

Universidade Federal de Juiz de Fora  
Instituto de Ciências Exatas  
Departamento de Estatística

**Victor Silveira Gomes**

**Detecção de Ponto de Mudança nos Parâmetros do Modelo Normal  
Assimétrico: Algoritmo EM**

Juiz de Fora

2016

**Victor Silveira Gomes**

**Detecção de Ponto de Mudança nos Parâmetros do Modelo Normal  
Assimétrico: Algoritmo EM**

Monografia apresentada ao Departamento de Estatística da Universidade Federal de Juiz de Fora, na área de concentração em Estatística, como requisito parcial para obtenção do título de Bacharel em Estatística.

Orientadora: Camila Borelli Zeller

Juiz de Fora

2016

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF  
com os dados fornecidos pelo(a) autor(a)

Gomes, Victor Silveira.

Detecção de Ponto de Mudança nos Parâmetros do Modelo Normal  
Assimétrico: Algoritmo EM / Victor Silveira Gomes. – 2016.

2 f.

Orientadora: Camila Borelli Zeller

Trabalho de Conclusão de Curso – Universidade Federal de Juiz de Fora,  
Instituto de Ciências Exatas. Departamento de Estatística, 2016.

1. Ponto de Mudança. 2. Algoritmo EM. 3. Seleção de Modelos. 4.  
Critério de Informação de Schwarz. 5. Família de Distribuição Normal  
Assimétrica. Zeller, Camila Borelli, orient. II. Título.

Victor Silveira Gomes

Detecção de Ponto de Mudança nos Parâmetros do Modelo Normal  
Assimétrico: Algoritmo EM

Monografia apresentada ao Departamento de Estatística da Universidade Federal de Juiz de Fora, na área de concentração em Estatística, como requisito parcial para obtenção do título de Bacharel em Estatística.

BANCA EXAMINADORA

---

Camila Borelli Zeller  
Doutora em Estatística - Universidade Estadual de  
Campinas

---

Clécio da Silva Ferreira  
Doutor em Estatística – Universidade de São Paulo

---

Lupércio França Bessegato  
Doutor em Estatística – Universidade Federal de Minas  
Gerais

# Agradecimentos

Primeiramente ao meu grande amigo Euler (*in memoriam*) que mesmo sendo um grande merecedor, infelizmente não pode compartilhar dessa mesma felicidade que eu.

Aos familiares, por terem sido uma base sólida para que pudesse chegar onde estou agora. Meus pais, meus irmãos Henrique e Clara, meus primos Adriano, Rafael, Savinho, Daniel Mocó, Marquinho, Gabriela, Sâmara e o mais novo membro dela, o Gabriel.

Aos professores do departamento de estatística por todo aprendizado ao longo do curso, em especial a professora Camila, que além das excelentes aulas foi minha orientadora, me ajudando e compreendendo os problemas que tive ao longo desse processo.

Aos meus vários amigos. Os de longa data, Lana, Leandro, Dedé, Maia, Gabriel, Beijo, Bertante e Daniel, o tempo não irá diminuir essa nossa amizade. Aos que fiz em todo esse tempo de faculdade Mario, Augusto, Alan, Chocolate, Walmir, Marcão e Gaby, sofremos mas conseguimos. Aos do mestrado, que em tão pouco tempo se tornaram pessoas tão importantes, Pagani, Pablo, Luis, Gisele, Thaís, Isabella e minha querida Larissa. Aos do futebol de onze horas do sábado, Fabiano, Visoná, Carlinho, Gabriel Aliane, Fubá, Cardoso, Wendell, Flávio Sereno, Vertigo, Will e os outros todos que não me lembro agora. E por fim aos do futebol de uma hora do sábado, Diógenes, Fernandinho Messi, Edu, Fernando, Luis Fernando, Wesley e o resto do pessoal. Ninguém é feliz sem amigos.

# Resumo

O problema de ponto de mudança surgiu inicialmente em controle de qualidade (PAGE, 1954) e antes da introdução da hipótese de ponto de mudança associada aos modelos, pesquisadores enfrentavam dificuldades em estabelecer um modelo para alguns conjuntos de dados. Neste trabalho, propomos estudar o problema de ponto de mudança no contexto de assimetria. Desenvolvemos um algoritmo EM simples para inferência por máxima verossimilhança dos parâmetros do modelo normal assimétrico com ponto de mudança. O critério de informação de Schwarz é proposto para localizar o ponto de mudança nos parâmetros do modelo normal assimétrico. Além disso, a matriz informação de Fisher é analiticamente derivada, alguns estudos de simulação e aplicações a conjunto de dados reais são finalmente apresentados com o propósito de ilustrar o modelo e os resultados inferenciais desenvolvidos aqui.

**Palavras-chave:** Algoritmo EM; Família de Distribuições Normal Assimétrica; Ponto de Mudança; Critério de Informação de Schwarz; Seleção de Modelos.

# ABSTRACT

The change point problem emerged initially in quality control (PAGE, 1954) and before the introduction of the change point hypothesis associated with models, researchers faced difficulties in establishing a model for some datasets. In this work, we propose to study the change point problem in asymmetry context. We have developed a simple EM algorithm for maximum likelihood inference of the parameters of the skew-normal change point model. The Schwarz information criterion is proposed to locate a change point in the skew-normal model parameters. Moreover, the observed Fisher information matrix is analytically derived and some simulation studies and applications to real data sets are finally presented for the propose of illustrating the model and inferential results developed here.

**Keywords:** EM Algorithm; Skew-Normal Family of Distributions; Change Point; Schwarz Information Criterion; Model Selection

# Lista de ilustrações

Figura 1 – Dados simulados. Gráficos das densidades da normal assimétrica, $\mu = 4$ e $\sigma = 5$ , para diferentes valores do parâmetro $\lambda$ . . . . .	11
Figura 2 – Frequências dos pontos de mudança estimados para os modelos no cenário 1. . . . .	26
Figura 3 – Frequências dos pontos de mudança estimados para os modelos no cenário 2. . . . .	27
Figura 4 – Frequências dos pontos de mudança estimados para os modelos no cenário 3. . . . .	27
Figura 5 – Vazão ( $m^3/s$ ) média anual do Rio Nilo na cidade de Assuã, no Egito, entre os anos de 1871 e 1970. . . . .	28
Figura 6 – Histogramas com as densidades estimadas sobrepostas para os dados do Rio Nilo. . . . .	30
Figura 7 – $SIC(k)$ vs $k$ para os casos da distribuições normal e normal assimétrica. . . . .	30
Figura 8 – Retornos semanais do índice <i>Dow Jones</i> de 1 de julho de 1971 até 2 de agosto de 1974. . . . .	32
Figura 9 – Histogramas com as densidades estimadas sobrepostas para os dados do índice <i>Dow Jones</i> . . . . .	34
Figura 10 – $SIC(k)$ vs $k$ para os casos da distribuições normal e normal assimétrica. . . . .	34
Figura 11 – Frequências dos pontos de mudança estimados para o caso de múltiplos pontos de mudança. . . . .	37

# Lista de tabelas

Tabela 1	– Cenário 1: média e desvio padrão (DP) para o algoritmo EM estimado baseado nas 500 amostras com ponto de mudança no modelo normal assimétrico. EP é a média do erro padrão estimado. . . . .	24
Tabela 2	– Cenário 2: média e desvio padrão (DP) para o algoritmo EM estimado baseado nas 500 amostras com ponto de mudança no modelo normal assimétrico. EP é a média do erro padrão estimado. . . . .	25
Tabela 3	– Cenário 3: média e desvio padrão (DP) para o algoritmo EM estimado baseado nas 500 amostras com ponto de mudança no modelo normal assimétrico. EP é a média do erro padrão estimado. . . . .	25
Tabela 4	– Resultados do procedimento SIC. . . . .	29
Tabela 5	– Estimativas dos parâmetros dos modelos com mudança em $\hat{k} = 28$ . . .	29
Tabela 6	– Estimativas dos parâmetros dos modelos sem o ponto de mudança . . .	29
Tabela 7	– Resultados do procedimento SIC. . . . .	32
Tabela 8	– Estimativas dos parâmetros dos modelos com mudança em $\hat{k} = 89$ . . .	33
Tabela 9	– Estimativas dos parâmetros dos modelos sem ponto de mudança . . .	33
Tabela 10	– Valores esperados e variâncias antes e depois do ponto de mudança $\hat{k} = 89$ . . . . .	33

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>7</b>
<b>1.1</b>	<b>Objetivos</b>	<b>8</b>
<b>1.2</b>	<b>Descrição dos Capítulos</b>	<b>8</b>
<b>2</b>	<b>DISTRIBUIÇÃO NORMAL ASSIMÉTRICA</b>	<b>10</b>
<b>2.1</b>	<b>Definição e Algumas Propriedades</b>	<b>10</b>
<b>2.2</b>	<b>Inferência pelo Método de Máxima Verossimilhança</b>	<b>12</b>
2.2.1	Algoritmo EM	12
<b>3</b>	<b>MODELO NORMAL ASSIMÉTRICO COM PONTO DE MUDANÇA</b>	<b>16</b>
<b>3.1</b>	<b>Descrição do Modelo com Ponto de Mudança</b>	<b>16</b>
<b>3.2</b>	<b>Inferência Estatística</b>	<b>17</b>
3.2.1	Detecção do ponto de mudança nos parâmetros do modelo normal assimétrico	17
3.2.2	Estimação dos Parâmetros via Algoritmo EM	18
3.2.3	Erros Padrão Aproximados	21
<b>4</b>	<b>EXEMPLOS NUMÉRICOS</b>	<b>23</b>
<b>4.1</b>	<b>Estudos de Simulação</b>	<b>23</b>
4.1.1	Experimento I: Estimação via Algoritmo EM	23
4.1.2	Experimento II: Detecção do Ponto de Mudança	26
<b>4.2</b>	<b>Dados Reais</b>	<b>27</b>
4.2.1	Vazão do Rio Nilo	28
4.2.2	Índice <i>Dow Jones</i>	31
<b>5</b>	<b>CONCLUSÕES E PERSPECTIVAS FUTURAS</b>	<b>35</b>
	<b>APÊNDICE A – ESBOÇO PARA MÚLTIPLOS PONTOS DE MUDANÇA</b>	<b>36</b>
	<b>APÊNDICE B – OUTRAS ESTRUTURAS DE MUDANÇA</b>	<b>38</b>
	<b>APÊNDICE C – CÓDIGO DA MATRIZ INFORMAÇÃO OBSERVADA NO SOFTWARE R</b>	<b>39</b>
	<b>REFERÊNCIAS</b>	<b>41</b>

# 1 INTRODUÇÃO

O problema do ponto de mudança tem sido um tema de constante interesse na literatura estatística. Na área de controle de qualidade por exemplo, existem diversos trabalhos que tratam de ponto de mudança como Lai (1995) e Hawkins, Qiu e Kang (2003). O problema de ponto de mudança pode ser encontrado também em outras áreas, tais como economia, finanças, medicina, psicologia, geologia, química, literatura, dentre outras. Do ponto de vista estatístico, o ponto de mudança é um local ou um ponto no tempo onde as observações seguem uma distribuição até esse ponto e após o mesmo seguem outra distribuição.

A importância do ponto de mudança pode ser notada pelo grande número de artigos que vem sendo publicados em vários periódicos. Muitos deles dissertam acerca de diferentes tipos de estruturas de ponto de mudança em relação à média, à variância e à média e variância, simultaneamente, de uma sequência de variáveis aleatórias independentes e normalmente distribuídas. Por exemplo, os problemas de ponto de mudança na média e/ou na variância, no contexto de normalidade dos dados, já foram examinados por Chen e Gupta (1996), Chen e Gupta (1997), Chen e Gupta (1998), Chen e Gupta (1999), Chen e Gupta (2003), Chen e Gupta (2011).

Agora, vamos voltar nossa atenção para além dos modelos gaussianos, e estudar outros importantes modelos. Poucos autores têm considerado o problema do ponto de mudança em outras distribuições, tais como a gamma (KANDER; ZACKS, 1966; HSU, 1979) e a exponencial (WORSLEY, 1986; HACCOU; MEELIS; GEER, 1988; HACCOU; MEELIS, 1988), por exemplo. Quando falamos da modelagem de dados no contexto assimétrico, a escassez de trabalhos sobre ponto de mudança é ainda maior. Quando a distribuição dos dados é assimétrica, esta pode ser modelada por meio de membros da classe de distribuições assimétricas misturas de escala normal assimétrica (BRANCO; DEY, 2001). Nesta classe, a família de distribuições normal assimétrica talvez seja a pioneira. Para nosso conhecimento, até a presente data, apenas os trabalhos de Arellano-Valle, Castro e Loschi (2013) e Ngunkeng e Ning (2014), sob o ponto de vista Bayesiano e frequentista, respectivamente, realizaram um estudo sobre ponto de mudança nos parâmetros do modelo normal assimétrico.

A identificação dos pontos de mudança desempenha um importante papel. Por exemplo, em um processo de produção contínuo, é esperado que a qualidade dos produtos se mantenha estável. Entretanto, por muitas razões, o processo pode falhar na produção de produtos com a mesma qualidade. Portanto, deseja-se encontrar se há um ponto em que a partir dele a qualidade do produto começa a se deteriorar, veja Chen e Gupta (2011) para mais detalhes. No contexto de finanças, veja, por exemplo, os trabalhos de Arellano-Valle, Castro

e Loschi (2013) e Ngunkeng e Ning (2014). Os métodos envolvidos para a identificação dos pontos de mudança são principalmente a (clássica) razão de verossimilhanças, Bayesiano e a abordagem por critérios de informação.

Neste trabalho, abordamos o problema de ponto de mudança, na estrutura da média e da variância simultaneamente, em dados com distribuição normal assimétrica. A família de distribuições normal assimétrica, introduzida por Azzalini (1985), é uma interessante classe de distribuições por conta de sua tratabilidade e flexibilidade em modelar dados reais que apresentam assimetria. Além disso, ela inclui a família normal como um caso especial. Em particular, estudaremos a detecção de ponto de mudança nos parâmetros do modelo normal assimétrico, sob uma visão frequentista. No contexto dos critérios de informação, o critério de informação de Schwarz (SIC) tem sido utilizado por vários autores para a detecção de ponto de mudança sob a suposição de simetria dos dados; veja Chen (1998), Chen e Gupta (2001), Osorio, e Galea (2005) e Contreras (2014) para mais detalhes. Neste trabalho, usaremos o SIC para localizar o ponto de mudança nos parâmetros do modelo normal assimétrico assim como em Ngunkeng e Ning (2014). Porém, o enfoque do trabalho é voltado para a estimação dos parâmetros do modelo via algoritmo EM. O algoritmo EM tem diversas vantagens sobre a maximização direta da função verossimilhança uma vez que é facilmente implementável, numericamente estável e bastante acurado.

