gamma\left(\frac{v+3}{2}\right)}{\left(\frac{\eta^2+v}{2}\right)^{\frac{v+3}{2}}} \int_0^\infty \frac{\left(\frac{\eta^2+v}{2}\right)^{\frac{v+3}{2}}}{\Gamma\left(\frac{v+3}{2}\right)} \zeta^{\frac{v+3}{2}-1} \exp\left(-\frac{\zeta}{2}(\eta^2+v)\right) \Phi(\lambda\eta\sqrt{\zeta}) d\zeta. \quad (1.48)$$

a integral acima é

$$E \left[\Phi(\lambda\eta\sqrt{\zeta}) \right].$$

Aplicando (1.8) e substituindo (1.43) em (1.48) temos

$$\begin{aligned} E[\zeta|Y] &= \frac{\left(\frac{\eta^2+v}{2}\right)^{\frac{v+1}{2}}}{\Gamma\left(\frac{v+1}{2}\right)T_{v+1}\left(\lambda\eta\sqrt{\frac{v+1}{\eta^2+v}}\right)} \frac{\Gamma\left(\frac{v+3}{2}\right)}{\left(\frac{\eta^2+v}{2}\right)^{\frac{v+3}{2}}} T_{2\left(\frac{v+3}{2}\right)} \left(\lambda\eta\sqrt{\frac{v+3}{\frac{\eta^2+v}{2}}} \right) \\ &= \frac{\Gamma\left(\frac{v+1}{2}+1\right)}{\Gamma\left(\frac{v+1}{2}\right)} \left(\frac{\eta^2+v}{2}\right)^{\frac{v+1}{2}-\frac{v+3}{2}} \frac{T_{v+3}\left(\lambda\eta\sqrt{\frac{v+3}{\eta^2+v}}\right)}{T_{v+1}\left(\lambda\eta\sqrt{\frac{v+1}{\eta^2+v}}\right)} \\ &= \left(\frac{v+1}{2}\right) \left(\frac{\eta^2+v}{2}\right)^{-1} \frac{T_{v+3}\left(\lambda\eta\sqrt{\frac{v+3}{\eta^2+v}}\right)}{T_{v+1}\left(\lambda\eta\sqrt{\frac{v+1}{\eta^2+v}}\right)} \\ &= \left(\frac{v+1}{\eta^2+v}\right) \frac{T_{v+3}\left(\lambda\eta\sqrt{\frac{v+3}{\eta^2+v}}\right)}{T_{v+1}\left(\lambda\eta\sqrt{\frac{v+1}{\eta^2+v}}\right)}. \end{aligned} \tag{1.49}$$

Para a obtenção de (1.45) precisamos substituir M dado em (1.46) que é imediata no denominador, em T_{v+1} , mas no numerador, em T_{v+3} , temos

$$\lambda\eta\sqrt{\frac{v+3}{\eta^2+v}} = \lambda\eta\sqrt{\frac{v+1}{\eta^2+v}}\sqrt{\frac{v+3}{v+1}} = \lambda\eta\sqrt{\frac{v+1}{\eta^2+v}}\sqrt{\frac{v+3}{v+1}} = M\sqrt{\frac{v+3}{v+1}}.$$

Fazendo as substituições obtemos o resultado desejado.

Para desenvolver as próximas esperanças condicionais do passo E precisaremos do seguinte lema.

Lema 1.4.1 Seja $X \sim NT(\mu, \sigma^2)I\{a_1 < x < a_2\}$ uma distribuição normal truncada. Então

$$\begin{aligned} 1. \quad E(X) &= \mu - \sigma \frac{\phi(\alpha_2) - \phi(\alpha_1)}{\Phi(\alpha_2) - \Phi(\alpha_1)} \\ 2. \quad E(X^2) &= \mu^2 + \sigma^2 - \sigma^2 \frac{\alpha_2\phi(\alpha_2) - \phi(\alpha_1)}{\Phi(\alpha_2) - \Phi(\alpha_1)} - 2\mu\sigma \frac{\phi(\alpha_2) - \phi(\alpha_1)}{\Phi(\alpha_2) - \Phi(\alpha_1)} \end{aligned}$$

Onde $\alpha_i = \frac{(a_i - \mu)}{\sigma}$ com $i = 1, 2$.

Demonstração: Inicialmente vamos obter a função geradora de momentos

$$\begin{aligned} M(t) = E[e^{tX}] &= \frac{\int_{a_1}^{a_2} e^{tx} f(x) dx}{\Phi\left(\frac{a_2-\mu}{\sigma}\right) - \Phi\left(\frac{a_1-\mu}{\sigma}\right)} \\ &= \frac{1}{\Phi\left(\frac{a_2-\mu}{\sigma}\right) - \Phi\left(\frac{a_1-\mu}{\sigma}\right)} \times \int_{a_1}^{a_2} e^{tx} f(x) dx. \end{aligned} \quad (1.50)$$

Vamos calcular separadamente a integral acima

$$\begin{aligned} \frac{1}{\sigma\sqrt{(2\pi)}} \int_{a_1}^{a_2} e^{tx} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx &= \frac{1}{\sigma\sqrt{(2\pi)}} \int_{a_1}^{a_2} e^{-\frac{1}{2\sigma^2}[x-(\sigma^2 t+\mu)]^2 - (\sigma^2 t+\mu)^2 + \mu^2} dx \\ &= e^{-\frac{1}{2\sigma^2}[\mu^2 - (\sigma^2 t+\mu)^2]} \frac{1}{\sigma\sqrt{2\pi}} \int_{a_1}^{a_2} e^{-\frac{1}{2}\left(\frac{x-\gamma}{\sigma}\right)^2} dx \\ &= e^{\mu t + \sigma^2 t^2} \int_{a_1}^{a_2} \frac{1}{\sigma} \phi\left(\frac{x-\gamma}{\sigma}\right) dx \\ &= e^{\mu t + \sigma^2 t^2/2} \left[\Phi\left(\frac{a_2-\gamma}{\sigma}\right) - \Phi\left(\frac{a_1-\gamma}{\sigma}\right) \right]. \end{aligned}$$

onde $\gamma = \sigma^2 t + \mu$.

Voltando em (1.50), temos

$$M(t) = e^{\mu t + \sigma^2 t^2/2} \frac{\Phi\left(\frac{a_2-\mu}{\sigma} - \sigma t\right) - \Phi\left(\frac{a_1-\mu}{\sigma} - \sigma t\right)}{\Phi\left(\frac{a_2-\mu}{\sigma}\right) - \Phi\left(\frac{a_1-\mu}{\sigma}\right)}. \quad (1.51)$$

Para mostrar a primeira parte do lema 1.4.1 vamos derivar (1.51) parcialmente em função de t , logo

$$\begin{aligned} M'(t) &= \mu + \sigma^2 t e^{\mu t + \sigma^2 t^2/2} \left[\frac{\Phi\left(\frac{a_2-\mu}{\sigma} - \sigma t\right) - \Phi\left(\frac{a_1-\mu}{\sigma} - \sigma t\right)}{\Phi\left(\frac{a_2-\mu}{\sigma}\right) - \Phi\left(\frac{a_1-\mu}{\sigma}\right)} \right] + \\ &+ e^{\mu t + \sigma^2 t^2/2} \left[\sigma \frac{\phi\left(\frac{a_2-\mu}{\sigma} - \sigma t\right) - \phi\left(\frac{a_1-\mu}{\sigma} - \sigma t\right)}{\Phi\left(\frac{a_2-\mu}{\sigma}\right) - \Phi\left(\frac{a_1-\mu}{\sigma}\right)} \right]. \end{aligned} \quad (1.52)$$

Fazendo $t = 0$, obtemos

$$E[X] = \mu - \sigma \frac{\phi(\alpha_2) - \phi(\alpha_1)}{\Phi(\alpha_2) - \Phi(\alpha_1)}.$$

Para obtermos a segunda parte da demonstração, vamos derivar (1.52) em relação a t

$$M''(t) = \left(\mu e^{\mu t + \frac{\sigma^2 t^2}{2}} + \sigma^2 t e^{\mu t + \frac{\sigma^2 t^2}{2}} \right) \left[\frac{\Phi(\alpha_2 - \sigma t) - \Phi(\alpha_1 - \sigma t)}{\Phi(\alpha_2) - \Phi(\alpha_1)} \right] -$$

$$\begin{aligned}
& -e^{\mu t + \frac{\sigma^2 t^2}{2}} \left[\sigma \frac{\phi(\alpha_2 - \sigma t) - \phi(\alpha_1 - \sigma t)}{\Phi(\alpha_2) - \Phi(\alpha_1)} \right] \\
= & \mu(\mu + \sigma^2 t)e^{\mu t + \frac{\sigma^2 t^2}{2}} + \sigma^2 \left[e^{\mu t + \frac{\sigma^2 t^2}{2}} + \right. \\
& \left. + t(\mu + \sigma^2 t)e^{\mu t + \frac{\sigma^2 t^2}{2}} \right] \left[\frac{\Phi(\alpha_2 - \sigma t) - \Phi(\alpha_1 - \sigma t)}{\Phi(\alpha_2) - \Phi(\alpha_1)} \right] + \\
& + \left(\mu e^{\mu t + \frac{\sigma^2 t^2}{2}} + \sigma^2 t e^{\mu t + \frac{\sigma^2 t^2}{2}} \right) (-\sigma) \left[\frac{\phi(\alpha_2 - \sigma t) - \phi(\alpha_1 - \sigma t)}{\Phi(\alpha_2) - \Phi(\alpha_1)} \right] - \\
& - (\mu + \sigma^2 t)e^{\mu t + \frac{\sigma^2 t^2}{2}} \left[\sigma \frac{\phi(\alpha_2 - \sigma t) - \phi(\alpha_1 - \sigma t)}{\Phi(\alpha_2) - \Phi(\alpha_1)} \right] + \\
& + \sigma^2 \frac{\phi'(\alpha_2 - \sigma t) - \phi'(\alpha_1 - \sigma t)}{\Phi(\alpha_2) - \Phi(\alpha_1)}.
\end{aligned}$$

Fazendo $t = 0$, temos

$$\begin{aligned}
E[X^2] &= \mu^2 + \sigma^2 \frac{\Phi(\alpha_2) - \Phi(\alpha_1)}{\Phi(\alpha_2) - \Phi(\alpha_1)} - \mu\sigma \frac{\phi(\alpha_2) - \phi(\alpha_1)}{\Phi(\alpha_2) - \Phi(\alpha_1)} - \\
& - \mu\sigma \frac{\phi(\alpha_2) - \phi(\alpha_1)}{\Phi(\alpha_2) - \Phi(\alpha_1)} \\
&= \sigma^2 + \mu^2 + \sigma^2 \frac{\phi'(\alpha_2) - \phi'(\alpha_1)}{\Phi(\alpha_2) - \Phi(\alpha_1)} - 2\mu\sigma \frac{\phi(\alpha_2) - \phi(\alpha_1)}{\Phi(\alpha_2) - \Phi(\alpha_1)} \\
&= \mu^2 + \sigma^2 - \sigma^2 \frac{\alpha_2 \phi(\alpha_2) - \alpha_1 \phi(\alpha_1)}{\Phi(\alpha_2) - \Phi(\alpha_1)} - 2\mu\sigma \frac{\phi(\alpha_2) - \phi(\alpha_1)}{\Phi(\alpha_2) - \Phi(\alpha_1)},
\end{aligned}$$

onde

$$\begin{aligned}
\phi'(\alpha_i) &= -\left(\frac{\alpha_i - \mu}{\sigma}\right) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\alpha_i - \mu}{\sigma}\right)^2} \\
&= -\alpha_i \phi(\alpha_i).
\end{aligned}$$

Voltando as esperanças condicionais do passo E

ii) Vamos mostrar que a esperança condicional de $\gamma\zeta$ dado $Y = y$ é dada por

$$E[\gamma\zeta|y] = \delta_\lambda(y - \xi)E[\zeta|Y] + \frac{\sqrt{1 - \delta_\lambda^2}}{\pi f(y)} \left(\frac{\eta^2}{v(1 - \delta_\lambda^2)} + 1 \right)^{-\left(\frac{v}{2} + 1\right)} \quad (1.53)$$

com efeito, das propriedades de esperança condicional podemos reescrever

$$E[\gamma\zeta|y] = E[\zeta E[\gamma|y, \zeta]|y]. \quad (1.54)$$

por (1.44) podemos usar a primeira parte do lema (1.4.1), temos

$$E[\gamma|y, \zeta] = \delta_\lambda(y - \xi) + \frac{\phi\left(\frac{\delta_\lambda(y-\xi)}{\sqrt{\frac{(1-\delta_\lambda^2)}{\zeta}}\sigma}\right)}{\Phi\left(\frac{\delta_\lambda(y-\xi)}{\sqrt{\frac{(1-\delta_\lambda^2)}{\zeta}}\sigma}\right)} \sqrt{\frac{(1-\delta_\lambda^2)}{\zeta}}\sigma. \quad (1.55)$$

Da proposição (1.2.1)

$$\delta_\lambda = \frac{\lambda}{\sqrt{1-\lambda^2}} \iff \lambda^2 = \delta_\lambda^2(1-\lambda^2) \iff \lambda = \frac{\delta_\lambda}{\sqrt{1-\delta_\lambda^2}}. \quad (1.56)$$

Substituindo (1.56) em (1.57), então

$$E[\gamma|y, \zeta] = \delta_\lambda(y - \xi) + \frac{\phi(\lambda\eta\sqrt{\zeta})}{\Phi(\lambda\eta\sqrt{\zeta})} \sqrt{\frac{(1-\delta_\lambda^2)}{\zeta}}\sigma. \quad (1.57)$$

