CAMILA BORELLI ZELLER

DISTRIBUIÇÕES MISTURAS DE ESCALA SKEW-NORMAL: ESTIMAÇÃO E DIAGNÓSTICO EM MODELOS LINEARES

Tese apresentada ao Instituto de Matemática, Estatística e Computação Científica, da Universidade Estadual de Campinas, para a obtenção do Título de Doutor em Estatística.

Orientador: Prof. Dr. Filidor E. Vilca Labra

Co-Orientador: Prof. Dr. Víctor Hugo Lachos Dávila

CAMPINAS

2009

---

† Este trabalho contou com apoio financeiro da CAPES.

Distribuições Misturas de Escala Skew-Normal: Estimação e Diagnóstico em Modelos Lineares

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida por Camila Borelli Zeller e aprovada pela comissão julgadora.

Campinas, 17 de dezembro de 2009

Prof. Dr. Filidor E. Vilca Labra
Orientador

Prof. Dr. Víctor Hugo Lachos Dávila
Co-orientador

Banca Examinadora:

1 Prof. Dra. Nancy Lopes Garcia (IMECC-UNICAMP)
2 Prof. Dr. Heleno Bolfarine (IME-USP)
3 Prof. Dra. Silvia Lopes de Paula Ferrari (IME-USP)
4 Prof. Dr. Manuel Galea-Rojas (Universidade de Valparaíso, Chile)

Tese apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP, como requisito parcial para obtenção do Título de DOUTOR em Estatística.

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IMECC DA UNICAMP
Bibliotecária: Miriam Cristina Alves – CRB8 / 5094

Zeller, Camila Borelli

Z38d Distribuições misturas de escala skew-normal: estimação e diagnóstico em modelos lineares / Camila Borelli Zeller -- Campinas, [S.P. : s.n.], 2009.

Orientadores : Filidor E. Vilca Labra; Víctor Hugo Lachos Dávila

Tese (Doutorado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Algoritmos de expectativa de maximização. 2. Distribuição normal assimetrica. 3. Influência local. 4. Misturas de escala. 5. Modelos lineares (Estatística). I. Labra, Filidor Edilfonso Vilca. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Título em inglês: Scale mixtures of skew-normal distributions: estimation and diagnostics for linear models.

Palavras-chave em inglês (Keywords): 1. EM algorithms. 2. Skew normal distribution. 3. Local influence. 4. Scale mixtures. 5. Linear models (Statistics).

Área de concentração: Métodos estatísticos

Titulação: Doutora em Estatística

Banca examinadora: Prof. Dr. Filidor Edilfonso Vilca Labra (IMECC-UNICAMP)
Profa. Dra. Nancy Lopes Garcia (IMECC-UNICAMP)
Prof. Dr. Heleno Bolfarine (IME-USP)
Profa. Dra. Silvia Lopes de Paula Ferrari (IME-USP)
Prof. Dr. Manuel Galea-Rojas (Universidade de Valparaíso, Chile)

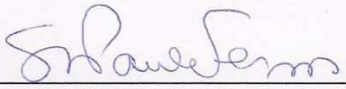
Data da defesa: 17/12/2009

Programa de Pós-Graduação: Doutorado em Estatística

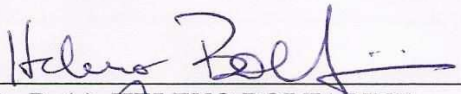
Tese de Doutorado defendida em 17 de dezembro de 2009 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.

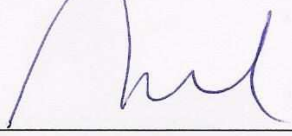
Prof(a). Dr(a). FILIDOR EDILFONSO VILCA LABRA



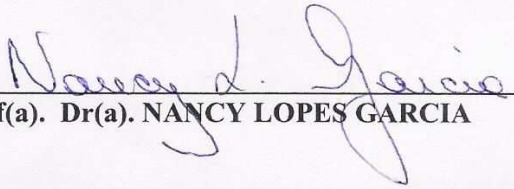
Prof(a). Dr(a). SILVIA LOPES DE PAULA FERRARI



Prof(a). Dr(a). HELENO BOLFARINE



Prof(a). Dr(a). MANUEL GALEA ROJAS



Prof(a). Dr(a). NANCY LOPES GARCIA

DEDICATÓRIA

Aos meus pais Carlos e Lourdes Helena com gratidão.
Aos meus inseparáveis irmãos Carlos e Caio pela compreensão.

Aos meus avós maternos Elóguia e Caio Duílio e à minha avó Celina com admiração.
Ao meu avô Charles (in memorian) com saudade.

Ao meu marido Marcelo e ao meu filho Pedro com eterna paixão.

Agradecimentos

- Ao professor Filidor pela confiança, sugestões, apoio e pela excelente orientação durante a elaboração deste trabalho.
- Ao professor Víctor Hugo pela co-orientação, sua experiência foi fundamental na obtenção dos resultados dessa tese. Admiro sua capacidade de pesquisa e sua presteza na colaboração com os outros.
- Aos professores do Departamento de Estatística pelos ensinamentos concedidos.
- Aos meus colegas da pós que dividiram esses anos de estudo e que de alguma forma colaboraram na minha formação acadêmica e na execução da tese.
- À Capes pelo apoio financeiro.

Resumo

Neste trabalho, estudamos alguns aspectos de estimação e diagnóstico de influência local (Cook, 1986) em modelos lineares, especificamente no modelo de regressão linear, no modelo linear misto e no modelo de Grubbs sob a classe de distribuições assimétricas misturas de escala skew-normal (SMSN) (Branco & Dey, 2001). Esta família de distribuições tem como membros particulares as versões simétrica e assimétrica das distribuições t-Student, slash e normal contaminada, todas com caudas mais pesadas que a distribuição normal. A estimação dos parâmetros será via o algoritmo EM (Dempster *et al.*, 1977) e a análise de diagnóstico será baseada na técnica de dados aumentados que usa a esperança condicional da função log-verossimilhança dos dados aumentados (função-Q) proveniente do algoritmo EM, como proposta por Zhu & Lee (2001) e Lee & Xu (2004).

Assim, pretendemos contribuir positivamente para desenvolvimento da àrea dos modelos lineares, estendendo alguns resultados encontrados na literatura, por exemplo, Pinheiro *et al.* (2001), Arellano-Valle *et al.* (2005), Osorio (2006), Montenegro *et al.* (2009a), Montenegro *et al.* (2009b), Osorio *et al.* (2009), Lachos *et al.* (2010), entre outros.

Palavras-chave: Algoritmo EM; Assimetria; Distância de Mahalanobis; Distribuições Misturas de Escala Skew-Normal; Influência Local; Modelos Lineares.

Abstract

In this work, we study some aspects of the estimation and the diagnostics based on the local influence (Cook, 1986) in linear models under the class of scale mixtures of the skew-normal (SMSN) distribution, as proposed by Branco & Dey (2001). Specifically, we consider the linear regression model, the linear mixed model and the Grubbs' measurement error model. The SMSN class of distributions provides a useful generalization of the normal and the skew-normal distributions since it covers both the asymmetric and heavy-tailed distributions such as the skew-t, the skew-slash, the skew-contaminated normal, among others. The local influence analysis will be based on the conditional expectation of the complete-data log-likelihood function (function-Q) from the EM algorithm (Dempster *et al.*, 1977)), as proposed by Zhu & Lee (2001) and Lee & Xu (2004).

We believe that the results of our work have contributed positively to the development of this area of linear models, since we have extended some results from the works of Pinheiro *et al.* (2001), Arellano-Valle *et al.* (2005), Osorio (2006), Montenegro *et al.* (2009a), Montenegro *et al.* (2009b), Osorio *et al.* (2009), Lachos *et al.* (2010), among others.

Key-words: EM Algorithm; Linear Models; Local Influence; Mahalanobis Distance; Scale Mixtures of Skew-Normal Distributions; Skewness.

Conteúdo

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

1.1 Motivação

Os modelos lineares são uma das técnicas mais populares em pesquisa, pois apresentam uma estrutura que permite aplicações em diversas áreas científicas, por exemplo, agricultura, biologia, ciências médicas, economia, entre outras. Na literatura, os estudos de inferência nos modelos lineares têm sido considerados sob a distribuição normal. Contudo, em muitas situações, inferências sob normalidade são impróprias, por exemplo, quando os dados provêem de uma distribuição com caudas mais ou menos pesadas que a distribuição normal ou ainda assimétrica. Modelos alternativos ao modelo normal que preservam a estrutura simétrica e que permitam reduzir a influência dos "outliers" têm sido sugeridos por muitos autores. Por exemplo, Lange *et al.* (1989) propõem o modelo t-Student, Yamaguchi (2001) sugere usar a distribuição normal contaminada e Osorio (2006) propõe as distribuições misturas de escala normal. Essas distribuições

são membros particulares de uma classe mais ampla e conhecida na literatura como *distribuições elípticas* (Fang *et al.*, 1990).

Do ponto de vista prático, muitos autores têm usado transformações de variáveis para alcançar normalidade ou pelo menos a simetria e em muitas situações, seus resultados são satisfatórios. Entretanto, Azzalini & Capitanio (1999) têm apontado alguns problemas. Por exemplo, as variáveis transformadas são mais difíceis de serem interpretadas, especialmente quando cada variável é transformada usando uma função diferente.

Embora a classe de distribuições elípticas represente uma boa alternativa à distribuição normal, ela não é adequada em situações nas quais a distribuição das observações é assimétrica. De fato, em muitos problemas práticos a distribuição empírica é unimodal com um certo grau de assimetria. Por exemplo, Hill & Dixon (1982) discutem e evidenciam numericamente a presença de assimetria em dados reais. Assim, é de interesse prático estudar distribuições que sejam menos sensíveis do que a distribuição gaussiana a certos desvios das suposições consideradas, construindo famílias paramétricas que sejam analiticamente tratáveis, que possam acomodar valores práticos de assimetria e curtose e que contenham estritamente a distribuição normal, como caso especial.

Quando a distribuição dos dados tem um comportamento assimétrico, esta pode ser modelada através de membros da classe de *distribuições assimétricas*, também conhecidas na literatura como *distribuições skew-elípticas*. Nesta classe, a distribuição skew-normal univariada é talvez a pioneira nesta idéia. Inicialmente, introduzida por O'hagan & Leornard (1976) como uma distribuição a priori em análise bayesiana e posteriormente, formalmente estudada por Azzalini (1985) (do ponto de vista clássico)

como uma extensão natural da distribuição normal para modelar a estrutura assimétrica presente nos dados. Usando esta idéia, outras distribuições assimétricas como skew-t, skew-cauchy, skew-slash, skew-normal contaminada e skew-exponencial potência foram desenvolvidas. Extensões para o caso multivariado podem ser encontradas em Azzalini & Dalla-Valle (1996), Azzalini & Capitanio (1999), Branco & Dey (2001), Sahu *et al.* (2003), Azzalini & Capitanio (2003), Lachos (2004), Genton (2004), Arellano-Valle & Genton (2005), Wang & Genton (2006), entre outros. Desses resultados, Arellano-Valle *et al.* (2005) consideram uma extensão multivariada da distribuição skew-normal proposta por Azzalini (1985) e a utilizam no contexto de modelos lineares mistos (LMM), mostrando que é possível obter uma forma fechada para a densidade marginal das quantidades observáveis de modo que a inferência clássica possa ser tratada usando técnicas padrões de otimização numérica. Lachos *et al.* (2007a) implementam o algoritmo EM para encontrar os estimadores de máxima verossimilhança dos parâmetros do modelo linear misto skew-normal (SN-LMM), sendo que o processo iterativo da etapa M tem forma fechada para todos os parâmetros envolvidos no modelo, estendendo, alguns resultados encontrados em Pinheiro & Bates (2000). Além disso, Montenegro *et al.* (2009a) desenvolvem o método de influência local no SN-LMM usando a metodologia proposta por Zhu & Lee (2001).

Estudos em modelos estatísticos sob a classe de distribuições assimétricas estão sendo aplicados com sucesso, tanto no contexto teórico quanto prático. Recentemente, Lachos *et al.* (2009) e Lachos *et al.* (2010) consideram as distribuições misturas de escala skew-normal no modelo de regressão multivariado com erros de medida e modelos lineares mistos, respectivamente. No contexto de aplicações dos modelos teóricos desenvolvidos, é importante estudar a sensibilidade dos resultados obtidos no processo de estimação dos parâmetros. Surge então a importância da metodologia de análise de

diagnóstico desses modelos ajustados.

Suponha que já assumimos um determinado modelo como correto. Dessa forma, temos interesse em analisar os dados com o intuito de observar se alguma observação particular controla propriedades importantes na estimação dos parâmetros. Um método para avaliar a influência de determinadas observações, sob certo esquema de perturbação, foi proposto por Cook (1986), conhecido na literatura como *método de influência local*. Esta técnica de diagnóstico tem sido aplicada com sucesso por muitos autores em modelos lineares sob a classe de distribuições simétricas misturas de escala normal, veja por exemplo, alguns resultados e referências recentes em Osorio (2006).

Alguns resultados sob a distribuição skew-normal podem ser encontrados, por exemplo, em Lachos *et al.* (2005) e Lachos *et al.* (2008) que aplicam a técnica de influência local no modelo de calibração comparativa e no modelo de regressão com erros nas variáveis com intercepto nulo, respectivamente. Apesar da abrangente aplicabilidade do método proposto por Cook (1986) em alguns casos, resulta difícil ou às vezes intratável aplicá-lo diretamente sendo necessário o uso de métodos alternativos. Uma alternativa interessante a abordagem de Cook (1986) foi proposta por Zhu & Lee (2001).

Na literatura estatística existem poucos trabalhos aplicados sobre estimação e análise de diagnóstico em modelos lineares sob a classe de distribuições assimétricas misturas de escala skew-normal, onde se baseia nosso objetivo. Na próxima seção, apresentamos uma breve descrição do algoritmo EM e em seguida mostraremos alguns critérios que serão utilizados para a selecionar distribuições dentro da classe misturas de escala skew-normal (SMSN). Posteriormente, apresentamos um resumo de métodos de diagnóstico de influência aplicados aos modelos lineares que serão considerados, especificamente,

modelo de regressão linear, modelo linear misto e modelo de Grubbs. Finalmente, descrevemos os objetivos específicos assim como a organização do trabalho.

1.2 Algoritmo EM

Neste trabalho, a obtenção do estimador de máxima verossimilhança (EMV) de $\boldsymbol{\theta}$ (parâmetro de interesse) será baseada no algoritmo EM. O algoritmo EM (Dempster *et al.*, 1977) é um enfoque amplamente aplicado no cálculo iterativo de estimativas 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 de estimativas de máxima verossimilhança. Os dados aumentados, também chamados dados completos, correspondem aos dados observados, referidos nesta formulação como dados incompletos, e dados adicionais conhecidos como dados perdidos ou não observáveis. Neste contexto, as funções de verossimilhança, baseadas nos dados completos e observados, são denominadas verossimilhança de dados completos e dados incompletos, respectivamente. É importante salientar que a parte aumentada dos dados não requer que eles sejam "perdidos" no sentido estrito da palavra, pois somente representam um mecanismo técnico. De fato, esta idéia é usada para descrever uma variedade de modelos estatísticos, tais como misturas, efeitos aleatórios, agrupamentos, censura e dados parcialmente observados (Liu, 1999).

Considere que $\mathbf{y}_{obs}$ denota os dados observados e $\mathbf{y}_{mis}$ denota os dados não observáveis, sendo os dados completos $\mathbf{y}_c = (\mathbf{y}_{obs}, \mathbf{y}_{mis})$, onde $\mathbf{y}_{obs}$ é acrescido de $\mathbf{y}_{mis}$. Denotamos $f(\mathbf{y}_c|\boldsymbol{\theta})$ a função de verossimilhança dos dados completos e $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) =$

$\log(f(\mathbf{y}_c|\boldsymbol{\theta}))$ a função log-verossimilhança dos dados completos. O algoritmo EM aborda problemas com dados incompletos indiretamente mediante a substituição da parte não observável em $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)$ por suas esperanças condicionais dado $\mathbf{y}_{obs}$, usando o ajuste atual para $\boldsymbol{\theta}$. Dessa forma, definimos $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ (função-Q) como a esperança da função log-verossimilhança dos dados completos, ou seja, $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = E\{\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}_{obs}, \widehat{\boldsymbol{\theta}}\}, \boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Cada iteração do algoritmo EM consiste de dois passos: esperança (passo E) e maximização (passo M). A $(t+1)$-ésima iteração do algoritmo EM é definida como

Passo E: Para $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}^{(t)}$, calcular $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(t)})$ como

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(t)}) = E\{\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}_{obs}, \widehat{\boldsymbol{\theta}}^{(t)}\};$$

Passo M: Obter $\widehat{\boldsymbol{\theta}}^{(t+1)}$ que maximize $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(t)})$, tal que

$$Q(\widehat{\boldsymbol{\theta}}^{(t+1)}|\widehat{\boldsymbol{\theta}}^{(t)}) > Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(t)}), \quad \forall \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Deve-se alternar os passos E e M repetidamente até atingir a convergência. Cada iteração do algoritmo EM incrementa o logaritmo da função de verossimilhança observada $\ell(\widehat{\boldsymbol{\theta}}|\mathbf{y}_{obs})$ de modo que $\ell(\widehat{\boldsymbol{\theta}}^{(t)}|\mathbf{y}_{obs}) \leq \ell(\widehat{\boldsymbol{\theta}}^{(t+1)}|\mathbf{y}_{obs})$ e o algoritmo tipicamente converge a um máximo local ou global da função de verossimilhança. Como critério de convergência, podemos utilizar $||\widehat{\boldsymbol{\theta}}^{(t+1)} - \widehat{\boldsymbol{\theta}}^{(t)}|| < \epsilon$, onde $||\mathbf{a}||$ indica a norma do vetor $\mathbf{a}$ e $\epsilon > 0$.

Quando o passo M do algoritmo EM é complicado, este pode ser amenizado realizando o processo de maximização condicional a alguma função dos parâmetros que estão sendo estimados. Este algoritmo EM generalizado proposto por Meng & Rubin (1993) é denominado algoritmo de maximização condicional de esperança (ECM). A

idéia neste caso é substituir o passo M do algoritmo EM por uma sequência de passos de maximização condicional (CM) computacionalmente mais simples, em que cada um deles maximiza a função-Q sujeita às restrições em $\boldsymbol{\theta}$. É importante notar que as propriedades de simplicidade, estabilidade e convergência monótona do algoritmo EM são compartilhadas pelo algoritmo ECM, porém com taxas de convergência mais velozes. Neste trabalho, denominados essa variante do algoritmo EM por algoritmo tipo-EM.