## 1.1 Objetivos

Neste trabalho, fornecemos alguns resultados adicionais para o problema de ponto de mudança no contexto de assimetria, em particular para o modelo normal assimétrico. O enfoque do trabalho é voltado para a estimação dos parâmetros do modelo via algoritmo EM. Podemos então relacionar os seguintes objetivos específicos: (i) detectar o ponto de mudança nos parâmetros do modelo normal assimétrico usando o critério de informação SIC; (ii) implementar e avaliar o algoritmo EM proposto computacionalmente; e (iii) aplicar esses resultados para analisar dados reais. Acreditamos que os resultados deste trabalho são um suplemento necessário de Arellano-Valle, Castro e Loschi (2013) e Ngunkeng e Ning (2014).

## 1.2 Descrição dos Capítulos

O restante do trabalho está organizado como segue. No Capítulo 2, apresentamos uma breve discussão sobre a família de distribuições normal assimétrica e algumas de suas propriedades. Especificamente, o algoritmo EM para estimação por máxima verossimilhança dos parâmetros do modelo normal assimétrico é apresentado na Seção 2.2.1. No Capítulo 3, abordamos o problema de ponto de mudança, na estrutura da média e da variância simultaneamente, em dados com distribuição normal assimétrica que são

observados sequencialmente. O algoritmo EM para estimação por máxima verossimilhança dos parâmetros do modelo normal assimétrico com ponto de mudança é dado na Seção 3.2 e discussões sobre a detecção do ponto de mudança e sobre a matriz informação observada também são apresentadas. No Capítulo 4, alguns estudos de simulação e aplicações a conjunto de dados reais são apresentados com o propósito de ilustrar o modelo e os resultados inferenciais desenvolvidos. O Capítulo 5 apresenta algumas conclusões e perspectivas futuras a serem exploradas na área. Alguns avanços podem ser vistos nos Apêndices A e B. O código no *software* R da matriz informação observada para a distribuição normal assimétrica presente no Apêndice C.

## 2 DISTRIBUIÇÃO NORMAL ASSIMÉTRICA

Em muitas análises na literatura estatística, existe uma tendência geral de se supor a normalidade dos dados, algo que nem sempre é o mais adequado. Por conta disso, nos últimos anos o estudo de modelos mais flexíveis ao gaussiano tem sido de grande interesse dos pesquisadores. Sob essa motivação é que será proposto, neste trabalho, um estudo baseado na distribuição normal assimétrica (em inglês, “skew-normal”). Esta distribuição inclui a distribuição normal, como caso especial, e fornece flexibilidade em capturar uma ampla variedade de comportamentos não normais, por simplesmente adicionar um parâmetro que controla o grau de assimetria. Neste capítulo, será apresentada uma revisão sobre algumas das propriedades e resultados referentes à distribuição normal assimétrica localização-escala univariada. É vasta a quantidade de trabalhos relevantes sobre a distribuição normal assimétrica na literatura, alguns deles encontram-se citados neste trabalho.

### 2.1 Definição e Algumas Propriedades

Algumas versões, extensões e unificações da distribuição normal assimétrica podem ser encontradas em Azzalini (1985), Azzalini e Capitanio (1999), Azzalini (2005), Arellano-Valle, Branco e Genton (2006), Arellano-Valle e Azzalini (2006) e Arellano-Valle e Genton (2010), por exemplo. Para os propósitos deste trabalho, diremos que uma variável aleatória  $Z$  segue uma distribuição normal assimétrica padrão com parâmetro de forma  $\lambda$  ( $\lambda \in \mathbb{R}$ ), e denotaremos por  $Z \sim \text{SN}(\lambda)$ , se sua densidade é dada por

$$f(z|\lambda) = 2\phi(z)\Phi(\lambda z), \quad z \in \mathbb{R}, \quad (1)$$

onde  $\phi(\cdot)$  e  $\Phi(\cdot)$  representam a função densidade e função distribuição acumulada da distribuição normal padrão. Parâmetros que regulam a localização e a dispersão podem ser introduzidos considerando a transformação  $Y = \mu + \sigma Z$ , onde  $\mu \in \mathbb{R}$  e  $\sigma > 0$ . Dessa forma, segue abaixo a definição da distribuição normal assimétrica localização-escala.

**Definição 1.** *Uma variável aleatória  $Y$  segue distribuição normal assimétrica com parâmetro de localização  $\mu$ , parâmetro de dispersão  $\sigma$  e parâmetro de forma  $\lambda$ , e denotaremos por  $Y \sim \text{SN}(\mu, \sigma^2, \lambda)$ , se sua função densidade é dada por*

$$f(y|\mu, \sigma^2, \lambda) = \frac{2}{\sigma} \phi\left(\frac{y - \mu}{\sigma}\right) \Phi\left(\lambda \frac{y - \mu}{\sigma}\right), \quad y \in \mathbb{R}. \quad (2)$$

É importante notar que o parâmetro de forma  $\lambda$  caracteriza a assimetria da densidade, sendo  $\lambda > 0$  assimetria positiva,  $\lambda < 0$  assimetria negativa e  $\lambda = 0$  simetria, coincidindo

com a densidade da normal. A Figura 2.1 apresenta alguns exemplos de densidades da normal assimétrica para diferentes valores do parâmetro  $\lambda$ .

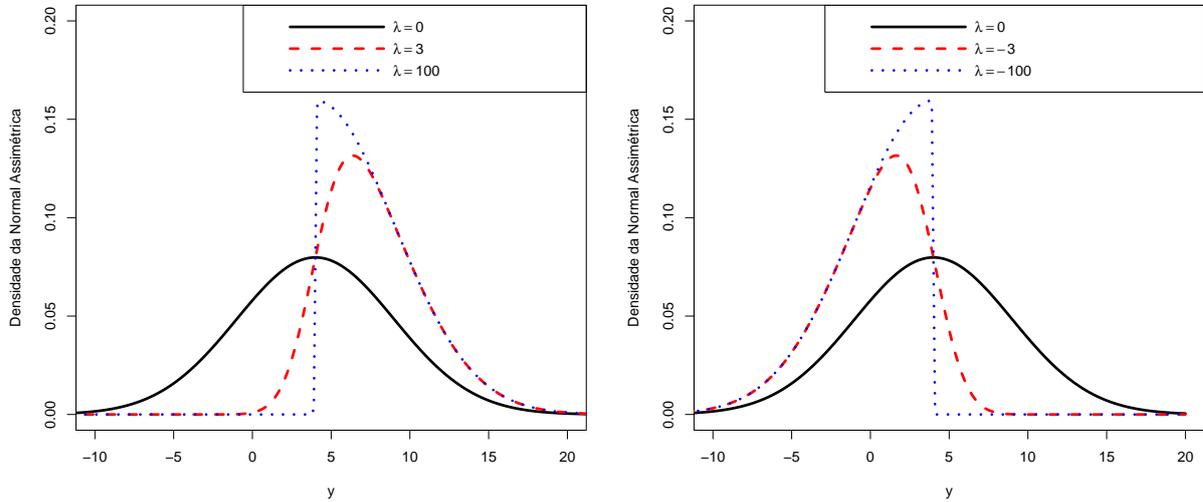


Figura 1 – Dados simulados. Gráficos das densidades da normal assimétrica,  $\mu = 4$  e  $\sigma = 5$ , para diferentes valores do parâmetro  $\lambda$ .

A variável aleatória  $Y$ , com função de densidade em (2), tem representação estocástica dada por

$$Y \stackrel{d}{=} \mu + \Delta T + \gamma T_1, \quad (3)$$

onde “ d ” significa equivalência em distribuição,  $\Delta = \sigma\delta$ ,  $\delta = \frac{\lambda}{\sqrt{1 + \lambda^2}}$ ,  $\gamma^2 = \sigma^2 - \Delta^2$ ,  $T = |T_0|$ ,  $|T_0|$  denota o valor absoluto de  $T_0$ ,  $T_0 \sim N(0, 1)$  e  $T_1 \sim N(0, 1)$  são todas variáveis independentes. A partir desta representação, podemos gerar números pseudo-aleatórios de uma distribuição normal assimétrica. Além disso, também podemos obter a média e a variância de uma variável aleatória  $Y \sim SN(\mu, \sigma^2, \lambda)$ , dadas por

$$E(Y) = \mu + \sqrt{\frac{2}{\pi}}\Delta \quad \text{e} \quad Var(Y) = \sigma^2 \left(1 - \frac{2}{\pi}\delta^2\right). \quad (4)$$

Note que, quando  $\lambda = 0$  e conseqüentemente  $\delta = 0$  ou  $\Delta = 0$ , temos a média e variância da distribuição normal (caso particular).

Antes de apresentar a representação hierárquica da normal assimétrica, será definida a distribuição normal truncada ( $NT$ ). Dizemos que  $T$  tem uma distribuição  $NT$  no intervalo  $(a, b)$  com parâmetros  $\mu$  e  $\sigma^2$ , denotada por  $T \sim NT(\mu, \sigma^2, (a, b))$ , quando  $T$  tiver a mesma distribuição de  $X|(a < X < b)$  para alguma variável aleatória  $X \sim N(\mu, \sigma^2)$ . Se  $a = 0, b = \infty, \mu = 0$  e  $\sigma^2 = 1$ , então a chamamos de half-normal padrão ( $HN(0, 1)$ ). Proveniente da representação estocástica (3), a distribuição normal assimétrica pode ser

reescreta hierarquicamente como segue

$$\begin{aligned} Y|T = t &\sim N(\mu + \Delta t, \gamma^2), \\ T &\sim \text{HN}(0, 1). \end{aligned} \quad (5)$$

Esta representação hierárquica é útil para realizar inferência estatística nos parâmetros do modelo normal assimétrico. Em particular, a representação em (5) facilita a implementação do algoritmo EM para obtenção dos estimadores de máxima verossimilhança dos parâmetros do modelo.

## 2.2 Inferência pelo Método de Máxima Verossimilhança

Proveniente da função densidade dada em (2), a função de log-verossimilhança para  $\boldsymbol{\theta} = (\mu, \sigma, \lambda)^\top$ , dada uma amostra  $Y_1, \dots, Y_n$ , é definida por

$$\ell(\boldsymbol{\theta}) = n \ln 2 - n \ln \sigma - \frac{n}{2} \ln 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 + \sum_{i=1}^n \ln \Phi \left( \lambda \frac{(y_i - \mu)}{\sigma} \right). \quad (6)$$

A função escore pode ser obtida derivando-se a função log-verossimilhança, com respeito a cada um dos parâmetros desconhecidos. As equações seguem abaixo.

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) - \frac{\lambda}{\sigma} \sum_{i=1}^n \frac{\phi \left( \lambda \frac{(y_i - \mu)}{\sigma} \right)}{\Phi \left( \lambda \frac{(y_i - \mu)}{\sigma} \right)}; \quad (7)$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (y_i - \mu)^2 - \frac{1}{\sigma^2} \sum_{i=1}^n \lambda (y_i - \mu) \left[ \frac{\phi \left( \lambda \frac{(y_i - \mu)}{\sigma} \right)}{\Phi \left( \lambda \frac{(y_i - \mu)}{\sigma} \right)} \right]; \quad (8)$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \lambda} = \sum_{i=1}^n \left( \frac{y_i - \mu}{\sigma} \right) \left[ \frac{\phi \left( \lambda \frac{(y_i - \mu)}{\sigma} \right)}{\Phi \left( \lambda \frac{(y_i - \mu)}{\sigma} \right)} \right]. \quad (9)$$

Pelas equações (7)-(9), nota-se que não existem soluções explícitas para o problema de maximização da função de log-verossimilhança dada em (6). Nesse trabalho, será utilizado o algoritmo EM para o cálculo dos estimadores de máxima verossimilhança nos valendo da representação hierárquica dada em (5) e introduzindo uma reparametrização de modo a obtermos formas fechadas para o algoritmo EM no passo M. Outras abordagens podem ser empregadas, tais como os métodos de Newton-Raphson, de Escore de Fisher ou também uma análise bayesiana (RODRÍGUEZ, 2005), por exemplo. A aplicação do algoritmo EM foi feita através do pacote *sn* no *software* R versão 2.11.1.

### 2.2.1 Algoritmo EM

O algoritmo EM (DEMPSTER; LAIRD; RUBIN, 1977) é um enfoque aplicado no cálculo iterativo de estimativas dos parâmetros do modelo ( $\boldsymbol{\theta}$ ) pelo método de máxima

verossimilhança, sendo bastante útil para problemas com dados incompletos. Muitos problemas em estatística podem ser considerados utilizando uma formulação de dados aumentados permitindo assim simplificar a obtenção das estimativas de máxima verossimilhança. Os dados aumentados, também chamados dados completos, correspondem aos dados observados ( $y$ ), referidos nesta formulação como dados incompletos, e dados adicionais conhecidos como dados perdidos (*missing*) ou não observáveis ( $t$ ), obtendo-se, deste modo, o vetor de dados completos  $\mathbf{y}_c = (\mathbf{y}, \mathbf{t})$ , de tal forma que a função de log-verossimilhança é representada por  $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)$ . Cada iteração do algoritmo EM consiste em dois passos:

- **Passo E (Esperança):** Calcule  $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = E(\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)})$ , onde  $\hat{\boldsymbol{\theta}}^{(k)}$  representa a estimativa de  $\boldsymbol{\theta}$  na  $k$ -ésima iteração;
- **Passo M (Maximização):** Encontre  $\hat{\boldsymbol{\theta}}^{(k+1)}$  que maximiza  $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)})$ .

Estes passos se alternam até convergir e geralmente, por exemplo, é considerado como critério de convergência  $\|\hat{\boldsymbol{\theta}}^{(k+1)} - \hat{\boldsymbol{\theta}}^{(k)}\| < \epsilon$ , onde  $\epsilon = 10^{-5}$ .

Em seguida, descrevemos o algoritmo EM para a obtenção das estimativas de máxima verossimilhança dos parâmetros do modelo normal assimétrico. Tais resultados uma vez compreendidos nesta seção, torna mais fácil o entendimento do processo de estimação dos parâmetros do modelo normal assimétrico no contexto de ponto de mudança.

Neste processo de estimação, considere  $\mathbf{y} = (y_1, \dots, y_n)^\top$  o vetor de respostas observáveis para  $n$  unidades amostrais e  $\mathbf{t} = (t_1, \dots, t_n)^\top$ . Então, utilizando a representação hierárquica dada em (5) segue que a função log-verossimilhança completa associada com  $\mathbf{y}_c = (\mathbf{y}, \mathbf{t})^\top$  é dada por

$$\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = c - \frac{n}{2} \ln \gamma^2 - \frac{1}{2\gamma^2} \sum_{i=1}^n (y_i - \mu - \Delta t_i)^2, \quad (10)$$

onde  $c$  é uma constante que não depende de  $\boldsymbol{\theta}$ .