Substituindo agora (1.57) em (1.54)

$$\begin{aligned} E[\gamma\zeta|y] &= E\left[\zeta\left(\delta_\lambda(y - \xi) + \frac{\phi(\lambda\eta\sqrt{\zeta})}{\Phi(\lambda\eta\sqrt{\zeta})} \sqrt{\frac{(1-\delta_\lambda^2)}{\zeta}}\sigma\right) | Y\right] \\ &= (\delta_\lambda(y - \xi)) E[\zeta|Y] + \sqrt{1-\delta_\lambda^2}\sigma E\left[\sqrt{\zeta} \frac{\phi(\lambda\eta\sqrt{\zeta})}{\Phi(\lambda\eta\sqrt{\zeta})} | Y\right]. \end{aligned} \quad (1.58)$$

Como $E[\zeta|Y]$ é conhecido, vamos calcular $E\left[\sqrt{\zeta} \frac{\phi(\lambda\eta\sqrt{\zeta})}{\Phi(\lambda\eta\sqrt{\zeta})} | Y\right]$ separadamente

$$E\left[\sqrt{\zeta} \frac{\phi(\lambda\eta\sqrt{\zeta})}{\Phi(\lambda\eta\sqrt{\zeta})} | Y\right] = \int_0^\infty \left(\sqrt{\zeta} \frac{\phi(\lambda\eta\sqrt{\zeta})}{\Phi(\lambda\eta\sqrt{\zeta})}\right) f(\zeta|y) d\zeta, \quad (1.59)$$

onde

$$f(\zeta|y) = \frac{f(\zeta, y)}{f(y)}. \quad (1.60)$$

Substituindo apenas (1.41) em (1.60)

$$f(\zeta|y) = \frac{\sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \zeta^{\frac{v-1}{2}} \frac{\left(\frac{v}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)} \exp\left(-\frac{\zeta}{2}(\eta^2 + v)\right) \Phi(\lambda\eta\sqrt{\zeta})}{f(y)}.$$

Substituindo em (1.59), temos

$$\begin{aligned}
E \left[\sqrt{\zeta} \frac{\phi(\lambda\eta\sqrt{\zeta})}{\Phi(\lambda\eta\sqrt{\zeta})} | Y \right] &= \int_0^\infty \zeta^{1/2} \frac{\phi(\lambda\eta\sqrt{\zeta})}{\Phi(\lambda\eta\sqrt{\zeta})} \frac{\sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \zeta^{\frac{v-1}{2}} \frac{(\frac{v}{2})^{\frac{v}{2}}}{\Gamma(\frac{v}{2})} \exp(-\frac{\zeta}{2}(\eta^2 + v)) \Phi(\lambda\eta\sqrt{\zeta})}{f(y)} d\zeta \\
&= \frac{(\frac{v}{2})^{\frac{v}{2}}}{\sigma \Gamma(\frac{v}{2}) f(y)} \int_0^\infty \zeta^{(\frac{v}{2}+1)-1} \exp\left\{-\frac{\zeta}{2}(\eta^2 + v)\right\} \sqrt{\frac{2}{\pi}} \phi(\lambda\eta\sqrt{\zeta}) d\zeta.
\end{aligned} \tag{1.61}$$

Como

$$\phi(\lambda\eta\sqrt{\zeta}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\lambda^2\eta^2\zeta\right), \tag{1.62}$$

substituindo (1.62) em (1.61), temos

$$\begin{aligned}
E \left[\sqrt{\zeta} \frac{\phi(\lambda\eta\sqrt{\zeta})}{\Phi(\lambda\eta\sqrt{\zeta})} | Y \right] &= \frac{(\frac{v}{2})^{\frac{v}{2}}}{\pi\sigma\Gamma(\frac{v}{2}) f(y)} \int_0^\infty \zeta^{(\frac{v}{2}+1)-1} \exp\left\{-\frac{\zeta}{2}[(\eta^2 + v) + \lambda^2\eta^2]\right\} d\zeta \\
&= \frac{(\frac{v}{2})^{\frac{v}{2}}}{\pi\sigma\Gamma(\frac{v}{2}) f(y)} \int_0^\infty \zeta^{(\frac{v}{2}+1)-1} \exp\left\{-\frac{\zeta}{2}[\eta^2(1 + \lambda^2) + v]\right\} d\zeta.
\end{aligned} \tag{1.63}$$

Por outro lado

$$1 + \lambda^2 = 1 + \frac{\delta_\lambda^2}{1 - \delta_\lambda^2} = \frac{1}{1 - \delta_\lambda^2}. \tag{1.64}$$

Substitua (1.64) em (1.63), então

$$\begin{aligned}
E \left[\sqrt{\zeta} \frac{\phi(\lambda\eta\sqrt{\zeta})}{\Phi(\lambda\eta\sqrt{\zeta})} | Y \right] &= \frac{(\frac{v}{2})^{\frac{v}{2}}}{\pi\sigma\Gamma(\frac{v}{2}) f(y)} \int_0^\infty \zeta^{(\frac{v}{2}+1)-1} \exp\left\{-\frac{\zeta}{2}\left[\frac{\eta^2}{1 - \delta_\lambda^2} + v\right]\right\} d\zeta \\
&= \frac{(\frac{v}{2})^{\frac{v}{2}}}{\pi\sigma f(y) \left[\frac{1}{2}\left(\frac{\eta^2}{1 - \delta_\lambda^2} + v\right)\right]^{v/2}} \\
&\quad \times \int_0^\infty \zeta \frac{\left[\frac{1}{2}\left(\frac{\eta^2}{1 - \delta_\lambda^2} + v\right)\right]^{v/2}}{\Gamma(\frac{v}{2})} \zeta^{\frac{v}{2}-1} \exp\left\{-\frac{1}{2}\left(\frac{\eta^2}{1 - \delta_\lambda^2} + v\right)\zeta\right\} d\zeta.
\end{aligned} \tag{1.65}$$

Na integral acima temos o produto de ζ pela densidade da gamma com parâmetros $\left(\frac{v}{2}, \frac{1}{2}\left[\frac{\eta^2}{1 - \delta_\lambda^2} + v\right]\right)$ em função de ζ , ou seja, temos $E[\zeta]$ que substituindo em (1.65), temos

$$E \left[\sqrt{\zeta} \frac{\phi(\lambda\eta\sqrt{\zeta})}{\Phi(\lambda\eta\sqrt{\zeta})} | Y \right] = \frac{(\frac{v}{2})^{\frac{v}{2}}}{\pi\sigma f(y) \left[\frac{1}{2}\left(\frac{\eta^2}{1 - \delta_\lambda^2} + v\right)\right]^{v/2}} \frac{\frac{v}{2}}{\frac{1}{2}\left(\frac{\eta^2}{1 - \delta_\lambda^2} + v\right)}$$

$$\begin{aligned}
&= \frac{v^{\frac{v}{2}}}{\pi \sigma f(y) \left[\left(\frac{\eta^2}{1-\delta_\lambda^2} + v \right) \right]^{v/2}} \frac{v}{\left(\frac{\eta^2}{1-\delta_\lambda^2} + v \right)} \\
&= \frac{v^{\frac{v}{2}+1}}{\pi \sigma f(y) \left(\frac{\eta^2}{1-\delta_\lambda^2} + v \right)^{\frac{v}{2}+1}} \\
&= \frac{1}{\pi \sigma f(y)} \left(\frac{\eta^2}{v(1-\delta_\lambda^2)} + 1 \right)^{-\left(\frac{v}{2}+1\right)}. \quad (1.66)
\end{aligned}$$

substituindo (1.66) em (1.58) obtemos o resultado desejado.

iii) Vamos mostrar que a esperança condicional de $\gamma^2\zeta$ dado $Y = y$ é dada por

$$E[\gamma^2\zeta|y] = \delta_\lambda^2(y-\xi)^2 E[\zeta|Y] + (1-\delta_\lambda^2)\sigma^2 + \frac{\delta_\lambda(y-\xi)\sqrt{1-\delta_\lambda^2}}{\pi f(y)} \left(\frac{\eta^2}{v(1-\delta_\lambda^2)} + 1 \right)^{-\left(\frac{v}{2}+1\right)} \quad (1.67)$$

De fato, como no item ii) temos

$$E[\gamma^2\zeta|y] = E[\zeta E[\gamma^2|y, \zeta]|y], \quad (1.68)$$

e ainda por (1.44)

$$\gamma|\zeta, y \sim NT \left(\delta_\lambda(y-\xi), \frac{(1-\delta_\lambda^2)}{\zeta}\sigma^2; (0, \infty) \right).$$

Logo pela segunda parte do lema (1.4.1)

$$E[\gamma^2|y, \zeta] = \delta_\lambda^2(y-\xi)^2 + \frac{(1-\delta_\lambda^2)}{\zeta}\sigma^2 + \frac{\phi(\lambda\eta\sqrt{\zeta})}{\Phi(\lambda\eta\sqrt{\zeta})} \delta_\lambda(y-\xi)\sigma\sqrt{\frac{(1-\delta_\lambda^2)}{\zeta}}.$$

Substituindo em (1.68)

$$\begin{aligned}
E[\gamma^2\zeta|y] &= E \left[\zeta \left(\delta_\lambda^2(y-\xi)^2 + \frac{(1-\delta_\lambda^2)}{\zeta}\sigma^2 + \frac{\phi(\lambda\eta\sqrt{\zeta})}{\Phi(\lambda\eta\sqrt{\zeta})} \delta_\lambda(y-\xi)\sigma\sqrt{\frac{(1-\delta_\lambda^2)}{\zeta}} \right) | y \right] \\
&= \delta_\lambda^2(y-\xi)^2 E[\zeta|y] + (1-\delta_\lambda^2)\sigma^2 + \delta_\lambda(y-\xi)\sigma\sqrt{(1-\delta_\lambda^2)} E \left[\left(\sqrt{\zeta} \frac{\phi(\lambda\eta\sqrt{\zeta})}{\Phi(\lambda\eta\sqrt{\zeta})} \right) | y \right]. \quad (1.69)
\end{aligned}$$

Substituindo (1.66) em (1.69) obtemos o resultado desejado.

Para finalizar a processo de reparametrização vamos fazer uma substituição dos parâmetros das esperanças obtidas acima usando a transformação (1.31).

Dados $\beta = \delta_\lambda \sigma$ e $\alpha^2 = (1 - \delta_\lambda^2) \sigma^2$ obtemos $\sigma^2(\alpha, \beta)$ e $\lambda(\alpha, \beta)$ resolvendo o seguinte sistema de equações:

$$\begin{aligned} \beta = \delta_\lambda \sigma & \Rightarrow \sigma^2 = \alpha^2 + \beta^2 \\ \alpha^2 = (1 - \delta_\lambda^2) \sigma^2 & \delta_\lambda = \frac{\beta}{\sqrt{\alpha^2 + \beta^2}} \end{aligned}$$

que substituindo em $\delta_\lambda = \frac{\lambda}{1 + \lambda^2}$, obtemos $\lambda = \frac{\beta}{\alpha}$. Logo

$$\sigma^2(\alpha, \beta) = \alpha^2 + \beta^2 \quad (1.70)$$

e

$$\lambda(\alpha, \beta) = \frac{\beta}{\alpha}. \quad (1.71)$$

i) Substituindo (1.70) e (1.71) em (1.45), temos

$$s_1 = E[\zeta|Y] = \left(\frac{v+1}{\eta_r^2 + v} \right) \frac{T_{v+3} \left(M_r \sqrt{\frac{v+3}{v+1}} \right)}{T_{v+1}(M_r)}. \quad (1.72)$$

ii) Substituindo (1.70) e (1.71) em (1.53), temos

$$s_2 = E[\gamma \zeta | y] = \beta \eta_r E[\zeta|Y] + \frac{\alpha}{\sqrt{\alpha^2 + \beta^2} \pi f_Y(y)} \left(\frac{1}{v} \rho^2 + 1 \right)^{-\left(\frac{v}{2} + 1\right)} \quad (1.73)$$

iii) Substituindo (1.70) e (1.71) em (1.67), temos

$$s_3 = E[\gamma^2 \zeta | y] = \beta^2 \eta_r^2 E[\zeta|Y] + \alpha^2 + \frac{\beta \eta_r \alpha}{\pi \sqrt{\alpha^2 + \beta^2} f_Y(y)} \left(\frac{1}{v} \rho^2 + 1 \right)^{-\left(\frac{v}{2} + 1\right)} \quad (1.74)$$

onde

$$\eta_r = \frac{(y - \xi)}{\sqrt{\alpha^2 + \beta^2}}, \quad (1.75)$$

$$M_r = \frac{\beta}{\alpha} \eta_r \sqrt{\frac{v+1}{\eta_r^2 + v}}, \quad (1.76)$$

$$\rho = \frac{(y - \xi)}{\alpha}. \quad (1.77)$$

Capítulo 2

Mistura Finita de Densidades t-Assimétricas

2.1 Mistura Finita de t-Assimétricas

Uma mistura de densidades t-assimétricas é uma distribuição definida pela densidade

$$\begin{aligned} \psi(y|\Theta) &= \sum_{i=1}^g w_i f_i(y|\theta_i), \quad y \in \mathfrak{R}, \\ w_i &\geq 0, \quad \sum_{i=1}^g w_i = 1, \end{aligned} \tag{2.1}$$

onde $f_i(\cdot|\theta_i)$ é a densidade t-assimétrica com vetor de parâmetros $\theta_i = (\xi_i, \alpha_i^2, \beta_i, \nu_i)$, os w_i 's são os pesos da mistura e $\Theta = (w_1, \dots, w_g, \theta_1, \dots, \theta_g)$ denota o vetor com todos os parâmetros.

No contexto de modelagem por mistura de densidades é introduzido para cada Y_j , $j = 1, \dots, n$ da amostra um conjunto de variáveis indicadoras $\mathbf{Z}_j = (Z_{1j}, \dots, Z_{gj})^T$, composto de variáveis binárias tais que

$$Z_{kj} = \begin{cases} 1 & \text{se } \mathbf{Y}_j \sim f_k, \\ 0 & \text{caso contrário.} \end{cases}$$

e $\sum_{i=1}^g Z_{ij} = 1$.