1.3 Seleção de Modelos

Apesar de não serem testes formais, como em Zhang & Davidian (2001), verificamos a adequação dos modelos SMSN aos dados inspecionando alguns critérios de informação. Três critérios foram selecionados: o critério de informação de Akaike (AIC, $-l(\hat{\boldsymbol{\theta}})+P$), o critério de informação bayesiano de Schawarz (BIC, $-l(\hat{\boldsymbol{\theta}})+0.5\log(n)P$) e o critério de Hannan-Quinn (HQ, $-l(\hat{\boldsymbol{\theta}})+\log(\log(n))P$), onde $l(\boldsymbol{\theta})$ é função log-verossimilhança, $\widehat{\boldsymbol{\theta}}$ é a estimativa de máxima verossimilhança de $\boldsymbol{\theta}$, n é o tamanho amostral e P é o número de parâmetros livres no modelo. Cada um desses critérios baseia-se numa penalização da verossimilhança na medida em que o modelo se torna mais complexo, isto é, modelos com um grande número de parâmetros. Dessa forma, o modelo que apresente o menor valor do critério de informação será o modelo selecionado.

1.4 Diagnóstico de Influência

Um dos principais objetivos da modelagem estatística é avaliar a qualidade do modelo postulado para representar o fenômeno sob estudo. Contudo, frequentemente existem observações que podem alterar substancialmente a inferência estatística quando são retiradas do conjunto de dados. Na literatura estatística tem havido portanto, grande

interesse na detecção de tais observações com o intuito de avaliar o impacto das mesmas no modelo ajustado.

Tais fatos têm motivado o desenvolvimento de duas principais linhas de pesquisa em diagnóstico de influência: identificação de observações influentes, mediante a eliminação de casos, também conhecida como técnica de influência global, veja Cook & Weisberg (1982) e Chatterjee & Hadi (1988) para mais detalhes, ou a técnica de infuência local (Cook, 1986) que permite avaliar a influência que pequenas perturbações podem exercer sobre os componentes do modelo (dados) usando uma medida de influência apropriada.

Embora a metodologia de influência local proposta por Cook (1986) venha sendo aplicada com sucesso em diferentes áreas da estatística, observamos que dependendo da complexidade do modelo, a aplicação dessa abordagem envolve extensas manipulações algébricas e em alguns casos um intenso trabalho computacional. Uma alternativa interessante à proposta de Cook (1986), e útil nos casos em que resulta difícil aplicar diretamente os métodos apresentados por Cook (1986), foi proposta por Zhu & Lee (2001) e é baseada na função de verossimilhança aumentada, resultante da implementação do algoritmo EM.

1.4.1 Influência Local

O objetivo do método de influência local é investigar o comportamento de alguma medida de influência $T(\boldsymbol{\omega})$ quando pequenas perturbações são introduzidas no modelo (ou dados) por meio de um vetor de perturbação $\boldsymbol{\omega} \in \boldsymbol{\Omega} \subset \Re^g$. Neste trabalho, usamos como medida de influência a função-Q (Zhu & Lee, 2001) para estudar a influência nos modelos lineares.

A seguir apresentaremos em forma resumida a abordagem de influência local de Zhu & Lee (2001) que é baseada na conhecida metodologia de influência local de Cook (1986). Considere o vetor de perturbação $\boldsymbol{\omega} = (w_1, ..., w_g)^\top$ variando numa região aberta em $\boldsymbol{\Omega} \subset \mathbb{R}^g$, seja $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_c)$ a função log-verossimilhança dos dados completos do modelo perturbado. Assumimos que existe um vetor de não perturbação $\boldsymbol{\omega}_o$, tal que $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}_o|\mathbf{Y}_c) = \ell_c(\boldsymbol{\theta}|\mathbf{Y}_c)$ para todo $\boldsymbol{\theta}$. Denotamos $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ como a função que maximiza $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = E[\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}]$. O gráfico de influência é definido como $\boldsymbol{\alpha}(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, f_Q(\boldsymbol{\omega}))^\top$, onde $f_Q(\boldsymbol{\omega})$ é a função Q-afastamento definida por

$$f_Q(\boldsymbol{\omega}) = 2\left[Q\left(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}\right) - Q\left(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})|\widehat{\boldsymbol{\theta}}\right)\right].$$

Segundo as metodologias desenvolvidas em Cook (1986), Zhu & Lee (2001) e Zhu & Lee (2003), a curvatura normal $C_{f_Q,\mathbf{d}}$ de $\boldsymbol{\alpha}(\boldsymbol{\omega})$ em $\boldsymbol{\omega}_o$ na direção do vetor unitário $\mathbf{d}$ é dada por

$$C_{f_Q,\mathbf{d}} = -2\mathbf{d}^\top \ddot{Q}_{\boldsymbol{\omega}_o}\mathbf{d}, \text{ onde } -\ddot{Q}_{\boldsymbol{\omega}_o} = \boldsymbol{\nabla}_{\boldsymbol{\omega}_o}^\top \left\{-\ddot{Q}_\theta(\widehat{\boldsymbol{\theta}})\right\}^{-1} \boldsymbol{\nabla}_{\boldsymbol{\omega}_o},$$

em que $\ddot{Q}_\theta(\widehat{\boldsymbol{\theta}}) = \left.\dfrac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ e $\boldsymbol{\nabla}_{\boldsymbol{\omega}} = \left.\dfrac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}$.

Como em Cook (1986), a expressão $-\ddot{Q}_{\boldsymbol{\omega}_o}$ é importante para detectar observações localmente influentes. A avaliação de casos influentes é baseada na inspeção visual de

$$\{M(0)_l = B_{f_Q,\mathbf{u}_l}; l = 1, \ldots, g\}$$

versus o índice l, onde $B_{f_Q,\mathbf{u}}(\boldsymbol{\theta}) = C_{f_Q,\mathbf{u}}(\boldsymbol{\theta})/tr[-2\ddot{Q}_{\boldsymbol{\omega}_o}]$ e $\mathbf{u}_l$ é um vetor de perturbação básica, tal que a $l-$ésima entrada é igual a 1 e as restantes iguais a 0. Veja Zhu & Lee (2001) para outras propriedades teóricas de $B_{f_Q,\mathbf{u}_l}$, tal como a invariância sob reparametrização de $\boldsymbol{\theta}$. Até agora, não há uma regra geral para julgar a magnitude da

influência de um caso específico nos dados. Recentemente, Lee & Xu (2004) propuseram usar $\overline{M(0)} + c^* SM(0)$ como marca de referência ("benchmark") para considerar as observações influentes, onde $\overline{M(0)}$ e $SM(0)$ são a média e o desvio padrão de $\{M(0)_l, l = 1, \ldots, g\}$, respectivamente e c^* é uma constante positiva arbitrária, veja Montenegro *et al.* (2009a) para mais detalhes, por exemplo.

1.5 "Outliers" e Robustez

Os termos "outliers" e robustez desempenham um papel importante nesta tese. O termo "outlier" muitas vezes é usado informalmente e Lucas (1997) ressalta que "outliers" são definidos com respeito aos modelos. Dessa forma, observações podem ser "outliers" em um modelo e serem perfeitamente regulares para outro modelo.

Um dos objetivos subjacentes às técnicas de estimação e diagnóstico consideradas nesta tese é o desenvolvimento de procedimentos sob a classe de distribuições SMSN que sejam robustos na presença de "outliers", ou seja, métodos estatísticos que não sejam afetados (ou menos afetados) por observações extremas, comumente chamadas de observações aberrantes ("outliers", anômalas ou atípicas). Uma das características dessa classe rica de modelos é que as distribuições SMSN podem naturalmente atribuir pesos diferentes para cada observação e consequentemente controlar a influência da observação no processo de estimação, por exemplo.

Para concluir a discussão com respeito aos termos "outliers" e robustez, discutiremos a seguir a relação de métodos robustos na presença de "outliers" com outros ramos da literatura estatística, especificamente diagnóstico de influência. As áreas de pesquisa relacionadas com métodos robustos na presença de "outliers" e diagnóstico (Cook &

Weisberg, 1982) consideram o problema de "outliers" sob perspectivas diferentes. Segundo Lucas (1997), talvez seja por este fato que, até recentemente, não exista muita interação entre esses campos de pesquisa.

Em diagnóstico, procedimentos são desenvolvidos para detectar "outliers" e observações influentes com o intuito de avaliar o impacto das mesmas no modelo ajustado. Definimos observações influentes como sendo os elementos do conjunto de dados que efetivamente alteram aspectos da análise. Por outro lado, a área de pesquisa em métodos robustos na presença de "outliers" considera um enfoque baseado na acomodação de observações aberrantes, possíveis observações influentes, através de procedimentos robustos, os quais têm sido desenvolvidos, por exemplo, modificando o método de máxima verossimilhança sob normalidade com o intuito de atenuar o efeito de tais observações nos resultados da estimação; veja Lucas (1997) para mais detalhes. Sob esse enfoque também podemos considerar as distribuições assimétricas com caudas mais pesadas do que as distribuições simétricas misturas de escala normal. Uma escolha interessante neste sentido corresponde à classe de distribuições SMSN.

Dessa forma, neste trabalho, consideramos uma interação entre os campos de pesquisa baseados em métodos robustos na presença de "outliers" e diagnóstico de influência, propondo extensões dos modelos lineares clássicos sob a classe de distribuições SMSN.

1.6 Definição dos Objetivos

O objetivo geral deste trabalho é apresentar um estudo de estimação e diagnóstico de influência nos modelos lineares sob a classe de distribuições assimétricas SMSN. A estimação dos parâmetros será via o algoritmo EM, a análise de influência local será

baseada na metodologia proposta por Zhu & Lee (2001) e os seguintes modelos lineares serão considerados:

- Modelo de Regressão Linear.
- Modelo Linear Misto.
- Modelo de Grubbs.

Podemos então relacionar os seguintes objetivos específicos:

1 Descrever a classe de distribuições assimétricas misturas de escala skew-normal, conforme discutida em Branco & Dey (2001), Lachos & Vilca (2007), Kim (2008) e Lachos *et al.* (2010), por exemplo. Em seguida, desenvolver um algoritmo para estimação dos parâmetros e imputação de valores em modelos SMSN para dados parcialmente observados ("missing values"), considerando as propriedades introduzidas para a classe SMSN. Esta linha de pesquisa está relacionada com os resultados encontrados em Lin *et al.* (2009a), por exemplo.

2 Discutir algumas extensões unificadas da classe de distribuições assimétricas misturas de escala skew-normal, generalizando as distribuições assimétricas introduzidas por Sahu *et al.* (2003) e Arellano-Valle *et al.* (2007), por exemplo, amplamente aplicadas em análises bayesianas.

3 Realizar um estudo de estimação e diagnóstico no modelo de regressão linear quando as observações seguem uma distribuição na classe SMSN, generalizando alguns resultados encontrados em Galea-Rojas *et al.* (2003) e Osorio *et al.* (2007), por exemplo.

4 Apresentar estudos de análise de diagnóstico no modelo linear misto sob a classe de distribuições SMSN, fornecendo um suplemento necessário para o trabalho de Lachos *et al.* (2010).

5 Desenvolver um estudo de estimação e diagnóstico no modelo de Grubbs sob a classe de distribuições SMSN, generalizando alguns resultados recentes encontrados em Osorio *et al.* (2009) e Montenegro *et al.* (2009b), por exemplo.

1.7 Apresentação dos Capítulos

No Capítulo 2, apresentamos a classe de distribuições misturas de escala skew-normal, originalmente introduzida por Branco & Dey (2001) e posteriormente estudada por Lachos & Vilca (2007) e Kim (2008), por exemplo, com a finalidade de descrever suas propriedades, tais como representação estocástica, momentos, formas quadráticas, entre outras. Adicionalmente, generalizamos alguns resultados encontrados em Lin & Lee (2008), úteis em aplicações com dados ausentes com estrutura SMSN. Em seguida, desenvolvemos um algoritmo simples para estimação dos parâmetros e imputação de valores em modelos SMSN quando os dados não são completamente observados, considerando as propriedades introduzidas para a classe de distribuições SMSN. Neste contexto, um estudo de simulação será considerado. Além disso, alguns resultados adicionais das distribuições SMSN são desenvolvidos com o intuito de introduzir uma nova distribuição que pode ser utilizada em finanças (Barndorff-Nielsen, 1997).

No Capítulo 3, propomos extensões unificadas da classe de distribuições SMSN e também desenvolvemos algumas propriedades, tais como momentos, transformações lineares, distribuição marginal, entre outras. Inicialmente, consideramos uma extensão da distribuição skew-normal proposta por Arellano-Valle *et al.* (2007) que faz da in-

ferência bayesiana uma alternativa viável para tratar os modelos lineares, estendendo em certo sentido alguns resultados desenvolvidos por Sahu *et al.* (2003), Arellano-Valle & Genton (2005) e Arellano-Valle *et al.* (2007). Um conjunto de dados reais será utilizado para ilustrar a utilidade dessa extensão proposta na modelagem de dados no contexto bayesiano. Posteriormente, realizamos um estudo inicial com a finalidade de introduzir algumas gene-ralizações das distribuições SMSN quando a variável de mistura U segue uma distribuição multivariada, estendendo alguns resultados encontrados em Lyu & Simoncelli (2008) e utilizados na modelagem de imagens.

No Capítulo 4, apresentamos o modelo de regressão linear sob a classe de distribuições misturas de escala skew-normal (SMSN-RM), incluindo o algoritmo EM para estimação por máxima verossimilhança e a matriz informação observada, útil no cálculo dos desvios padrões das estimativas dos parâmetros do modelo. A seguir desenvolvemos a metodologia de influência local pertinente ao SMSN-RM, sendo que quatro esquemas de perturbação são considerados. Exemplos numéricos considerando dados reais e simulados são apresentados para ilustrar a metodologia proposta.

No Capítulo 5, o modelo linear misto sob a classe de distribuições misturas de escala skew-normal (SMSN-LMM) é definido e um algoritmo EM para estimação por máxima verossimilhança é discutido, conforme descrito em Lachos *et al.* (2010). Posteriormente, desenvolvemos a metodologia de influência local pertinente ao SMSN-LMM, sendo que quatro esquemas de perturbação são considerados. A metodologia é ilustrada usando o conjunto de dados "Framingham Cholesterol", analisado inicialmente por Zhang & Davidian (2001).

No Capítulo 6, discutimos a formulação do modelo de Grubbs sob a classe de dis-

tribuições SMSN, denotado por modelo SMSN-G, bem como alguns aspectos inferenciais, incluindo o algoritmo tipo-EM e o cálculo da matriz informação observada. A metodologia de influência local para o modelo SMSN-G, considerando três esquemas de perturbação, é apresentada. Os resultados desenvolvidos são ilustrados usando o famoso conjunto de dados de Barnett (1969). O modelo de Grubbs pode ser entendido como um caso particular do modelo linear com efeitos mistos, contudo este modelo também pode ser considerado um caso particular do modelo de Barnett (1969). Salientamos que tal modelo não é um caso particular do modelo linear misto usual.

O Capítulo 7 finaliza esta tese com conclusões e diretrizes para trabalhos futuros.

Capítulo 2

Distribuições Misturas de Escala Skew-Normal

A distribuição normal é uma suposição rotineira na análise de dados, inclusive no contexto de dados parcialmente observados, mas tal condição pode ser irreal especialmente, para dados com forte assimetria e/ou caudas pesadas. Na prática, há um grande número de dados com assimetria e/ou caudas pesadas, por exemplo, dados de renda familiar, dados de contagem de CD4 provenientes de estudos associados à AIDS, dentre outros (Hill & Dixon, 1982). Dessa forma, este capítulo é motivado por uma considerável pesquisa que tem sido feita para introduzir famílias paramétricas flexíveis que acomodem desvios da suposição de normalidade e assim, amenizem a necessidade de transformações dos dados. Nessa linha de pesquisa, desenvolvemos alguns resultados adicionais para a classe de distribuições misturas de escala skew-normal. Além disso, propomos um algoritmo simples para a estimação dos parâmetros e imputação

de valores em modelos SMSN quando valo-res ausentes ("missing values") ocorrem nos dados, estendendo em certo sentido alguns resultados recentes desenvolvidos por Lin *et al.* (2009a) e Lin & Lin (2009) para as distribuições skew-normal propostas por Sahu *et al.* (2003) e Arellano-Valle *et al.* (2007), respectivamente.

2.1 Introdução

Modelagem estatística com dados parcialmente observados em análise multivariada (Beale & Little, 1975) é um problema importante que muitos pesquisadores encontram na prática. Métodos de estimação para o modelo normal multivariado a partir de dados parcialmente observados (Liu, 1999) estão amplamente desenvolvidos e estudados na literatura, uma vez que a distribuição normal desempenha um papel proeminente na análise estatística multivariada. Vários programas estatísticos como o SAS contém procedimentos disponíveis para a imputação de valores ausentes a partir do modelo normal e outras abordagens recentes também estão disponíveis.

A classe de distribuições misturas de escala normal (SMN) (Andrews & Mallows, 1974) tem sido usada em diversas aplicações estatísticas para inferência robusta, inclusive no contexto de dados parcialmente observados; veja, por exemplo, Little & Rubin (1987), Liu (1999), Lin *et al.* (2006) e Lin *et al.* (2009b). Esta classe compreende as distribuições normal, t-Student, normal contaminada, slash, entre outras. Apesar de em muitas situações a suposição de normalidade ou simetria ser considerada com sucesso, tal suposição pode ser irreal especialmente, para dados com forte assimetria e/ou caudas pesadas.

Branco & Dey (2001) generalizam a classe de distribuições SMN combinando as-

simetria com caudas pesadas e denotam esta classe por misturas de escala skew-normal (SMSN); veja também Lachos & Vilca (2007) e Kim (2008). Esta extensão resulta uma classe flexível de distribuições para modelos multivariados, uma vez que tem como casos especiais a distribuição skew-normal (SN) (Azzalini & Dalla-Valle, 1996) e as distribuições simétricas definidas por Andrews & Mallows (1974). Além disso, esta classe contém as distribuições skew-t, skew-slash e skew-normal contaminada, por exemplo, todas com caudas mais pesadas que a distribuição skew-normal (e normal) e dessa forma, podem ser usadas para inferência robusta em muitos tipos de modelos.