### Passo E:

No passo E do algoritmo EM, toma-se o valor esperado da log-verossimilhança completa dada em (10) condicionada no vetor de dados observados e no  $\boldsymbol{\theta}$  em estado corrente. Assim, o valor esperado condicional da log-verossimilhança completa é

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = c - \frac{n}{2} \ln \gamma^2 - \frac{1}{2\gamma^2} \sum_{i=1}^n (y_i - \mu)^2 + \frac{1}{\gamma^2} \sum_{i=1}^n \hat{t}_i^{(k)} (y_i - \mu) \Delta \quad (11)$$

$$- \frac{1}{2\gamma^2} \sum_{i=1}^n \hat{t}_i^{(k)2} \Delta^2. \quad (12)$$

Note que as seguintes quantidades devem ser obtidas

$$\hat{t}_i^{(k)} = E(T_i|y_i, \hat{\boldsymbol{\theta}}^{(k)}) \quad (13)$$

e

$$\hat{t}_i^2{}^{(k)} = E(T_i^2|y_i, \hat{\boldsymbol{\theta}}^{(k)}). \quad (14)$$

Para o cálculo das quantidades (13) e (14), é necessário obter a distribuição condicional  $T_i|Y_i = y_i$ . A obtenção de tal distribuição condicional e consequentemente de seus momentos é baseada nos Lemas 1 e 2 a seguir, cujas demonstrações podem ser encontradas em Lachos (2004), páginas 22 e 44, respectivamente.

**Lema 1.** *Seja  $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  e  $\mathbf{X} \sim N_q(\boldsymbol{\eta}, \boldsymbol{\Omega})$ , então*

$$\begin{aligned} \phi_p(\mathbf{y}|\boldsymbol{\mu} + \boldsymbol{\Psi}\mathbf{x}, \boldsymbol{\Sigma})\phi_q(\mathbf{x}|\boldsymbol{\eta}, \boldsymbol{\Omega}) &= \phi_p(\mathbf{y}|\boldsymbol{\mu} + \boldsymbol{\Psi}\boldsymbol{\eta}, \boldsymbol{\Sigma} + \boldsymbol{\Psi}\boldsymbol{\Omega}\boldsymbol{\Psi}^\top) \\ &\times \phi_q(\mathbf{x}|\boldsymbol{\eta}, \boldsymbol{\Lambda}\boldsymbol{\Psi}^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu} - \boldsymbol{\Psi}\boldsymbol{\eta}), \boldsymbol{\Lambda}), \end{aligned} \quad (15)$$

onde  $\boldsymbol{\Lambda} = (\boldsymbol{\Omega}^{-1} + \boldsymbol{\Psi}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi})^{-1}$  e  $\phi_k(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma})$  é a função densidade da normal  $k$ -variada com vetor de médias  $\boldsymbol{\mu}$  e matriz de covariâncias  $\boldsymbol{\Sigma}$ .  $\square$

Tem-se que

$$f(y_i, t_i|\boldsymbol{\theta}) = f(y_i|\boldsymbol{\theta})f(t_i|y_i, \boldsymbol{\theta}). \quad (16)$$

Portanto, da representação hierárquica em (5)

$$f(y_i, t_i|\boldsymbol{\theta}) = 2\phi(y_i|\mu, \gamma^2 + \Delta^2)\phi\left(t_i\left|\frac{\Delta}{\gamma^2 + \Delta^2}(y_i - \mu), \frac{\gamma^2}{\gamma^2 + \Delta^2}\right.\right)I_{(0, +\infty)}(t_i), \quad (17)$$

sendo que  $I_A(\cdot)$  é a função indicadora de um conjunto  $A$  e por convenção,  $\phi(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \phi_1(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Considerando  $\mu_{T_i} = \frac{\Delta}{\gamma^2 + \Delta^2}(y_i - \mu)$  e  $M_T^2 = \frac{\gamma^2}{\gamma^2 + \Delta^2}$ , tem-se que

$$T_i|Y_i = y_i \sim NT(\mu_{T_i}, M_T^2, (0; +\infty)). \quad (18)$$

O próximo resultado está relacionado com as propriedades da distribuição normal truncada e será útil para o cálculo dos valores esperados (13) e (14).

**Lema 2.** *Seja  $X \sim N(\eta, \tau)$ . Então, para  $b$  uma constante real, segue que*

$$\begin{aligned} E(X|X > b) &= \eta + \frac{\phi\left(\frac{b - \eta}{\tau}\right)}{1 - \Phi\left(\frac{b - \eta}{\tau}\right)}\tau; \\ E(X^2|X > b) &= \eta^2 + \tau^2 + \frac{\phi\left(\frac{b - \eta}{\tau}\right)}{1 - \Phi\left(\frac{b - \eta}{\tau}\right)}(\eta + b)\tau. \quad \square \end{aligned}$$

Assim, temos que

$$\hat{t}_i^{(k)} = \hat{\mu}_{T_i}^{(k)} + \widehat{M}_T^{(k)} \hat{\tau}_{T_i}^{(k)}; \quad (19)$$

$$\hat{t}_i^2{}^{(k)} = \widehat{\mu}_{T_i}^2{}^{(k)} + \widehat{M}_T^2{}^{(k)} + \widehat{M}_T^{(k)} \hat{\mu}_{T_i}^{(k)} \hat{\tau}_{T_i}^{(k)}, \quad (20)$$

onde  $\hat{\tau}_{T_i}^{(k)} = W_\phi \left( \frac{\hat{\mu}_{T_i}^{(k)}}{\widehat{M}_T^{(k)}} \right)$ , sendo  $W_\phi = \frac{\phi(\cdot)}{\Phi(\cdot)}$ .

### Passo M:

Atualizar  $\hat{\theta}^{(k)}$ , maximizando  $Q(\theta|\hat{\theta}^{(k)})$  sobre  $\theta$ , que resulta nas seguintes expressões fechadas:

$$\hat{\mu}^{(k+1)} = \sum_{i=1}^n \frac{(y_i - \hat{\Delta}^{(k)} \hat{t}_i^{(k)})}{n}, \quad (21)$$

$$\hat{\Delta}^{(k+1)} = \frac{\sum_{i=1}^n \hat{t}_i^{(k)} (y_i - \hat{\mu}^{(k)})}{\sum_{i=1}^n \hat{t}_i^2{}^{(k)}}, \quad (22)$$

$$\hat{\gamma}^2{}^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \left[ (y_i - \hat{\mu}^{(k)})^2 - 2\hat{\Delta}^{(k)} \hat{t}_i^{(k)} (y_i - \hat{\mu}^{(k)}) + \hat{\Delta}^2{}^{(k)} \hat{t}_i^2{}^{(k)} \right]. \quad (23)$$

Note que as estimativas para  $\sigma$  e  $\lambda$  podem ser obtidas de (22) e (23), usando a propriedade de invariância dos estimadores de máxima verossimilhança. Dessa forma,

$$\hat{\sigma}^{(k+1)} = \sqrt{\hat{\Delta}^2{}^{(k)} + \hat{\gamma}^2{}^{(k)}}, \quad \hat{\lambda}^{(k+1)} = \frac{\hat{\Delta}^{(k)}}{\sqrt{\hat{\gamma}^2{}^{(k)}}}. \quad (24)$$

As interações são repetidas até que uma regra de convergência adequada seja satisfeita. Valores iniciais são necessários para implementar esse algoritmo. Entretanto, por conta da reparametrização e com a finalidade de verificar que a estimativa de máxima verossimilhança foi encontrada, recomenda-se rodar o algoritmo EM usando uma amplitude de diferentes valores iniciais. Note que, quando  $\lambda = 0$  (ou  $\delta = 0$ ), as equações do passo M se reduzem às equações quando assumimos a distribuição normal.

# 3 MODELO NORMAL ASSIMÉTRICO COM PONTO DE MUDANÇA

Em geral, os modelos de ponto de mudança podem ser divididos em dois grupos. Modelos com uma mudança descontínua no ponto de mudança (indicados como modelos de ponto de mudança descontínua) e modelos com uma mudança contínua no ponto de mudança (indicados como modelos de ponto de mudança contínua). Independentemente desta classificação, os modelos de ponto de mudança envolvem três questões: primeiro, a escolha da localização dos pontos de mudança. Em segundo lugar, a escolha de um número adequado de pontos de mudança e em terceiro lugar, a escolha de um modelo adequado para cada subconjunto de dados. Neste trabalho, iremos considerar o modelo normal assimétrico com um ponto de mudança descontínua.

## 3.1 Descrição do Modelo com Ponto de Mudança

Nesta seção, consideramos os resultados anteriormente introduzidos para descrever um ponto de mudança nos parâmetros de locação, escala e forma quando os dados são sequencialmente observados e seguem uma distribuição normal assimétrica. Na literatura estatística, existem poucos estudos aplicados considerando ponto de mudança no contexto de assimetria, que é o nosso objetivo.

Seja  $Y_1, \dots, Y_n$  uma sequência de observações obtidas em uma situação prática e suponha que cada

$$Y_i \sim \text{SN}(\mu_i, \sigma_i^2, \lambda_i), \quad i = 1, \dots, n.$$

Geralmente, assume-se que o mesmo modelo é válido para todo o conjunto de dados, mas isso nem sempre é o caso. Em contraste, o modelo pode mudar depois de um ponto específico, que pode ou não ser conhecido. Na literatura estatística, este ponto é conhecido como ponto de mudança (CHEN; GUPTA, 2011). Os dados devem ter alguma ordenação natural para ser capaz de encontrar tal ponto de mudança. Esta ordem pode ser determinada por alguma ordenação natural para ser capaz de encontrar tal ponto de mudança, por exemplo, o tempo de observação. A alteração na estrutura pode ser na média, na variância, ou em ambas. O interesse aqui é sobre mudança nos parâmetros da distribuição normal assimétrica, podendo acarretar mudança na estrutura da média e/ou da variância.

Portanto, o seguinte modelo normal assimétrico com ponto de mudança é interessante investigar. Isto é,

$$Y_i \sim \text{SN}(\mu_1, \sigma_1^2, \lambda_1), \quad i = 1, \dots, k \quad \text{e} \quad (25)$$

$$Y_i \sim \text{SN}(\mu_2, \sigma_2^2, \lambda_2) \quad i = k + 1, \dots, n, \quad (26)$$

onde  $k$  é a posição desconhecida do ponto de mudança.

Note que quando  $k = n$ , o modelo não apresenta qualquer mudança nos parâmetros do modelo normal assimétrico. Por outro lado, se  $k < n$  o modelo apresenta então um ponto de mudança na posição  $k$ , tal que  $3 \leq k \leq n - 3$ . No modelo definido em (25)-(26),  $p = 3$ . Neste contexto, a detecção do ponto de mudança pode ser considerada como um procedimento de seleção de modelos.

## 3.2 Inferência Estatística

Nesta seção, vamos discutir alguns aspectos inferenciais do modelo normal assimétrico com ponto de mudança. Discussões sobre a detecção do ponto de mudança são dadas na Seção 3.2.1. Nas seções seguintes, o algoritmo EM (DEMPSTER; LAIRD; RUBIN, 1977) para obter o estimador de máxima verossimilhança de  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ , onde  $\boldsymbol{\theta}_j = (\mu_j, \sigma_j, \lambda_j)^\top$ ,  $j = 1, 2$ , e a estimação do erro padrão são também apresentados. Todos os procedimentos aqui propostos podem ser facilmente implementados em *softwares* estatísticos, tal como o *software* R (R Development Core Team, 2011). O código R pode ser disponibilizado, mediante solicitação.

### 3.2.1 Detecção do ponto de mudança nos parâmetros do modelo normal assimétrico

Os métodos envolvidos para a identificação dos pontos de mudança são principalmente a (clássica) razão de verossimilhanças, Bayesiano e a abordagem por critérios de informação. Em particular, estudaremos a detecção de ponto de mudança nos parâmetros do modelo normal assimétrico, sob uma visão frequentista. No contexto dos critérios de informação, o critério de informação de Schwarz (SIC) tem sido utilizado por vários autores para a detecção de ponto de mudança sob a suposição de simetria dos dados; veja Chen (1998), Chen e Gupta (2001), Osorio, e Galea (2005) e Contreras (2014) para mais detalhes. Neste trabalho, usaremos o SIC para localizar o ponto de mudança nos parâmetros do modelo normal assimétrico assim como em Ngunkeng e Ning (2014). Porém, o enfoque do trabalho é voltado para a estimação dos parâmetros do modelo via algoritmo EM.

O critério de informação SIC (SCHWARZ, 1978) é definido por

$$SIC = -2\ell(\hat{\boldsymbol{\theta}}) + p^* \log(n), \quad (27)$$

onde  $\ell(\hat{\boldsymbol{\theta}})$  corresponde à log-verossimilhança avaliada na estimativa de máxima verossimilhança dos parâmetros,  $\hat{\boldsymbol{\theta}}$ ,  $p^*$  é o número de parâmetros do modelo e  $n$  é o tamanho amostral. Vale ressaltar que para o modelo normal assimétrico sem ponto de mudança, o SIC será denotado por  $SIC(n)$  e  $p^* = 3$ . Já para o modelo normal assimétrico com

ponto de mudança, o SIC será denotado por  $SIC(k)$  e  $p^* = 6$ . Note que maximizar a log-verossimilhança é equivalente à minimizar o SIC. O modelo que minimiza o SIC é considerado o que melhor se ajusta aos dados.

O critério de informação de Schwarz sob o modelo normal assimétrico sem ponto de mudança, isto é, para  $k = n$ , denotado por  $SIC(n)$ , é dado por:

$$\begin{aligned} SIC(n) &= -2\ell(\hat{\boldsymbol{\theta}}) + 3 \log(n), \\ &= -2n \log(2) + n \log(2\pi) - 2 \sum_{i=1}^n b_i(\hat{\boldsymbol{\theta}}) + 3 \log(n), \end{aligned} \quad (28)$$

onde  $b_i(\hat{\boldsymbol{\theta}}) = -\frac{1}{2} \log(\hat{\sigma}^2) + \log(\hat{\kappa}_i)$ , com  $\hat{\kappa}_i = \exp\left\{-\frac{(y_i - \hat{\mu})^2}{2\hat{\sigma}^2}\right\} \Phi\left(\hat{\lambda}\left(\frac{y_i - \hat{\mu}}{\hat{\sigma}}\right)\right)$ .