Dadas as probabilidades da mistura w_1, \dots, w_g os vetores de variáveis indicadoras $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ são independentes com densidade multinomial, dada por

$$f(\mathbf{z}_j) = \omega_1^{z_{1j}} \omega_2^{z_{2j}} \dots (1 - \omega_1 - \dots - \omega_{g-1})^{z_{gj}}. \tag{2.2}$$

Usaremos $\mathbf{Z}_j \sim \mathcal{M}(1; \omega_1, \dots, \omega_g)$ para denotar a distribuição das \mathbf{Z}_j com $j = 1, \dots, n$. A partir dessa inclusão faremos a modelagem dos dados ditos completos $x_j = (y_j, z_j)$.

2.2 Função log-Verossimilhança para os dados completos

Considere que dispomos de uma amostra aleatória $\{y_1, y_2, \dots, y_n\}$ da distribuição (2.1). A densidade conjunta de $x_j = (y_j, z_j)$ é dada por

$$\begin{aligned} f(x_j|\Theta) &= f(y_j|z_j; \Theta)f(z_j; \Theta) \\ &= \prod_{i=1}^g (f_i(y_j; \theta_i))^{(z_{ij})} f(z_j; \Theta). \end{aligned} \quad (2.3)$$

Em (2.3), pela distribuição assumida por z_j , temos

$$f(z_j|\Theta) = f(z_j; \mathbf{w}) = \prod_{i=1}^g w_i^{z_{ij}}. \quad (2.4)$$

Substituindo (2.4) em (2.3) obtemos

$$f(x_j|\Theta) = \prod_{i=1}^g w_i^{z_{ij}} (f_i(y_j; \theta_i))^{(z_{ij})}. \quad (2.5)$$

Considerando agora a independência dos dados incompletos e (2.5), temos que a função de verossimilhança é dada por

$$\begin{aligned} L_c(\Theta) &= \prod_{j=1}^n f(x_j; \Theta) \\ &= \prod_{j=1}^n \prod_{i=1}^g w_i^{z_{ij}} (f_i(y_j; \theta_i))^{z_{ij}}. \end{aligned} \quad (2.6)$$

De (2.5) vemos que a função de log-verossimilhança para dados completos é dada por

$$l_c(\Theta) = \sum_{j=1}^n \sum_{i=1}^g Z_{ij} \log w_i + \sum_{j=1}^n \sum_{i=1}^g Z_{ij} \log f_i(y_j; \theta_i). \quad (2.7)$$

Vimos no capítulo anterior uma representação estocástica do modelo t-assimétrico. Com a inclusão das variáveis indicadoras Z'_{ij} s temos uma representação estocástica

da mistura de densidades t-assimétricas dada por

$$\begin{aligned} Y_j | \gamma_j, \zeta_j, z_{ij} = 1 &\sim N\left(\xi_i + \beta_i \gamma_j, \frac{\alpha_i^2}{\zeta_j}\right), \\ \gamma_j | \zeta_j, z_{ij} = 1 &\sim TN\left(0, \frac{1}{\zeta_j}; (0, \infty)\right), \\ \zeta_j, z_{ij} = 1 &\sim \Gamma(v_i/2, v_i/2), \\ Z_j &\sim M(1; w_1, \dots, w_g). \end{aligned}$$

Note que usamos a representação reparametrizada.

A densidade conjunta baseada nas observações y_j é dada por

$$\begin{aligned} f_i(y_j; \theta_i) &= \frac{1}{2\pi\alpha_i} \frac{\left(\frac{v_i}{2}\right)^{v_i/2}}{\Gamma\left(\frac{v_i}{2}\right)} \zeta_j^{(v_i/2)} \exp\left\{-\frac{v_i}{2}\zeta_j\right\} \exp\left\{-\frac{\zeta_j}{2}\rho_{ij}^2\right\} \\ &\times \exp\left\{\frac{\rho_{ij}\beta_i\gamma_j\zeta_j}{\alpha_i} - \frac{\gamma_j^2\zeta_j}{2}\left(1 + \frac{\beta_i^2}{\alpha_i^2}\right)\right\} \end{aligned}$$

para cada $i = 1, \dots, n$ e $j = 1, \dots, g$.

Logo a log-verossimilhança para os dados completos em (2.7) fica assim definida

$$\begin{aligned} l_c(\Theta) &= \sum_{j=1}^n \sum_{i=1}^g Z_{ij} \log w_i \\ &- \sum_{j=1}^n \sum_{i=1}^g Z_{ij} \left\{ -\frac{v_i}{2}\zeta_j - \left(\frac{\zeta_j}{2}\rho_{ij}^2\right) + \left(\frac{\rho_{ij}\beta_i\gamma_j\zeta_j}{\alpha_i}\right) - \left[\frac{\gamma_j^2\zeta_j}{2}\left(1 + \frac{\beta_i^2}{\alpha_i^2}\right)\right] - n \log 2\pi \right. \\ &\quad \left. - \frac{n}{2} \log \alpha_i^2 + \frac{nv_i}{2} \log\left(\frac{v_i}{2}\right) - n \log \Gamma\left(\frac{v_i}{2}\right) + \frac{v_i}{2} \log \zeta_j \right\}. \end{aligned}$$

onde $\rho_{ij} = \frac{y_j - \xi_i}{\beta_i}$.

2.3 Estimação via algoritmo ECME

Nesta seção trataremos da elaboração de um algoritmo EM para misturas. Como no capítulo anterior definiremos os passos da extensão ECME. Inicialmente vamos determinar a Q-função e calcular as esperanças condicionais. Temos que

$$Q(\Theta | \Theta^{(k)}) = E_z[l_c(\Theta) | Y_{(n)}, \Theta^{(k)}]$$

$$= \sum_{j=1}^n \sum_{i=1}^g E \left[Z_{ij} \mid Y_{(n)}, \Theta^{(k)} \right] \log w_i + \sum_{j=1}^n \sum_{i=1}^g E \left[Z_{ij} \log f_i(y_j; \theta_i) \mid Y_{(n)}, \theta^{(k)} \right], \quad (2.8)$$

onde $\Theta^{(k)} = (w_1^{(k)}, w_2^{(k)}, \dots, w_g^{(k)}, \theta_1^{(k)}, \theta_2^{(k)}, \dots, \theta_g^{(k)})$, representa uma dada aproximação de Θ .

Aplicando as esperanças, temos

$$\begin{aligned} Q(\Theta | \hat{\Theta}^{(k)}) &= \sum_{j=1}^n \sum_{i=1}^g \left(E \left[Z_{ij} | y_j; \hat{\Theta}^{(k)} \right] \log w_i \right) \\ &+ \sum_{i=1}^g \sum_{j=1}^n \left(-\frac{v_i}{2} E \left[Z_{ij} \zeta_j | y_j; \hat{\Theta}^{(k)} \right] \right) - \sum_{i=1}^g \sum_{j=1}^n \left(\frac{\rho_{ij}^2}{2} E \left[Z_{ij} \zeta_j | y_j; \hat{\Theta}^{(k)} \right] \right) \\ &+ \sum_{i=1}^g \sum_{j=1}^n \left(\frac{\beta_i \rho_{ij}}{\alpha_i} E \left[Z_{ij} \gamma_j \zeta_j | y_j; \hat{\Theta}^{(k)} \right] \right) \\ &- \sum_{i=1}^g \sum_{j=1}^n \left(2^{-1} \left(1 + \frac{\beta_i^2}{\alpha_i^2} \right) E \left[Z_{ij} \gamma_j^2 \zeta_j | y_j; \hat{\Theta}^{(k)} \right] \right) \\ &- \sum_{i=1}^g Z_{ij} \left(\frac{n}{2} \log \alpha_i^2 + n \log 2\pi - \frac{nv_i}{2} \log \left(\frac{v_i}{2} \right) + n \log \Gamma \left(\frac{v_i}{2} \right) \right) \\ &+ \sum_{i=1}^g \sum_{j=1}^n \left(\frac{v_i}{2} E \left[Z_{ij} \log \zeta_j | y_j; \hat{\Theta}^{(k)} \right] \right). \end{aligned}$$

Tomemos os termos $\hat{z}_{ij}^{(k)}$, $\hat{s}_{1ij}^{(k)}$, $\hat{s}_{2ij}^{(k)}$, $\hat{s}_{3ij}^{(k)}$ e $\hat{s}_{4ij}^{(k)}$ que correspondem as esperanças condicionais

$$\begin{aligned} &E[Z_{ij} | y_j; \hat{\Theta}^{(k)}], \quad E[Z_{ij} \zeta_j | y_j; \hat{\Theta}^{(k)}], \quad E[Z_{ij} \gamma_j \zeta_j | y_j; \hat{\Theta}^{(k)}], \\ &E[Z_{ij} \gamma_j^2 \zeta_j | y_j; \hat{\Theta}^{(k)}], \quad E[Z_{ij} \log \zeta_j | y_j; \hat{\Theta}^{(k)}], \end{aligned} \quad (2.9)$$

respectivamente, tal que (2.8) pode ser escrita por

$$\begin{aligned} Q(\Theta | \hat{\Theta}^{(k)}) &= \sum_{j=1}^n \sum_{i=1}^g \left(\hat{z}_{ij}^{(k)} \log w_i \right) \\ &+ \sum_{i=1}^g \sum_{j=1}^n \left(-\frac{v_i}{2} \hat{s}_{1ij}^{(k)} \right) - \sum_{i=1}^g \sum_{j=1}^n \left(\frac{\rho_{ij}^2}{2} \hat{s}_{1ij}^{(k)} \right) + \sum_{i=1}^g \sum_{j=1}^n \left(\frac{\beta_i \rho_{ij}}{\alpha_i} \hat{s}_{2ij}^{(k)} \right) \\ &- \sum_{i=1}^g \sum_{j=1}^n \left(2^{-1} \left(1 + \frac{\beta_i^2}{\alpha_i^2} \right) \hat{s}_{3ij}^{(k)} \right) \end{aligned}$$

$$\begin{aligned}
& - \sum_{i=1}^g Z_{ij} \left(\frac{n}{2} \log \alpha_i^2 + n \log 2\pi - \frac{v_i}{2} \log \left(\frac{v_i}{2} \right) + n \log \Gamma \left(\frac{v_i}{2} \right) \right) \\
& + \sum_{i=1}^g \sum_{j=1}^n \left(\frac{v_i}{2} \hat{s}_{4ij}^{(k)} \right). \tag{2.10}
\end{aligned}$$

Das esperanças condicionais dadas em (2.9) vamos calcular apenas as que usaremos no Passo E.

$$\begin{aligned}
\mathbf{i)} \quad \hat{z}_{ij}^{(k)} &= E_r[Z_{ij}|Y_j; \hat{\Theta}^{(k)}] = E_r[z_{ij}|Y_{(n)}, \Theta^{(k)}] \\
&= Pr[z_{ij} = 1|Y_{(n)}, \Theta^{(k)}] = Pr[z_{ij} = 1|y_j, \Theta^{(k)}], \tag{2.11}
\end{aligned}$$

onde a terceira igualdade decorre da definição de z_{ij} e, a última, devido a independência dos z'_j s. Considerando a distribuição dada em (2.4), temos que

$$Pr[z_{ij} = 1|\Theta^{(k)}] = Pr[z_{ij} = 1|z_{il} = 0 \quad \forall l \neq j|\theta^{(k)}] = w_i^{(k)}. \tag{2.12}$$

Usando (2.3), vemos que

$$f(y_j|z_j; \Theta^{(k)}) = \prod_{i=1}^g \left(f_i(y_j; \theta_i^{(k)}) \right)^{z_{ij}}, \tag{2.13}$$

e, portanto, temos que

$$f(y_j|z_{ij} = 1; \Theta^{(k)}) = f_i(y_j; \theta_i^{(k)}). \tag{2.14}$$

Empregando o teorema de Bayes, vemos que

$$\hat{z}_{ij}^{(k)} = Pr[z_{ij} = 1|y_j, \Theta^{(k)}] = \frac{Pr[z_{ij} = 1|\Theta^{(k)}]f(Y_j|z_{ij} = 1; \Theta^{(k)})}{p(y_j; \Theta^{(k)})}. \tag{2.15}$$

Usando (2.1), (2.14) e (2.12) em (2.15), obtemos

$$\hat{z}_{ij}^{(k)} = \frac{w_i^{(k)} f_i(y_j; \theta_i^{(k)})}{\sum_{t=1}^{(s)} w_t^{(k)} f_t(y_j; \theta_t^{(k)})} \tag{2.16}$$

de modo que

$$\hat{z}_{ij}^{(k)} = \frac{w_i^{(k)} f(y_j|\xi_i^{(k)}, \alpha_i^{2(k)}, \beta_i^{(k)}, v_i^{(k)})}{\psi(y_j|\Theta^{(k)})}. \tag{2.17}$$

Da expressão acima, vemos que $\hat{z}_{ij}^{(k)}$ representa uma estimativa de probabilidade de y_j pertencer a uma população cuja a distribuição é dada por $f_i(\cdot; \theta_i^{(k)})$, com base em uma dada estimativa $\Theta^{(k)}$ do vetor de parâmetros Θ .

$$\text{ii) } \hat{s}_{1ij}^{(k)} = E \left[Z_{ij} \zeta_j | y_j, \hat{\Theta}^{(k)} \right] = E \left[\zeta_j E \left[Z_{ij} | y_j, \hat{\Theta}^{(k)} \right] | y_j, \hat{\Theta}^{(k)} \right]. \quad (2.18)$$

Mas $E \left[Z_{ij} | y_j, \hat{\Theta}^{(k)} \right]$ fica conhecido quando é dado valor a y_j , então (2.18) fica

$$\hat{s}_{1ij}^{(k)} = E \left[Z_{ij} | y_j, \hat{\Theta}^{(k)} \right] E \left[\zeta_j | y_j, \hat{\Theta}^{(k)} \right]. \quad (2.19)$$