De uma forma geral, neste capítulo, descrevemos as distribuições misturas de escala skew-normal multivariadas, destacamos algumas de suas propriedades discutidas em Branco & Dey (2001), Lachos & Vilca (2007), Kim (2008), Lachos *et al.* (2009) e Lachos *et al.* (2010), por exemplo e generalizamos alguns resultados encontrados em Lin & Lee (2008), úteis em aplicações com dados ausentes com estrutura SMSN. Além disso, alguns resultados adicionais das distribuições SMSN foram propostos com o objetivo de introduzir uma nova distribuição. Distribuição atraente, pois pode ser utilizada no contexto de finanças; veja, por exemplo, Barndorff-Nielsen (1997).

Este capítulo está organizado como segue. Na próxima seção, definimos as distribuições misturas de escala skew-normal e destacamos algumas propriedades, tais como momentos e representação estocástica. Além disso, apresentamos o algoritmo EM para estimação dos parâmetros do modelo SMSN, como discutido em Lachos & Vilca (2007). Na Seção 2.3, definimos o modelo SMSN no contexto de dados parcialmente observados e propomos um algoritmo para estimação dos parâmetros e para a imputação de valores ausentes. Na Seção 2.4, um estudo de simulação é considerado. Em seguida, na Seção 2.5, apresentamos alguns resultados adicionais para a classe de

distribuições misturas de escala skew-normal. Finalmente, algumas observações finais são dadas na Seção 2.6. Provas técnicas dos resultados desenvolvidos neste capítulo estão no Apêndice A.

2.2 Distribuições Misturas de Escala Skew-Normal

Nesta seção, definimos as distribuições SMSN e destacamos algumas propriedades interessantes que são compartilhadas por todos os elementos dessa classe. Como discutido por Arellano-Valle & Genton (2005), existem diversas definições de distribuição skew-normal. Dessa forma, inicialmente, definiremos a distribuição skew-normal (SN) que será conside-rada na construção da classe de distribuições SMSN.

Um vetor aleatório p-dimensional $\mathbf{W}$ segue uma distribuição skew-normal com vetor de locação $\boldsymbol{\mu} \in \mathbb{R}^p$, matriz de dispersão $\boldsymbol{\Sigma}$ (definida positiva $p \times p$) e vetor de assimetria $\boldsymbol{\lambda} \in \mathbb{R}^p$ se sua função densidade de probabilidade (pdf) é dada por

$$f(\mathbf{w}) = 2\phi_p(\mathbf{w}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi(\mathrm{A}), \tag{2.1}$$

onde $\mathrm{A} = \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{w} - \boldsymbol{\mu})$, $\phi_p(.|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ é a pdf da normal p-variada com vetor de médias $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$ ($N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$) e $\Phi(.)$ representa a função distribuição acumulada (cdf) da normal padrão, tal que $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1}$. Neste caso, denotamos por $\mathbf{W} \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ quando $\mathbf{W}$ tem pdf dada em (2.1). Exceto por uma simples diferença na parametrização considerada em (2.1), este modelo corresponde ao introduzido por Azzalini & Dalla-Valle (1996), cujas propriedades estão extensivamente estudadas em Azzalini & Capitanio (1999) e em Arellano-Valle & Genton (2005).

Considere $\mathbf{Z} = \mathbf{W} - \boldsymbol{\mu}$ e como $a\mathbf{Z} \sim SN_p(\mathbf{0}, a^2\boldsymbol{\Sigma}, \boldsymbol{\lambda})$, para todo $a > 0$, a classe de dis-

tribuições misturas de escala skew-normal pode ser definida como segue. Considerando a construção da classe de distribuições normal/independente introduzida por Lange & Sinsheimer (1993), temos que um vetor aleatório p-dimensional com distribuição SMSN é construído a partir da mistura escala de uma variável aleatória skew-normal e uma variável aleatória positiva, i.e.,

$$\mathbf{Y} = \boldsymbol{\mu} + \kappa^{1/2}(U)\mathbf{Z}, \tag{2.2}$$

onde $\boldsymbol{\mu}$ é um vetor locação, $\mathbf{Z} \sim SN_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, $\kappa(.)$ é uma função de pesos e U é a variável aleatória de mistura positiva com cdf $H(u;\boldsymbol{\nu})$ e pdf $h(u;\boldsymbol{\nu})$, independente de $\mathbf{Z}$, cuja distribuição está indexada por um vetor de parâmetros ou escalar $\boldsymbol{\nu}$ conhecido (Lucas, 1997) ou desconhecido que controla as caudas das distribuições. Neste trabalho, usamos as seguintes notações equivalentes $H(u;\boldsymbol{\nu})$, $H(u)$ ou H para nos referir a cdf da variável aleatória de mistura U.

De (2.2), temos que dado $U = u$, $\mathbf{Y}$ segue uma distribuição $SN_p(\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma}, \boldsymbol{\lambda})$. Assim, integrando em u, a pdf de $\mathbf{Y}$ é dada por

$$f(\mathbf{y}) \;=\; 2\int_0^\infty \phi_p(\mathbf{y}|\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma})\Phi(\kappa^{-1/2}(u)\mathrm{A})dH(u;\boldsymbol{\nu}), \quad \mathbf{y} \in \mathbb{R}^p. \tag{2.3}$$

Para uma vetor aleatório com pdf como em (2.3), denotaremos por $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$. Note que quando $\kappa(u) = 1/u$, a distribuição de $\mathbf{Y}$ pertence à classe de distribuições skew-normal/independente discutida em Lachos & Vilca (2007) e Lachos *et al.* (2010), por exemplo. Por sua vez, quando $\boldsymbol{\lambda} = \mathbf{0}$, a distribuição $SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ corresponde à classe simétrica de distribuições misturas de escala normal, denotada por $SMN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$; veja Andrews & Mallows (1974) para mais detalhes no contexto univariado. Além disso, se $\kappa(u) = 1/u$, a distribuição de $\mathbf{Y}$ pertence à classe de distribuições normal/independente discutida em Lange & Sinsheimer (1993).

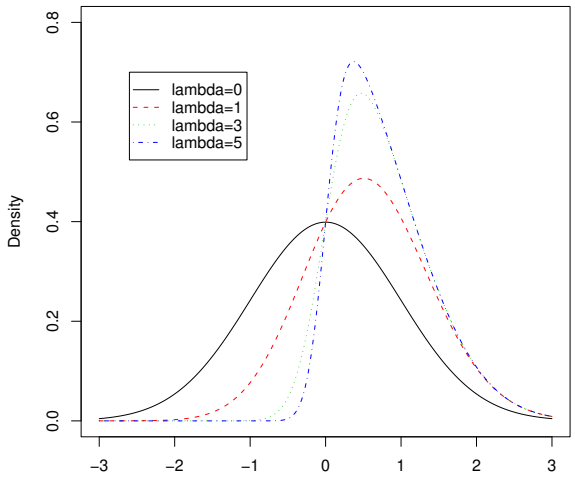


Figura 2.1: Dados simulados. Densidade da skew-normal univariada padrão para diferentes assimetrias.

Note que a densidade dada em (2.3) se reduz para a densidade da skew-normal, definida em (2.1), quando assumimos $\kappa(U) = 1$. Sem perda de generalidade, exemplos da densidade da skew-normal univariada padrão ($\mu = 0$ e $\Sigma = \sigma^2 = 1$), denotada por $SN_1(\lambda)$ ou $SN(\lambda)$, para $\lambda = 0, 1, 3$ e 5 são dados na Figura 2.1. Quando $\lambda = 0$, a densidade da skew-normal padrão coincide com a densidade da normal padrão. Como pode ser visto, valores positivos de λ introduzem assimetria positiva na densidade. De modo análogo, valores negativos de λ introduzem assimetria negativa na densidade.

Algumas propriedades das distribuições SMSN são:

- Permite combinar assimetria com caudas pesadas;
- Admite uma representação estocástica que permite uma fácil implementação do

algoritmo EM e estudos de suas propriedades;

- Devido às propriedades interessantes que tem a classe de distribuições SMSN, esta família de distribuições permite a acomodação de observações aberrantes e consequentemente a obtenção de procedimentos estatísticos robustos, considerando tanto assimetria e caudas mais ou menos pesadas que a distribuição skew-normal;

- Contém como casos particulares as distribuiçõs skew-normal, skew-t, skew-slash, skew-normal contaminada, skew-exponencial potência e outras que podem ser geradas mudando a distribuição da variável aleatória positiva;

- Permite estender os modelos desenvolvidos sob normalidade considerando distribuições simétricas ou assimétricas com caudas mais pesadas do que a normal.

De acordo com Lachos & Vilca (2007), um vetor aleatório p-dimensional com pdf dada em (2.3) possui representação estocástica dada por

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \kappa^{1/2}(U)\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\delta}|T_0| + (\mathbf{I}_n - \boldsymbol{\delta}\boldsymbol{\delta}^\top)^{1/2}\mathbf{T}_1), \quad \text{com} \quad \boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}}, \tag{2.4}$$

onde " $\stackrel{d}{=}$ " significa "equivalência em distribuição", $|T_0|$ denota o valor absoluto de T_0, $U \sim H(\cdot;\boldsymbol{\nu})$, $T_0 \sim \mathrm{N}_1(0,1)$ e $\mathbf{T}_1 \sim \mathrm{N}_p(\mathbf{0},\mathbf{I}_p)$ são todas variáveis independentes. A representação em (2.4) além de facilitar a implementação do algoritmo EM pode ser usada também para derivar muitas propriedades da distribuição de $\mathbf{Y}$, por exemplo, da forma quadrática distância de Mahalanobis $d_\lambda = (\mathbf{Y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})$, útil para verificar o ajuste e detectar "outliers" (Pinheiro *et al.*, 2001).

Além disso, segundo Lachos *et al.* (2010), apresentamos alguns resultados relacionados aos momentos condicionais de funções de U dado $\mathbf{y}$, denotados por u_r e η_r. Resultados que serão essenciais para a implementação do algoritmo EM e encontram-se resumidos a seguir.

Proposição 2.2.1 *Considere que* $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ *e* $U \sim H$ *é a variável de mistura. Então,*

$$\begin{aligned} u_r &= E[\kappa^{-r}(U)|\mathbf{y}] = \frac{2f_0(\mathbf{y})}{f(\mathbf{y})} E[\kappa^{-r}(U_\mathbf{y})\Phi(\kappa^{-1/2}(U_\mathbf{y})A)], \\ \eta_r &= E[\kappa^{-r/2}(U)W_\Phi(\kappa^{-1/2}(U)A)|\mathbf{y}] = \frac{2f_0(\mathbf{y})}{f(\mathbf{y})} E[\kappa^{-r/2}(U_\mathbf{y})\phi_1(\kappa^{-1/2}(U_\mathbf{y})A)], \end{aligned}$$

onde f_0 *é a pdf de* $\mathbf{Y}_0 \sim SMN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$, $U_\mathbf{y} \overset{d}{=} U|\mathbf{Y}_0 = \mathbf{y}$, $W_\Phi(x) = \phi_1(x)/\Phi(x)$, $x \in \mathbb{R}$ *e* $A = \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$.

Prova A prova deste resultado segue diretamente do Lema A.1.1 dado no Apêndice A, considerando $g(u) = \kappa^{-r}(u)$ e $g(u) = \kappa^{-r/2}(u)W_\Phi(\kappa^{-1/2}(u)\mathrm{A})$, respectivamente.

Segundo Lachos & Vilca (2007), temos que o vetor aleatório SMSN é invariante sob transformações lineares o que implica que a distribuição marginal de $\mathbf{Y}$ ainda pertence à classe de distribuições SMSN.

Proposição 2.2.2 *Considere que* $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$. *Então, para quaisquer vetor fixo* $\mathbf{b} \in \mathbb{R}^m$ *e matriz* $\mathbf{B} \in \mathbb{R}^{m \times p}$ *de posto completo,*

$$\mathbf{V} = \mathbf{b} + \mathbf{B}\mathbf{Y} \sim SMSN_m(\mathbf{b} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^\top, \boldsymbol{\lambda}^*; H),$$

onde $\boldsymbol{\lambda}^* = \boldsymbol{\delta}^*/(1 - \boldsymbol{\delta}^{*\top}\boldsymbol{\delta}^*)^{1/2}$, *com* $\boldsymbol{\delta}^* = (\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^\top)^{-1/2}\mathbf{B}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$. *Além disso, se* $m = p$ *a matriz* $\mathbf{B}$ *é não singular, então* $\boldsymbol{\lambda}^* = \boldsymbol{\lambda}$

Prova A prova segue da Proposição 5.4 em Branco & Dey (2001). Quando $\mathbf{B}$ é uma matriz não singular, é fácil ver que $\boldsymbol{\delta}^* = \boldsymbol{\delta}$.

Corolário 2.2.3 *Considere que* $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ *e que* $\mathbf{Y}$ *está particionado como* $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ *com dimensões* p_1 *e* p_2 $(p_1 + p_2 = p)$, *respectivamente. Considere*

$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top$ *e* $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^\top, \boldsymbol{\lambda}_2^\top)^\top$ *sendo as correspondentes partições de* $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$ *e* $\boldsymbol{\lambda}$. *Então, a densidade marginal de* $\mathbf{Y}_1$ *é* $SMSN_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{11}^{1/2}\widetilde{\boldsymbol{\upsilon}}; H)$, *com* $\widetilde{\boldsymbol{\upsilon}} = \dfrac{\boldsymbol{\upsilon}_1 + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\upsilon}_2}{\sqrt{1 + \boldsymbol{\upsilon}_2^\top \boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2}}$, *onde* $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ *e* $\boldsymbol{\upsilon} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda} = (\boldsymbol{\upsilon}_1^\top, \boldsymbol{\upsilon}_2^\top)^\top$.

Prova A prova segue da Proposição 2.2.2, considerando $\mathbf{B} = [\mathbf{I}_{p_1}, \mathbf{0}_{p_2}]$ e $\mathbf{b}$=$\mathbf{0}$.

De acordo com Lachos *et al.* (2010), o próximo resultado pode ser útil em aplicações com modelos lineares. Por exemplo, quando o modelo linear depende de um vetor não-observável de efeitos aleatórios e de um vetor de erros aleatórios, onde os efeitos aleatórios seguem distribuição SMSN e os erros seguem distribuição SMN (modelo linear misto).

Proposição 2.2.4 *Sob a notação definida no Corolário 2.2.3, considerando* $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21} = \mathbf{0}$ *e* $\boldsymbol{\lambda}_2 = \mathbf{0}$ *segue que dado* $U = u$, $\mathbf{Y}_1$ *e* $\mathbf{Y}_2$ *são independentes. Além disso, se* $\boldsymbol{\mu}_2 = \mathbf{0}$ *então* $\mathbf{Y}_1$ *e* $\mathbf{Y}_2$ *são não correlacionados.*

Prova Dado $U = u$, temos que $\mathbf{Y} \sim SN_p(\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma}, \boldsymbol{\lambda})$ e segue da Proposição 6 em Azzalini & Capitanio (1999) que $\mathbf{Y}_1$ e $\mathbf{Y}_2$ são independentes. O caso em que $\boldsymbol{\mu}_2 = 0$ segue da Proposição 2.2.2, tal que $\mathbf{Y}_1 \sim SMSN_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\lambda}_1; H)$ e $\mathbf{Y}_2 \sim SMN_{p_2}(\mathbf{0}, \boldsymbol{\Sigma}_{22}; H)$. Dessa forma, $Cov[\mathbf{Y}_1, \mathbf{Y}_2] = E[\mathbf{Y}_1\mathbf{Y}_2^\top] = E[E[\mathbf{Y}_1\mathbf{Y}_2^\top|U]] = E[E[\mathbf{Y}_1|U]E[\mathbf{Y}_2^\top|U]] = \mathbf{0}$ o que conclui a prova.

Considerando os momentos condicionais de $\mathbf{Y}_2$ dado $\mathbf{Y}_1 = \mathbf{y}_1$ desenvolvidos por Lin & Lee (2008) para a distribuição skew-normal proposta por Azzalini & Dalla-Valle (1996), propomos uma extensão desses resultados no contexto das distribuições SMSN.

Proposição 2.2.5 *Considere a notação do Corolário 2.2.3. Se* $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ *então,*

(a) O primeiro momento de $\mathbf{Y}_2$ *condicionado a* $\mathbf{Y}_1 = \mathbf{y}_1$ *e* $U = u$ *é dado por*

$$E[\mathbf{Y}_2|\mathbf{y}_1, u] = \boldsymbol{\mu}_{2.1} + \kappa^{1/2}(u) W_\Phi(\kappa^{-1/2}(u)\widetilde{\boldsymbol{\upsilon}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))\boldsymbol{G},$$

onde $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$, $\boldsymbol{G} = \dfrac{\boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2}{\sqrt{1 + \boldsymbol{\upsilon}_2^\top \boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2}}$ *e* W_Φ *como na Proposição 2.2.1.*

(b) O segundo momento de $\mathbf{Y}_2$ *condicionado a* $\mathbf{Y}_1 = \mathbf{y}_1$ *e* $U = u$ *é dado por*

$$\begin{aligned} E[\mathbf{Y}_2\mathbf{Y}_2^\top|\mathbf{y}_1, u] &= \boldsymbol{\mu}_{2.1}\boldsymbol{\mu}_{2.1}^\top + \kappa(u)\boldsymbol{\Sigma}_{22.1} \\ &+ \kappa^{1/2}(u) W_\Phi(\kappa^{-1/2}(u)\widetilde{\boldsymbol{\upsilon}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1)) \left\{\boldsymbol{\mu}_{2.1}\boldsymbol{G}^\top + \boldsymbol{G}\boldsymbol{\mu}_{2.1}^\top - \boldsymbol{G}\boldsymbol{G}^\top\right\}. \end{aligned}$$

Prova Veja Apêndice A.2.

O resultado a seguir pode ser útil em aplicações com dados que não são completamente observados, por exemplo, especificando $\mathbf{Y}_2 = \mathbf{Y}^m$ e $\mathbf{Y}_1 = \mathbf{Y}^o$, onde $\mathbf{Y}^m$ e $\mathbf{Y}^o$ denotam as partes não observadas e observadas de $\mathbf{Y}$, respectivamente. Pelo Corolário 2.2.6, um preditor para a parte não observada é dado por (2.5).

Corolário 2.2.6 *Se* $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ *então, o primeiro momento de* $\mathbf{Y}_2$ *condicionado a* $\mathbf{Y}_1 = \mathbf{y}_1$ *é dado por*

$$E[\mathbf{Y}_2|\mathbf{y}_1] = \boldsymbol{\mu}_{2.1} + \frac{\boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2}{\sqrt{1 + \boldsymbol{\upsilon}_2^\top \boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2}}\eta_{-1}, \tag{2.5}$$

com η_{-1} *definido na Proposição 2.2.1.*

Prova A prova segue diretamente da Proposição 2.2.5 e do fato que $E[\mathbf{Y}_2|\mathbf{y}_1] = E_U[E[\mathbf{Y}_2|\mathbf{y}_1, U]|\mathbf{y}_1]$.