Analogamente, o critério de informação de Schwarz no modelo normal assimétrico com ponto de mudança, isto é, para  $k < n$ , denotado por  $SIC(k)$ , é dado por:

$$\begin{aligned} SIC(k) &= -2\ell(\hat{\boldsymbol{\theta}}) + 2 \times 6 \log(n), \\ &= -2n \log(2) + n \log(2\pi) - 2 \sum_{i=1}^n b_i(\hat{\boldsymbol{\theta}}) + 2 \times 6 \log(n), \end{aligned} \quad (29)$$

onde

$$b_i(\hat{\boldsymbol{\theta}}) = \begin{cases} -\frac{1}{2} \log(\hat{\sigma}_1^2) + \log(\hat{\kappa}_i), & i = 1, \dots, k, \\ -\frac{1}{2} \log(\hat{\sigma}_2^2) + \log(\hat{\kappa}_i), & i = k + 1, \dots, n, \end{cases} \quad (30)$$

com

$$\hat{\kappa}_i = \begin{cases} \exp\left\{-\frac{(y_i - \hat{\mu}_1)^2}{2\hat{\sigma}_1^2}\right\} \Phi\left(\hat{\lambda}_1\left(\frac{y_i - \hat{\mu}_1}{\hat{\sigma}_1}\right)\right), & i = 1, \dots, k, \\ \exp\left\{-\frac{(y_i - \hat{\mu}_2)^2}{2\hat{\sigma}_2^2}\right\} \Phi\left(\hat{\lambda}_2\left(\frac{y_i - \hat{\mu}_2}{\hat{\sigma}_2}\right)\right), & i = k + 1, \dots, n. \end{cases} \quad (31)$$

De acordo com o princípio do critério de informação, estimamos a posição do ponto de mudança  $k$  por  $\hat{k}$ , tal que  $SIC(\hat{k})$  seja minimizado. Então, o estimador de máxima verossimilhança do ponto de mudança denotado por  $\hat{k}$ , deve satisfazer,

$$SIC(\hat{k}) = \{\min SIC(k) : 3 \leq k \leq n - 3\},$$

tal que caso  $SIC(\hat{k}) < SIC(n)$ , o modelo com ponto de mudança em  $\hat{k}$  será o escolhido, caso contrário, o modelo sem ponto de mudança é considerado o que melhor se ajusta aos dados.

### 3.2.2 Estimação dos Parâmetros via Algoritmo EM

Nesta subseção, derivamos um algoritmo EM para obter o estimador de máxima verossimilhança de  $\boldsymbol{\theta}$ . A estimação dos parâmetros relacionados com o modelo definido em (25)-(26), com uma mudança numa dada posição  $k$ , é um processo iterativo, tal como descrito abaixo.

Para explorar o algoritmo EM, apresentamos o modelo normal assimétrico com ponto de mudança no contexto de dados incompletos, utilizando os resultados apresentados no Capítulo 2. Assim, a partir de (3), o modelo normal assimétrico com ponto de mudança pode ser reescrito hierarquicamente do seguinte modo

$$Y_i|T_i = t_i \sim N(\mu_1 + \Delta_1 t_i, \gamma_1^2), \quad i = 1, \dots, k, \quad (32)$$

$$Y_i|T_i = t_i \sim N(\mu_2 + \Delta_2 t_i, \gamma_2^2), \quad i = k + 1, \dots, n, \quad (33)$$

$$T_i \sim \text{HN}(0, 1), \quad (34)$$

para  $i = 1, \dots, n$ , todas independentes, onde  $\Delta_j = \sigma_j \delta_j$ ,  $\delta_j = \frac{\lambda_j}{\sqrt{1 + \lambda_j^2}}$  e  $\gamma_j^2 = \sigma_j^2 - \Delta_j^2$ ,  $j = 1, 2$ . Seja  $\mathbf{y} = (y_1, \dots, y_n)^\top$  e  $\mathbf{t} = (t_1, \dots, t_n)^\top$ . Então, sob a representação hierárquica (32)-(34), segue que a função log-verossimilhança completa associada com  $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{t}^\top)^\top$  é

$$\begin{aligned} \ell_c(\boldsymbol{\theta}|\mathbf{y}_c) &= c - \frac{k}{2} \log(\gamma_1^2) - \frac{1}{2\gamma_1^2} \sum_{i=1}^k (y_i - \mu_1 - \Delta_1 t_i)^2 \\ &\quad - \frac{(n-k)}{2} \log(\gamma_2^2) - \frac{1}{2\gamma_2^2} \sum_{i=k+1}^n (y_i - \mu_2 - \Delta_2 t_i)^2, \end{aligned}$$

onde  $c$  é uma constante que independe do vetor de parâmetros  $\boldsymbol{\theta}$ . Considerando  $\hat{\boldsymbol{\theta}}_j^{(h)} = (\hat{\mu}_j^{(h)}, \hat{\sigma}_j^{(h)}, \hat{\lambda}_j^{(h)})^\top$ , a estimativa de  $\boldsymbol{\theta}$  na  $h$ -ésima iteração. Segue, após algumas álgebras simples, que a esperança condicional da função log-verossimilhança completa tem a forma

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(h)}) = c + \sum_{i=1}^k Q_{i1}(\boldsymbol{\theta}_1|\hat{\boldsymbol{\theta}}^{(h)}) + \sum_{i=k+1}^n Q_{i2}(\boldsymbol{\theta}_2|\hat{\boldsymbol{\theta}}^{(h)}), \quad (35)$$

onde

$$Q_{i1}(\boldsymbol{\theta}_1|\hat{\boldsymbol{\theta}}^{(h)}) = -\frac{1}{2} \log(\gamma_1^2) - \frac{1}{2\gamma_1^2} \left( (y_i - \mu_1)^2 - 2\hat{t}_i^{(h)}(y_i - \mu_1)\Delta_1 + \hat{t}_i^{2(h)} \Delta_1^2 \right),$$

para  $i = 1, \dots, k$  e

$$Q_{i2}(\boldsymbol{\theta}_2|\hat{\boldsymbol{\theta}}^{(h)}) = -\frac{1}{2} \log(\gamma_2^2) - \frac{1}{2\gamma_2^2} \left( (y_i - \mu_2)^2 - 2\hat{t}_i^{(h)}(y_i - \mu_2)\Delta_2 + \hat{t}_i^{2(h)} \Delta_2^2 \right),$$

para  $i = k + 1, \dots, n$ , onde  $\hat{t}_i^{(h)} = E[T_i|y_i, \hat{\boldsymbol{\theta}}^{(h)}]$  e  $\hat{t}_i^{2(h)} = E[T_i^2|y_i, \hat{\boldsymbol{\theta}}^{(h)}]$ . Usando propriedades conhecidas da esperança condicional, obtemos

$$\hat{t}_i^{(h)} = \begin{cases} \hat{m}_{i1}^{(h)} + \hat{M}_1^{(h)} \hat{\tau}_{i1}^{(h)}, & i = 1, \dots, k, \\ \hat{m}_{i2}^{(h)} + \hat{M}_2^{(h)} \hat{\tau}_{i2}^{(h)}, & i = k + 1, \dots, n \end{cases} \quad (36)$$

e

$$\hat{t}_i^{2(h)} = \begin{cases} \hat{m}_{i1}^{2(h)} + \hat{M}_1^{2(h)} + \hat{M}_1^{(h)} \hat{m}_{i1}^{(h)} \hat{\tau}_{i1}^{(h)}, & i = 1, \dots, k, \\ \hat{m}_{i2}^{2(h)} + \hat{M}_2^{2(h)} + \hat{M}_2^{(h)} \hat{m}_{i2}^{(h)} \hat{\tau}_{i2}^{(h)}, & i = k + 1, \dots, n, \end{cases} \quad (37)$$

com  $\widehat{m}_{ij} = \frac{\widehat{\Delta}_j}{\widehat{\Delta}_j^2 + \widehat{\gamma}_j^2} (y_i - \mu_j)$ ,  $\widehat{M}_j^2 = \frac{\widehat{\gamma}_j^2}{\widehat{\Delta}_j^2 + \widehat{\gamma}_j^2}$  e  $\widehat{\tau}_{ij} = W_\Phi \left( \frac{\widehat{m}_{ij}}{\widehat{M}_j} \right)$ ,  $j = 1, 2$ , tal que  $W_{\Phi(\cdot)} = \frac{\phi(\cdot)}{\Phi(\cdot)}$ , com todas essas quantidades avaliadas em  $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(h)}$ . As demonstrações desses resultados podem ser encontradas em Zeller, Lachos e Vilca-Labra (2011).

O primeiro termo do lado direito de (35) é a esperança condicional da função log-verossimilhança completa das primeiras  $k$  observações e o segundo termo é a esperança condicional da função log-verossimilhança completa das últimas  $n - k$  observações. Para um dado  $k$ , ambos os termos são independentes entre si. Assim, as estimativas de máxima verossimilhança de  $\boldsymbol{\theta}_1$  e  $\boldsymbol{\theta}_2$  são as estimativas de máxima verossimilhança obtidas através de dois modelos separados. Dessa forma, temos o seguinte algoritmo EM:

**Passo E:** Dado  $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(h)}$ , calcule  $\widehat{t}_i^{(h)}$  e  $\widehat{t}_i^{2(h)}$ , para  $i = 1, \dots, n$ , usando (36)-(37).

**Passo M:** Atualize  $\widehat{\boldsymbol{\theta}}^{(h+1)}$  maximizando  $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(h)})$  sobre  $\boldsymbol{\theta}$  e então, obtemos as seguintes expressões fechadas

$$\begin{aligned} \widehat{\mu}_1^{(h+1)} &= \sum_{i=1}^k \frac{(y_i - \widehat{\Delta}_1^{(h)} \widehat{t}_i^{(h)})}{k}, & \widehat{\mu}_2^{(h+1)} &= \sum_{i=k+1}^n \frac{(y_i - \widehat{\Delta}_2^{(h)} \widehat{t}_i^{(h)})}{n-k}, \\ \widehat{\Delta}_1^{(h+1)} &= \frac{\sum_{i=1}^k \widehat{t}_i^{(h)} (y_i - \widehat{\mu}_1^{(h)})}{\sum_{i=1}^k \widehat{t}_i^{2(h)}}, & \widehat{\Delta}_2^{(h+1)} &= \frac{\sum_{i=k+1}^n \widehat{t}_i^{(h)} (y_i - \widehat{\mu}_2^{(h)})}{\sum_{i=k+1}^n \widehat{t}_i^{2(h)}}, \\ \widehat{\gamma}_1^{2(h+1)} &= \frac{1}{k} \sum_{i=1}^k \left( (y_i - \widehat{\mu}_1^{(h)})^2 - 2\widehat{\Delta}_1^{(h)} \widehat{t}_i^{(h)} (y_i - \widehat{\mu}_1^{(h)}) + \widehat{\Delta}_1^{2(h)} \widehat{t}_i^{2(h)} \right), \\ \widehat{\gamma}_2^{2(h+1)} &= \frac{1}{n-k} \sum_{i=k+1}^n \left( (y_i - \widehat{\mu}_2^{(h)})^2 - 2\widehat{\Delta}_2^{(h)} \widehat{t}_i^{(h)} (y_i - \widehat{\mu}_2^{(h)}) + \widehat{\Delta}_2^{2(h)} \widehat{t}_i^{2(h)} \right), \\ \widehat{\sigma}_j^{(h+1)} &= \sqrt{\widehat{\Delta}_j^{2(h)} + \widehat{\gamma}_j^{2(h)}}, & \widehat{\lambda}_j^{(h+1)} &= \frac{\widehat{\Delta}_j^{(h)}}{\sqrt{\widehat{\gamma}_j^{2(h)}}}, \quad j = 1, 2. \end{aligned}$$

As iterações são repetidas até que uma regra de convergência adequada seja satisfeita, por exemplo,

$$\left| \frac{\ell(\widehat{\boldsymbol{\theta}}^{(h+1)})}{\ell(\widehat{\boldsymbol{\theta}}^{(h)})} - 1 \right| < 10^{-5}. \quad (38)$$

Valores iniciais necessários para implementar este algoritmo são obtidos sob a suposição de normalidade quando  $\widehat{\lambda}_j^{(0)} = 3\text{sin}(\widehat{\rho}_j)$ , onde  $\widehat{\rho}_j$  é o coeficiente de assimetria amostral do grupo  $j$ , para  $j = 1, 2$ , tal que o grupo 1 consiste nas primeiras  $k$  observações. No entanto, a fim de garantir que as verdadeiras estimativas de máxima verossimilhança sejam identificadas, recomendamos a execução do algoritmo EM utilizando diferentes valores iniciais. Note que quando  $\lambda_j = 0$  (ou  $\Delta_j = 0$ ) as equações do passo M reduzem para as equações obtidas assumindo distribuição normal. Particularmente, este algoritmo generaliza os resultados encontrados em Chen e Gupta (1999).

### 3.2.3 Erros Padrão Aproximados

Nesta subseção, derivamos a matriz de informação observada associada ao vetor de parâmetros  $\boldsymbol{\theta}$ . A matriz de informação observada será usada para calcular os erros padrão da estimativa  $\hat{\boldsymbol{\theta}}$ . Seguindo o mesmo procedimento utilizado em Zeller, Lachos e Vilca-Labra (2011) e Contreras (2014), derivamos a matriz de covariância de  $\hat{\boldsymbol{\theta}}$  através da inversão da matriz de informação observada.

Suponha que temos observações sobre  $n$  indivíduos independentes  $Y_1, \dots, Y_n$ , onde

$$\begin{aligned} Y_i &\sim \text{SN}(\mu_1, \sigma_1^2, \lambda_1), \quad i = 1, \dots, k \quad \text{e} \\ Y_i &\sim \text{SN}(\mu_2, \sigma_2^2, \lambda_2) \quad i = k + 1, \dots, n. \end{aligned}$$

Então, a função log-verossimilhança de  $\boldsymbol{\theta}$ , dada a amostra observada  $y_1, \dots, y_n$ , é da forma  $\ell(\boldsymbol{\theta}) = n \log(2) - \frac{n}{2} \log(2\pi) + \sum_{i=1}^n \ell_i(\boldsymbol{\theta})$ , com  $\ell_i(\boldsymbol{\theta})$  definida em (30). Assim, a matriz hessiana  $\mathbf{H} = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \begin{pmatrix} \mathbf{H}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 \end{pmatrix}$ , onde  $\mathbf{H}_1 = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1^\top}$  e  $\mathbf{H}_2 = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2^\top}$  tem os elementos dados por