Substituindo (2.17) e aplicando (1.72) na componente i temos

$$\hat{s}_{1ij}^{(k)} = \hat{z}_{ij}^{(k)} \left(\frac{\hat{v}_i^{(k)} + 1}{\hat{\eta}_{rij}^{2(k)} + \hat{v}_i^{(k)}} \right) \frac{T_{\hat{v}_i^{(k)}+3} \left(\hat{M}_{rij}^{(k)} \sqrt{\frac{\hat{v}_i^{(k)}+3}{\hat{v}_i^{(k)}+1}} \right)}{T_{\hat{v}_i^{(k)}+1} \left(\hat{M}_{rij}^{(k)} \right)}; \quad (2.20)$$

$$\text{iii) } \hat{s}_{2ij}^{(k)} = E \left[Z_{ij} \gamma_j \zeta_j | y_j, \hat{\Theta}^{(k)} \right] = E \left[\gamma_j \zeta_j E \left[Z_{ij} | y_j, \hat{\Theta}^{(k)} \right] | y_j, \hat{\Theta}^{(k)} \right]. \quad (2.21)$$

Como $E \left[Z_{ij} | y_j, \hat{\Theta}^{(k)} \right]$ é conhecido, temos (2.21) fica

$$\hat{s}_{1ij}^{(k)} = E \left[Z_{ij} | y_j, \hat{\Theta}^{(k)} \right] E \left[\gamma_j \zeta_j | y_j, \hat{\Theta}^{(k)} \right] \quad (2.22)$$

substituindo (2.17) e aplicando (1.73) na componente i temos

$$\begin{aligned} \hat{s}_{2ij}^{(k)} &= E \left[Z_{ij} | y_j, \hat{\Theta}^{(k)} \right] E \left[\gamma_j \zeta_j | y_j, \hat{\Theta}^{(k)} \right] \\ &= E \left[Z_{ij} | y_j, \hat{\Theta}^{(k)} \right] E \left[\zeta_j | y_j, \hat{\Theta}^{(k)} \right] \beta_i^{(k)} \eta_{rij}^{(k)} \\ &+ E \left[Z_{ij} | y_j, \hat{\Theta}^{(k)} \right] \left(\frac{\alpha_i^{(k)}}{\sqrt{\alpha_i^{2(k)} + \beta_i^{2(k)}} \pi \Psi(y_j | \hat{\Theta}^{(k)})} \left(\frac{1}{v_i^{(k)}} \rho_{ij}^{2(k)} + 1 \right)^{-\left(\frac{v_i^{(k)}}{2} + 1\right)} \right) \\ &= \beta_i^{(k)} \eta_{rij}^{(k)} E \left[Z_{ij} | y_j, \hat{\Theta}^{(k)} \right] E \left[\zeta_j | y_j, \hat{\Theta}^{(k)} \right] \\ &+ E \left[Z_{ij} | y_j, \hat{\Theta}^{(k)} \right] \frac{\alpha_i^{(k)}}{\sqrt{\alpha_i^{2(k)} + \beta_i^{2(k)}} \pi \Psi(y_j | \hat{\Theta}^{(k)})} \left(\frac{1}{v_i^{(k)}} \rho_{ij}^{2(k)} + 1 \right)^{-\left(\frac{v_i^{(k)}}{2} + 1\right)} \\ &= \beta_i^{(k)} \eta_{rij}^{(k)} \hat{s}_{1ij}^{(k)} + \hat{z}_{ij}^{(k)} \left\{ \frac{\alpha_i^{(k)}}{\sqrt{\alpha_i^{2(k)} + \beta_i^{2(k)}} \pi \Psi(y_j | \hat{\Theta}^{(k)})} \left(\frac{1}{v_i^{(k)}} \rho_{ij}^{2(k)} + 1 \right)^{-\left(\frac{v_i^{(k)}}{2} + 1\right)} \right\} \end{aligned} \quad (2.23)$$

De modo análogo, agora usando a equação (1.74) na componente i mostra

$$\text{iv) } \hat{s}_{3ij}^{(k)} = E \left[Z_{ij} \gamma_j^2 \zeta_j | y_j, \hat{\Theta}^{(k)} \right] = E \left[\gamma_j^2 \zeta_j E \left[Z_{ij} | y_j, \hat{\Theta}^{(k)} \right] | y_j, \hat{\Theta}^{(k)} \right] \quad (2.24)$$

logo

$$\begin{aligned} \hat{s}_{3ij}^{(k)} &= \beta_i^{2(k)} \eta_{rij}^{2(k)} \hat{s}_{1ij}^{(k)} \\ &+ \hat{z}_{ij}^{(k)} \left\{ \alpha_i^{(k)} + \frac{\beta_i^{(k)} \eta_{rij}^{(k)} \alpha_i^{(k)}}{\sqrt{\alpha_i^{2(k)} - \beta_i^{2(k)} \pi \Psi(y_j | \hat{\Theta}^{(k)})}} \left(\frac{1}{v_i^{(k)} \rho_{ij}^{2(k)}} + 1 \right)^{-\left(\frac{v_i^{(k)}}{2} + 1\right)} \right\} \end{aligned}$$

2.3.1 O Algoritmo ECME para Mistura Finita de Densidades t-Assimétricas

Nesta subseção vamos apresentar os passos do algoritmo ECME para estimar os parâmetros da mistura finita de densidades t-assimétricas. Para este estudo escolhemos tomar como iguais, embora desconhecidos, os parâmetros de graus de liberdade. Todos os procedimentos algébricos para obtenção desses resultados foram vistos nas seções anteriores.

O algoritmo ECME para Mistura Finita de Densidades t-Assimétricas, é definido por:

Passo E: Dado $\Theta = \hat{\Theta}^{(k)}$, um valor inicial. Calcular

$$\hat{z}_{ij}^{(k)}, \hat{s}_{1ij}^{(k)}, \hat{s}_{2ij}^{(k)} \text{ e } \hat{s}_{3ij}^{(k)} \quad (2.25)$$

para $i = 1, \dots, g$ e $j = 1, \dots, n$.

Passo CM 1: Calcular $\hat{\omega}^{(k+1)} = n^{-1} \sum_{j=1}^n \hat{z}_{ij}^{(k)}$

Passo CM 2: Obter o estimador $\hat{\xi}_i^{(k)}$

$$\hat{\xi}_i^{(k+1)} = \frac{\sum_{j=1}^n \left(\hat{s}_{1ij}^{(k)} y_j - \beta_i^k \hat{s}_{2ij}^{(k)} \right)}{\sum_{j=1}^n \hat{s}_{1ij}^{(k)}} \quad (2.26)$$

Passo CM 3: Fixado $\xi_i = \hat{\xi}_i^{k+1}$, calcular

$$\hat{\alpha}_i^{2(k+1)} = \frac{\sum_{j=1}^n \left((y_j - \hat{\xi}_i^{(k+1)})^2 \hat{s}_{1ij}^{(k)} - 2(y_j - \hat{\xi}_i^{(k+1)}) \beta_i^{(k)} \hat{s}_{2ij}^{(k)} + \beta_i^{(k)2} \hat{s}_{3ij}^{(k)} \right)}{n} \quad (2.27)$$

Passo CM 4: Fixados $\xi_i = \hat{\xi}_i^{k+1}$ e $\alpha_i = \hat{\alpha}_i^{2(k+1)}$, calcular

$$\hat{\beta}_i^{(k+1)} = \frac{\sum_{j=1}^n \left((y_j - \hat{\xi}_i^{(k+1)})^2 \hat{s}_{2ij}^{(k)} \right)}{\sum_{j=1}^n \hat{s}_{3ij}^{(k)}} \quad (2.28)$$

Passo CML: Fixados $\hat{\omega}^{(k+1)}$, $\xi_i = \hat{\xi}_i^{k+1}$, $\alpha_i = \hat{\alpha}_i^{2(k+1)}$ e $\hat{\beta}_i^{(k+1)}$, calcular

$$\hat{v}^{(k+1)} = \underset{v}{\operatorname{argmax}} \sum_{j=1}^n \log \left\{ \sum_i^g \hat{w}_i^{(k+1)} f_i \left(y_j | (\hat{\xi}_i^{(k+1)}, \hat{\alpha}_i^{2(k+1)}, \hat{\beta}_i^{(k+1)}, v) \right) \right\}. \quad (2.29)$$

Capítulo 3

Consistência das Estimativas e Uma Aplicação

Neste capítulo vamos apresentar um estudo de simulação para avaliar a consistência das estimativas dos parâmetros da mistura finita de densidades t-assimétricas dado em (2.1) obtido pelo algoritmo ECME. Apresentaremos também uma aplicação da teoria desenvolvida em uma modelagem com dados reais.

3.1 Estudo de Simulação

Com o objetivo de avaliar a consistência dos estimadores dos parâmetros da mistura finita de densidades t-assimétricas, obtidos no capítulo anterior, apresentaremos um estudo de simulação no qual foram consideradas misturas de duas e três densidades t-assimétricas.

Para avaliar a convergência dessas estimativas para os seu verdadeiros valores, em grandes amostras, calculamos, neste estudo de simulação, o vício (Bias) e o erro quadrático médio (MSE) das estimativas de cada coordenada do vetor de estimativas e para cada componente das misturas separadamente.

A escolha dos parâmetros para essas misturas seguiram dois casos de agrupamento, cujas características são definidas de acordo com a distância:

$$D = \sum_{i=1}^g w_i ((\xi_i - \bar{\xi})^2 + (\alpha_i^2 - \bar{\alpha}^2)^2 + (\beta_i - \bar{\beta})^2 + (v_i - \bar{v})^2)$$

onde

$$\bar{\xi} = \sum_{i=1}^g w_i \xi_i, \bar{\alpha}^2 = \sum_{i=1}^g w_i \alpha_i, \bar{\beta} = \sum_{i=1}^g w_i \beta_i \text{ e } \bar{v} = \sum_{i=1}^g w_i v_i.$$

Intuitivamente, consideramos

1. D Pequeno

Chamamos “D Pequeno” o agrupamento de misturas de duas e três densidades t-assimétricas com modas próximas, nas quais apenas pelo histograma dessas misturas não é possível avaliar facilmente a quantidade de componentes existentes em cada mistura;

2. D Grande

Chamamos “D grande”, o agrupamento de misturas de duas e três densidades t-assimétricas com modas afastadas. Neste, é possível visualizar, através dos histogramas dessas misturas, a quantidade de componentes existentes em cada mistura.

Histogramas
Misturas de Duas Densidades t-Assimétricas

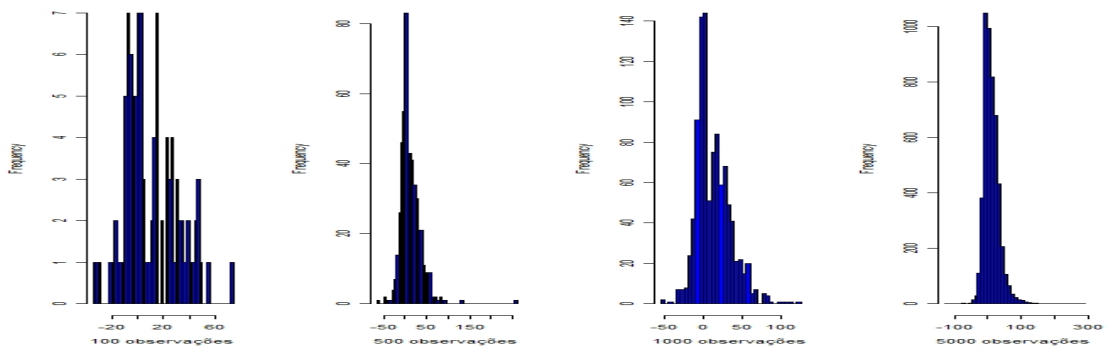


Figura 3.1: D Pequeno

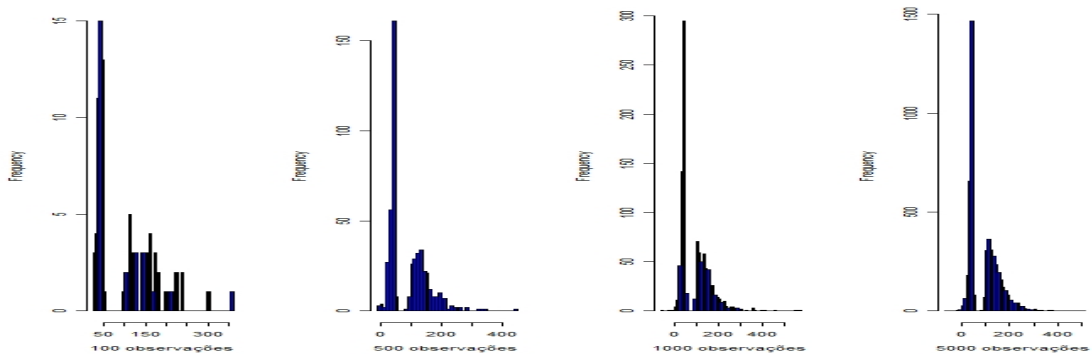


Figura 3.2: D Grande

As figuras 3.1 e 3.2 mostram histogramas para amostras com 100, 500, 1000 e 5000 observações geradas a partir de misturas de duas densidades t-assimétricas, nos casos de modas próximas (D Pequeno) e modas afastadas (D Grande), respectivamente. Em ambos os casos foram fixados valores para os parâmetros que são apresentados na tabela 3.1. O mesmo critério de escolha dos parâmetros foi atribuído à mistura de três densidades t-assimétricas.

Observe que na figura 3.1, para amostra com 5000 observações, o histograma não indica que se trata de mistura de densidades t-assimétricas. Já na figura 3.2, para o mesmo número de observações, a mistura de duas densidades t-assimétricas está bem evidenciada.