Na proposição seguinte, apresentamos o vetor de médias e a matriz de covariâncias de um vetor aleatório SMSN, conforme descritos em Lachos & Vilca (2007). Adicionalmente, desenvolvemos o coeficiente de curtose multidimensional para o vetor aleatório SMSN, representando uma extensão do coeficiente de curtose proposto por Azzalini & Capitanio (1999).

Proposição 2.2.7 *Suponha que* $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$. *Então,*

a) Se $E[\kappa^{1/2}(U)] < \infty$, *temos que* $E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}} E[\kappa^{1/2}(U)]\boldsymbol{\Delta}$;

b) Se $E[\kappa(U)] < \infty$, *temos que* $Var[\mathbf{Y}] = \boldsymbol{\Sigma}_y = E[\kappa(U)]\boldsymbol{\Sigma} - \frac{2}{\pi} E^2[\kappa^{1/2}(U)]\boldsymbol{\Delta}\boldsymbol{\Delta}^\top$;

c) Se $E[\kappa^2(U)] < \infty$, *temos que o coeficiente de curtose multidimensional é*

$$\gamma_2(\mathbf{Y}) = \frac{E[\kappa^2(U)]}{E^2[\kappa(U)]} a_{1y} - 4\frac{E[\kappa^{3/2}(U)]}{E^2[\kappa(U)]} a_{2y} + a_{3y} - p(p+2),$$

onde $a_{1y} = p(p+2)+2(p+2)\boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1}\boldsymbol{\mu}_y + 3(\boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1}\boldsymbol{\mu}_y)^2$, $a_{2y} = \left(p + \frac{2}{E[\kappa^{1/2}(U)]}\right)\boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1}\boldsymbol{\mu}_y + \left(1 + \frac{2}{E[\kappa^{1/2}(U)]} - \frac{\pi}{2}\frac{E[\kappa(U)]}{E^2[\kappa^{1/2}(U)]}\right)(\boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1}\boldsymbol{\mu}_y)^2$ *e* $a_{3y} = 2(p+2)\boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1}\boldsymbol{\mu}_y + 3(\boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1}\boldsymbol{\mu}_y)^2$, *com* $\boldsymbol{\mu}_y = E[\mathbf{Y} - \boldsymbol{\mu}] = \sqrt{\frac{2}{\pi}} E[\kappa^{1/2}(U)]\boldsymbol{\Delta}$ *e* $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$.

Prova A prova de a) e b) segue de (2.2). Para obter a expressão em c), usamos a definição do coeficiente de curtose multivariado introduzido por Mardia (1970). Sem perda de generalidade, consideramos que $\boldsymbol{\mu} = \mathbf{0}$, assim $\boldsymbol{\mu}_y = E[\mathbf{Y}] = \sqrt{\frac{2}{\pi}} E[\kappa^{1/2}(U)]\boldsymbol{\Delta}$. Inicialmente, note que a curtose é definida por $\gamma_2(\mathbf{Y}) = E[\{(\mathbf{Y} - \boldsymbol{\mu}_y)^\top \boldsymbol{\Sigma}_y^{-1}\mathbf{Y} - \boldsymbol{\mu}_y)\}^2]$. Considerando a representação estocástica de $\mathbf{Y}$ dada em (2.2), temos que

$$(\mathbf{Y} - \boldsymbol{\mu}_y)^\top \boldsymbol{\Sigma}_y^{-1}\mathbf{Y} - \boldsymbol{\mu}_y) \stackrel{d}{=} \kappa(U)\mathbf{Z}^\top \boldsymbol{\Sigma}_y^{-1}\mathbf{Z} - 2\kappa^{1/2}(U)\mathbf{Z}^\top \boldsymbol{\Sigma}_y^{-1}\boldsymbol{\mu}_y + \boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1}\boldsymbol{\mu}_y,$$

onde $\mathbf{Z} \sim SN_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. De acordo com a definição de $\gamma_2(\mathbf{Y})$ e após algumas manipulações algébricas, a prova segue usando os dois primeiros momentos da forma quadrática

(Genton *et al.*, 2001) e o Lema A.1.2 dado no Apêndice A. Note que sob a distribuição skew-normal, i.e., quando $\kappa(U) = 1$, o coeficiente de curtose multidimensional reduz para $\gamma_2(\mathbf{Y}) = 2(\pi - 3)(\boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1} \boldsymbol{\mu}_y)^2$, justamente o coeficiente de curtose para um vetor aleatório skew-normal; veja, por exemplo, Azzalini & Capitanio (1999).

A seguir apresentamos alguns casos especiais das distribuições SMSN e para cada um desses elementos, destacamos os momentos definidos nas Proposições 2.2.1 e 2.2.7, conforme descritos em Lachos *et al.* (2009) e Lachos *et al.* (2010).

- **Distribuição Skew-t Multivariada**

  A distribuição skew-t multivariada (Azzalini & Capitanio, 2003) com ν graus de liberdade, $ST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; \nu)$, pode ser obtida de (2.3) considerando $\kappa(u) = 1/u$, com $U \sim Gamma(\nu/2, \nu/2)$, $u > 0$, $\nu > 0$. A pdf de $\mathbf{Y}$ é dada por

  $$f(\mathbf{y}) = 2t_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu) T\left(\sqrt{\frac{p+\nu}{d_\lambda + \nu}} \mathrm{A}; \nu + p\right), \quad \mathbf{y} \in \mathbb{R}^p,$$

  onde $t_p(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$ e $T(\cdot; \nu)$ especificam, respectivamente, a pdf da distribuição t-Student p-variada e a cdf da distribuição t-Student padrão univariada. Um caso particular da distribuição skew-t é a distribuição skew-cauchy quando $\nu = 1$. Além disso, quando $\nu \uparrow \infty$, obtemos a distribuição skew-normal como caso limite. Neste caso, da Proposição 2.2.7, o vetor de médias e a matriz de covariâncias são dadas por, respectivamente,

  $$\begin{aligned} E[\mathbf{Y}] &= \boldsymbol{\mu} + \sqrt{\frac{\nu}{\pi}} \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \boldsymbol{\Delta}, \quad \nu > 1 \text{ e} \\ Var[\mathbf{Y}] &= \frac{\nu}{\nu - 2} \boldsymbol{\Sigma} - \frac{\nu}{\pi} \left(\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \right)^2 \boldsymbol{\Delta}\boldsymbol{\Delta}^\top, \quad \nu > 2. \end{aligned}$$

  Além disso, segue da Proposição 2.2.1 que $\mathbf{Y}_0 \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$, i.e. $\mathbf{Y}_0|U = u \sim N_p(\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})$ e $U \sim Gamma(\nu/2, \nu/2)$. Dessa forma, considerando o fato bem

conhecido que $U_{\mathbf{y}} \stackrel{d}{=} U|\mathbf{Y}_0 = \mathbf{y} \sim Gamma((\nu+p)/2, (\nu+\mathrm{d}_\lambda)/2)$, após algumas manipulações algébricas, temos que

$$\begin{aligned} u_r &= \frac{f_0(\mathbf{y})}{f(\mathbf{y})} \frac{2^{r+1}\Gamma(\frac{\nu+p+2r}{2})(\nu+\mathrm{d}_\lambda)^{-r}}{\Gamma(\frac{\nu+p}{2})} T\left(\sqrt{\frac{\nu+p+2r}{\nu+\mathrm{d}_\lambda}}\mathrm{A}; \nu+p+2r\right) \quad \text{e} \\ \eta_r &= \frac{f_0(\mathbf{y})}{f(\mathbf{y})} \frac{2^{(r+1)/2}\Gamma(\frac{\nu+p+r}{2})}{\pi^{1/2}\Gamma(\frac{\nu+p}{2})} \frac{(\nu+\mathrm{d}_\lambda)^{(\nu+p)/2}}{(\nu+\mathrm{d}_\lambda+\mathrm{A}^2)^{(\nu+p+r)/2}}, \end{aligned}$$

onde $\Gamma(.)$ é a função gama. Aplicações da distribuição skew-t em procedimentos de estimação robustos também podem ser encontradas em Lin & Lee (2008), Zhou & He (2008) e Azzalini & Genton (2008).

- **Distribuição Skew-Slash Multivariada**

Outra distribuição SMSN é a distribuição skew-slash, denotada por $SSL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; \nu)$, quando $\kappa(u) = 1/u$ e a distribuição de U é $Beta(\nu, 1)$, $0 < u < 1$ e $\nu > 0$. De (2.3), a pdf de $\mathbf{Y}$ é dada por

$$f(\mathbf{y}) = 2\nu \int_0^1 u^{\nu-1} \phi_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}) \Phi(u^{1/2}\mathrm{A}) du, \quad \mathbf{y} \in \mathbb{R}^p.$$

A distribuição skew-slash reduz para a distribuição skew-normal quando $\nu \uparrow \infty$. Neste caso, da Proposição 2.2.7, temos que

$$\begin{aligned} E[\mathbf{Y}] &= \boldsymbol{\mu} + \sqrt{\frac{2}{\nu}} \frac{2\nu}{2\nu-1} \boldsymbol{\Delta}, \quad \nu > 1/2 \quad \text{e} \\ Var[\mathbf{Y}] &= \frac{\nu}{\nu-1}\boldsymbol{\Sigma} - \frac{2}{\pi}\left(\frac{2\nu}{2\nu-1}\right)^2 \boldsymbol{\Delta}\boldsymbol{\Delta}^\top, \quad \nu > 1. \end{aligned}$$

Os momentos condicionais u^r e η^r para a distribuição skew-slash são dados a seguir. A prova segue considerando que $U_{\mathbf{y}} \sim Gamma((2\nu+p+2r)/2, d_\lambda/2)\mathbb{I}_{(0,1)}$

na Proposição 2.2.1 e $\mathbb{I}_A$ denota a função indicadora do conjunto A. Assim,

$$u_r = \frac{f_0(\mathbf{y})}{f(\mathbf{y})}\frac{2\Gamma(\frac{2\nu+p+2r}{2})}{\Gamma(\frac{2\nu+p}{2})}\left(\frac{2}{\mathrm{d}_\lambda}\right)^r \frac{P_1\left(\frac{2\nu+p+2r}{2}, \frac{\mathrm{d}_\lambda}{2}\right)}{P_1\left(\frac{p+2\nu}{2}, \frac{d_\lambda}{2}\right)} E\{\Phi(S^{1/2}\mathrm{A})\} \quad \text{e}$$

$$\eta_r = \frac{f_0(\mathbf{y})}{f(\mathbf{y})}\frac{2^{r/2+1/2}\Gamma\left(\frac{2\nu+p+r}{2}\right)}{\Gamma\left(\frac{2\nu+p}{2}\right)\pi^{1/2}}\frac{\mathrm{d}_\lambda^{(2\nu+p)/2}}{(\mathrm{d}_\lambda+\mathrm{A}^2)^{(2\nu+p+r)/2}}\frac{P_1\left(\frac{2\nu+p+r}{2}, \frac{\mathrm{d}_\lambda+\mathrm{A}^2}{2}\right)}{P_1\left(\frac{p+2\nu}{2}, \frac{d_\lambda}{2}\right)},$$

onde $P_x(a,b)$ denota a cdf da distribuição $Gamma(a,b)$ avaliada em x e $S \sim Gamma((2\nu+p+2r)/2, d_\lambda/2)\mathbb{I}_{(0,1)}$. Note que $\mathrm{E}\{\Phi(S^{1/2}\mathrm{A})\}$ pode ser calculada via integração de Monte Carlo, gerando observações independentes $S_1, \ldots, S_L$ de S e aproximando o valor esperado pela média $\frac{1}{L}\sum_{i=1}^{L}\Phi(S_i^{1/2}A)$. Observações provenientes da distribuição gama truncada podem ser geradas, por exemplo, usando o programa estatístico R (library `Runuran`) através da função `urgamma`. Aplicações da distribuição skew-slash podem ser encontradas em Wang & Genton (2006).

- **Distribuição Skew-Normal Contaminada Multivariada**

  Esta distribuição será denotada por $SCN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; \nu, \gamma)$, onde $0 \leq \nu \leq 1$, $0 < \gamma \leq 1$. Considerando $\kappa(u) = 1/u$ e U uma variável aleatória discreta assumindo dois estados, tal que sua função densidade de probabilidade dado $\boldsymbol{\nu} = (\nu, \gamma)^\top$ é denotada por

$$h(u; \boldsymbol{\nu}) = \nu\mathbb{I}_{(u=\gamma)} + (1-\nu)\mathbb{I}_{(u=1)}, \quad 0 < \nu < 1,\ 0 < \gamma \leq 1.$$

  Segue de (2.3) que

$$f(\mathbf{y}) = 2\{\nu\phi_p(\mathbf{y}|\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Sigma})\Phi(\gamma^{1/2}A) + (1-\nu)\phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi(A)\}, \quad \mathbf{y} \in \mathbb{R}^p.$$

A distribuição skew-normal contaminada reduz para a distribuição skew-normal quando $\gamma = 1$. Assim, da Proposição 2.2.7, o vetor de médias e matriz de covariâncias são dados por, respectivamente,

$$\begin{aligned} E[\mathbf{Y}] &= \boldsymbol{\mu} + \sqrt{\frac{2}{\nu}}\left(\frac{\nu}{\gamma^{1/2}} + 1 - \nu\right)\boldsymbol{\Delta} \quad \text{e} \\ Var[\mathbf{Y}] &= \left(\frac{\nu}{\gamma} + 1 - \nu\right)\boldsymbol{\Sigma} - \frac{2}{\pi}\left(\frac{\nu}{\gamma^{1/2}} + 1 - \nu\right)^2 \boldsymbol{\Delta}\boldsymbol{\Delta}^\top. \end{aligned}$$

Neste caso, considerando que $U_{\mathbf{y}}$ é uma variável aleatória discreta com função de probabilidade condicional $h_0(u|\mathbf{y}) = (1/f_0(\mathbf{y}))\{\nu\phi_p(\mathbf{y};\boldsymbol{\mu},\gamma^{-1}\boldsymbol{\Sigma})\mathbb{I}_{(u=\gamma)} + (1-\nu)\phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma})\mathbb{I}_{(u=1)}\}$. Da Proposição 2.2.1, temos que

$$\begin{aligned} u_r &= \frac{2}{f(\mathbf{y})}\left\{\nu\gamma^r\phi_p(\mathbf{y}|\boldsymbol{\mu},\gamma^{-1}\boldsymbol{\Sigma})\Phi(\gamma^{1/2}\mathrm{A}) + (1-\nu)\phi_p(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\Sigma})\Phi(\mathrm{A})\right\} \quad \text{e} \\ \eta_r &= \frac{2}{f(\mathbf{y})}\left\{\nu\gamma^{r/2}\phi_p(\mathbf{y}|\boldsymbol{\mu},\gamma^{-1}\boldsymbol{\Sigma})\phi_1(\gamma^{1/2}\mathrm{A}) + (1-\nu)\phi_p(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\Sigma})\phi_1(\mathrm{A})\right\}. \end{aligned}$$

Na Figura 2.2, mostramos as densidades da $SN_1(3)$, da $ST_1(3;2)$, da $SSL_1(3;1)$ e da $SCN_1(3;0.5,0.5)$ padrões univariadas, i.e., $\mu = 0$, $\Sigma = \sigma^2 = 1$ e $p = 1$. Alteramos a escala das densidades para que tivessem mesmo valor na origem. Note que as quatro distribuições são assimétricas positivas, sendo que as distribuições skew-slash e skew-t têm caudas mais pesadas que a distribuição skew-normal.

Em seguida, descrevemos algumas propriedades da forma quadrática distância de Mahalanobis $d_\lambda = (\mathbf{Y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})$. Mais detalhes sobre propriedades das formas quadráticas podem ser encontradas em Lange & Sinsheimer (1993), Lachos & Vilca (2007) e Kim (2008), por exemplo.

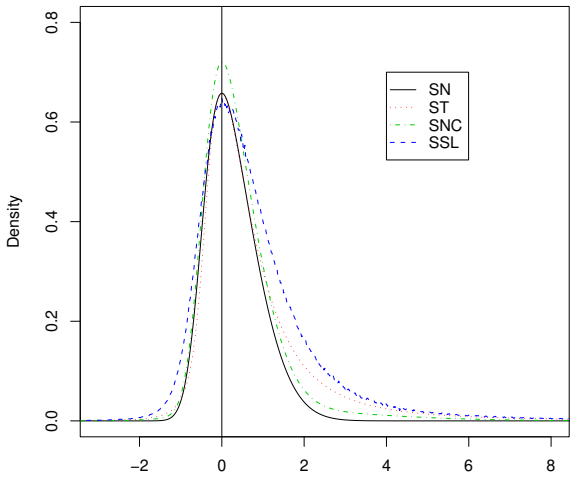


Figura 2.2: Dados simulados. Densidades da skew-normal (SN), da skew-t (ST), da skew-slash (SSL) e da skew-normal contaminada (SCN) padrões univariadas.

Proposição 2.2.8 *Considere que* $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$. *Desse modo, a forma quadrática* $d_\lambda = (\mathbf{Y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu})$ *tem a mesma distribuição de* $d = (\mathbf{Y}_0 - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y}_0 - \boldsymbol{\mu})$, *onde* $\mathbf{Y}_0 \sim SMN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$. *Além disso, para qualquer* $m > 0$

$$E[d_\lambda^m] = \frac{2^m \Gamma(m + p/2)}{\Gamma(p/2)} E[\kappa^m(U)].$$

Prova A prova segue da Proposição 5.2 em Branco & Dey (2001) junto com os resultados encontrados na Seção 2 em Lange & Sinsheimer (1993).

Dessa forma, a Proposição 2.2.8 afirma que $d_\lambda \sim \chi^2_p$ para o caso skew-normal, $d_\lambda \sim pF(p, \nu)$ para o caso skew-t, $Pr(d_\lambda \leq \upsilon) = Pr(\chi^2_p \leq \upsilon) - \frac{2^\nu \Gamma(\nu + p/2)}{\upsilon^\nu \Gamma(p/2)} Pr(\chi^2_{2\nu+p} \leq \upsilon)$ para a skew-slash e $Pr(d_\lambda \leq \upsilon) = \nu Pr(\chi^2_p \leq \gamma\upsilon) + (1 - \nu) Pr(\chi^2_p \leq \upsilon)$ para a skew-normal contaminada. Este resultado é interessante, pois permite avaliar os modelos

estatísticos na prática. Substituindo as estimativas de máxima verossimilhança de $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ na distância de Mahalanobis d_λ, podemos avaliar os ajustes dos modelos através da construção de envelopes (Montenegro *et al.*, 2009b). Além disso, através de gráficos da distância de Mahalanobis e considerando como marca de referência o quantil v da distribuição da forma quadrática d_λ, podemos identificar "outliers" (Pinheiro *et al.*, 2001). Por exemplo, para o caso skew-normal, temos que $v = \chi^2_p(\xi)$, onde $0 < \xi < 1$. Nos próximos capítulos, iremos considerar por simplicidade $d_\lambda = d$.