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mu_1^2} &= -\frac{k}{\sigma_1^2} - \frac{\lambda_1^2}{\sigma_1^2} \sum_{i=1}^k a_i W_\Phi(a_i) - \frac{\lambda_1^2}{\sigma_1^2} \sum_{i=1}^k W_\Phi^2(a_i), \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma_1^2} &= \frac{k}{\sigma_1^2} - \frac{3}{\sigma_1^2} \sum_{i=1}^k a_i^2 - \frac{\lambda_1^3}{\sigma_1^2} \sum_{i=1}^k a_i^3 W_\Phi(a_i) + \frac{2}{\sigma_1^2} \sum_{i=1}^k a_i W_\Phi(a_i) - \frac{\lambda_1^2}{\sigma_1^2} \sum_{i=1}^k a_i^2 W_\Phi^2(a_i), \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \lambda_1^2} &= -\frac{1}{\lambda_1^2} \sum_{i=1}^k a_i^3 W_\Phi(a_i) - \frac{1}{\lambda_1^2} \sum_{i=1}^k a_i^2 W_\Phi^2(a_i), \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mu_1 \partial \sigma_1} &= -\frac{2}{\sigma_1^2 \lambda_1} \sum_{i=1}^k a_i - \frac{\lambda_1}{\sigma_1^2} \sum_{i=1}^k a_i^2 W_\Phi(a_i) + \frac{\lambda_1}{\sigma_1^2} \sum_{i=1}^k W_\Phi(a_i) - \frac{\lambda_1}{\sigma_1^2} \sum_{i=1}^k a_i W_\Phi^2(a_i), \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mu_1 \partial \lambda_1} &= -\frac{1}{\sigma_1} \sum_{i=1}^k W_\Phi(a_i) + \frac{1}{\sigma_1} \sum_{i=1}^k a_i^2 W_\Phi(a_i) + \frac{1}{\sigma_1} \sum_{i=1}^k a_i W_\Phi^2(a_i), \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \sigma_1 \partial \lambda_1} &= \frac{1}{\sigma_1 \lambda_1} \sum_{i=1}^k a_i^3 W_\Phi(a_i) - \frac{1}{\sigma_1 \lambda_1} \sum_{i=1}^k a_i W_\Phi(a_i) + \frac{1}{\sigma_1 \lambda_1} \sum_{i=1}^k a_i^2 W_\Phi^3(a_i), \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mu_2^2} &= -\frac{(n-k)}{\sigma_2^2} - \frac{\lambda_2^2}{\sigma_2^2} \sum_{i=k+1}^n a_i W_\Phi(a_i) - \frac{\lambda_2^2}{\sigma_2^2} \sum_{i=k+1}^n W_\Phi^2(a_i), \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma_2^2} &= \frac{(n-k)}{\sigma_2^2} - \frac{3}{\sigma_2^2} \sum_{i=k+1}^n a_i^2 - \frac{\lambda_2^3}{\sigma_2^2} \sum_{i=k+1}^n a_i^3 W_\Phi(a_i) + \frac{2}{\sigma_2^2} \sum_{i=k+1}^n a_i W_\Phi(a_i) - \frac{\lambda_2^2}{\sigma_2^2} \sum_{i=k+1}^n a_i^2 W_\Phi^2(a_i), \end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \lambda_2^2} &= -\frac{1}{\lambda_2^2} \sum_{i=k+1}^n a_i^3 W_{\Phi}(a_i) - \frac{1}{\lambda_2^2} \sum_{i=k+1}^n a_i^2 W_{\Phi}^2(a_i), \\
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mu_2 \partial \sigma_2} &= -\frac{2}{\sigma_2^2 \lambda_2} \sum_{i=k+1}^n a_i - \frac{\lambda_2}{\sigma_2^2} \sum_{i=k+1}^n a_i^2 W_{\Phi}(a_i) + \frac{\lambda_2}{\sigma_2^2} \sum_{i=k+1}^n W_{\Phi}(a_i) - \frac{\lambda_2}{\sigma_2^2} \sum_{i=k+1}^n a_i W_{\Phi}^2(a_i), \\
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mu_2 \partial \lambda_2} &= -\frac{1}{\sigma_2} \sum_{i=k+1}^n W_{\Phi}(a_i) + \frac{1}{\sigma_2} \sum_{i=k+1}^n a_i^2 W_{\Phi}(a_i) + \frac{1}{\sigma_2} \sum_{i=k+1}^n a_i W_{\Phi}^2(a_i), \\
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \sigma_2 \partial \lambda_2} &= \frac{1}{\sigma_2 \lambda_2} \sum_{i=k+1}^n a_i^3 W_{\Phi}(a_i) - \frac{1}{\sigma_2 \lambda_2} \sum_{i=k+1}^n a_i W_{\Phi}(a_i) + \frac{1}{\sigma_2 \lambda_2} \sum_{i=k+1}^n a_i^2 W_{\Phi}^3(a_i),
\end{aligned}$$

onde

$$a_i = \begin{cases} \frac{\lambda_1(y_i - \mu_1)}{\sigma_1}, & i = 1, \dots, k, \\ \frac{\lambda_2(y_i - \mu_2)}{\sigma_2}, & i = k + 1, \dots, n. \end{cases}$$

Assim, se  $\mathbf{J} = -\mathbf{H}$  denota a matriz de informação observada para a log-verossimilhança  $\ell(\boldsymbol{\theta})$ , então a matriz de covariância aproximada  $\hat{\boldsymbol{\theta}}$  é dada como  $\widehat{Var}_{approx}(\hat{\boldsymbol{\theta}}) = \mathbf{J}^{-1}|_{\hat{\boldsymbol{\theta}}}$ .

## 4 EXEMPLOS NUMÉRICOS

Neste capítulo, aplicações em dados reais e estudos de simulação são apresentados a fim de ilustrar o modelo e os resultados inferenciais desenvolvidos. Neste trabalho, utilizamos o *software* R versão 2.11.1 (R Development Core Team, 2011) e com as devidas adaptações para o problema de ponto de mudança, trabalhamos também com as funções *rsn* e *sn.em* presentes no pacote *sn*. Entretanto, a função *sn.em* não retorna os erros padrão, então a matriz informação observada apresentada na Seção 3.2.3 foi implementada.

### 4.1 Estudos de Simulação

Nesta seção, realizamos uma ilustração numérica com dados simulados. Os métodos de estimação dos parâmetros do modelo e de detecção do ponto de mudança, propostos nas Seções 3.2.2 e 3.2.1, respectivamente, são avaliados usando dados simulados obtidos pelo procedimento de Monte Carlo.

#### 4.1.1 Experimento I: Estimação via Algoritmo EM

Neste experimento, consideramos três cenários para a simulação, a fim de verificar se podemos estimar os verdadeiros valores dos parâmetros com acurácia utilizando o método de estimação proposto. Este é o primeiro passo para assegurar que o procedimento de estimação funciona satisfatoriamente. Ajustamos o modelo normal assimétrico com ponto de mudança aos dados que foram artificialmente gerados a partir do modelo estudado. Neste caso, de acordo com trabalhos anteriores tal como Arellano-Valle, Bolfarine e Lachos (2005), que também realizaram simulações do modelo normal assimétrico utilizando o procedimento de Monte Carlo, geramos 500 amostras aleatórias de tamanho  $n = 25, 50, 100$  e 200 do modelo normal assimétrico com ponto de mudança com os seguintes valores para os parâmetros:  $\mu_1 = 4, \sigma_1 = 5, \lambda_1 = 3, \mu_2 = 2, \sigma_2 = 1$  e  $\lambda_2 = 1$ . Além disso, consideramos os seguintes cenários variando  $n_1$  (número de observações anteriores ao ponto de mudança) e  $n_2$  (número de observações posteriores ao ponto de mudança), tal que  $n_1$  corresponde no cenário 1: 50%, no cenário 2: 70% e no cenário 3: 90% de  $n = n_1 + n_2$ . Desenvolvemos a estimação por máxima verossimilhança via algoritmo EM para cada amostra, usando o critério de parada (38). Os valores médios e os correspondentes desvios padrão (DP) das estimativas obtidas via algoritmo EM em todas as amostras foram calculados via *bootstrap* paramétrico e os resultados estão apresentados nas Tabelas 1-3, que também contém os erros padrão aproximados (EP) calculados através da matriz de informação observada. Note que todas as estimativas pontuais são bastante acuradas em todos os cenários considerados. Pode-se observar pelas Tabelas 1-3 que o viés (diferença entre

a estimativa e o verdadeiro valor do parâmetro) torna-se menor à medida que o tamanho amostral  $n$  aumenta, como esperado. Além disso, é possível notar uma maior discrepância entre os valores de DP e EP quando se altera a posição do ponto de mudança fazendo com que se tenha menos observações após este ponto. Assim, observe que a precisão das estimativas é maior para os parâmetros de índice 1 (antes do ponto de mudança) do que para os parâmetros de índice 2 (após o ponto de mudança). O cenário 3 apresentado na Tabela 3 ilustra bem essa discrepância entre os erros em pequenas amostras, onde os parâmetros estimados após o ponto de mudança contém apenas 10% do tamanho total da amostra. Neste mesmo cenário também é possível ver que para os parâmetros estimados antes do ponto de mudança, onde há 90% do tamanho total da amostra, a diferença entre DP e EP é muito pequena, sendo significativamente menor em  $n = 200$  do que em  $n = 25$ . Os cenários 1 e 2 apresentados nas Tabelas 2 e 1 nos sugerem essa mesma conclusão. Assim, os resultados sugerem que o algoritmo EM proposto é eficiente para a estimação dos parâmetros do modelo normal assimétrico com ponto de mudança e o cálculo dos erros padrões via matriz informação observada é eficaz (no sentido que se aproxima do DP) para amostras grandes, pois assim teremos muitas observações antes e após o ponto de mudança.

Tabela 1 – Cenário 1: média e desvio padrão (DP) para o algoritmo EM estimado baseado nas 500 amostras com ponto de mudança no modelo normal assimétrico. EP é a média do erro padrão estimado.

		n=25, k=12					
Parâmetros		$\mu_1$	$\sigma_1$	$\lambda_1$	$\mu_2$	$\sigma_2$	$\lambda_2$
Média		5.6495	4.1032	2.2098	2.3729	1.0283	0.6806
DP		2.1993	1.2041	2.2878	0.7276	0.2684	2.0629
EP		4.4384	2.0993	4.5162	1.2294	0.5801	3.5414
		n=50, k=25					
Parâmetros		$\mu_1$	$\sigma_1$	$\lambda_1$	$\mu_2$	$\sigma_2$	$\lambda_2$
Média		5.0417	4.4032	2.6116	2.3837	1.0722	0.5530
DP		1.8436	1.0108	2.1434	0.7400	0.2153	2.0386
EP		2.4327	1.3359	2.8492	0.7952	0.4067	2.3725
		n=100, k=50					
Parâmetros		$\mu_1$	$\sigma_1$	$\lambda_1$	$\mu_2$	$\sigma_2$	$\lambda_2$
Média		4.4921	4.7197	3.0497	2.3341	1.0658	0.6303
DP		1.3255	0.7787	1.7502	0.6774	0.1762	1.7260
EP		1.1410	0.8621	1.9278	0.6043	0.3051	1.6369
		n=200, k=100					
Parâmetros		$\mu_1$	$\sigma_1$	$\lambda_1$	$\mu_2$	$\sigma_2$	$\lambda_2$
Média		4.1031	4.9063	3.2609	2.2239	1.0460	0.7696
DP		0.6167	0.5534	1.3109	0.5635	0.1392	1.2604
EP		0.5756	0.5511	1.3033	0.4360	0.2231	1.0806

Tabela 2 – Cenário 2: média e desvio padrão (DP) para o algoritmo EM estimado baseado nas 500 amostras com ponto de mudança no modelo normal assimétrico. EP é a média do erro padrão estimado.

		n=25, k=17					
Parâmetros		$\mu_1$	$\sigma_1$	$\lambda_1$	$\mu_2$	$\sigma_2$	$\lambda_2$
Média		5.3279	4.2550	2.3707	2.5213	0.9631	0.2882
DP		2.1014	1.1026	2.1947	0.7223	0.3262	2.1263
EP		3.2691	1.6708	3.4716	1.9578	0.7561	5.1960
		n=50, k=35					
Parâmetros		$\mu_1$	$\sigma_1$	$\lambda_1$	$\mu_2$	$\sigma_2$	$\lambda_2$
Média		4.6772	4.6670	2.9616	2.4235	1.0396	0.4284
DP		2.1014	0.9433	2.0509	0.7514	0.2528	2.2049
EP		1.7465	1.1112	2.5487	1.0997	0.5260	3.3399
		For n=100, k=70					
Parâmetro		$\mu_1$	$\sigma_1$	$\lambda_1$	$\mu_2$	$\sigma_2$	$\lambda_2$
Média		4.2583	4.8407	3.1325	2.4041	1.0732	0.5325
DP		0.9891	0.6671	1.4528	0.7229	0.2005	2.0321
EP		0.8549	0.7058	1.5892	0.7022	0.3733	2.1147
		For n=200, k=140					
Parâmetro		$\mu_1$	$\sigma_1$	$\lambda_1$	$\mu_2$	$\sigma_2$	$\lambda_2$
Média		4.0849	4.9289	3.0637	2.3119	1.0374	0.6657
DP		0.4310	0.4147	0.8698	0.6197	0.1801	1.4679
EP		0.4385	0.4302	0.9260	0.8165	0.3672	1.9551

Tabela 3 – Cenário 3: média e desvio padrão (DP) para o algoritmo EM estimado baseado nas 500 amostras com ponto de mudança no modelo normal assimétrico. EP é a média do erro padrão estimado.

		n=25, k=22					
Parâmetros		$\mu_1$	$\sigma_1$	$\lambda_1$	$\mu_2$	$\sigma_2$	$\lambda_2$
Média		4.6086	4.4969	2.6369	2.5252	0.6048	-0.0123
DP		1.1202	0.9975	1.4084	0.4936	0.3126	0.3572
EP		2.1695	1.3916	2.6498	4.8164	1.1711	10.8596
		n=50, k=45					
Parâmetros		$\mu_1$	$\sigma_1$	$\lambda_1$	$\mu_2$	$\sigma_2$	$\lambda_2$
Média		4.3008	4.7478	3.0095	2.4367	0.8360	0.4931
DP		0.8196	0.7345	1.3115	0.6382	0.3534	1.3191
EP		1.1291	0.8838	1.9559	3.1997	1.0081	7.3306
		For n=100, k=90					
Parâmetro		$\mu_1$	$\sigma_1$	$\lambda_1$	$\mu_2$	$\sigma_2$	$\lambda_2$
Média		4.1406	4.8779	3.0916	2.3276	0.9350	0.7063
DP		0.5683	0.5562	1.1006	0.6130	0.2810	1.5103
EP		0.6128	0.5816	1.2496	1.6804	0.6805	4.2185
		For n=200, k=180					
Parâmetro		$\mu_1$	$\sigma_1$	$\lambda_1$	$\mu_2$	$\sigma_2$	$\lambda_2$
Média		4.0574	4.9747	3.1294	2.2904	1.0386	0.7470
DP		0.3997	0.4089	0.8494	0.6520	0.2268	1.5585
EP		0.4077	0.4040	0.8687	1.1573	0.5145	2.8632

### 4.1.2 Experimento II: Detecção do Ponto de Mudança

Geralmente o ponto de mudança é desconhecido. Dessa forma, o propósito deste estudo de simulação é avaliar o critério SIC como ferramenta para detectar o ponto de mudança no contexto assimétrico. Neste caso, sem perda de generalidade, os mesmos cenários e valores para os parâmetros usados na seção anterior foram considerados com  $n = 25, 50, 100$  e  $200$ . Seguindo Hofrichter (2007), a performance do método de localização de ponto de mudança via SIC é avaliada através da distribuição de frequências dos pontos de mudança estimados obtidos nas amostras Monte Carlo. As Figuras 2-4 mostram que para todos os cenários a moda das distribuições simuladas dos pontos de mudança estimados ocorre na posição definida nos dados simulados. Isto nos leva a concluir que o critério SIC pode corretamente localizar o ponto de mudança. Isto pode ser visto nas frequências mais elevadas em  $\hat{k}$  nas Figuras 2-4.