Histogramas Misturas de Três Densidades t-Assimétricas

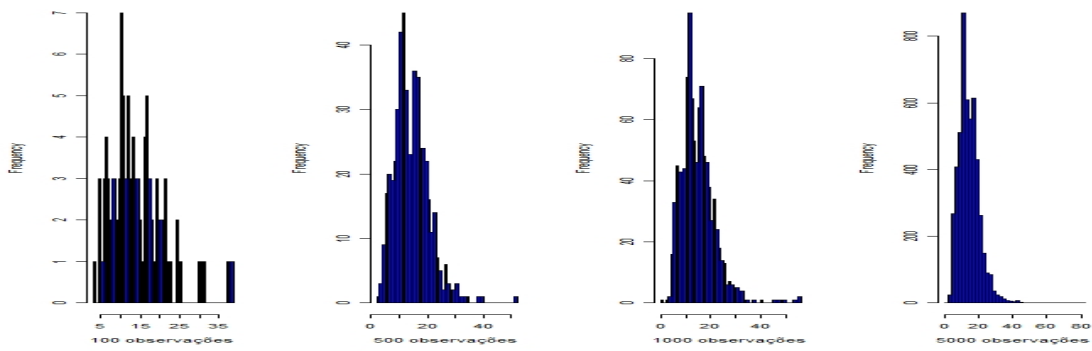


Figura 3.3: D Pequeno

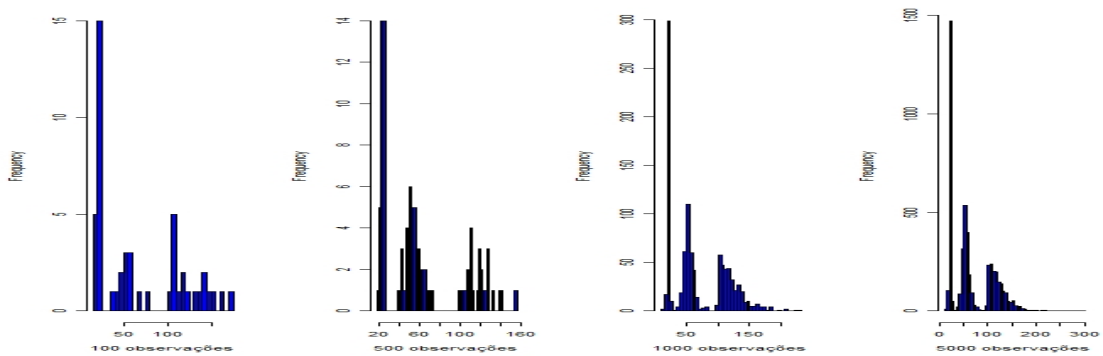


Figura 3.4: D Grande

As figuras 3.3 e 3.4 mostram histogramas de amostras com 100, 500, 1000 e 5000 observações geradas de misturas de três densidades t- assimétricas, nos casos D Pequeno e D Grande, respectivamente.

Na figura 3.3 os histogramas indicam, para todos o tamanhos de amostra, que se trata de mistura de densidades t-assimétricas, mas não é possível identificar que temos misturas de três densidades t-assimétricas. Para a amostra com 5000 observações, por exemplo, aparentemente o histograma trata de mistura de duas densidades t-assimétricas. Na figura 3.4 é possível identificar para todo tamanho de amostra que trata-se de mistura de três densidades t-assimétricas, observe que a partir de 500 observações esta quantidade já está bem evidenciada.

Para este desenvolvimento escolhemos tomar como iguais, embora desconhecidos, os valores dos graus de liberdade de todas as componentes da mistura. Ou seja, a expressão da distância D reduz-se a

$$D = \sum_{i=1}^g w_i \left((\xi_i - \bar{\xi})^2 + (\alpha_i^2 - \bar{\alpha}^2)^2 + (\beta_i - \bar{\beta})^2 \right).$$

A tabela a baixo mostra a combinação dos parâmetros para os modelos de misturas de duas e três densidades t-assimétricas distribuídas nos casos de D pequeno (P) e D grande (G).

| Nc | D | w_1 | ξ_1 | α_1^2 | β_1 | v_1 | w_2 | ξ_2 | α_2^2 | β_2 | v_2 | w_3 | ξ_3 | α_1^3 | β_3 | v_3 |
|----|---|-------|---------|--------------|-----------|-------|-------|---------|--------------|-----------|-------|-------|---------|--------------|-----------|-------|
| 2 | P | 0.5 | 5 | 10 | -5 | 4 | 0.5 | 10 | 20 | 10 | 4 | | | | | |
| 2 | G | 0.5 | 50 | 10 | -10 | 4 | 0.5 | 100 | 50 | 10 | 4 | | | | | |
| 3 | P | 1/3 | 5 | 3 | -5 | 4 | 1/3 | 10 | 4 | 3 | 4 | 1/3 | 15 | 5 | 10 | 4 |
| 3 | G | 1/3 | 25 | 2 | -10 | 4 | 1/3 | 50 | 7 | 1 | 4 | 1/3 | 100 | 25 | 20 | 4 |

Tabela 3.1: Combinação dos Parâmetros

O termo Nc indica número de componentes em cada mistura e os termos w_i , ξ_i , α_i^2 , β_i e v_i indicam os parâmetros de peso, locação, escala, assimetria e graus de liberdade da componente i, respectivamente, com $i = 1, 2, 3$.

As figuras 3.1, 3.2, 3.3 e 3.4 foram geradas com os parâmetros descritos na tabela acima.

3.1.1 Processo de Simulação

Para iniciar o processo de simulação vamos fazer variar o tamanho n da amostra de 50 a 10 000. Faremos então $n = 50, 100, 200, 500, 1000, 2000, 5000$ e 10 000 para cada mistura.

O processo de simulação segue os passos definidos abaixo:

Para cada n são fixados os valores iniciais, os “chutes” do algoritmo ECME, iguais aos verdadeiros valores dos parâmetros dados na tabela (3.1).

1. **Passo 1** Gerar n valores observados independentes, provenientes de uma mistura de densidades t-assimétricas;
2. **Passo 2** Obter via algoritmo ECME uma estimativa $\hat{\Theta}_i$;
3. **Passo 3** Repetir os passos 1-2, $r = 1000$ vezes, obtendo uma seqüência de estimativas $\{\hat{\Theta}_i^{(m)}\}$ e calcular:

3.1 O vício (Bias) do estimador para cada componente da mistura,

$$Bias(\hat{\Theta}_i) = \frac{1}{r} \sum_{m=1}^r \hat{\Theta}_i^{(m)} - \Theta_i;$$

3.2 O Erro Quadrático Médio (MSE) $\hat{\Theta}_i$ para cada componente da mistura,

$$MSE(\hat{\Theta}_i) = \frac{1}{r} \sum_{m=1}^r \left(\hat{\Theta}_i^{(m)} - \Theta_i \right)^2 .$$

Onde

$\Theta_i = (w_i, \xi_i, \alpha_i^2, \beta_i, v_i)$, r indica o número de interações da simulação,

$\hat{\Theta}_i^{(m)} = (\hat{w}_i^{(m)}, \hat{\xi}_i^{(m)}, \hat{\alpha}_i^{2(m)}, \hat{\beta}_i^{(m)}, \hat{v}_i^{(m)})$ é o estimador da m -ésima interação com $m = 1, \dots, r$ e $i = 1, \dots, g$.

3.1.2 Resultados

Podemos avaliar a consistência das estimativas obtidas via algoritmo ECME, analisando os seguintes gráficos gerados a partir do estudo de simulação definido anteriormente. Usamos as cores preto, vermelho e azul para representar o Bias e MSE da primeira, segunda e terceira componente em cada mistura, respectivamente.

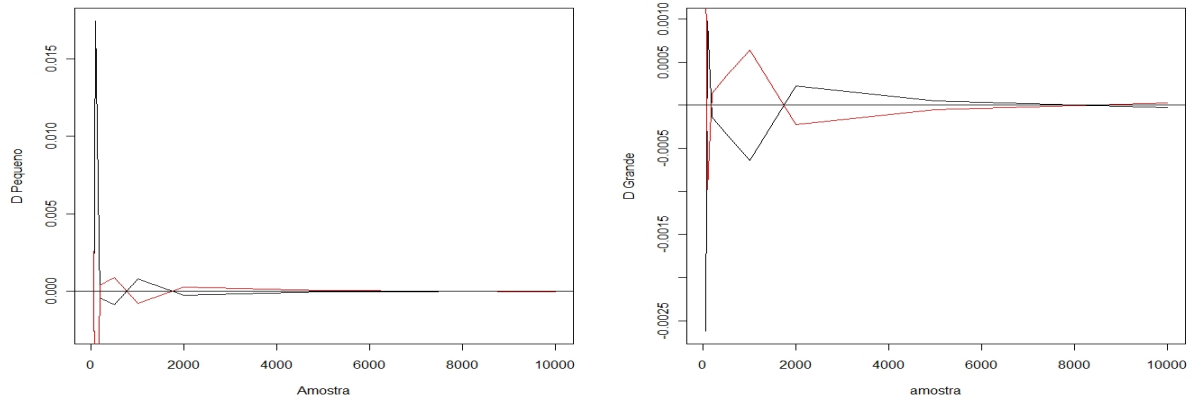


Figura 3.5: Bias do \hat{w}_i com $i = 1, 2$.

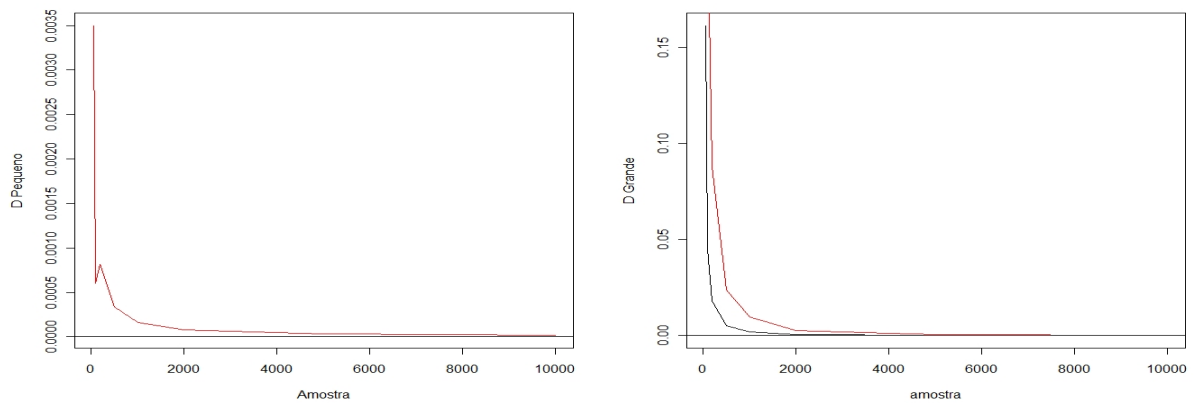


Figura 3.6: MSE do \hat{w}_i com $i = 1, 2$.

- Bias e MSE do estimador do parâmetro PESO do modelo de mistura de duas densidades t-Assimétricas nos casos **D Pequeno** e **D Grande**, respectivamente. As cores preto e vermelho indicam a primeira e segunda componente de cada mistura.

Os vícios das estimativas dos parâmetros pesos das componentes dessas misturas diminuem mais rapidamente no primeiro caso com $n \cong 4000$. No segundo caso, observe que para $n \cong 6000$ observações esses vícios começam a ser insignificantes, fig. 3.5.

Os MSEs dessas estimativas convergem simultaneamente no primeiro caso, sendo possível visualizar apenas uma linha. Já no segundo, a convergência para zero do MSE da estimativa \hat{w}_2 ocorre mais lentamente, fig. 3.6.

- Biais e MSE do estimador do parâmetro LOCAÇÃO do modelo de mistura de duas densidades t-Assimétricas nos casos **D Pequeno** e **D Grande**, respectivamente. As cores preto e vermelho indicam a primeira e segunda componente de cada mistura.

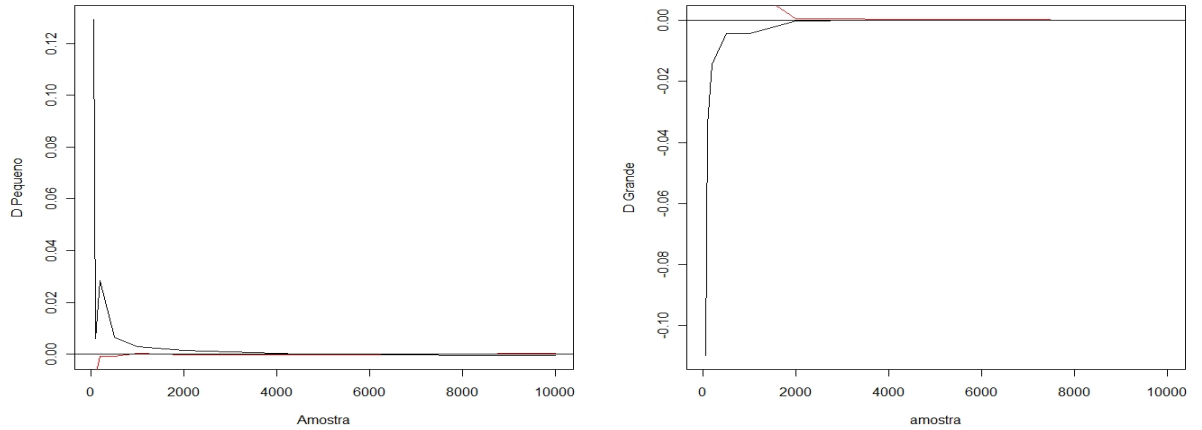


Figura 3.7: Biais do $\hat{\xi}_i$ com $i = 1, 2$.