Na próxima seção, apresentamos o algoritmo EM para obter as estimativas de máxima verossimilhança dos parâmetros do modelo SMSN, conforme descrito em Lachos & Vilca (2007). Este algoritmo será desenvolvido em conjunto com as propriedades dadas acima, em particular a representação estocástica do modelo SMSN definida em (2.4) e os momentos condicionais apresentados na Proposição 2.2.1.

2.2.1 Algoritmo EM

Suponha que temos n observações independentes, $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, onde

$$\mathbf{Y}_i \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H), i = 1, \ldots, n. \tag{2.6}$$

O vetor de parâmetros é dado por $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\alpha}^\top, \boldsymbol{\lambda}^\top)^\top$, onde $\boldsymbol{\alpha}$ denota o conjunto minimal de parâmetros, tal que a matriz $\boldsymbol{\Sigma}$ esteja bem definida. Como já mencionado, a cdf H está indexada por $\boldsymbol{\nu}$ e neste trabalho, assumimos $\boldsymbol{\nu}$ conhecido, como discutido em Lange & Sinsheimer (1993) e Lucas (1997). Os parâmetros do modelo SMSN serão estimados via algoritmo EM e para explorá-lo, usamos os resultados desenvolvidos na Seção 2.2. Dessa forma, de (2.4), o modelo SMSN pode ser reescrito hierarquicamente

como

$$\mathbf{Y}_i|T_i = t_i, U_i = u_i, \overset{\text{ind}}{\sim} N_p(\boldsymbol{\mu} + \boldsymbol{\Delta} t_i, \kappa(u_i)\boldsymbol{\Gamma}), \tag{2.7}$$

$$T_i|U_i = u_i \overset{\text{ind}}{\sim} HN_1(0, \kappa(u_i)), \tag{2.8}$$

$$U_i \overset{\text{iid}}{\sim} H(\boldsymbol{\nu}), \tag{2.9}$$

$i = 1, \ldots, n$, todos independentes, onde $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$, $\boldsymbol{\Gamma} = \boldsymbol{\Sigma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top$ e $HN_1(0, \sigma^2)$ denota a distribuição half-$N(0, \sigma^2)$ (Johnson *et al.*, 1994).

Neste processo de estimação, considere $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, o vetor de respostas observáveis para n unidades amostrais, $\mathbf{u} = (u_1, \ldots, u_n)^\top$ e $\mathbf{t} = (t_1, \ldots, t_n)^\top$. Então, sob a representação hierárquica (2.7)–(2.9), segue que a função log-verossimilhança completa associada com $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{u}^\top, \mathbf{t}^\top)^\top$ é

$$\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = c - \frac{n}{2}\log|\boldsymbol{\Gamma}| - \frac{1}{2}\sum_{i=1}^{n}\kappa^{-1}(u_i)(\mathbf{y}_i - \boldsymbol{\mu} - \boldsymbol{\Delta} t_i)^\top\boldsymbol{\Gamma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu} - \boldsymbol{\Delta} t_i),$$

onde c é uma constante que independe do vetor de parâmetros $\boldsymbol{\theta}$. Considere $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\boldsymbol{\mu}}^{(k)\top}, \widehat{\boldsymbol{\alpha}}^{(k)\top}, \widehat{\boldsymbol{\lambda}}^{(k)\top})^\top$ a estimativa de $\boldsymbol{\theta}$ na k-ésima iteração.

Após manipulações algébricas, a esperança condicional da função log-verossimilhança completa é dada por $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$, onde

$$\begin{aligned} Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) \;=\; & c - \frac{n}{2}\log|\boldsymbol{\Gamma}| - \frac{1}{2}\sum_{i=1}^{n}\widehat{u}_i^{(k)}(\mathbf{y}_i - \boldsymbol{\mu})^\top\boldsymbol{\Gamma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}) + \sum_{i=1}^{n}\widehat{ut_i}^{(k)}(\mathbf{y}_i - \boldsymbol{\mu})^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta} \\ & -\frac{1}{2}\sum_{i=1}^{n}\widehat{ut^2}_i^{(k)}\boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta}, \end{aligned}$$

com $\widehat{u}_i = \mathrm{E}[\kappa^{-1}(U_i)|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i]$, $\widehat{ut}_i = \mathrm{E}[\kappa^{-1}(U_i)T_i|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i]$ e $\widehat{ut_i^2} = \mathrm{E}[\kappa^{-1}(U_i)T_i^2|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i]$. Usando propriedades conhecidas da esperança condicional, as quantidades $\widehat{u}_i$, $\widehat{ut}_i$ e $\widehat{ut_i^2}$

podem ser facilmente obtidas dos resultados apresentados na Proposição 2.2.1. Veja mais detalhes em Lachos & Vilca (2007). Portanto, temos o seguinte algoritmo EM:

Passo E: Dado $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, calcule $\widehat{ut^2}_i^{(k)}$, $\widehat{ut}_i^{(k)}$ e $\widehat{u}_i^{(k)}$, $i = 1, \ldots, n$.

Passo M: Atualize $\widehat{\boldsymbol{\theta}}^{(k+1)}$ maximizando $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ sobre $\boldsymbol{\theta}$, o que leva às seguintes expressões fechadas:

$$
\begin{aligned}
\widehat{\boldsymbol{\mu}}^{(k+1)} &= \frac{\sum_{i=1}^{n}(\widehat{u}_i^{(k)}\mathbf{y}_i - \widehat{ut}_i^{(k)}\boldsymbol{\Delta}^{(k)})}{\sum_{i=1}^{n}\widehat{u}_i^{(k)}}, \\
\widehat{\boldsymbol{\Gamma}}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^{n}\Big[\widehat{u}_i^{(k)}(\mathbf{y}_i - \boldsymbol{\mu}^{(k)})(\mathbf{y}_i - \boldsymbol{\mu}^{(k)})^\top - \widehat{ut}_i^{(k)}\boldsymbol{\Delta}^{(k)}(\mathbf{y}_i - \boldsymbol{\mu})^\top \\
&\quad - \widehat{ut}_i^{(k)}(\mathbf{y}_i - \boldsymbol{\mu}^{(k)})\boldsymbol{\Delta}^{\top(k)} + \widehat{ut^2}_i^{(k)}\boldsymbol{\Delta}^{(k)}\boldsymbol{\Delta}^{(k)\top}\Big], \\
\widehat{\boldsymbol{\Delta}}^{(k+1)} &= \frac{\sum_{i=1}^{n}\widehat{ut}_i^{(k)}(\mathbf{y}_i - \boldsymbol{\mu}^{(k)})}{\sum_{i=1}^{n}\widehat{ut^2}_i^{(k)}} \\
\widehat{\mathbf{D}}^{(k+1)} &= \widehat{\boldsymbol{\Gamma}}^{(k+1)} + \widehat{\boldsymbol{\Delta}}^{(k+1)}\widehat{\boldsymbol{\Delta}}^{\top(k+1)} \\
\widehat{\boldsymbol{\lambda}}^{(k+1)} &= \widehat{\mathbf{D}}^{(k+1)-1/2}\widehat{\boldsymbol{\Delta}}^{(k+1)}/(1 - \widehat{\boldsymbol{\Delta}}^{\top(k+1)}\widehat{\mathbf{D}}^{(k+1)-1}\widehat{\boldsymbol{\Delta}}^{(k+1)})^{1/2}.
\end{aligned}
$$

Note que quando $\boldsymbol{\lambda} = \mathbf{0}$ (ou $\boldsymbol{\Delta} = \mathbf{0}$) as equações do passo M se reduzem às equações obtidas quando assumimos as distribuições SMN.

2.2.2 Matriz Informação Observada

A função de log-verossimilhança para $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\alpha}^\top, \boldsymbol{\lambda}^\top)^\top \in \mathbb{R}^q$ dada a amostra observada $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, tal que $\mathbf{Y}_i$ está definido em (2.6), é dada por $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n}\ell_i(\boldsymbol{\theta})$, onde $\ell_i(\boldsymbol{\theta}) = \log 2 - \frac{p}{2}log2\pi - \frac{1}{2}\log|\boldsymbol{\Sigma}| + \log K_i$, com

$$K_i = K_i(\boldsymbol{\theta}) = \int_0^\infty \kappa^{-p/2}(u_i)\exp\{-\frac{1}{2}\kappa^{-1}(u_i)d_i\}\Phi(\kappa^{-1/2}(u_i)A_i)dH(u_i;\boldsymbol{\nu}),$$

$d_i = (\mathbf{y}_i - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})$ e $A_i = \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})$. Usando a seguinte notação $I_i^F(w) = \int_0^\infty \kappa^{-w}(u_i)\exp\{-\frac{1}{2}\kappa^{-1}(u_i)d_i\}F(\kappa^{-1/2}(u_i)A_i)dH(u_i;\boldsymbol{\nu})$, onde $F(.)$ é a função

$\Phi(.)$ ou $\phi(.) = \phi_1(.)$, tal que $\phi(.)$ denota a pdf da normal padrão, podemos escrever $K_i(\boldsymbol{\theta}) = I_i^{\Phi}(\frac{p}{2})$, $i = 1, \ldots, n$. Dessa forma, a matriz de segundas derivadas com respeito a $\boldsymbol{\theta}$ é dada por

$$\mathbf{L} = \sum_{i=1}^{n} \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = -\frac{n}{2} \frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} - \sum_{i=1}^{n} \frac{1}{K_i^2} \frac{\partial K_i}{\partial \boldsymbol{\theta}} \frac{\partial K_i}{\partial \boldsymbol{\theta}^\top} + \sum_{i=1}^{n} \frac{1}{K_i} \frac{\partial^2 K_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top},$$

onde

$$\frac{\partial K_i}{\partial \boldsymbol{\theta}} = I_i^{\phi}(\frac{p+1}{2}) \frac{\partial A_i}{\partial \boldsymbol{\theta}} - \frac{1}{2} I_i^{\Phi}(\frac{p+2}{2}) \frac{\partial d_i}{\partial \boldsymbol{\theta}}$$

e

$$\begin{aligned} \frac{\partial^2 K_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} &= \frac{1}{4} I_i^{\Phi}(\frac{p+4}{2}) \frac{\partial d_i}{\partial \boldsymbol{\theta}} \frac{\partial d_i}{\partial \boldsymbol{\theta}^\top} - \frac{1}{2} I_i^{\Phi}(\frac{p+2}{2}) \frac{\partial^2 d_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \\ &\quad -\frac{1}{2} I_i^{\phi}(\frac{p+3}{2}) (\frac{\partial A_i}{\partial \boldsymbol{\theta}} \frac{\partial d_i}{\partial \boldsymbol{\theta}^\top} + \frac{\partial d_i}{\partial \boldsymbol{\theta}} \frac{\partial A_i}{\partial \boldsymbol{\theta}^\top}) - I_i^{\phi}(\frac{p+3}{2}) A_i \frac{\partial A_i}{\partial \boldsymbol{\theta}} \frac{\partial A_i}{\partial \boldsymbol{\theta}^\top} \\ &\quad +I_i^{\phi}(\frac{p+1}{2}) \frac{\partial^2 A_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}. \end{aligned}$$

Para os casos particulares skew-t, skew-slash e skew-normal contaminada, as integrais acima podem ser obtidas.

- *Skew-t.*

$$\begin{aligned} I_i^{\Phi}(w) &= \frac{2^w \nu^{\nu/2} \Gamma(w + \nu/2)}{\Gamma(\nu/2)(\nu + d_i)^{\nu/2+w}} T_1\left(\frac{A_i}{(d_i+\nu)^{1/2}} \sqrt{2w+\nu}; 2w+\nu\right) \quad \text{e} \\ I_i^{\phi}(w) &= \frac{2^w \nu^{\nu/2}}{\sqrt{2\pi}\Gamma(\nu/2)} \left(\frac{1}{d_i + A_i^2 + \nu}\right)^{\frac{\nu+2w}{2}} \Gamma(\frac{\nu+2w}{2}). \end{aligned}$$

- *Skew-Slash.*

$$\begin{aligned} I_i^{\Phi}(w) &= \frac{\nu 2^{w+\nu} \Gamma(w+\nu)}{d_i^{w+\nu}} P_1(w+\nu, \frac{d_i}{2}) E[\Phi(S_i^{1/2} A_i)] \quad \text{e} \\ I_i^{\phi}(w) &= \frac{\nu 2^{w+\nu} \Gamma(w+\nu)}{\sqrt{2\pi}(d_i + A_i^2)^{w+\nu}} P_1(w+\nu, \frac{d_i + A_i^2}{2}), \end{aligned}$$

onde $S_i \sim Gamma(w+\nu, \frac{d_i}{2}) \mathbb{I}_{(0,1)}$.

- *Skew-Normal Contaminada.*

$$I_i^{\Phi}(w) = \sqrt{2\pi}\{\nu\gamma^{w-1/2}\phi_1(\sqrt{d_i}|0,\frac{1}{\gamma})\Phi(\gamma^{1/2}A_i) + (1-\nu)\phi_1(\sqrt{d_i}|0,1)\Phi(A_i)\} \quad \text{e}$$
$$I_i^{\phi}(w) = \nu\gamma^{w-1/2}\phi_1(\sqrt{d_i+A_i^2}|0,\frac{1}{\gamma}) + (1-\nu)\phi_1(\sqrt{d_i+A_i^2}).$$

As derivadas de $\log\boldsymbol{\Sigma}$, d_i e A_i envolvem algumas manipulações algébricas e podem ser encontradas em Lachos & Vilca (2007). Dessa forma, $\mathbf{J} = -\mathbf{L}$ denota a matriz de informação observada para a log-verossimilhança marginal $\ell(\boldsymbol{\theta})$. Na prática, $\mathbf{J}$ é usualmente desconhecida e a substituímos por $\widehat{\mathbf{J}}$ que é a matriz $\mathbf{J}$ avaliada na estimativa de máxima verossimilhança de $\boldsymbol{\theta}$.

Na próxima seção, propomos estudar os modelos SMSN no contexto de dados ausentes, considerando um algoritmo para estimação dos parâmetros e para a imputação de valores ausentes a partir do modelo SMSN será discutido.

2.3 Dados Ausentes com Estrutura Misturas de Escala Skew-Normal

Inferência estatística com dados ausentes é um importante problema aplicado uma vez que valores parcialmente observados (planejados ou não planejados) são comumente encontrados na prática. Os recentes avanços computacionais e teóricos fazem desse tópico uma área de pesquisa estatística ativa. Nesta seção, consideramos dados ausentes com estrutura misturas de escala skew-normal.

Suponha que $\mathbf{Y} = (\mathbf{Y}_1^\top, \ldots, \mathbf{Y}_n^\top)^\top$ é um vetor aleatório $np \times 1$, tal que $n > p$ e $\mathbf{Y}_i \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$. Desejamos desenvolver ferramentas estatísticas para ana-

lisar os modelos SMSN no contexto de dados ausentes, no sentido que observações de cada $\mathbf{Y}_i$ podem não ser completamente observadas.

De acordo com a notação definida em Lin *et al.* (2009a), considere que $\mathbf{Y}_i$ pode ser particionado como $(\mathbf{Y}_i^o, \mathbf{Y}_i^m)$, onde $\mathbf{Y}_i^o(p_i^o \times 1)$ e $\mathbf{Y}_i^m\ ((p - p_i^o) \times 1)$ denotam as porções observadas e não observadas de $\mathbf{Y}_i$, respectivamente. Para tal especificação, temos que os resultados discutidos nas Seções 2.2, 2.2.1 e 2.2.2 são válidos e em particular, o Corolário 2.2.6 será útil em aplicações com dados que não são completamente observados.

2.3.1 Aspectos Computacionais: Algoritmo

Note que do Corolário 2.2.6, sob mecanismos de dados ausentes ao acaso (MAR: "missing at random"), podemos formular um algoritmo simples para o cálculo das estimativas de máxima verossimilhança dos parâmetros e realizar imputações para valores ausentes com estrutura SMSN. Dessa forma, considerando que $\mathbf{Y}_2 = \mathbf{Y}^m$ é o componente não observado de $\mathbf{Y}$, um algoritmo baseado no método de estimação proposto na Seção 2.2.1 e no Corolário 2.2.6 é formulado a seguir.

Considere $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\boldsymbol{\mu}}^{(k)\top}, \widehat{\boldsymbol{\alpha}}^{(k)\top}, \widehat{\boldsymbol{\lambda}}^{(k)\top})^\top$ a estimativa de $\boldsymbol{\theta}$ na k-ésima iteração para $k = 0, 1, 2, \ldots$. Com $k = 0$, para um conjunto de dados parcialmente observado, simplesmente preenchemos os dados ausentes pela $E[\mathbf{Y}^m|\mathbf{y}^o]$, definida em (2.5), das variáveis correspondentes. Então,

Passo 1: Com os dados preenchidos e $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, obtemos uma nova estimativa $\widehat{\boldsymbol{\theta}}^{(k+1)}$ através do algoritmo EM para dados completamente observados; veja Seção 2.2.1 e

Passo 2: Com $\widehat{\boldsymbol{\theta}}^{(k+1)}$, os valores ausentes podem ser preditos através $E[\mathbf{Y}^m|\mathbf{y}^o]$.

Os passos são repetidos até que uma regra de convergência seja satisfeita, por exemplo, $||\widehat{\boldsymbol{\theta}}^{(k+1)}/\widehat{\boldsymbol{\theta}}^{(k)} - 1||$. Valores iniciais são necessários para implementar este procedimento e os momentos amostrais podem ser considerados.

Este algoritmo permite a realização de inferências para dados ausentes com estrutura misturas de escala skew-normal e considera as incertezas dos parâmetros causadas pelos dados parcialmente observados. Por outro lado, uma crítica frequente é que as incertezas associadas às imputações de valores ausentes não são fornecidas na análise. Contornando este problema, os desvios padrões das predições dos valores ausentes podem ser derivados dos resultados desenvolvidos na Proposição 2.2.5. De acordo com Nielsen (1997), a normalidade assintótica ainda é válida quando os valores ausentes ocorrem ao acaso (MAR). Dessa forma, os desvios padrões das estimativas dos parâmetros podem ser obtidos invertendo a matriz informação observada, definida na Seção 2.2.2, considerando os dados ausentes preenchidos pela última iteração do algoritmo.