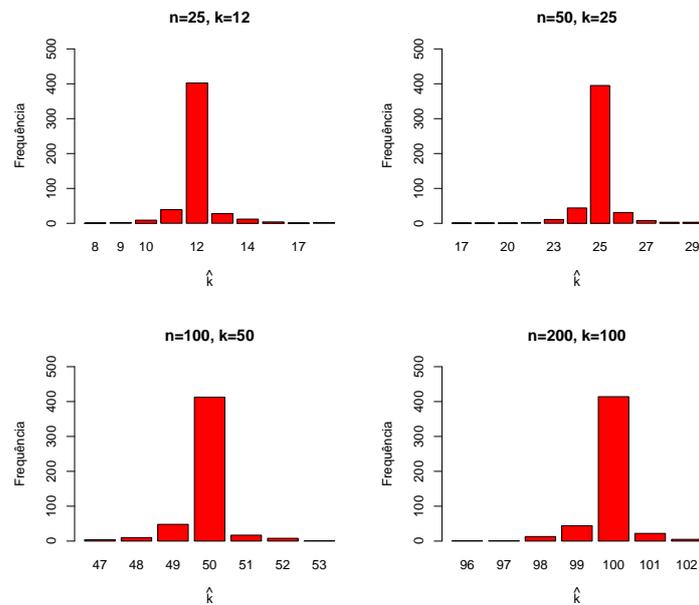


Figura 2 – Frequências dos pontos de mudança estimados para os modelos no cenário 1.

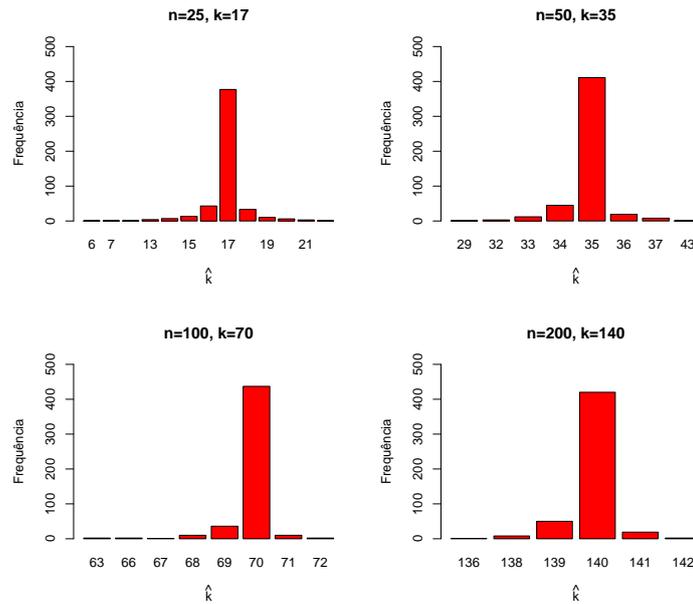


Figura 3 – Frequências dos pontos de mudança estimados para os modelos no cenário 2.

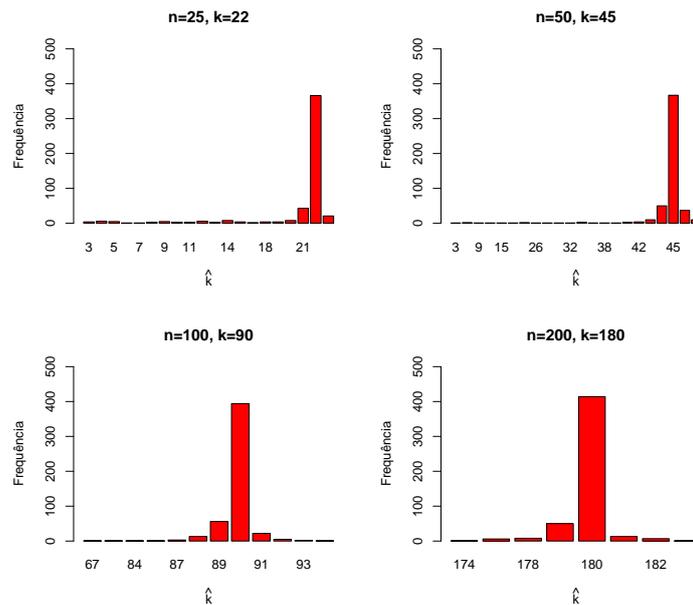


Figura 4 – Frequências dos pontos de mudança estimados para os modelos no cenário 3.

## 4.2 Dados Reais

Vamos considerar dois conjuntos de dados reais que já foram analisados no contexto de ponto de mudança sob o modelo normal. Nesta seção, expandimos a análise desses dados reais, no contexto de ponto de mudança, sob o modelo normal assimétrico. Realizamos estimação por máxima verossimilhança dos parâmetros do modelo, via algoritmo EM,

detectamos o ponto de mudança (quando houver) via critério SIC e por fim, comparamos os ajustes dos modelos normal e normal assimétrico com ponto de mudança via critério SIC também. Através do critério SIC, procura-se verificar qual dos modelos se adequa melhor aos dados.

#### 4.2.1 Vazão do Rio Nilo

O Rio Nilo é o principal rio que banha vários países africanos como Uganda, Tanzânia, Ruanda, Quênia, República Democrática do Congo, Burundi, Sudão, Etiópia e principalmente o Egito, em que cerca de 90% da sua população vive às suas margens. Há mais de 5000 anos o Rio Nilo proporciona riquezas às diversas civilizações e culturas que vivem às suas margens, pois durante o seu período de cheia, ele transborda e suas águas levam uma grande quantidade de sedimentos que são distribuídos ao longo de suas margens. Assim, quando chega a época das vazantes, as águas abaixam e deixam no solo uma enorme quantidade de nutrientes importantes para sua fertilidade. Com o intuito de evitar catástrofes durante o período de cheia, quando o Rio Nilo transbordava, os egípcios construíram diques e posteriormente barragens ao longo de vários pontos de suas margens (COBB, 1978).

Os dados da Figura 5, disponíveis em [www.ssfpack.com/DKbook.html](http://www.ssfpack.com/DKbook.html), se referem a vazão média anual do Rio Nilo na cidade de Assuã, no sul do Egito, entre 1871 e 1970, sendo que o interesse é investigar se ao longo desses 100 anos aproximadamente, a vazão deste rio foi alterada por alguma interferência humana ou natural. Note que a Figura 5 indica uma mudança na estrutura da média (principalmente) e da variância na sequência de observações.

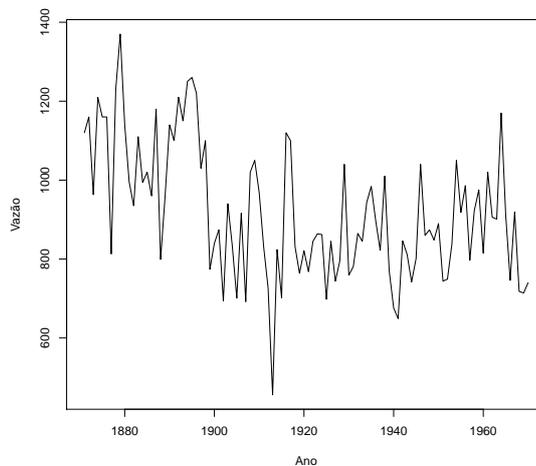


Figura 5 – Vazão ( $m^3/s$ ) média anual do Rio Nilo na cidade de Assuã, no Egito, entre os anos de 1871 e 1970.

Os dados do Rio Nilo são uma referência clássica na literatura estatística no contexto de ponto de mudança. Neste contexto, os dados têm sido utilizados para ilustração numérica por um grande número *softwares* e autores. Por exemplo, de um ponto de vista Bayesiano, podemos mencionar Moreno, Casella e García-Ferrer (2005) e GirÓN, Moreno e Casella (2007). Portanto, nesta seção revisitamos os dados do Rio Nilo com o intuito de fornecer inferências adicionais usando o modelo normal assimétrico com ponto de mudança. Espera-se que a vazão tenha uma distribuição assimétrica devido a ocorrência das cheias.

Tabela 4 – Resultados do procedimento SIC.

Modelo	$SIC(n)$	$\min SIC(k)$	$\hat{k}$
Normal	1318.242	1279.107	28
Normal Assimétrico	1319.542	1277.445	28

O resultado apresentado na Tabela 4 mostra que os modelos com ponto de mudança em  $k = 28$  se adequam melhor aos dados do que o modelo sem mudança, sendo o modelo normal assimétrico ligeiramente superior ao modelo normal. As estimativas dos parâmetros para os modelos normal e normal assimétrico com ponto de mudança na posição 28 estão apresentadas na Tabela 5. A fim de compararmos os resultados obtidos, também apresentamos as estimativas dos parâmetros para os modelos normal e normal assimétrico sem ponto de mudança na Tabela 6. Espera-se que a razão seja assimétrica devido a ocorrência das cheias.

Tabela 5 – Estimativas dos parâmetros dos modelos com mudança em  $\hat{k} = 28$ .

Modelo	Estimativas					
	$\mu_1$	$\sigma_1$	$\lambda_1$	$\mu_2$	$\sigma_2$	$\lambda_2$
Normal (DP)	1097.7500 (27.3555)	132.5636 (17.3388)	- (-)	849.9722 (14.7568)	123.9069 (9.9872)	- (-)
Normal Assimétrico (DP)	1243.5671 (78.6940)	197.0554 (39.6131)	-2.4576 (1.9764)	760.0383 (60.7304)	153.0937 (23.3562)	1.0896 (1.4593)

Tabela 6 – Estimativas dos parâmetros dos modelos sem o ponto de mudança

Modelo	Estimativas		
	$\mu$	$\sigma$	$\lambda$
Normal (DP)	919.3500 (16.1747)	168.3792 (11.7810)	- (-)
Normal Assimétrico (DP)	736.8624 (39.3759)	248.2862 (30.5969)	2.3024 (1.0115)

A Figura 6 mostra os histogramas com as densidades estimadas sobrepostas para os dados do Rio Nilo. Podemos notar que os histogramas com densidades estimadas

sobrepostas para os modelos com uma mudança na posição 28 não revelam diferenças notáveis entre o modelo normal e o modelo normal assimétrico quando consideramos as observações subsequentes à 28<sup>a</sup> posição.

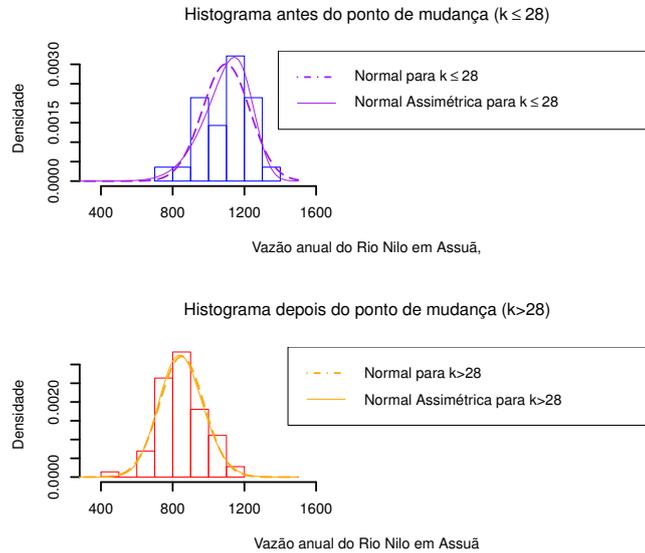


Figura 6 – Histogramas com as densidades estimadas sobrepostas para os dados do Rio Nilo.

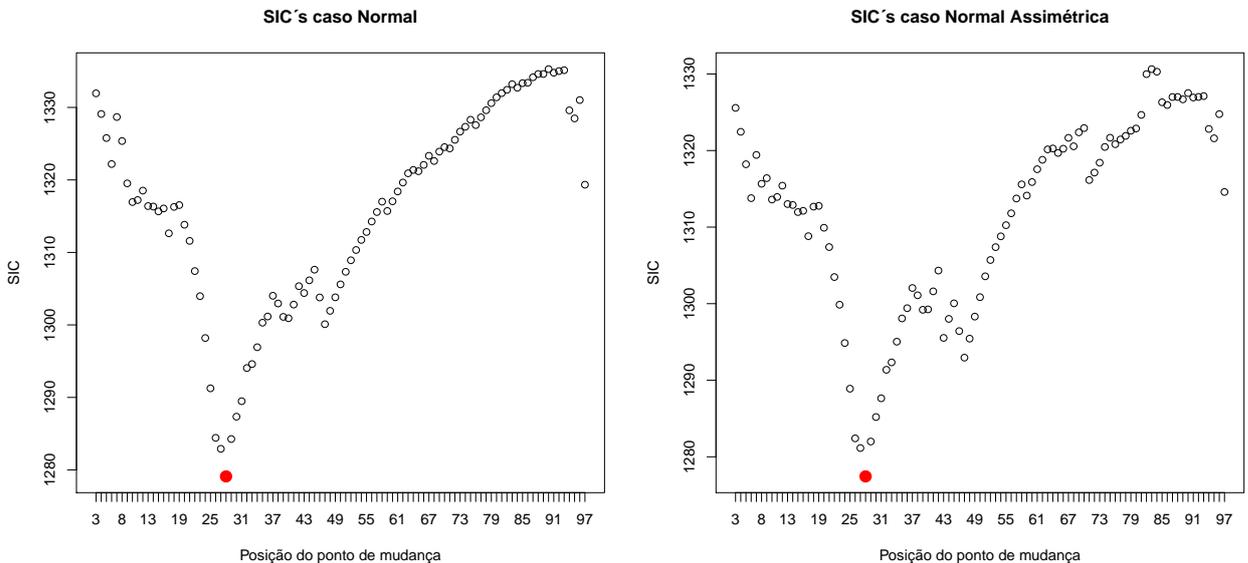


Figura 7 –  $SIC(k)$  vs  $k$  para os casos da distribuições normal e normal assimétrica.

O gráfico do  $SIC(k)$  vs  $k$  é considerado, pois é possível visualizar a estimativa de máxima verossimilhança de  $k$ . Em ambos os diagramas a posição 28 é a estimativa de máxima verossimilhança de  $k$ , como citado anteriormente, podendo ser visto na Figura

7. Note que, de acordo com a Tabela 4, para o modelo normal, os valores de  $SIC(k)$  são maiores do que para o modelo normal assimétrico. Dessa forma, para este conjunto de dados, sugerimos usar o modelo normal assimétrico como alternativa para o modelo normal. O período  $\hat{k} = 28$  nos dados se refere ao ano de 1902. Pelos registros históricos, neste ano foi construída a represa de Assuã que seria a principal responsável pela queda na vazão média do Rio Nilo conforme visto nas Figuras 5 e 6. O modelo normal assimétrico é de grande importância, neste caso, pois através dele é possível ver que após a construção dessa barragem a vazão do rio foi regulada, visto de sinal entre as estimativas dos parâmetros de forma  $\lambda_1$  e  $\lambda_2$ , ou seja, assimetrias em diferentes direções, apresentadas na Tabela 5.