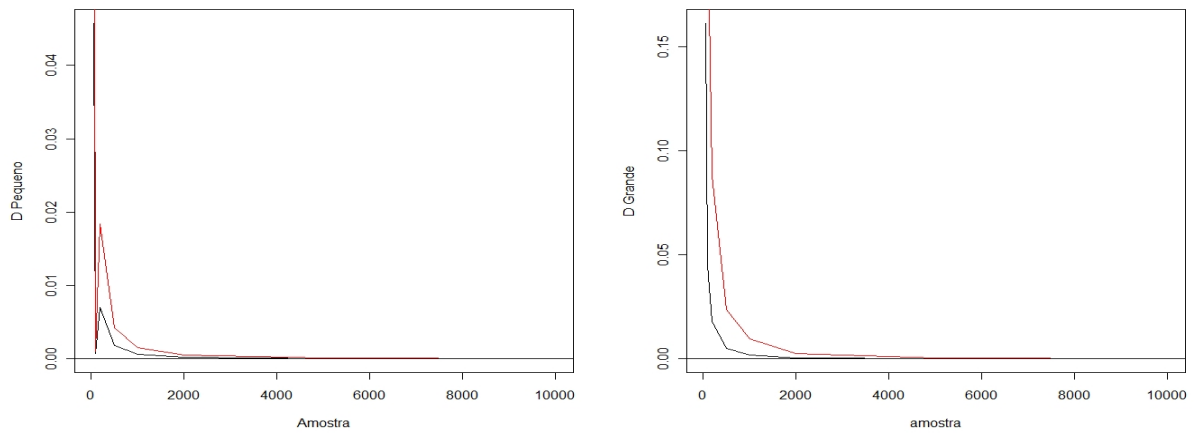


Figura 3.8: MSE do $\hat{\xi}_i$ com $i = 1, 2$.

Os vícios das estimativas dos parâmetros de locação das componentes dessas misturas passam a ser insignificantes para amostra com aproximadamente 2000 observações em ambos os casos de agrupamento, fig. 3.7. Este número é considerado pequeno, se comparado ao ponto de convergência dos vícios das do peso. Por outro lado, o MSE do $\hat{\xi}_1$ converge para zero mais rapidamente que o do $\hat{\xi}_2$ em ambos os casos, fig. 3.8.

- Bias e MSE do estimador do parâmetro ESCALA do modelo de mistura de duas densidades t-Assimétricas nos casos **D Pequeno** e **D Grande**, respectivamente. As cores preto e vermelho indicam a primeira e segunda componente de cada mistura.

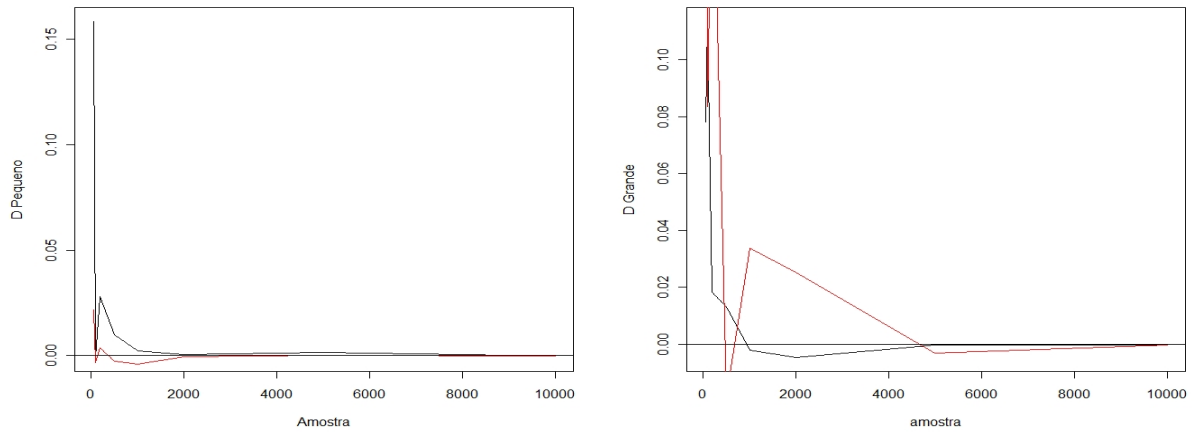


Figura 3.9: Bias do $\hat{\alpha}_i^2$ com $i = 1, 2$.

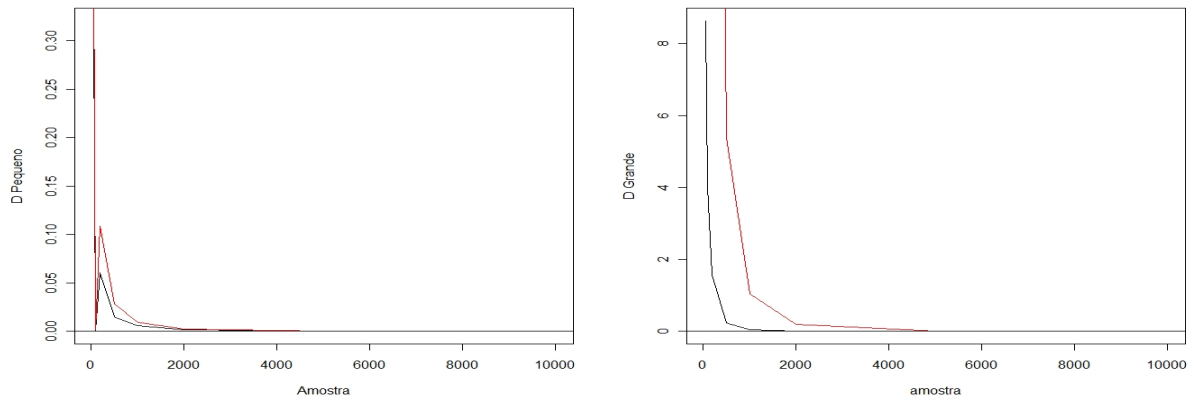


Figura 3.10: MSE do $\hat{\alpha}_i^2$ com $i = 1, 2$.

No caso D Pequeno, tanto os vícios quanto os MSEs das estimativas dos parâmetros de escala das componentes dessas misturas têm comportamentos parecidos quanto a convergência, o mesmo não ocorre no segundo caso. Observe que para D Grande o vício da estimativa $\hat{\alpha}_1^2$ converge mais rapidamente que o da estimativa $\hat{\alpha}_2^2$. O mesmo ocorre com o MSE dessas estimativas, enquanto $\hat{\alpha}_1^2$ converge para zero com $n \cong 2000$ a estimativa $\hat{\alpha}_2^2$ precisa de mais que o dobro de observações para atingir a convergência, figs. 3.9 - 3.10.

- Bias e MSE do estimador do parâmetro ASSIMETRIA do modelo de mistura de duas densidades t-Assimétricas nos casos **D Pequeno** e **D Grande**, respectivamente. As cores preto e vermelho indicam a primeira e segunda componente de cada mistura.

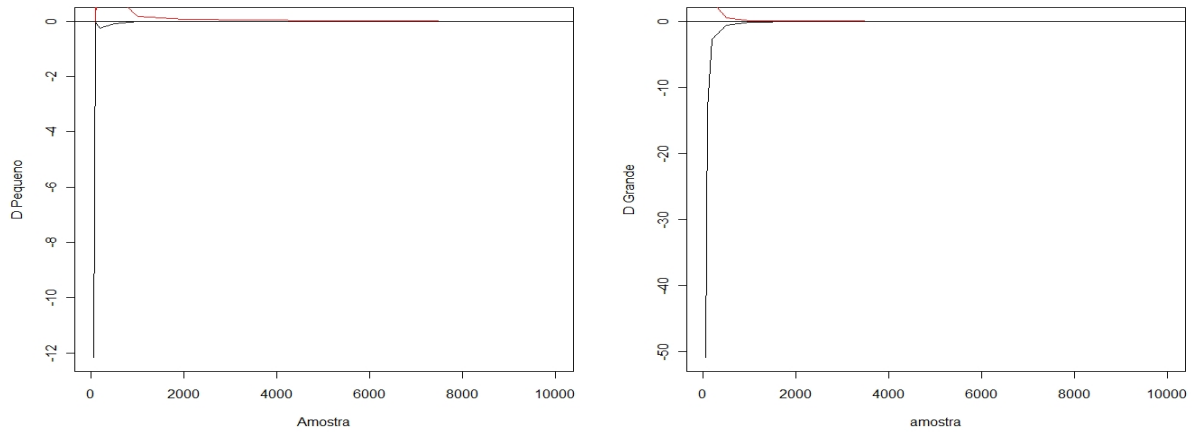


Figura 3.11: Bias $\hat{\beta}_i$ com $i = 1, 2$.

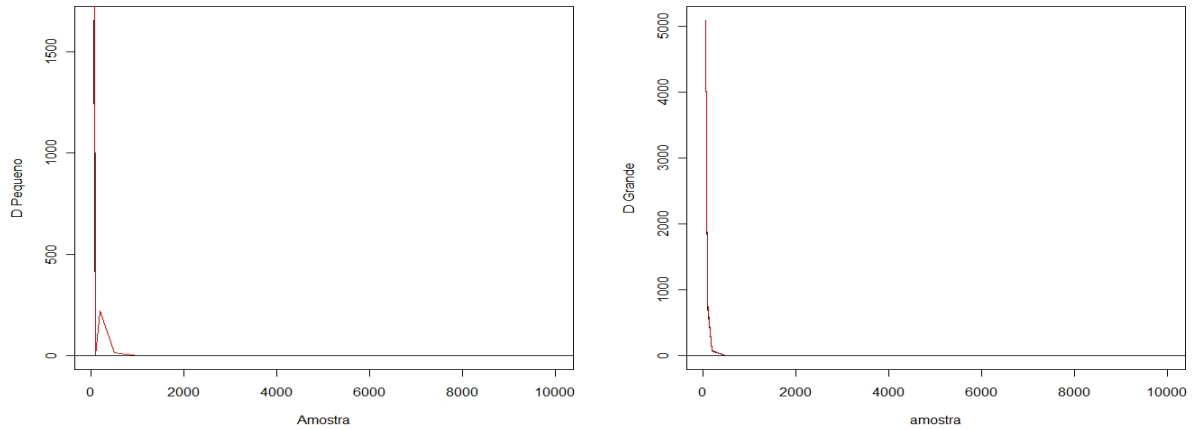


Figura 3.12: MSE do $\hat{\beta}_i$ com $i = 1, 2$.

Os vícios das estimativas dos parâmetros de assimetria das componentes dessas misturas atingem insignificância em aproximadamente 1000 observações em ambos os casos de agrupamento, fig. 3.11.

Os MSEs dessas estimativas convergem para zero simultaneamente sendo possível visualizar apenas uma linha, fig. 3.12.

- Biais e MSE do estimador do parâmetro GRAUS DE LIBERDADE do modelo de mistura de duas densidades t-Assimétricas nos casos **D Pequeno** e **D Grande**, respectivamente. As cores preto e vermelho indicam a primeira e segunda componente de cada mistura.

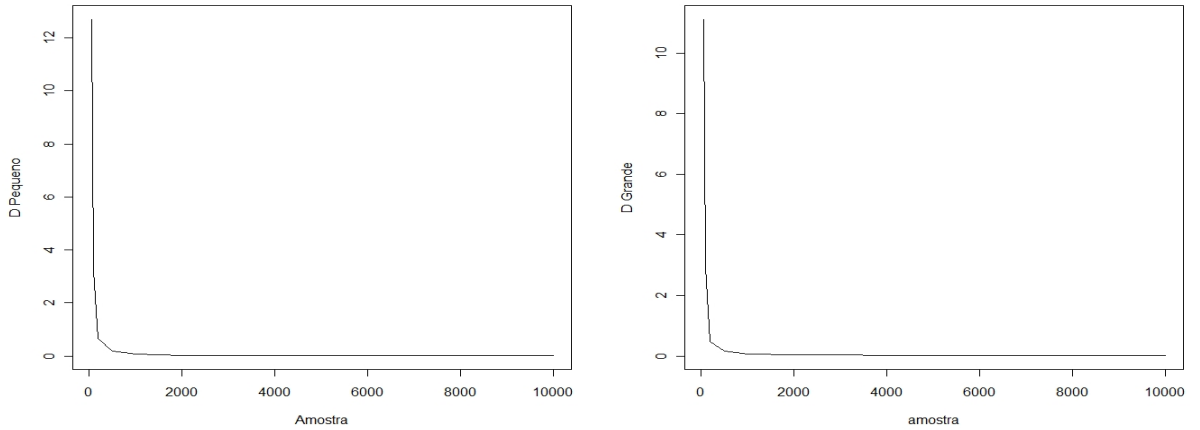


Figura 3.13: Biais \hat{v}_i com $i = 1, 2$.

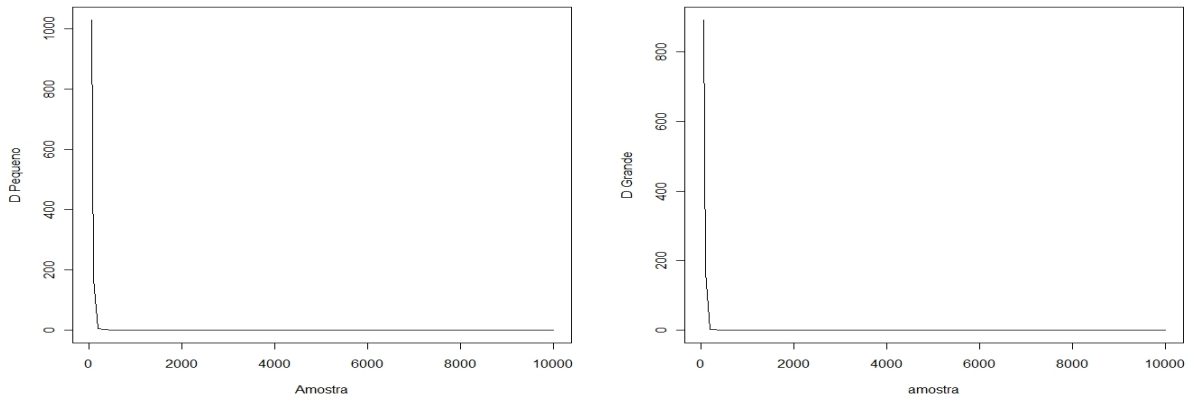


Figura 3.14: MSE do \hat{v}_i com $i = 1, 2$.