2.4 Estudos de Simulação

Com o intuito de examinar o desempenho dos métodos propostos, apresentamos alguns estudos de simulação. Os dados simulados são obtidos facilmente usando a representação estocástica dada em (2.4) com $\kappa(U) = 1/U$ e $U \sim H$, sendo H especificada conforme o interesse; veja Seção 2.2 para mais detalhes.

2.4.1 Estudo I: Algoritmo EM

Nesta seção, para avaliar o desempenho do algoritmo EM apresentado na Seção 2.2.1, conduzimos um estudo de simulação. Neste caso, consideramos apenas o modelo ST,

tal que $\mathbf{Y}_i \sim ST_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; \nu)$, com $\boldsymbol{\mu} = (\mu_1, \mu_2)^\top = (-1, 2)^\top$, $\boldsymbol{\Sigma}(\boldsymbol{\alpha}) = \mathbf{I}_2$, onde $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^\top = (1, 0, 1)^\top$, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^\top = (10, 30)^\top$ e $\nu = 3$. Tamanhos amostrais foram fixados em $n = 20, 100, 500, 1000$ e 5000. Para cada combinação de parâmetros e tamanho amostral, 300 amostras do modelo ST foram geradas artificialmente e então, calculamos o erro quadrático médio amostral (mse). Para μ_1, por exemplo, esta medida é calculada como $mse_1 = \dfrac{1}{300}\displaystyle\sum_{k=1}^{300}(\hat{\mu_1}^{(k)} - \mu_1)^2$, onde $\hat{\mu_1}^{(k)}$ é a estimativa de μ_1 em cada amostra k. Definições para os outros parâmetros podem ser obtidas por analogia.

A Tabela 2.1 apresenta os resultados das simulações. Como sugerido por Lange *et al.* (1989), a função log-verossimilhança perfilada foi usada na obtenção do valor de $\boldsymbol{\nu}$. Observe o padrão de convergência para zero. Dessa forma, os estimadores de máxima verossimilhança derivados do algoritmo EM são consistentes.

Tabela 2.1: Dados simulados. Erros quadráticos médios amostrais.

		Tamanho Amostral				
Medida	Parâmetro	20	100	500	1000	5000
mse	μ_1	1.7090e-01	2.9500e-02	5.2000e-03	2.1000e-03	1.8760e-04
	μ_2	2.0300e-02	3.4000e-03	6.0000e-04	2.1000e-04	2.1000e-05
	α_1	4.1940e-01	4.3000e-02	1.0100e-02	4.3000e-03	6.3300e-04
	α_2	1.5880e-01	2.4800e-02	4.5000e-03	1.7000e-03	1.8540e-04
	α_3	7.3800e-02	1.3400e-02	2.2000e-03	8.0000e-04	7.4200e-05
	λ_1	0.7278e+02	0.1159e+02	0.1806e+01	5.8910e-01	4.1700e-02
	λ_2	0.7495e+02	0.2911e+02	0.2620e+01	6.5260e-01	1.2600e-02
	ν	0.19159e+01	6.5900e-01	9.9600e-02	4.4300e-02	9.2000e-03

2.4.2 Estudo II: Algoritmo Proposto

Nesta seção, inicialmente, avaliamos o desempenho do algoritmo proposto na Seção 2.3.1 no contexto de estimação dos parâmetros dos modelos SMSN para dados parcialmente observados. Dessa forma, conduzimos um estudo de simulação sob dois cenários. Para cada cenário, simulamos 1000 conjuntos de dados provenientes dos modelos bivariados normal (cenário 1) e skew-normal (cenário 2) com tamanhos amostrais $n = 100$, $\boldsymbol{\mu} = (10, 20)^\top$, $\boldsymbol{\Sigma}(\boldsymbol{\alpha}) = \begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix}$ e $\boldsymbol{\lambda} = (-2, 4)^\top$ para o caso skew-normal. Para conduzir este estudo de simulação no contexto de dados parcialmente observados, valores ausentes artificiais foram considerados (Zio *et al.*, 2007). Assumimos uma taxa de valores ausentes de 4%.

A Tabela 2.2 apresenta os resultados numéricos para as estimativas dos parâmetros quando normalidade e distribuição skew-normal são consideradas. Note que o algoritmo proposto é eficiente para estimação dos parâmetros tanto no contexto simétrico ($\boldsymbol{\lambda} = \mathbf{0}$) quanto assimétrico, quando valores ausentes estão presentes nos dados.

Imputações de Valores Ausentes

Posteriormente, para avaliar o desempenho do algoritmo proposto na Seção 2.3.1, no contexto de imputação de valores ausentes com estrutura SMSN, conduzimos outro estudo de simulação. Neste caso, simulamos 1000 conjuntos de dados provenientes do modelo bivariado skew-normal com tamanhos amostrais $n = 100$, $\boldsymbol{\mu} = (10, 20)^\top$, $\boldsymbol{\Sigma}(\boldsymbol{\alpha}) = \begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix}$, $\boldsymbol{\lambda} = (-2, 4)^\top$ e 4% de valores ausentes. Para cada conjunto de dados, ajustamos os modelos skew-normal e normal. As Figuras 2.3 e 2.4 apresentam os resultados numéricos relacionados com as imputações dos valores ausentes para nor-

Tabela 2.2: Dados simulados. Resultados Monte Carlo (MC) baseados em 1000 conjuntos de dados e 4% de valores ausentes. Média MC é a média das estimativas obtidas via o algoritmo proposto. Os verdadeiros valores estão entre parênteses. SE MC é o desvio padrão das estimativas.

	Cenários			
	Normal		Skew-Normal	
Parâmetro	Média MC	SE MC	Média MC	SE MC
μ_1 (10)	9.9993	0.0893	9.8931	0.4212
μ_2 (20)	19.9962	0.1985	20.2173	0.3801
α_1 (2)	1.9706	0.2813	1.9915	0.5206
α_2 (-1)	-0.9919	0.3340	-0.5935	0.6267
α_3 (5)	4.8017	0.8575	4.3425	0.9941
λ_1 (-2)	-	-	-1.4099	0.8863
λ_2 (4)	-	-	2.8790	0.8141

malidade e distribuição skew-normal. Dessa forma, as Figuras 2.3 e 2.4 apresentam gráficos de caixa para o viés das predições sobre os valores ausentes nos casos normal e skew-normal, respectivamente. O viés está definido como $bias = (y_{ij} - \widehat{y_{ij}})$, onde y_{ij} é o valor atual e $\widehat{y_{ij}}$ é o respectivo valor predito. Note que o algoritmo proposto é eficiente para predizer valores ausentes no contexto assimétrico; veja Tabela 2.3.

Tabela 2.3: Dados simulados. Resultados Monte Carlo (MC) baseados em 1000 conjuntos de dados e 4% de valores ausentes. Média MC é a média do viés das predições obtidas via o algoritmo proposto. SE MC é o desvio padrão do viés das predições.

	Normal		Skew-Normal	
Observação	Média MC	SE MC	Média MC	SE MC
1	-1.3151	1.6170	-0.0001	1.4334
2	-1.3504	1.7049	-0.0124	1.5669
3	-1.3999	1.8263	-0.1008	1.6236
4	-1.3247	1.6362	0.0061	1.5181

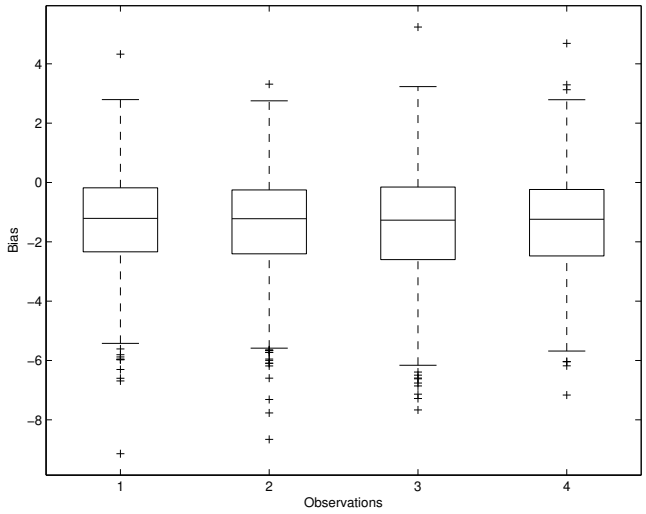


Figura 2.3: Dados simulados. Gráficos de caixa para o viés das predições sobre os valores ausentes no caso normal.

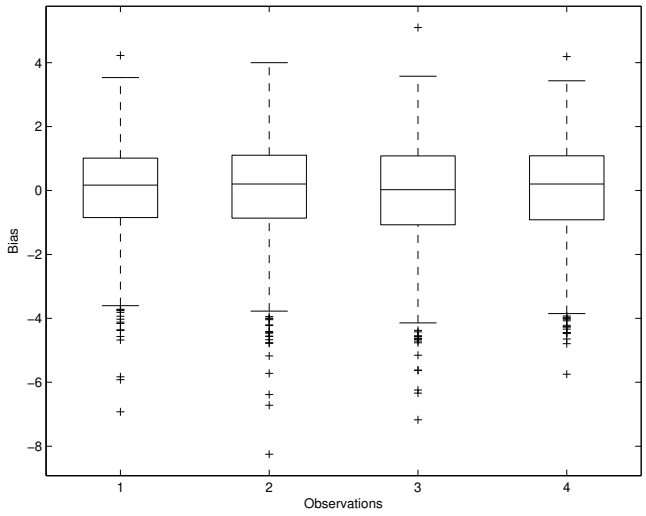


Figura 2.4: Dados simulados. Gráficos de caixa para o viés das predições sobre os valores ausentes no caso skew-normal.

Como outra ilustração, de acordo com Lin & Lin (2009), a precisão das predições sobre os valores ausentes será avaliada através do erro absoluto médio (MAE) e do erro

relativo absoluto médio (MARE). Estas medidas são dadas por

$$MAE = \frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{2}|y_{ij} - \widehat{y_{ij}}| \quad MARE = \frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{2}|\frac{y_{ij} - \widehat{y_{ij}}}{y_{ij}}|,$$

onde m denota o número de observações ausentes. Comparações de resultados para os casos normal e skew-normal, considerando as medidas MAE e MARE, estão listadas na Tabela 2.4 para diversas taxas de dados ausentes. Na Tabela 2.4, a porcentagem de melhora relativa (RIP) é definida como o decréscimo em porcentagem do erro de predição relativo quando comparamos os preditores nos casos normal e skew-normal. Por exemplo, considerando a medida MAE, temos RIP = {[MAE(normal)-MAE(skew-normal)]/MAE(normal)}*100. Neste estudo, observamos que ambos os modelos (normal e skew-normal) apresentam um bom desempenho preditivo, mas o preditor baseado no modelo skew-normal exibe uma considerável precisão das predições quando comparamos diversas taxas de valores ausentes.

Tabela 2.4: Dados simulados. Comparação das medidas MAE e MARE variando as taxas de valores ausentes para os casos normal e skew-normal.

Taxa (%)	MAE			MARE		
	Normal	Skew-Normal	RIP	Normal	Skew-Normal	RIP
10	0.1800	0.1207	32.9444	0.0080	0.0056	30.0000
20	0.1001	0.0606	39.4605	0.0045	0.0028	37.7778
30	0.0728	0.0405	44.3681	0.0033	0.0019	42.4242
40	0.0710	0.0304	57.1831	0.0032	0.0014	56.2500

Dessa forma, os estudos de simulação, desenvolvidos nesta seção, confirmam a necessidade do desenvolvimento de procedimentos para realizar inferência no contexto de dados ausentes com estrutura assimétrica. É importante ressaltar que estudos de simulação sob diferentes cenários podem ser considerados e seria uma linha interessante de

pesquisa.

A seguir apresentamos alguns resultados adicionais para a classe de distribuições misturas de escala skew-normal.

2.5 Resultados Adicionais das Distribuições Misturas de Escala Skew-Normal

Nesta seção, propomos uma nova distribuição pertencente à classe SMSN, onde a variável de mistura segue uma distribuição gaussiana inversa generalizada (GIG).

Antes de discutirmos nossa proposta metodológica, apresentamos as distribuições GIG (Jorgensen, 1982) e a gaussiana inversa normal (NIG) (Barndorff-Nielsen, 1997). Para a distribuição NIG, descrevemos propriedades, tais como momentos, transformações line-ares, densidades marginais e condicionais, formas quadráticas e casos particulares (normal, t-Student, slash, normal contaminada e laplace, por exemplo) que serão úteis para a obtenção de resultados para a nova distribuição SMSN.

Ressaltamos que alguns dos resultados descritos para a NIG encontram-se na literatura apenas para o caso univariado. Neste trabalho, descrevemos a distribuição NIG no contexto multivariado como um membro da classe SMN. Esta distribuição será descrita em termos de uma normal multivariada e uma variável aleatória positiva (variável de mistura) com distribuição gaussiana inversa generalizada.

2.5.1 Distribuição Gaussiana Inversa Generalizada

A função densidade de probabilidade de uma variável aleatória com distribuição gaussiana inversa generalizada é definida por

$$h(u;\lambda_*,\delta_*,\gamma_*)=\left(\frac{\gamma_*}{\delta_*}\right)^{\lambda_*}\frac{u^{\lambda_*-1}}{2K_{\lambda_*}(\delta_*\gamma_*)}exp(-\frac{1}{2}(\frac{\delta_*^2}{u}+\gamma_*^2u)),\quad u>0, \tag{2.10}$$

onde $K_{\lambda_*}(x)$ denota a função de Bessel modificada do terceiro tipo de índice λ, avaliada em x. Assim, consideramos $GIG(\lambda_*,\delta_*,\gamma_*)$ para indicar que U tem densidade dada em (2.10). Nesta distribuição, os espaços paramétricos para δ_*, γ_* e λ_* são especificados como segue:

i) $\delta_*\geq 0$, $\gamma_*>0$ se $\lambda_*>0$;

ii) $\delta_*>0$, $\gamma_*\geq 0$ se $\lambda_*<0$;

iii) $\delta_*>0$, $\gamma_*>0$ se $\lambda_*=0$.

Dessa forma, a família de distribuições $GIG(\lambda_*,\delta_*,\gamma_*)$ contém várias distribuições como casos particulares. Por exemplo, as distribuições gamma, gamma inversa, gaussiana inversa (IG), entre outras. A seguir descrevemos alguns elementos:

i) Quando $\delta_*=0$, $\gamma_*>0$, $\lambda_*>0$, obtemos a distribuição $Gamma(\lambda_*,\gamma_*^2/2)$ com função densidade dada por

$$h(u;\lambda_*,\gamma_*)=\left(\frac{\gamma_*^2}{2}\right)^{\lambda_*}\frac{u^{\lambda_*-1}}{\Gamma(\lambda_*)}exp(-\frac{1}{2}\gamma_*^2u),\ u>0;$$

ii) Quando $\delta_*>0$, $\gamma_*=0$ e $\lambda_*<0$, obtemos a distribuição $I\Gamma(\lambda_*,\gamma_*^2/2)$ (Gamma Inversa) com densidade dada por

$$h(u;\lambda_*,\delta_*)=\left(\frac{2}{\delta_*^2}\right)^{\lambda_*}\frac{u^{\lambda_*-1}}{\Gamma(-\lambda_*)}exp(-\frac{\delta_*^2}{2u}),\ u>0;$$

iii) Quando $\lambda_* = -1/2$, obtemos a distribuição $IG(\delta_*, \gamma_*)$ (Gaussiana Inversa) com densidade dada por

$$h(u; \delta_*, \gamma_*) = \left(\frac{\delta_*^2}{2\pi}\right)^{1/2} u^{-3/2} exp(-(\gamma_*^2/2u)(u - (\delta_*/\gamma_*))^2), \ u > 0.$$

Os momentos centrais de ordem m de uma variável aleatória U com distribuição $GIG(\lambda_*, \delta_*, \gamma_*)$ são dados por

$$E[U^m] = \left(\frac{\delta_*}{\gamma_*}\right)^m \frac{K_{\lambda_*+m}(\delta_*\gamma_*)}{K_{\lambda_*}(\delta_*\gamma_*)}, \quad \delta_*\gamma_* > 0, \ m \in \mathbb{Z}. \tag{2.11}$$

Note que os momentos de ordem m dependem da função $K_{\lambda_*}(.)$. Assim, na sequência apresentamos alguns resultados sobre a função de Bessel modificada do terceiro tipo:

i) A função Bessel $K_{\lambda_*}(u)$ é uma função positiva e contínua de $\lambda_* > 0$ e $u > 0$;

ii) Para qualquer $\lambda_* \geq 0$ fixo, a função $K_{\lambda_*}(u)$ é positiva e decrescente em u no intervalo $(0, \infty)$;

iii) Para qualquer $u > 0$ fixado, a função $K_{\lambda_*}(u)$ é positiva e crescente em u no intervalo $(0, \infty)$;

iv) Para qualquer $\lambda_* \geq 0$ e $u > 0$ fixos, a função $K_{\lambda_*}(u)$ satisfaz as relações $K_{\lambda_*}(u) = K_{-\lambda_*}(u)$, $K_{\lambda_*+1}(u) = \frac{2\lambda_*}{u} K_{\lambda_*}(u) + K_{\lambda_*-1}(u)$ e $K_{\lambda_*-1}(u) + K_{\lambda_*+1}(u) = -2K'_{\lambda_*}(u)$;

v) Para um inteiro não-negativo r $K_{r+1/2}(u) = \sqrt{\frac{\pi}{2u}} e^{-u} \sum_{k=0}^{r} \frac{(r+k)!(2u)^{-k}}{(r-k)!k!}$. Em particular, $K_{1/2}(u) = \sqrt{\frac{\pi}{2u}} e^{-u}$.

Veja Jorgensen (1982) para mais detalhes da distribuição gaussiana inversa generalizada. Além disso, a distribuição gaussiana inversa generalizada tem sido de grande aplicabilidade em pesquisas de diversas áreas; veja, por exemplo, Embrechts (1983) e Thabane & Haq (1999).

2.5.2 Distribuição Gaussiana Inversa Normal

Nesta seção, definimos a distribuição gaussiana inversa normal (NIG) multivariada como um membro da classe SMN e desenvolvemos algumas propriedades.

Considerando que $\mathbf{Z} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$ e $U \sim GIG(\lambda_*, \delta_*, \gamma_*)$ são independentes, propomos estudar a distribuição da nova variável aleatória $\mathbf{Y} = \boldsymbol{\mu} + U^{1/2}\mathbf{Z}$. De acordo com esta representação pode-se ver facilmente que a distribuição condicional de $\mathbf{X}$ dado $U = u$ é $N_p(\boldsymbol{\mu}, u\mathbf{\Sigma})$, útil para obter pdf do vetor $\mathbf{Y}$ que é apresentado a seguir.