#### 4.2.2 Índice *Dow Jones*

O índice *Dow Jones* é conhecido no mundo inteiro como um dos principais e mais antigos índices de Nova York, que tem como função estabelecer o cálculo do movimento de mercado das 30 maiores ações das mais importantes companhias norte americanas. Desde que foi criado em 1896, este índice teve grandes oscilações em períodos marcantes da história norte americana, como, por exemplo, após a primeira guerra mundial, durante a crise de 1929, no atentado de 11 de setembro e mais recentemente, na crise financeira mundial em 2008. Os dados da Figura 8, disponíveis em Hsu (1979), se referem aos retornos semanais do índice *Dow Jones* de 1 de julho de 1971 até 2 de agosto de 1974. Adotando  $P_t$  como o valor do índice *Dow Jones* durante o período  $t$ , num total de 162 valores, o interesse principal nesses dados é pelo retorno deste índice, denotado por  $R_t$ , que é calculado pela equação (39).

$$R_t = \frac{P_{t+1} - P_t}{P_t}, \quad t = 1, \dots, 161. \quad (39)$$

Hsu (1979) através do teste de Portmanteau mostrou que a hipótese nula de correlação dos dados igual a zero não é rejeitada, portanto podemos pressupor independência entre os retornos nesse conjunto de dados.

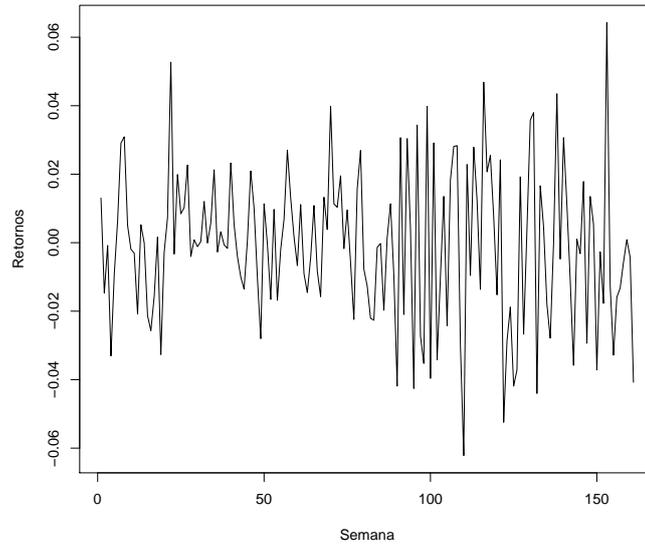


Figura 8 – Retornos semanais do índice *Dow Jones* de 1 de julho de 1971 até 2 de agosto de 1974.

Note que a Figura 8 indica uma mudança na estrutura da média e da variância (principalmente) na sequência de observações. Este conjunto de dados já foi analisado por Hsu (1979), no contexto de ponto de mudança, considerando a distribuição gama. Portanto, nesta seção revisitamos os dados do índice *Dow Jones* com o intuito de fornecer inferências adicionais usando os modelos normal e normal assimétrico com ponto de mudança. O procedimento SIC foi aplicado para detectar mudanças, no modelo normal assimétrico, bem como para o modelo normal. Os resultados estão resumidos na Tabela 7. Para ambos os modelos, um ponto de mudança é detectado no período  $\hat{k} = 89$ , e este resultado está de acordo com a conclusão obtida por Hsu (1979).

Tabela 7 – Resultados do procedimento SIC.

Modelo	$SIC(n)$	$\min SIC(k)$	$\hat{k}$
Normal	-760.753	-768.459	89
Normal Assimétrico	-755.671	-769.923	89

As estimativas dos parâmetros para os modelos normal e normal assimétrico com ponto de mudança na posição 89 estão apresentadas na Tabela 8. A fim de compararmos os resultados obtidos, também apresentamos as estimativas dos parâmetros para os modelos normal e normal assimétrico sem ponto de mudança na Tabela 9. Do ponto de vista financeiro, através da Tabela 10 nota-se que após o ponto de mudança, as médias de retorno se tornaram negativa nos dois modelos e a variância praticamente dobrou no modelo normal assimétrico e triplicou no modelo normal, mostrando assim que após o

ponto de mudança os retornos em média caíram e se tornaram mais instáveis.

Tabela 8 – Estimativas dos parâmetros dos modelos com mudança em  $\hat{k} = 89$ .

Modelo	Estimativas					
	$\mu_1$	$\sigma_1$	$\lambda_1$	$\mu_2$	$\sigma_2$	$\lambda_2$
Normal (DP)	0.00101 (0.00166)	0.01559 (0.00118)	- (-)	-0.00303 (0.00310)	0.02794 (0.00226)	- (-)
Normal Assimétrico (DP)	-0.01218 (0.00936)	0.02042 (0.00319)	1.38831 (1.28589)	-0.01748 (0.02153)	0.03146 (0.00513)	0.70345 (1.46974)

Tabela 9 – Estimativas dos parâmetros dos modelos sem ponto de mudança

Modelo	Estimativas		
	$\mu$	$\sigma$	$\lambda$
Normal (DP)	-0.00070 (0.00169)	0.02208 (0.00127)	- (-)
Normal Assimétrico (DP)	-0.00297 (0.01543)	0.02219 (0.00277)	0.12348 (1.11162)

Tabela 10 – Valores esperados e variâncias antes e depois do ponto de mudança  $\hat{k} = 89$ .

Modelo	$E(Y_1)$	$Var(Y_1)$	$E(Y_2)$	$Var(Y_2)$
Normal	0.00101	0.00024	0.01559	0.00078
Normal Assimétrico	0.09053	0.00042	-0.00304	0.00099

A Figura 9 mostra os histogramas com as densidades estimadas sobrepostas para os dados do índice *Dow Jones*. Podemos notar que os histogramas com densidades estimadas sobrepostas para os modelos com uma mudança na posição 89 não revelam diferenças notáveis entre os modelos normal e normal assimétrico.

O gráfico do  $SIC(k)$  vs  $k$  é considerado, pois é possível visualizar a estimativa de máxima verossimilhança de  $k$ . Em ambos os diagramas a posição 89 é a estimativa de máxima verossimilhança de  $k$ , como citado anteriormente, e que pode ser visto na Figura 10. Note que, no entanto, para o modelo normal, os valores de  $SIC(k)$  são um pouco maiores do que para o modelo normal assimétrico, conforme mostra a Tabela 7. Dessa forma, para este conjunto de dados, sugerimos usar o modelo normal assimétrico como alternativa para o modelo normal. Retornando à equação (39), o retorno  $R_{89}$  é referente ao  $P_{90}$ , cujo período é na semana de 19 a 23 em março de 1973. Hsu (1979) justifica que este período refere-se à instabilidade política decorrente do escândalo de *Watergate*, dito como o maior escândalo político da história norte americana, que culminou com a renúncia do presidente da época, Richard Nixon.

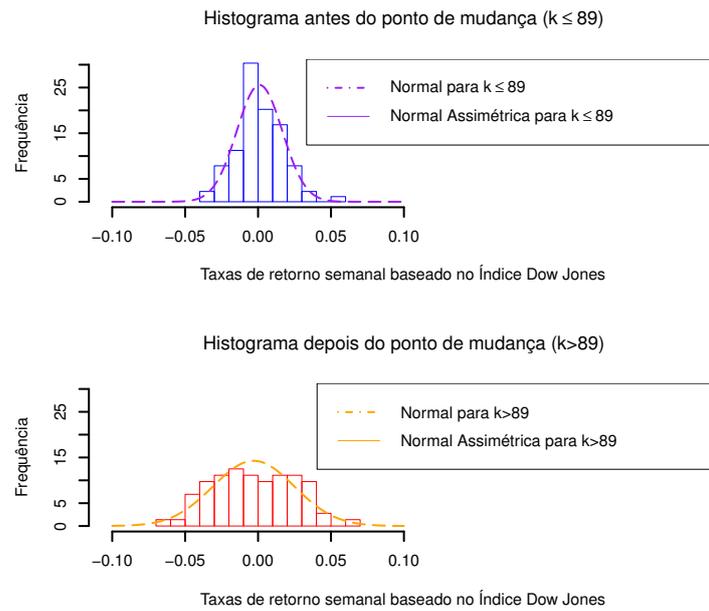


Figura 9 – Histogramas com as densidades estimadas sobrepostas para os dados do índice *Dow Jones*.

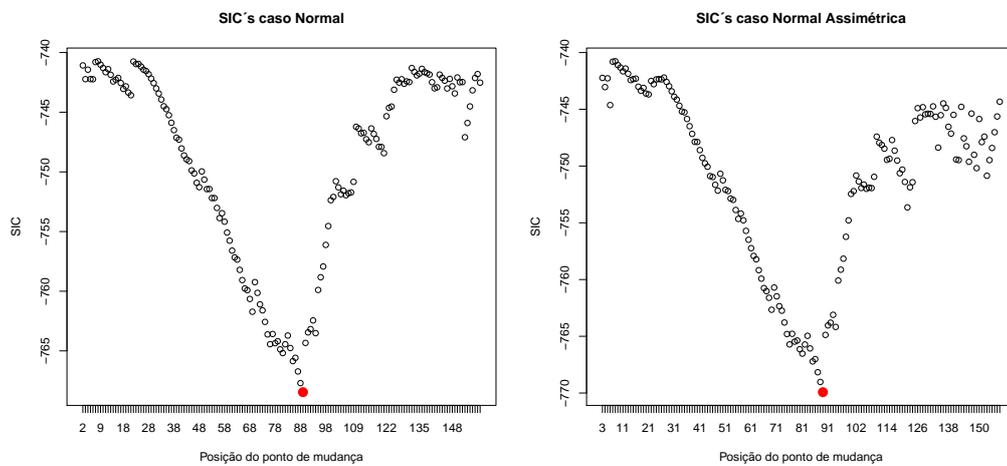


Figura 10 –  $SIC(k)$  vs  $k$  para os casos da distribuições normal e normal assimétrica.

## 5 CONCLUSÕES E PERSPECTIVAS FUTURAS

Em resumo, neste trabalho abordamos a identificação de ponto de mudança, na estrutura da média e da variância simultaneamente, em dados com distribuição normal assimétrica que são observados sequencialmente, complementando e estendendo os resultados encontrados em Arellano-Valle, Castro e Loschi (2013) e Ngunkeng e Ning (2014). O algoritmo EM do pacote *sn* foi adaptado para o contexto de ponto de mudança, fornecendo uma solução analítica para os estimadores de máxima verossimilhança dos parâmetros do modelo. O algoritmo EM tem diversas vantagens sobre a maximização direta da função verossimilhança uma vez que é facilmente implementável, numericamente estável e bastante acurado, como demonstrado através dos estudos de simulação. A matriz de informação observada pode ser usada para calcular o erro padrão dos parâmetros estimados. Para determinar o ponto de mudança, usamos o critério de informação de Schwarz. Quando a inferência sob um ponto de mudança é vista como um problema de seleção de modelo, o SIC, ou outro critério de informação, apresenta-se como uma maneira bastante eficaz para se localizar o ponto de mudança. Fica evidente que a detecção de ponto de mudança através do SIC oferece uma importante simplificação, tornando simples a implementação computacional. Alguns estudos de simulação e aplicações a conjunto de dados reais são apresentados com o propósito de ilustrar o modelo e os resultados inferenciais desenvolvidos aqui. Os resultados desse trabalho podem ser úteis para determinar ponto de mudança em muitas outras aplicações práticas.

Verifica-se que o método de estimação, via algoritmo EM, apresentado neste trabalho, deva produzir resultados satisfatórios no contexto de múltiplos pontos de mudança aplicando o procedimento de segmentação binária implementado em Chen e Gupta (1997), Ngunkeng e Ning (2014) e Hasan, Ning e Gupta (2014), por exemplo. Trata-se de um tema de pesquisa futura. Alguns avanços podem ser vistos no Apêndice A. Além disso, vários trabalhos de pesquisa podem ser derivados e/ou direcionados a partir dos resultados desta monografia, entre eles podemos mencionar os seguintes: estender o problema de ponto de mudança para a classe de distribuições assimétricas misturas de escala normal assimétrica e considerar mudanças na estrutura da média ou da variância. Alguns avanços podem ser vistos no Apêndice B.

# APÊNDICE A – Esboço para Múltiplos Pontos de Mudança

De fato, o problema de múltiplos pontos de mudança pode ser tratado como um problema de único ponto de mudança, utilizando o método de segmentação binária (CHEN; GUPTA, 1997). Uma descrição geral da técnica de segmentação binária na detecção das mudanças pode ser resumida nas seguintes etapas.

- **Passo 1:** Avaliar se há ou não um ponto de mudança via o critério SIC, descrito na Seção 3.2.1. Se não há um ponto de mudança, então pare. Caso contrário, vá para o passo 2.
- **Passo 2:** Avalie as duas subsequências antes e depois do ponto de mudança, encontrado no passo 1, separadamente para uma mudança.
- **Passo 3:** Repita o processo até que não haja mais subsequências com pontos de mudança.
- **Passo 4:** A coleção de posições de pontos de mudança encontrada nas etapas 1-3 é denotada por  $\hat{k}_1, \hat{k}_2, \dots, \hat{k}_q$  e o número total estimado de pontos de mudança é  $q$ .

Note que este procedimento de segmentação binária tem o mérito de detectar o número de pontos de mudança e as suas posições ao mesmo tempo, economizando tempo computacional.

Para ilustrar o procedimento de segmentação binária na detecção das mudanças, via critério SIC, juntamente com as estimativas de máxima verossimilhança dos parâmetros do modelo, obtidas via algoritmo EM, realizamos um pequeno estudo de simulação. Ajustamos o modelo normal assimétrico com ponto de mudança aos dados que foram artificialmente gerados a partir do modelo estudado. Neste caso, geramos 500 amostras aleatórias de tamanho  $n = 100$  do modelo normal assimétrico com três pontos de mudança nas posições  $k_1 = 25$ ,  $k_2 = 50$  e  $k_3 = 75$ , com os seguintes valores para os parâmetros:  $\mu_1 = 3$ ,  $\sigma_1 = 1$ ,  $\lambda_1 = 3$ ,  $\mu_2 = 8$ ,  $\sigma_2 = 7$ ,  $\lambda_2 = 3$ ,  $\mu_3 = 6$ ,  $\sigma_3 = 1$ ,  $\lambda_3 = 2$ ,  $\mu_4 = 1$ ,  $\sigma_4 = 2$  e  $\lambda_4 = 1$ . Seguindo Hofrichter (2007), a performance do método de localização de ponto de mudança via SIC é avaliada através da distribuição de frequências dos pontos de mudança estimados obtidos nas amostras Monte Carlo. A Figura 11 mostra uma distribuição multimodal com as modas localizadas nas posições dos pontos de mudança definidas nos dados simulados. Trata-se de um estudo inicial e outros cenários devem ser considerados para avaliarmos a eficiência do procedimento de segmentação binária na detecção das mudanças, via critério

SIC, juntamente com as estimativas de máxima verossimilhança dos parâmetros do modelo, obtidas via algoritmo EM.