Como os parâmetros de graus de liberdade foram tomados iguais em todas as componentes dessas misturas, as linhas que representam os vícios e os MSEs de suas estimativas são iguais. Observe que, tanto os vícios quanto os MSEs dessas estimativas têm comportamentos semelhantes em ambos os casos de agrupamento, figs. 3.13 - 3.14.

Para os próximos resultados, o processo de simulação foi encerrado em $n = 5000$.

- Biais e MSE do estimador do parâmetro PESO do modelo de mistura de três densidades t-Assimétricas nos casos **D Pequeno** e **D Grande**, respectivamente. As cores preto, vermelho e azul indicam a primeira, segunda e terceira componente de cada mistura.

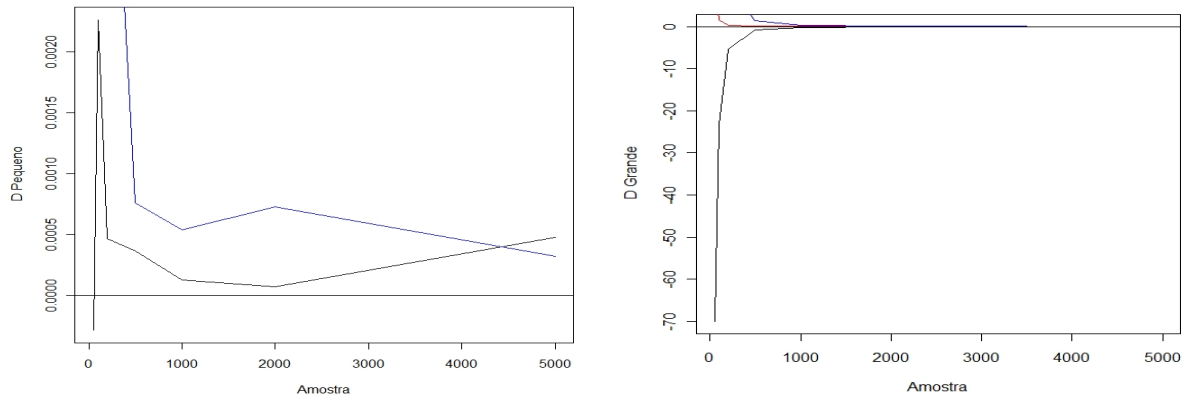


Figura 3.15: Biais do \hat{w}_i com $i = 1, 2, 3$.

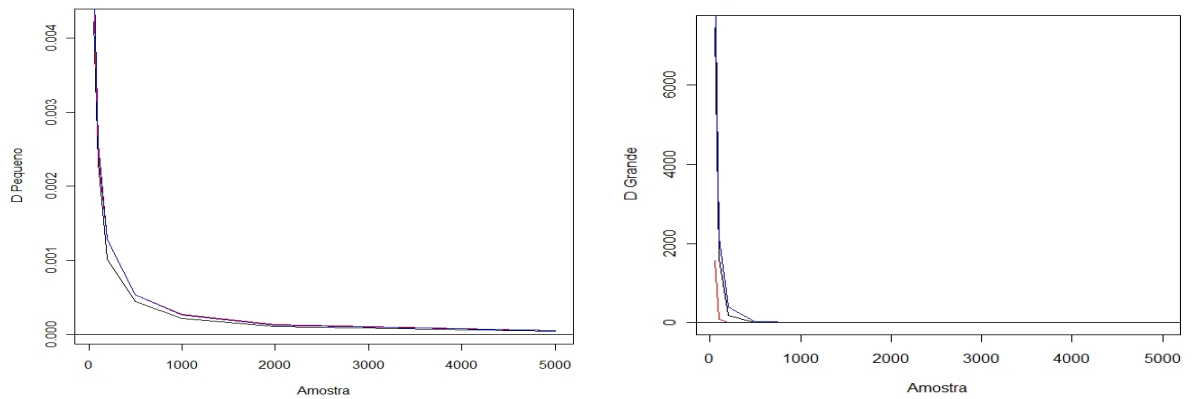


Figura 3.16: MSE do \hat{w}_i com $i = 1, 2, 3$.

Os vícios das estimativas dos parâmetros pesos das componentes dessas misturas assumem valores bem próximos de zero no primeiro caso, fig. 3.15. Ainda neste caso, observe que a linha vermelha que indica o vício da estimativa $\hat{\xi}_2$ não aparece. Já no segundo caso é possível observar a convergência dos vícios das três estimativas avaliadas. Por outro lado, observando a figura fig. 3.16, note que os MSEs dessas estimativas convergem para zero mais rapidamente no segundo caso de agrupamento.

- Bias e MSE do estimador do parâmetro LOCAÇÃO do modelo de mistura de três densidades t-Assimétricas nos casos "D Pequeno e D Grande, respectivamente. As cores preto, vermelho e azul indicam a primeira, segunda e terceira componente de cada mistura.

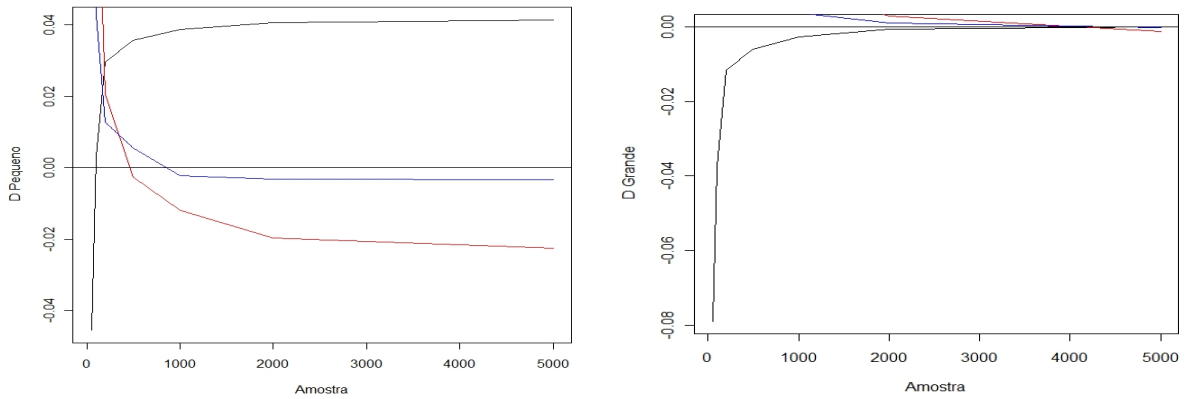


Figura 3.17: Bias do $\hat{\xi}_i$ com $i = 1, 2, 3$.

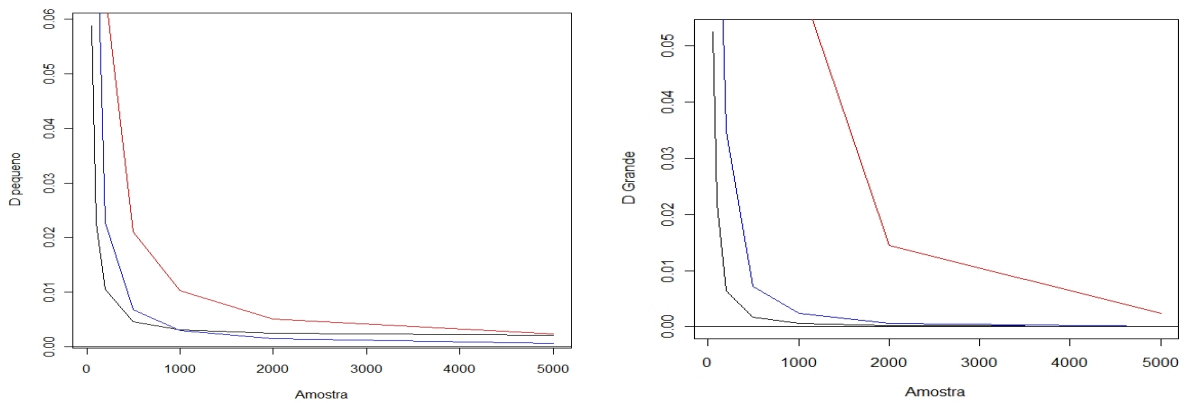


Figura 3.18: MSE do $\hat{\xi}_i$ com $i = 1, 2, 3$.

Os vícios das estimativas dos parâmetros de locação das componentes dessas misturas, no caso D Grande ficam insignificantes para um número aproximado de 4000 observações. Já no caso D Pequeno até o número máximo de observações simuladas, não foi possível atingir a insignificância esperada, fig. 3.17. Acredita-se pelo MSE desta mistura, fig.3.18, para um número maior de observações na simulação teríamos a tal convergência.

- Biais e MSE do estimador do parâmetro ESCALA do modelo de mistura de três densidades t-Assimétricas nos casos **D Pequeno** e **D Grande**, respectivamente. As cores preto, vermelho e azul indicam a primeira, segunda e terceira componente de cada mistura.

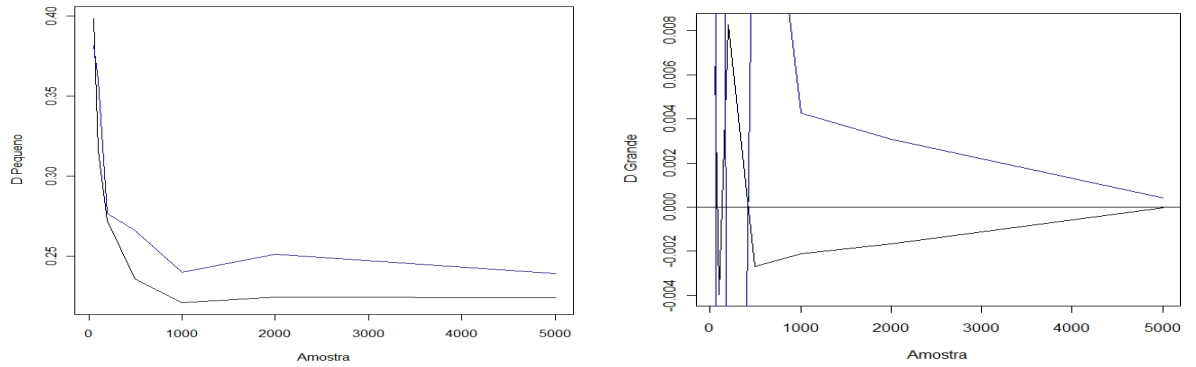


Figura 3.19: Biais do $\hat{\alpha}_i^2$ com $i = 1, 2, 3$.

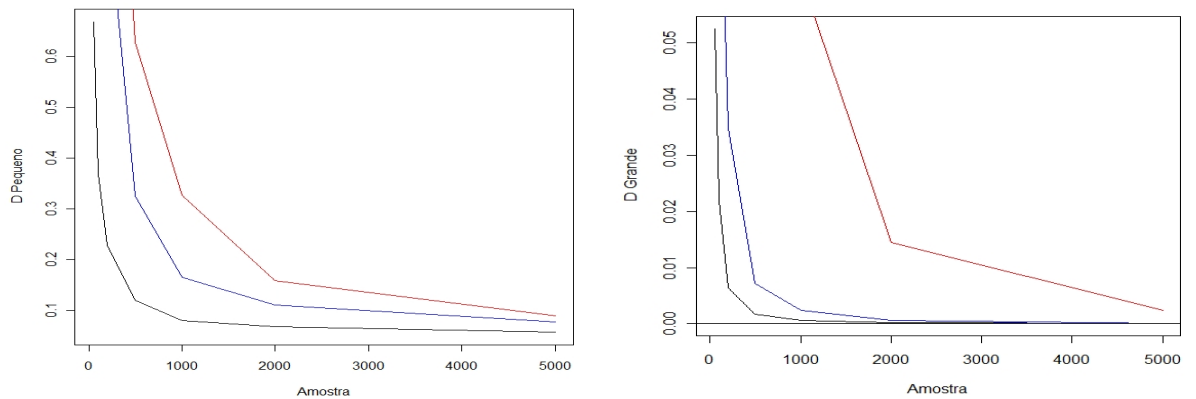


Figura 3.20: MSE do $\hat{\alpha}_i^2$ com $i = 1, 2, 3$.

Observe que em ambos os casos de agrupamento o vício da estimativa $\hat{\alpha}_2^2$ não aparece, fig. 3.19. Apenas no segundo caso os vícios das $\hat{\alpha}_1^2$ e $\hat{\alpha}_3^2$ convergem para zero. Ainda neste caso, agora analisando a fig.3.20, podemos perceber que o MSE da $\hat{\alpha}_2^2$ precisa de um número maior de observações na simulação para atingir a convergência para zero. No primeiro caso, os MSEs das estimativas dos parâmetros de escala não atingiram a convergência esperada.

- Bias e MSE do estimador do parâmetro ASSIMETRIA do modelo de mistura de três densidades t-Assimétricas nos casos **D Pequeno** e **D Grande**, respectivamente. As cores preto, vermelho e azul indicam a primeira, segunda e terceira componente de cada mistura.

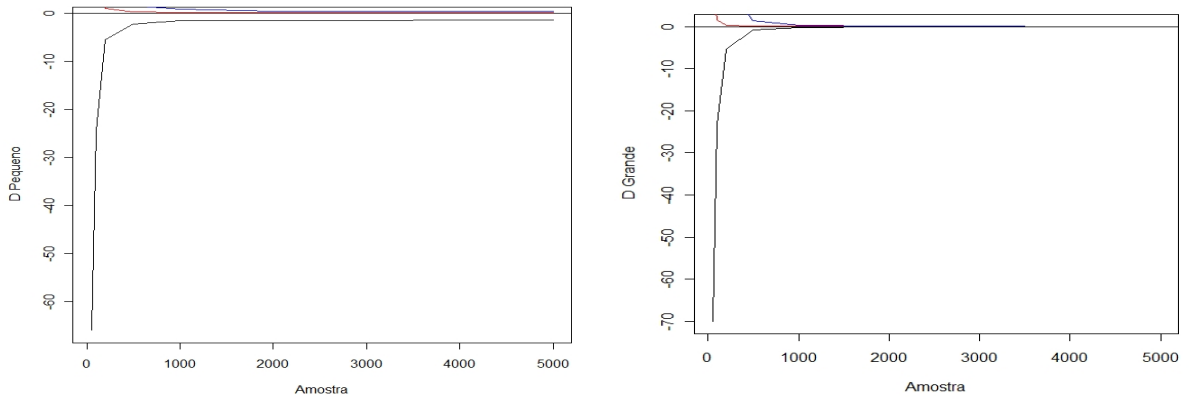


Figura 3.21: Bias do $\hat{\beta}_i$ com $i = 1, 2, 3$.