Proposição 2.5.1 *Seja* $\mathbf{Z} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$ *e* $U \sim GIG(\lambda_*, \delta_*, \gamma_*)$ *independentes, tal que* $\mathbf{Y}$ *é definida por* $\mathbf{Y} = \boldsymbol{\mu} + U^{1/2}\mathbf{Z}$, *onde* $\boldsymbol{\mu} \in \mathbb{R}^p$ *e* $\mathbf{\Sigma}$ *é uma matriz simétrica* $p \times p$. *Então,* $\mathbf{Y}$ *tem pdf dada por*

$$f(\mathbf{y}) = a(\boldsymbol{\nu}_*)|\mathbf{\Sigma}|^{-1/2}(\sqrt{\delta_*^2 + d_y})^{-p/2+\lambda_*} K_{-p/2+\lambda_*}((\sqrt{\delta_*^2 + d_y})\gamma_*), \tag{2.12}$$

onde

$$a(\boldsymbol{\nu}_*) = (\frac{\gamma_*}{2\pi})^{p/2} \frac{1}{\delta_*^{\lambda_*} K_{\lambda_*}(\delta_*\gamma_*)} \quad \text{e} \quad d_y = (\mathbf{y} - \boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}),$$

com $\boldsymbol{\nu}_* = (\lambda_*, \delta_*, \gamma_*)^\top$.

Prova A prova decorre do fato que $\mathbf{Y}|U = u \sim \mathrm{N}_p(\boldsymbol{\mu}, u\mathbf{\Sigma})$.

Note que a densidade dada em (2.12) depende da distância de Mahalanobis $d = d_y = (\mathbf{y} - \boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ e do vetor de parâmetros $\boldsymbol{\nu}_* = (\lambda_*, \delta_*, \gamma_*)^\top$ associado com a distribuição GIG.

Para um vetor aleatório p-dimensional $\mathbf{Y} = (Y_1, ..., Y_p)^\top$ com pdf dada em (2.12), dizemos que $\mathbf{Y}$ tem distribuição gaussiana inversa normal e usamos a seguinte notação

$\mathbf{Y} \sim NIG_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \lambda_*, \delta_*, \gamma_*)$. Neste caso, $\boldsymbol{\mu} \in \mathbb{R}^p$ representa o parâmetro de locação e $\boldsymbol{\Sigma}$ representa a matriz de escala (definida positiva $p \times p$). Por sua vez, se $\boldsymbol{\mu} = \mathbf{0}$ e $\boldsymbol{\Sigma} = \mathbf{I}_p$, obtemos a distribuição NIG padrão.

Note que a função densidade dada em (2.12) é justamente a densidade de um vetor com distribuição elíptica, tal que $g(u) = (\sqrt{\delta_*^2 + u})^{-p/2+\lambda_*} K_{-p/2+\lambda_*}((\sqrt{\delta_*^2 + u})\gamma_*)$. Neste caso, denotamos por $EC_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g)$; veja Fang *et al.* (1990) para mais detalhes sobre as distribuições elípticas. Dessa forma, o vetor aleatório $\mathbf{Y}$ com distribuição $NIG_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \lambda_*, \delta_*, \gamma_*)$ pode ser representado por

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + R\mathbf{A}^\top \mathbf{u}^{(p)},$$

em que $\mathbf{u}^{(p)}$ é um vetor aleatório uniformemente distribuído na esfera unitária $\mathbb{R}^p$, R é uma variável aleatória não negativa independente de $\mathbf{u}^{(p)}$, cuja função densidade será definida posteriormente e $\mathbf{A}$ é uma matriz quadrada $p \times p$, tal que $\mathbf{A}^\top \mathbf{A} = \boldsymbol{\Sigma}$.

Na próxima proposição, apresentamos a função geradora de momentos (mgf) e a função característica de um vetor aleatório com distribuição NIG. Resultados similares para a distribuição gaussiana inversa normal univariada podem ser encontrados em Barndorff-Nielsen (1997) e Anderson (2001), por exemplo.

Proposição 2.5.2 *Considere que* $\mathbf{Y} \sim NIG_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \lambda_*, \delta_*, \gamma_*)$. *Então,*

a) A função caracterítica de $\mathbf{Y}$ *é dada por*

$$\varphi_{\mathbf{Y}}(\mathbf{t}) = e^{i\mathbf{t}^\top \boldsymbol{\mu}} \left(\frac{\gamma_*}{\sqrt{\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t} + \gamma_*^2}} \right)^{\lambda_*} \frac{K_{\lambda_*}(\delta_* \sqrt{\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t} + \gamma_*^2})}{K_{\lambda_*}(\delta_* \gamma_*)}, \quad \mathbf{t} \in \mathbb{R}^p$$

e

b) A mgf de $\mathbf{Y}$ *é dada por*

$$M_{\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{t}^\top \boldsymbol{\mu}} \left(\frac{\gamma_*}{\sqrt{\gamma_*^2 - \mathbf{t}^\top \boldsymbol{\Sigma}_* \mathbf{t}}} \right)^{\lambda_*} \frac{K_{\lambda_*}(\delta \sqrt{\gamma_*^2 - \mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}})}{K_{\lambda_*}(\delta_* \gamma_*)}, \quad \mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t} < \gamma_*^2, \quad \mathbf{t} \in \mathbb{R}^p.$$

Prova: Da Preposição 2.5.1, temos que $\mathbf{Y}|U = u \sim N_p(\boldsymbol{\mu}, u\boldsymbol{\Sigma})$. Assim, $E[e^{\mathbf{t}^\top \mathbf{Y}}|U = u] = e^{i\mathbf{t}^\top \boldsymbol{\mu} - (u/2)\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}}$. Das propriedades de esperança condicional, temos que

$$\varphi_{\mathbf{Y}}(\mathbf{t}) = E_U[E[e^{\mathbf{t}^\top \mathbf{Y}}|U]] = \int_0^\infty e^{i\mathbf{t}^\top \boldsymbol{\mu} - (u/2)\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}} h(u; \lambda_*, \delta_*, \gamma_*) du$$

e da definição da função de Bessel modificada do terceiro tipo $K_{\lambda_*}(.)$, concluímos a prova do resultado a). A prova de b) é similar ao item a).

Os resultados desenvolvidos na Proposição 2.5.2 podem ser usados para obter os momentos de $\mathbf{Y}$ e para derivar a distribuição de funções de $\mathbf{Y}$. Além disso, note que a mgf de $\mathbf{Y}$ pode ser escrita como

$$M_{\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{t}^\top \boldsymbol{\mu}} M_{GIG}(\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}/2), \quad \mathbf{t} \in \mathbb{R}^p,$$

onde

$$M_{GIG}(s) = \left(\frac{\gamma_*}{\sqrt{\gamma_*^2 - 2s}} \right)^{\lambda_*} \frac{K_{\lambda_*}(\delta_* \sqrt{\gamma_*^2 - 2s})}{K_{\lambda_*}(\delta_* \gamma_*)}, \quad 2s < \gamma_*^2, \quad s \in \mathbb{R},$$

é a mgf de $U \sim GIG(\lambda_*, \delta_*, \gamma_*)$.

Na próxima preposição, obtemos o vetor de médias e a matriz de covariâncias de um vetor aleatório com distribuição NIG. A prova segue dos resultados da Proposição 2.5.1 e do fato de U e $\mathbf{Z}$ serem independentes.

Proposição 2.5.3 *Suponha que* $\mathbf{Y} \sim NIG_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \lambda_*, \delta_*, \gamma_*)$. *Então,*

a) Se $E[U^{1/2}] < \infty$, *temos que* $E[\mathbf{Y}]$ *existe e*

$$E[\mathbf{Y}] = \boldsymbol{\mu}.$$

b) Se $E[U] < \infty$, *temos que o segundo momento de* $\mathbf{Y}$ *existe e*

$$Var[\mathbf{Y}] = E[U]\boldsymbol{\Sigma}.$$

De acordo com resultados desenvolvidos em Wang *et al.* (2004), a seguir derivamos a distribuição de uma função de um vetor aleatório com distribuição NIG.

Proposição 2.5.4 *Seja* $g : \mathbb{R}^p \longrightarrow \mathbb{R}^q$ *uma função mensurável. Suponha que* $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *e denote a pdf de* $\mathbf{W}_1 = g(\mathbf{X})$ *por* $\phi_p^g(\mathbf{w_1}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$. *Então, se* $\mathbf{Y} \sim NIG_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \lambda_*, \delta_*, \gamma_*)$, *temos que a pdf de* $\mathbf{W} = g(\mathbf{Y})$ *é dada por*

$$f_{\mathbf{W}}(\mathbf{w}) = \int_0^\infty \phi_p^g(\mathbf{w}|\boldsymbol{\mu}, u\boldsymbol{\Sigma}) h(u; \lambda_*, \delta_*, \gamma_*) du.$$

Prova A prova segue da Proposição 2.5.2 e de resultados desenvolvidos neste contexto em Wang *et al.* (2004).

Como um subproduto da Proposição 2.5.4, temos os próximos corolários. Resultados equivalentes para a distribuição gaussiana inversa normal univariada podem ser encontrados em Salberg *et al.* (2001).

Corolário 2.5.5 *Considere que* $\mathbf{Y} \sim NIG_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \lambda_*, \delta_*, \gamma_*)$. *Então, a pdf da forma quadrática* $R = (\mathbf{Y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu})$ *é dada por*

$$f_R(r) = \frac{r^{p/2-1}}{2^{p/2}\Gamma(p/2)} \left(\frac{\gamma_*}{\sqrt{\delta_*^2 + r}} \right)^{p/2} \left(\frac{\sqrt{\delta_*^2 + r}}{\delta_*} \right)^{\lambda_*} \frac{K_{-p/2+\lambda_*}(\gamma_* \sqrt{\delta_*^2 + r})}{K_{\lambda_*}(\delta_* \gamma_*)}.$$

Este resultado é interessante porque nos permite na prática verificar o ajuste do modelo. Por outro lado, o Corolário 2.5.5 juntamente com o resultado desenvolvido na Seção 2 em Lange & Sinsheimer (1993) nos permite obter os momentos de ordem m de R (distância de Mahalanobis, denotada neste trabalho por d).

Corolário 2.5.6 *Considere que* $\mathbf{Y} \sim NIG_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \lambda_*, \delta_*, \gamma_*)$. *Então, para qualquer* $m > 0$

$$E[R^m] = \frac{2^m \Gamma(m + p/2)}{\Gamma(p/2)} (\frac{\delta_*}{\gamma_*})^m \frac{K_{\lambda_*+m}(\delta_*\gamma_*)}{K_{\lambda_*}(\delta_*\gamma_*)}.$$

De acordo com os resultados encontrados em Anderson (2001) para o caso univariado e usando o Corolário 2.5.6, obtemos os coeficientes de assimetria e curtose (Mardia, 1970), respectivamente, dados por

$$\gamma_1(\mathbf{Y}) = 0 \text{ e } \gamma_2(\mathbf{Y}) = p(p+2)\frac{K_{\lambda_*+2}(\delta_*\gamma_*)}{K_{\lambda_*+1}(\delta_*\gamma_*)}\frac{K_{\lambda_*}(\delta_*\gamma_*)}{K_{\lambda_*+1}(\delta_*\gamma_*)} - p(p+2).$$

Em seguida, mostramos que o vetor aleatório NIG é invariante sob transformações lineares e consequentemente a distribuição marginal ainda tem distribuição NIG, generalizando alguns resultados encontrados em Anderson (2001), por exemplo.

Proposição 2.5.7 *Considere que* $\mathbf{Y} \sim NIG_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \lambda_*, \delta_*, \gamma_*)$. *Então, para quaisquer vetor* $\mathbf{b} \in \mathbb{R}^m$ *e matriz* $\mathbf{A} \in \mathbb{R}^{m\times p}$ *de posto completo, temos que*

$$\mathbf{X} = \mathbf{b} + \mathbf{A}\mathbf{Y} \sim NIG_m(\mathbf{b} + \mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top; \lambda_*, \delta_*, \gamma_*).$$

Prova A prova deste resultado segue diretamente da Proposição 2.5.2, já que $\varphi_{\mathbf{b}+\mathbf{A}\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{s}^\top \mathbf{b}} \varphi_{\mathbf{Y}}(\mathbf{A}^\top \mathbf{t})$.

Inspirados por Fotopoulos *et al.* (2007) e Cambanis *et al.* (2000), obtemos a densidade condicional de $\mathbf{Y}_1$ dado $\mathbf{Y}_2 = \mathbf{y}_2$ e momentos condicionais de funções de U (por exemplo, $g(U)$) dado $\mathbf{Y}$. Estes resultados encontram-se resumidos a seguir.

Proposição 2.5.8 *Considere que* $\mathbf{Y} \sim NIG_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \lambda_*, \delta_*, \gamma_*)$, *tal que*

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \text{e } \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

onde $\mathbf{Y}_1 : m \times 1$, $\boldsymbol{\mu}_1 : m \times 1$, $\boldsymbol{\Sigma}_{11} : m \times m$, *com* $0 < m < p$. *Então,*

$$\mathbf{Y}_1|\mathbf{Y}_2 = \mathbf{y}_2 \sim NIG_m(\boldsymbol{\mu}_{1.2}, \boldsymbol{\Sigma}_{11.2}; \lambda_1, \delta_{y_2}, \gamma_*),$$

onde $\boldsymbol{\mu}_{1.2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_2 - \boldsymbol{\mu}_2)$, $\boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$, $\lambda_1 = \lambda_* - (p-m)/2$ *e* $\delta_{y_2} = \sqrt{\delta_*^2 + q(y_2)}$, *com* $q(y_2) = (\mathbf{y}_2 - \boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_2 - \boldsymbol{\mu}_2)$ *e sua pdf é dada por*

$$f(\mathbf{y}_1|\mathbf{y}_2) \quad = \quad a(\lambda_1, \delta_{y_2}, \gamma)|\boldsymbol{\Sigma}_{11.2}|^{-1/2}(\sqrt{\delta_{y_2}^2 + d_{y_{12}}})^{-m/2+\lambda_1} K_{-m/2+\lambda_1}(\gamma_* \sqrt{\delta_{y_2}^2 + d_{y_{12}}}),$$

onde $d_{y_{12}} = (\mathbf{y}_1 - \boldsymbol{\mu}_{1.2})^\top \boldsymbol{\Sigma}_{11.2}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_{1.2})$.

Proposição 2.5.9 *Seja* $g(.)$ *uma funcão mensurável não negativa e* $\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + U^{1/2}\mathbf{Z} \sim NIG_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \lambda_*, \delta_*, \gamma_*)$. *Então,* $U|\mathbf{Y} = \mathbf{y} \sim GIG(-p/2 + \lambda_*, \sqrt{\delta_*^2 + d_y}, \gamma_*)$ *e*

$$E[g(U)|\mathbf{Y} = \mathbf{y}] = E_{GIG}[g(U_\mathbf{y})],$$

onde $U_\mathbf{y} \sim GIG(-p/2 + \lambda_*, \sqrt{\delta_*^2 + d_y}, \gamma_*)$, *com* $d_y = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})$. *Em particular, se* $g(u) = u^m$, *temos que*

$$E[U^m|\mathbf{Y} = \mathbf{y}] = \left(\frac{\sqrt{\delta_*^2 + d_y}}{\gamma_*}\right)^m \frac{K_{m-p/2+\lambda_*}(\gamma_* \sqrt{\delta_*^2 + d_y})}{K_{-p/2+\lambda_*}(\gamma_* \sqrt{\delta_*^2 + d_y})}.$$

Exemplos

Nesta seção, obtemos alguns casos especias da distribuição NIG. A maioria dos exemplos serão obtidos como caso limite da distribuição $NIG_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \lambda_*, \delta_*, \gamma_*)$ quando $\boldsymbol{\nu}_* = (\lambda_*, \delta_*, \gamma_*)^\top \to \boldsymbol{\nu}_\infty$, para algum $\boldsymbol{\nu}_\infty = (\lambda_\infty, \delta_\infty, \gamma_\infty)^\top$. Para deriva-los, usaremos algumas propriedades assintóticas da função Bessel K_{λ_*} que podem ser encontradas em Abramowitz & Stegun (1972).