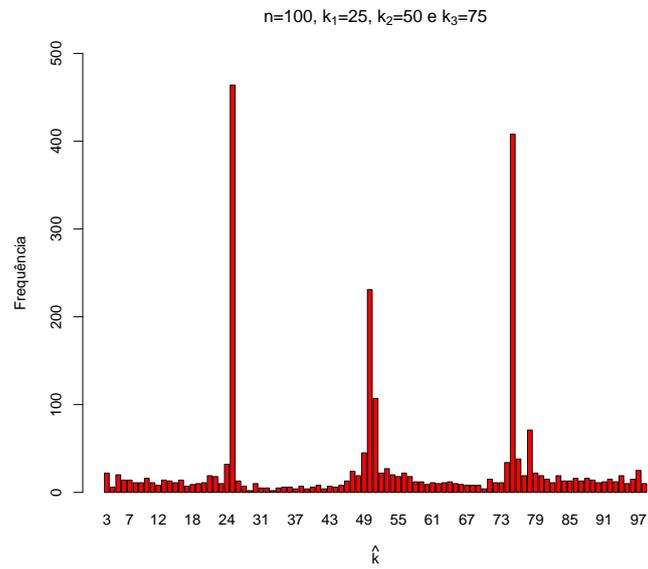


Figura 11 – Frequências dos pontos de mudança estimados para o caso de múltiplos pontos de mudança.

# APÊNDICE B – Outras Estruturas de Mudança

Seguem abaixo outras possibilidades de mudança nos parâmetros da distribuição normal assimétrica, podendo acarretar mudança na estrutura da média e/ou da variância.

- Mudança no parâmetro de locação.

$$Y_i \sim \text{SN}(\mu_1, \sigma^2, \lambda), \quad i = 1, \dots, k \quad \text{e}$$

$$Y_i \sim \text{SN}(\mu_2, \sigma^2, \lambda) \quad i = k + 1, \dots, n,$$

onde  $k$  é a posição desconhecida do ponto de mudança.

- Mudança nos parâmetros de escala e forma.

$$Y_i \sim \text{SN}(\mu, \sigma_1^2, \lambda_1), \quad i = 1, \dots, k \quad \text{e}$$

$$Y_i \sim \text{SN}(\mu, \sigma_2^2, \lambda_2) \quad i = k + 1, \dots, n,$$

onde  $k$  é a posição desconhecida do ponto de mudança.

## APÊNDICE C – Código da Matriz Informação Observada no *software* R

```

Inf.Fisher.sn.cp=function(x,mu.b,sigma.b,lambda.b,mu.a,sigma.a,
lambda.a,k){
n=length(x)
y=x[1:k]
w=x[(k+1):n]
n1=length(y)
n2=length(w)
#####Matriz Informacao de Fisher antes do ponto de mudanca#####
zsn.b=lambda.b*((y-mu.b)/sigma.b)
z.b=(y-mu.b)/sigma.b
d2.mu.b2=(-n1/sigma.b^2)-((lambda.b^2/sigma.b^2)*sum(zsn.b*(dnorm(zsn.b)
/pnorm(zsn.b))))-((lambda.b^2/sigma.b^2)*sum(dnorm(zsn.b)^2/pnorm(zsn.b)^2))
d2.sigma.b2=(n1/sigma.b^2)+((-3/sigma.b^2)*sum(z.b^2))
-((lambda.b^3/sigma.b^2)
*sum(z.b^3*(dnorm(zsn.b)/pnorm(zsn.b))))+(((2*lambda.b)/sigma.b^2)
*sum(z.b*(dnorm(zsn.b)/pnorm(zsn.b))))-((lambda.b^2/sigma.b^2)
*sum(z.b^2*(dnorm(zsn.b)^2/pnorm(zsn.b)^2)))
d2.lambda.b2=(-lambda.b*sum(z.b^3*(dnorm(zsn.b)/pnorm(zsn.b)))
-sum(z.b^2*(dnorm(zsn.b)^2/pnorm(zsn.b)^2)))
d2.mu.b.sigma.b=(-2/sigma.b^2)*sum(z.b)-((lambda.b^3/sigma.b^2)
*sum(z.b^2*(dnorm(zsn.b)/pnorm(zsn.b))))+((lambda.b/sigma.b^2)
*sum(dnorm(zsn.b)/pnorm(zsn.b)))-((lambda.b^2/sigma.b^2)
*sum(z.b*(dnorm(zsn.b)^2/pnorm(zsn.b)^2)))
d2.mu.b.lambda.b=(-1/sigma.b)*sum(dnorm(zsn.b)/pnorm(zsn.b))
+((lambda.b^2/sigma.b)*sum(z.b^2*(dnorm(zsn.b)/pnorm(zsn.b))))
+((lambda.b/sigma.b)*sum(z.b*(dnorm(zsn.b)^2/pnorm(zsn.b)^2)))
d2.sigma.b.lambda.b=((lambda.b^2/sigma.b)
*sum(z.b^3*(dnorm(zsn.b)/pnorm(zsn.b))))
-((1/sigma.b)*sum(z.b*(dnorm(zsn.b)/pnorm(zsn.b))))+((lambda.b/sigma.b)
*sum(z.b^2*(dnorm(zsn.b)^2/pnorm(zsn.b)^2)))
I.b=matrix(c(d2.mu.b2,d2.mu.b.sigma.b,d2.mu.b.lambda.b,d2.mu.b.sigma.b,
d2.sigma.b2,
d2.sigma.b.lambda.b,d2.mu.b.lambda.b,d2.sigma.b.lambda.b,d2.lambda.b2),3,3)

```

```

#####Matriz Informacao de Fisher depois do ponto de mudanca#####
zsn.a=lambda.a*((w-mu.a)/sigma.a)
z.a=(w-mu.a)/sigma.a
d2.mu.a2=(-n2/sigma.a^2)-((lambda.a^2/sigma.a^2)*sum(zsn.a*(dnorm(zsn.a)
/pnorm(zsn.a))))-((lambda.a^2/sigma.a^2)*sum(dnorm(zsn.a)^2/pnorm(zsn.a)^2))
d2.sigma.a2=(n2/sigma.a^2)+((-3/sigma.a^2)*sum(z.a^2))-((lambda.a^3/sigma.a^2)
*sum(z.a^3*(dnorm(zsn.a)/pnorm(zsn.a))))+(((2*lambda.a)/sigma.a^2)
*sum(z.a*(dnorm(zsn.a)/pnorm(zsn.a))))-((lambda.a^2/sigma.a^2)
*sum(z.a^2*(dnorm(zsn.a)^2/pnorm(zsn.a)^2)))
d2.lambda.a2=(-lambda.a*sum(z.a^3*(dnorm(zsn.a)/pnorm(zsn.a)))
-sum(z.a^2*(dnorm(zsn.a)^2/pnorm(zsn.a)^2)))
d2.mu.a.sigma.a=(-2/sigma.a^2)*sum(z.a))-((lambda.a^3/sigma.a^2)
*sum(z.a^2*(dnorm(zsn.a)/pnorm(zsn.a))))+((lambda.a/sigma.a^2)
*sum(dnorm(zsn.a)/pnorm(zsn.a)))-((lambda.a^2/sigma.a^2)
*sum(z.a*(dnorm(zsn.a)^2/pnorm(zsn.a)^2)))
d2.mu.a.lambda.a=(-1/sigma.a)*sum(dnorm(zsn.a)/pnorm(zsn.a))
+((lambda.a^2/sigma.a)
*sum(z.a^2*(dnorm(zsn.a)/pnorm(zsn.a))))
+((lambda.a/sigma.a)*sum(z.a*(dnorm(zsn.a)^2/pnorm(zsn.a)^2)))
d2.sigma.a.lambda.a=((lambda.a^2/sigma.a)
*sum(z.a^3*(dnorm(zsn.a)/pnorm(zsn.a))))
-((1/sigma.a)*sum(z.a*(dnorm(zsn.a)/pnorm(zsn.a))))
+((lambda.a/sigma.a)*sum(z.a^2*(dnorm(zsn.a)^2/pnorm(zsn.a)^2)))
I.a=matrix(c(d2.mu.a2,d2.mu.a.sigma.a,d2.mu.a.lambda.a,d2.mu.a.sigma.a,
d2.sigma.a2,
d2.sigma.a.lambda.a,d2.mu.a.lambda.a,d2.sigma.a.lambda.a,d2.lambda.a2),3,3)
I=list(I.b,I.a)
return(I)
}

```

# Referências

- ARELLANO-VALLE, R.; BOLFARINE, H.; LACHOS, V. Skew-normal linear mixed models. *Journal of Data Science*, v. 3, n. 4, p. 415–438, 2005.
- ARELLANO-VALLE, R. B.; AZZALINI, A. On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, v. 33, p. 561–574, 2006.
- ARELLANO-VALLE, R. B.; BRANCO, M. D.; GENTON, M. G. A unified view on skewed distributions arising from selections. *The Canadian Journal of Statistics*, v. 34, p. 581–601, 2006.
- ARELLANO-VALLE, R. B.; CASTRO, L. M.; LOSCHI, R. H. Change point detection in the skew-normal model parameters. *Communications in Statistics-Theory and Methods*, v. 42, p. 603–618, 2013.
- ARELLANO-VALLE, R. B.; GENTON, M. G. Multivariate unified skew-elliptical distributions. *Chilean Journal of Statistics*, v. 1, p. 17–33, 2010.
- AZZALINI, A. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, v. 12, p. 171–178, 1985.
- AZZALINI, A. The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, v. 32, p. 159–188, 2005.
- AZZALINI, A.; CAPITANIO, A. Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society*, v. 61, p. 579–602, 1999.
- BRANCO, M. D.; DEY, D. K. A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, v. 79, p. 99–113, 2001.
- CHEN, J. Testing for a change point in linear regression models. *Communications in Statistics - Theory and Methods*, v. 27, n. 10, p. 2481–2493, 1998.
- CHEN, J.; GUPTA, A. K. Detecting changes of mean in multidimensional normal sequences with application to literature and geology. *Computational Statistics*, v. 11, 1996. 211-221.
- CHEN, J.; GUPTA, A. K. Testing and locating variance change points with application to stock prices. *Journal of the American Statistical Association*, v. 92, 1997. 739-747.
- CHEN, J.; GUPTA, A. K. Information criterion and change point problem in regular models. *Department of Mathematics and Statistics, Bowling Green State University, Technical Report*, n. 98-05, 1998.
- CHEN, J.; GUPTA, A. K. Change point analysis of a gaussian model. *Statistical Papers*, v. 40, 1999. 323-333.
- CHEN, J.; GUPTA, A. K. On change point detection and estimation. *Communications in Statistics - Simulation and Computation*, v. 30, n. 3, p. 665–697, 2001.

- CHEN, J.; GUPTA, A. K. Information-theoretic approach for detecting change in the parameters of a normal model. *Mathematical Methods of Statistics*, v. 12, 2003. 116-130.
- CHEN, J.; GUPTA, A. K. *Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance*. New York: Springer Science & Business Media, 2011.
- COBB, G. W. The problem of the Nile: conditional solution to a changepoint problem. *Biometrika*, Biometrika Trust, v. 65, n. 2, p. 243–251, 1978.
- CONTRERAS, C. A. H. *Modelo de Regressão Linear Mistura de Escala Normal com Ponto de Mudança: Estimção e Diagnóstico*. Tese (Tese de Mestrado) — Departamento de Estatística, IMECC-UNICAMP, 2014.
- DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, v. 39, p. 1–38, 1977.
- GIRÓN, F. J.; MORENO, E.; CASELLA, G. *Objective bayesian analysis of multiple changepoints for linear models*. In *Bayesian Statistics 8*. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West (Eds). London: Oxford University Press, 2007.
- HACCOU, P.; MEELIS, E. Testing for the number of change points in a sequence of exponential random variables. *Journal of Statistical Computation and Simulation*, v. 30, 1988. 285-298.
- HACCOU, P.; MEELIS, E.; GEER, S. The likelihood ratio test for the change point problem for exponentially distributed random variables. *Stochastic Processes and Their Applications*, v. 27, 1988. 121-139.
- HASAN, A.; NING, W.; GUPTA, A. K. An information-based approach to the change-point problem of the noncentral skew t distribution with applications to stock market data. *Sequential Analysis*, Taylor & Francis, v. 33, n. 4, p. 458–474, 2014.
- HAWKINS, D. M.; QIU, P.; KANG, C. W. The changepoint model for statistical process control. *Journal of quality technology*, [Milwaukee]: American Society for Quality Control., v. 35, n. 4, p. 355–366, 2003.
- HOFRICHTER, J. *Change Point Detection in Generalized Linear Models*. Tese (Dissertation zur Erlangung des akademischen Grades Doktor der technischen Wissenschaften) — Graz University of Technology, 2007.
- HSU, D. Detecting shifts of parameter in gamma sequences with applications to stock price and air traffic flow analysis. *Journal of the American Statistical Association*, Taylor & Francis, v. 74, n. 365, p. 31–40, 1979.
- KANDER, Z.; ZACKS, S. Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points. *Annals of Mathematical Statistics*, v. 37, 1966. 11961210.
- LACHOS, V. H. *Modelos lineares mistos assimétricos*. Tese (Tese de Doutorado) — Departamento de Estatística, USP, 2004.

- LAI, T. L. Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 613–658, 1995.
- MORENO, E.; CASELLA, G.; GARCÍA-FERRER, A. Objective bayesian analysis of the changepoint problem. *Stochastic Environ. Res. Risk Assessment*, v. 19, p. 191–204, 2005.
- NGUNKENG, G.; NING, W. Information approach for the change-point detection in the skew-normal distribution and its applications. *Sequential Analysis*, v. 33, 2014. 475-490.
- OSORIO, F.; ; GALEA, M. Detection of a change-point in student-t linear regression models. *Statistical Papers*, v. 45, p. 31–48, 2005.
- PAGE, E. Continuous inspection schemes. *Biometrika*, JSTOR, v. 41, n. 1/2, p. 100–115, 1954.
- R Development Core Team. *R: A language and environment for statistical computing*. Vienna, Austria, 2011. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>.
- RODRÍGUEZ, C. L. B. *Inferência bayesiana no modelo normal assimétrico*. Tese (Doutorado) — Instituto de Matemática e Estatística da Universidade de São Paulo, 18/05/2005., 2005.
- SCHWARZ, G. Estimating the dimension of a model. *Ann. Statist.*, v. 6, p. 461–464, 1978.
- WORSLEY, K. J. Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika*, v. 73, 1986. 91-104.
- ZELLER, C. B.; LACHOS, V. H.; VILCA-LABRA, F. Local influence analysis for regression models with skew-normal independent distributions. *Journal of Applied Statistics*, v. 38, p. 343–368, 2011.