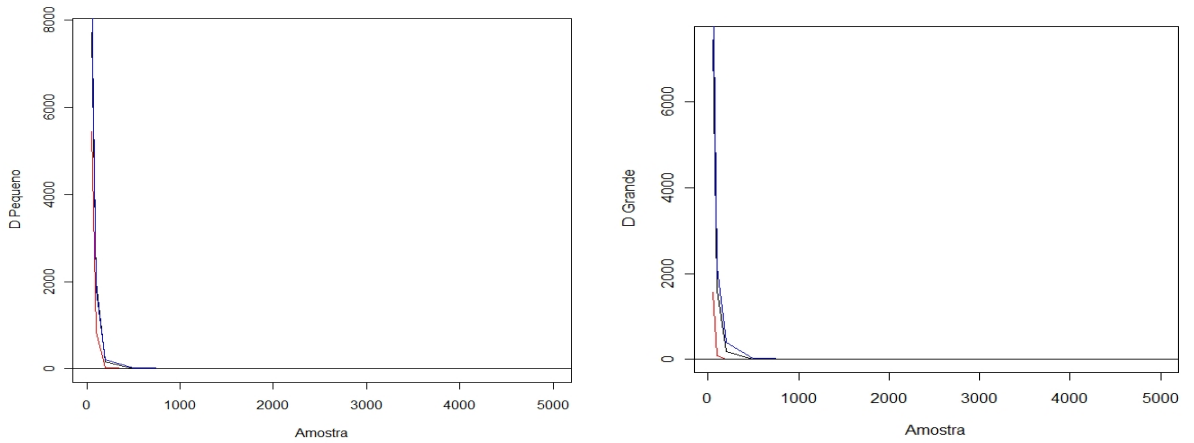


Figura 3.22: MSE do $\hat{\beta}_i$ com $i = 1, 2, 3$.

Os vícios das estimativas dos parâmetros de assimetria das componentes dessas misturas, no segundo caso se tornam insignificante com aproximadamente 1000 observações. Já no primeiro caso o vício da $\hat{\beta}_1$ não atinge a insignificância esperada, fig. 3.21. Em relação aos MSEs, essas estimativas têm comportamentos parecidos, em ambos os casos de agrupamento fig. 3.22.

- Bias e MSE do estimador do parâmetro GRAUS DE LIBERDADE do modelo de mistura de três densidades t-Assimétricas, nos casos **D Pequeno** e **D Grande**, respectivamente. As cores preto, vermelho e azul indicam a primeira, segunda e terceira componente de cada mistura.

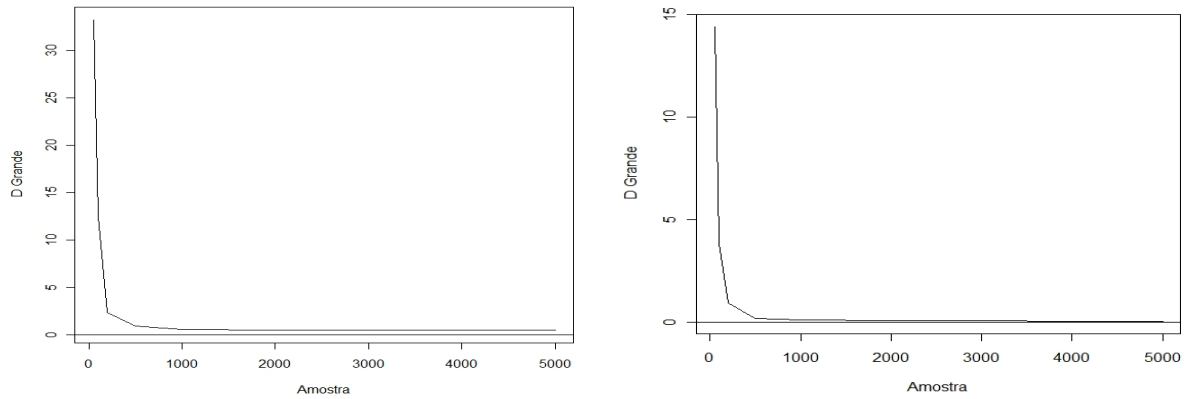


Figura 3.23: Bias do \hat{v}_i com $i = 1, 2, 3$.

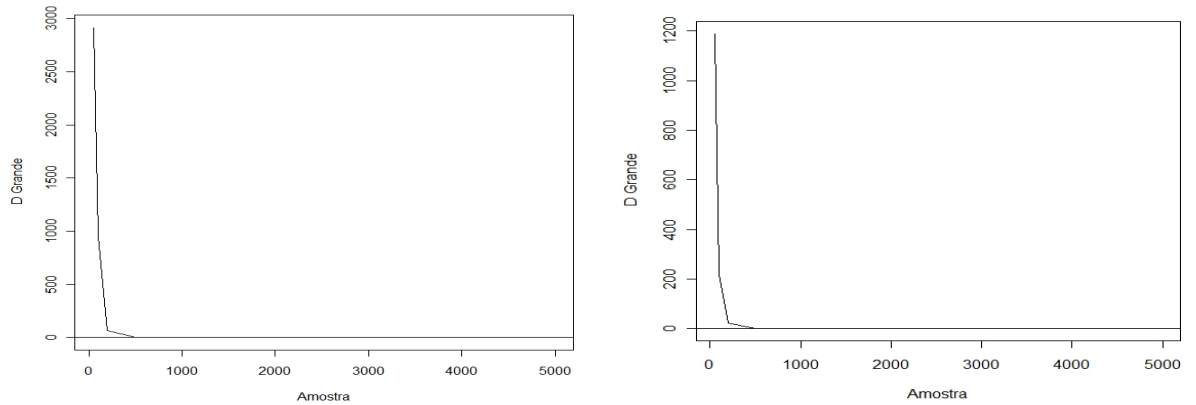


Figura 3.24: MSE do \hat{v}_i com $i = 1, 2, 3$.

Assim com na mistura de duas densidades t-assimétricas, os vícios e os MSEs das estimativas dos parâmetros graus de liberdade dessas misturas são representados por uma única linha. Observe que tanto os vícios quanto os MSEs dessas estimativas têm comportamentos semelhantes em ambos os casos de agrupamento figs. 3.23 - 3.24.

Tanto a implementação do algoritmo EM quanto a avaliação da eficiência das estimativas geradas pelo mesmo, foram obtidas pelo programa *R*. Devido as limitações desse programa, o estudo de simulação para mistura de três densidades t-assimétricas teve que ser interrompido quando n foi igual a 5000 observações. Mesmo com essa limitação, foi possível avaliar as propriedades assintóticas das estimativas geradas pelo algoritmo, uma vez que esse tamanho de observações é considerado muito grande para o contexto de modelagem.

Para todas as situações investigadas, no estudo de simulação, vimos que o algoritmo EM fornece estimativas assintoticamente consistentes para os parâmetros da mistura de densidades t-assimétricas.

Vimos também que, para a análise das propriedades assintóticas das estimativas geradas pelo algoritmo EM a relação de distância D é irrelevante para misturas de duas densidades t-assimétricas. Por outro lado, para mistura de três densidades t-assimétricas, esta relação de distância tem grande influência na convergência para zero dos vícios e erros quadráticos médios dessas estimativas. Foram comparados os dois casos de agrupamento (D Pequeno e D Grande) e para essas misturas, é possível obter a convergência do Bias e do MSE com um número bem menor de observações quando a relação de distância D for grande.

3.2 Aplicação

Para a ilustrar a aplicabilidade da metodologia deste trabalho, consideremos o conjunto de dados reais analisado em Dias & Wedel (2004), que está relacionado com o produto interno bruto (PIB) por capital (PPP US\$) em 1998 de 174 países (UNDP, 2000). Estes autores modelaram estes dados aplicando primeiramente uma transformação logarítmica e, depois disso, ajustando uma mistura de duas normais.

Vamos fazer a modelagem desses dados usando uma mistura de duas densidades t-assimétricas, dada por (2.1), na qual vamos considerar os graus de liberdade igualmente em cada componente da mistura. Inicialmente, mudaremos o conjunto de dados dividindo por 10 000 cada valor observado, em seguida vamos aplicar o algoritmo EM para obter uma estimativa dos parâmetros da mistura.

Para as iterações do algoritmo EM vamos tomar os seguintes valores iniciais:

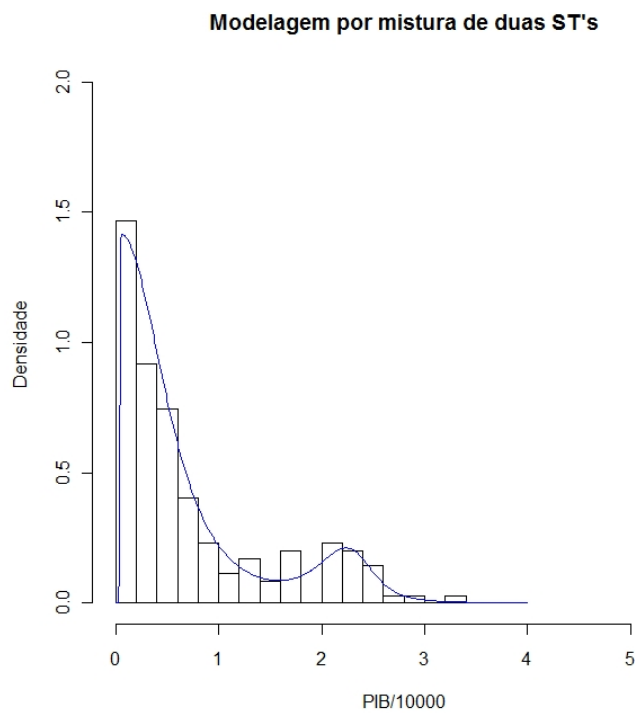
$$\begin{aligned}(\xi_1, \xi_2) &= (0, 3), \\(\alpha_1^2, \alpha_2^2) &= (0.2, 0.1), \\(\beta_1, \beta_2) &= (60, -10), \\(v_1, v_2) &= (3, 3), \\(w_1, w_2) &= (0.7, 0.3).\end{aligned}$$

Na tabela abaixo estão as estimativas geradas pelo algoritmo EM para a modelagem do conjunto de dados PIB/10 000, por uma mistura de duas densidades t-assimétricas.

| Parâmetros | Estimativas | Parâmetros | Estimativas |
|--------------|-------------|--------------|-------------|
| ξ_1 | 0.04361287 | ξ_2 | 2.38010153 |
| α_1^2 | 0.1969954 | α_2^2 | 0.1091982 |
| β_1 | 171.6752367 | β_2 | -0.9588135 |
| v_1 | 3.245425 | v_2 | 3.245425 |
| w_1 | 0.8507964 | w_2 | 0.1492036 |

Tabela 3.2: Estimativas geradas pelo EM

Na figura abaixo, apresentamos o histograma do conjunto de dados PIB/10 000 e a curva da densidade da mistura de duas densidades t-assimétricas. A densidade foi gerada a partir das estimativas fornecidas pelo algoritmo EM, vistos na tabela 3.2.



Considerações Finais

Mostramos neste trabalho um estudo detalhado do procedimento utilizado para a modelagem de dados heterogêneos, dotados de assimetria e com valores discrepantes, por uma mistura de densidades t-assimétricas definida por Lin *et al.* (2007a).

Para esta modelagem foi aplicada a teoria do algoritmo EM desenvolvida por Dempster *et al.* (1977) do qual usamos as extensões ECM e ECME introduzidas por Meng & Rubin (1993) e Liu & Rubin (1994), respectivamente .

Avaliamos as propriedades assintóticas das estimativas geradas pelo algoritmo através de um estudo de simulação, onde evidenciou-se-se que essas estimativas são assintoticamente consistentes.

A aplicação dessa teoria em uma modelagem com dados reais, mostrou que esta é uma alternativa adequada para a modelagem de dados assimétricos, heterogêneos e com observações discrepantes, evitando uma desnecessária transformação.

Referências Bibliográficas

- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171–178.
- Azzalini, A. & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society, Series B*, **65**, 367–389.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Statist. Soc.*, **B 39**, 1–38.
- Dias, J. G. & Wedel, M. (2004). An empirical comparison of EM, SEM and MCMC performance for problematic gaussian mixture likelihoods. *Statistics and Computing*, **14**, 323–332.
- Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer Verlag.
- Henze, N. (1986). A probabilistic representation of the skew-normal distribution. *Scandinavian Journal of Statistics*, **13**, 271–275.
- Lachos, V. H., Ghosh, P. & Arellano-Valle, R. B. (2009). Likelihood based inference for skew normal independent linear mixed models. *Statistica Sinica*. To appear.
- Lin, T. I., Lee, J. C. & Hsieh, W. J. (2007a). Robust mixture modelling using the skew t distribution. *Statistics and Computing*, **17**, 81–92.
- Lin, T. I., Lee, J. C. & Yen, S. Y. (2007b). Finite mixture modelling using the skew normal distribution. *Statistica Sinica*, **17**, 909–927.
- Liu, C. H. & Rubin, D. B. (1994). The ECME algorithm: a simple extension of em and ecm with faster monotone convergence. *Biometrika*, **81**, 633–648.
- McLachlan, G. J. & Peel, G. J. (2000). *Finite Mixture Models*. John Wiley and Sons.
- Meng, X. L. & Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika*, **80**, 267–278.