- **Distribuição t-Student Multivariada**

Denotada por $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$, esta distribuicão é obtida considerando $\lambda < 0$ fixo, $q(y) = \sqrt{\delta_*^2 + d_y}$ e $\gamma_* \to 0$. Então, de (2.12), temos que

$$\frac{\gamma_*^{\lambda_*}}{\delta_*^{\lambda_*} K_{\lambda_*}(\delta_*\gamma_*)} \sim \frac{2^{\lambda_*+1}}{\delta_*^{2\lambda_*}\Gamma(-\lambda_*)} \quad \text{e} \quad \frac{q(y)^{-p/2+\lambda_*}}{\gamma_*^{-p/2+\lambda_*}} K_{-p/2+\lambda_*}(\gamma_* q(y)) \sim \frac{q(y)^{2\lambda_*-p}\Gamma(p/2-\lambda_*)}{\delta_*^{-p/2+\lambda_*+1}}.$$

Note que

$$lim_{\gamma_*\to 0} f_Y(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-1/2} \frac{\Gamma(p/2-\lambda_*)}{(\pi\delta_*^2)^{p/2}\Gamma(-\lambda_*)} (1 + \frac{(\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})}{\delta_*^2})^{\lambda_*-p/2}.$$

Assim, a densidade da distribuição t-Student com ν graus de liberdade é obtida considerando $\nu = \delta_*^2 = -2\lambda_*$. Para $\lambda_* = -1/2$, i.e., quando U tem distribuição gaussiana inversa (IG), obtemos a distribuição cauchy multivariada, cuja pdf é dada por

$$lim_{\gamma\to 0} f_Y(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-1/2} \frac{\Gamma((p+1)/2)}{\pi^{(p+1)/2}} \left(1 + (\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)^{-(p+1)/2}.$$

É bem conhecido que a distribuição t-Student reduz para a distribuição normal quando $\nu \uparrow \infty$. Por outro lado, podemos obter a distribuição normal através da $NIG_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \lambda_*, \delta_*, \gamma_*)$. Considere $\gamma_* \to \infty$ e $\delta_* \to \infty$, tal que $\delta_*/\gamma_* \to 1$. Sob estas condições, temos que

$$K_{-p/2+\lambda_*}(\gamma_* q(y)) \sim \frac{\pi e^{-\gamma_* q(y)}}{\sqrt{2\gamma_* q(y)}}, \quad K_{\lambda_*}(\delta_*\gamma_*) \sim \frac{\pi e^{-\delta_*\gamma_*}}{\sqrt{2\delta_*\gamma_*}}, \quad lim_{\substack{\delta_*\to\infty \\ \gamma_*\to\infty \\ \delta_*/\gamma_*\to 1}} \frac{(\delta_*\gamma_*)^{1/2}}{q(y)^{(p+1)/2}} = 1$$

e consequentemente

$$f_Y(\mathbf{y}) \sim |\boldsymbol{\Sigma}|^{-1/2} \frac{1}{(2\pi)^{p/2}} \exp(\delta_*\gamma_* - \gamma_* q(y)). \tag{2.13}$$

Usando aproximação em série de Taylor, temos que $\sqrt{1 \pm y^2} = 1 \pm y^2/2 + o(y^2)$. Quando $y \to 0$, o exponente na expressão (2.13) converge para $-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ e concluímos que

$$lim_{\substack{\delta_*\to\infty \\ \gamma_*\to\infty \\ \delta_*/\gamma_*\to 1}} f_Y(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-1/2} \frac{1}{(2\pi)^{p/2}} \exp(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})).$$

- **Distribuição Normal Contaminada Multivariada**

  Denotada por $CN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu, \eta)$, $0 \leq \nu \leq 1$, $0 < \eta < 1$. Para obter esta distribuição, consideramos a mistura de $NIG_p(\boldsymbol{\mu}, (1/\eta)\boldsymbol{\Sigma}; \lambda_*, \delta_*, \gamma_*)$ e $NIG_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \lambda_*, \delta_*, \gamma_*)$ com pdf dadas por f_1 e f_2, respectivamente. Dessa forma, considere a pdf dada por

$$f(\mathbf{y}) \quad = \quad \nu f_1(\mathbf{y}|\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{\eta}; \lambda_*, \delta_*, \gamma_*) + (1-\nu) f_2(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}; \lambda_*, \delta_*, \gamma_*).$$

  Fazendo $\gamma_* \to \infty$ e $\delta_* \to \infty$, tal que $\delta_*/\gamma_* \to 1$ (como no caso normal), obtemos a distribuição normal contaminada multivariada, cuja pdf é dada por

$$lim_{(\delta_*, \gamma_*) \to (\infty, \infty)} f(\mathbf{y}) \quad = \quad \nu \phi_p(\mathbf{y}|\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{\eta}) + (1-\nu)\phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

- **Distribuição Laplace Multivariada**

  Denotada por $L_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$, fixando $\lambda_* > 0$ e $\delta_* \to 0$, obtemos

$$lim_{\delta_* \to 0} f_Y(\mathbf{y}) = (\frac{\gamma_*}{2\pi})^{p/2} \frac{\gamma_*^{\lambda_*}}{\Gamma(\lambda_*) 2^{\lambda_* - 1}} |\boldsymbol{\Sigma}|^{-1/2} (\sqrt{d_y})^{-p/2+\lambda_*} K_{-p/2+\lambda_*}((\sqrt{d_y})\gamma_*)$$

  que corresponde à pdf da distribuição variance-gamma multivariada; veja Madan *et al.* (1998) para mais detalhes. Por outro lado, considerando $\lambda_* = 1$, obtemos a pdf

$$lim_{\delta_* \to 0} f_Y(\mathbf{y}) = (\frac{\gamma_*}{2\pi})^{p/2} \gamma_* |\boldsymbol{\Sigma}|^{-1/2} (\sqrt{d_y})^{-p/2+1} K_{-p/2+1}((\sqrt{d_y})\gamma_*)$$

  que corresponde à função densidade da distribuição laplace multivariada. No caso especial de $p = 1$, a densidade acima reduz para a densidade da laplace univariada dada por $lim_{\delta_* \to 0} f_Y(y) = (\gamma_*/\sigma) e^{-(\gamma_*/\sigma)|y-\mu|}$.

- **Distribuição Slash Multivariada**

  Denotada por $SL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$. Neste caso, não derivamos a distribuição slash multivariada diretamente da pdf dada em (2.12), mas através do vetor aleatório $\mathbf{Y} =$

$\boldsymbol{\mu}+k^{1/2}(U)\mathbf{Z}$, onde $k(U)=e^{-U}$ e $U\sim GIG(1,0,\gamma_*)$, i.e., $U\sim Gamma(1,\gamma_*^2/2)=exp(\gamma_*^2/2)$. Note que

$$lim_{\delta_*\to 0} f_Y(\mathbf{y}) = f(\mathbf{y}) = (\gamma_*^2/2)\int_0^1 u^{(\gamma_*^2/2)-1}\phi_p(\mathbf{y}|\boldsymbol{\mu}, e^{-u}\boldsymbol{\Sigma})du.$$

Dessa forma, obtemos a distribuição $SL_p(\boldsymbol{\mu},\boldsymbol{\Sigma};\nu)$, onde $\nu=\gamma_*^2/2$. Como já mencionado, a distribuição slash multivariada reduz para a distribuição normal multivariada quando $\nu\uparrow\infty$.

Em geral, a distribuição gaussiana inversa normal (NIG) é usada para modelar dados simétricos e com caudas mais pesadas que a distribuição normal. Particularmente é atraente porque tem uma representação que permite o estudo de muitas propriedades. Barndorff-Nielsen (1997), Anderson (2001) e Fotopoulos *et al.* (2007), por exemplo, propuseram esta distribuicao em aplicações financeiras. De fato, os log-retornos de ativos do mercado financeiro medidos em intervalos de tempo curtos, por exemplo, diário e semanal são caracterizados pela sua não normalidade. A distribuição empírica dos log-retornos têm caudas mais pesadas do que uma distribuição normal o que implica que existem grandes variações nos log-retornos com maior frequência do que numa distribuição normal. Uma das ditribuições muito utilizada na literatura para modelar os log-retornos é a já conhecida distribuição hiperbólica generalizada (Teixeira, 2006) da qual se destaca a NIG devido às suas características teóricas.

Dessa forma, na próxima seção, vamos discutir uma nova distribuição SMSN denominada de gaussiana inversa skew-normal (SNIG). Especificamente, vamos considerar um vetor aleatório $\mathbf{Y}$ com representação $\mathbf{Y}=\boldsymbol{\mu}+U^{1/2}\mathbf{Z}$, onde a variável aleatória U segue uma distribuição gaussiana inversa generalizada independente de $\mathbf{Z}$ que segue uma distribuição skew-normal proposta por Azzalini & Capitanio (1999).

2.5.3 Distribuição Gaussiana Inversa Skew-Normal

Nesta seção, definimos e estudamos algumas propriedades da distribuição gaussiana inversa skew-normal que representa uma generalização da distribuição NIG.

Considere $\mathbf{Z} \sim SN_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ e $U \sim GIG(\lambda_*, \delta_*, \gamma_*)$ independentes. Para simplificar a notação, seja $\boldsymbol{\nu}_* = (\lambda_*, \delta_*, \gamma_*)^\top$ e assim, a pdf do vetor aleatório $\mathbf{Y} = \boldsymbol{\mu} + U^{1/2}\mathbf{Z}$ é dada por

$$\begin{aligned} f(\mathbf{y}) &= 2\int_0^\infty \phi_p(\mathbf{y}|\boldsymbol{\mu}, u\boldsymbol{\Sigma})\Phi_1(u^{-1/2}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu}))h(u)du \qquad (2.14)\\ &= a(\boldsymbol{\nu}_*)|\boldsymbol{\Sigma}|^{-1/2}(q(y))^{-p/2+\lambda_*}K_{-p/2+\lambda_*}(q(y)\gamma_*)E[2\Phi_1(V^{-1/2}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu}))], \end{aligned}$$

onde $q(y) = \sqrt{\delta_*^2 + d_y}$, $a(\boldsymbol{\nu}_*)$ é como em (2.12), $V \sim GIG(-p/2 + \lambda_*, q(y), \gamma_*)$ e $d_y = (\mathbf{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})$.

Dizemos que um vetor aleatório p-dimensional $\mathbf{Y}$ com pdf dada em (2.14) tem distribuição gaussiana inversa skew-normal e neste caso, denotamos por $\mathbf{Y} \sim SNIG_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; \boldsymbol{\nu}_*)$, onde $\boldsymbol{\mu} \in \mathbb{R}^p$ representa o parâmetro de locação e $\boldsymbol{\Sigma}$ representa a matriz de escala (definida positiva $p \times p$). Se $\boldsymbol{\mu} = \mathbf{0}$ e $\boldsymbol{\Sigma} = \mathbf{I}_p$, diremos que $\mathbf{Y}$ segue uma distribuição SNIG padrão e usaremos a seguinte notação $\mathbf{Y} \sim SNIG_p(\boldsymbol{\lambda}, \boldsymbol{\nu}_*)$.

Note que quando$\boldsymbol{\lambda} = \mathbf{0}$, obtemos a distribuição gaussiana inversa normal discutida na Seção 2.5.2. Essa nova distribuição também depende de $d = d_y = (\mathbf{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})$ que tem a mesma distribuição que d no contexto da NIG (veja a Proposição 2.2.8 e o Corolário 2.5.5). O próximo resultado apresenta a representação estocástica da distribuição SNIG.

Proposição 2.5.10 *Se* $\boldsymbol{Y} \sim SNIG_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; \boldsymbol{\nu}_*)$. *Então, temos que*

a)

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + U^{1/2}\boldsymbol{\Sigma}^{1/2}\{\boldsymbol{\delta}|Y_0| + (\mathbf{I}_n - \boldsymbol{\delta}\boldsymbol{\delta}^T)^{1/2}\mathbf{Y}_1\}, \tag{2.15}$$

onde $\boldsymbol{\delta} = \boldsymbol{\lambda}/\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}$, U, $Y_0 \sim N_1(0,1)$ *e* $\mathbf{Y}_1 \sim N_p(\mathbf{0}, \mathbf{I}_p)$ *são independentes.*

b)

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\{\boldsymbol{\delta}|W_0| + (\mathbf{I}_p - \boldsymbol{\delta}\boldsymbol{\delta}^T)^{1/2}\mathbf{W}_1\},$$

onde $W_0 \sim NIG(0,1;\boldsymbol{\nu}_*)$ *e* $\mathbf{W}_1 \sim NIG_p(\mathbf{0}, \mathbf{I}_p; \boldsymbol{\nu}_*)$.

A representação estocástica dada em (2.15) é útil para gerar números aleatórios, obtenção de propriedades e pode ser usada para obter o estimador de máxima verossimilhança dos parâmetros da distribuição SNIG via o algoritmo EM; veja, por exemplo, Karlis (2002) e Chang *et al.* (2005) no contexto da distribuição NIG.

Proposição 2.5.11 *Considere que* $\boldsymbol{Y} \sim SNIG_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; \boldsymbol{\nu}_*)$. *Então, a mgf de* $\boldsymbol{Y}$ *é dada por*

$$M_{\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{t}^\top\boldsymbol{\mu}}\left(\frac{\gamma_*}{\sqrt{\gamma_*^2 - \mathbf{t}^\top\boldsymbol{\Sigma}\mathbf{t}}}\right)^{\lambda_*}\frac{K_{\lambda_*}(\delta_*\sqrt{\gamma_*^2 - \mathbf{t}^\top\boldsymbol{\Sigma}\mathbf{t}})}{K_{\lambda_*}(\delta_*\gamma_*)}E[2\Phi_1(V^{1/2}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu}))],$$

onde $V \sim GIG(\lambda, \delta_*, \gamma_*^2 - \mathbf{t}^\top\boldsymbol{\Sigma}\mathbf{t})$, *com* $\mathbf{t}^\top\boldsymbol{\Sigma}\mathbf{t} < \gamma_*^2$, $\mathbf{t} \in \mathbb{R}^p$.

Da equação acima, temos que se $\boldsymbol{\lambda} = \mathbf{0}$, $M_{\mathbf{Y}}(\mathbf{t})$ reduz para a mgf da distribuição NIG dada na Proposição 2.5.2. Além disso, da Proposição 2.5.11, obtemos o vetor de médias e matriz de covariâncias do vetor com distribuição SNIG dados por, respectivamente,

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}E[U^{1/2}] \text{ e } Var[\mathbf{Y}] = E[U]\boldsymbol{\Sigma} - \frac{2}{\pi}E^2[U^{1/2}]\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^{1/2},$$

onde $E[U^{1/2}]$ e $E[U]$ são como em (2.11).

O vetor aleatório SNIG também é invariante sob transformações lineares e isso pode ser demonstrado a partir da mgf definida na Proposição 2.5.11.

Proposição 2.5.12 *Considere que* $\mathbf{Y} \sim SNIG_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \boldsymbol{\lambda}; \boldsymbol{\nu}_*)$. *Então, para quaisquer vetor* $\mathbf{b} \in \mathbb{R}^m$ *e matriz* $\mathbf{A}(m \times p)$ *de posto completo, temos que*

$$\mathbf{X} = \mathbf{b} + \mathbf{A}\mathbf{Y} \sim SNIG_m(\mathbf{b} + \mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top; \boldsymbol{\nu}_*).$$

2.6 Observações Finais

Neste capítulo, desenvolvemos propriedades relevantes para a classe de distribuições SMSN. Estas distribuições proporcionam flexibilidade de modelagem na prática para dados parcialmente observados que contém assimetria e caudas pesadas simultaneamente. Neste contexto, propomos um algoritmo simples, comparado aos procedimentos propostos por Lin *et al.* (2009a) e Lin & Lin (2009), para estimação dos parâmetros e imputação de dados ausentes com estrutura SMSN, pois consideramos diretamente o algoritmo EM desenvolvido no contexto de dados completamente observados. É importante ressaltar que a família de distribuições SMSN, considerada nesta tese, difere das distribuições utilizadas por Ferreira (2008) no contexto de inferência e dignóstico em modelos assimétricos. Em particular, a classe SMSN considerada neste trabalho é computacionalmente mais atrativa. Além disso, Ferreira (2008) considerou apenas análises univariadas.

Capítulo 3

Extensões das Distribuições Misturas de Escala Skew-Normal

A construção de famílias paramétricas de distribuições assimétricas que sejam analiticamente tratáveis e que possam acomodar valores práticos de assimetria e curtose, incluindo a distribuição normal como caso particular, pode ser útil para a modelagem de dados, análises estatísticas e estudos de métodos robustos. Branco & Dey (2001) discutem a classe de distribuições assimétricas misturas de escala skew-normal, mostrando que tal família de distribuições inclui a skew-normal e a normal como casos particulares e podem ser usadas para modelagem de dados. Nesta linha de pesquisa, propomos estudar extensões das distribuições SMSN e examinar propriedades de tais modelos. Inicialmente, consideramos uma extensão da distribuição skew-normal proposta por Arellano-Valle *et al.* (2007) que faz da inferência bayesiana uma alternativa viável para tratar os modelos lineares, estendendo alguns resultados desenvolvidos por Sahu *et al.*

(2003), Arellano-Valle & Genton (2005) e Arellano-Valle *et al.* (2007). Um conjunto de dados reais será utilizado para ilustrar a utilidade dessa extensão proposta para modelagem de dados no contexto bayesiano. Posteriormente, introduzimos algumas generalizações das distribuições SMSN quando a variável de mistura U segue uma distribuição multivariada, estendendo os resultados desenvolvidos por Lyu & Simoncelli (2008), úteis na modelagem de imagens.

3.1 Introdução

Durante a última década, há um interesse crescente em encontrar modelos flexíveis que representem as características dos dados de forma mais adequada possível e que reduzam suposições irreais. A motivação originou-se a partir de conjuntos de dados que muitas vezes não satisfazem algum padrão pressuposto como a normalidade.

Uma possível abordagem para a modelagem de dados consiste na construção de classes paramétricas flexíveis de distribuições multivariadas que exibam curtose diferente da distribuição normal. Embora os modelos elípticos forneçam alternativas para o modelo normal, estes só podem ser aplicados em situações práticas onde a simetria pareça razoável. Portanto, a construção de famílias paramétricas de distribuições assimétricas que sejam analiticamente tratáveis e que possam acomodar valores práticos de assimetria e curtose, incluindo a distribuição normal como caso particular, pode ser útil para a modelagem de dados, análises estatísticas e estudos de métodos robustos.

Há uma considerável linha de pesquisa nessa direção, por exemplo, Sahu *et al.* (2003) introduziram uma classe de distribuições assimétricas com caudas pesadas e Arellano-Valle & Genton (2005) propuseram a classe fundamental de distribuições assimétricas,

ambas úteis para implementar análises bayesianas. A análise bayesiana sob distribuições com caudas pesadas tem recebido considerável atenção na literatura estatística; veja, por exemplo, Sahu *et al.* (2003), Rosa *et al.* (2003), Rosa *et al.* (2004), Arellano-Valle *et al.* (2007), De La Cruz (2008), entre outros. Um trabalho pioneiro nesta área é Zellner (1976) que considerou um estudo bayesiano na distribuição t-Student multivariada.

Seguindo estas idéias, neste capítulo, introduzimos algumas generalizações das distribuições SMSN, por exemplo, extensões quando a variável de mistura U segue uma distribuição multivariada. Adicionalmente, propomos uma extensão da distribuição skew-normal proposta por Arellano-Valle *et al.* (2007) e também analisamos algumas propriedades de tais extensões.

Este capítulo está organizado como segue. Na Seção 3.2, propomos uma classe alternativa de distribuições SMSN, denominada classe de distribuições misturas de escala skew-normal estendidas (SSMSN) que será útil para inferência bayesiana. Na Seção 3.3, introduzimos o modelo de regressão sob a classe de distribuições SSMSN. Em seguida, na Seção 3.4, apresentamos um exemplo indicando a utilidade da metodologia proposta no contexto bayesiano. Posteriormente, na Seção 3.5, introduzimos algumas genera-lizações das distribuições SMSN. Finalmente, algumas observações finais são dadas na Seção 3.6 e provas técnicas dos resultados desenvolvidos neste capítulo estão no Apêndice B.

3.2 Classe Alternativa de Distribuições Misturas de Escala Skew-Normal

Nesta seção, propomos uma classe alternativa de distribuições SMSN e que faz da inferência bayesiana uma alternativa viável para tratar os modelos lineares. Esta classe alternativa de distribuições foi motivada pela Definição 1 de Arellano-Valle *et al.* (2007).

Antes de discutirmos nossa proposta metodológica, apresentamos uma breve introdução da distribuição skew-normal definida por Arellano-Valle *et al.* (2007), denotada por SSN. Como considerado em Arellano-Valle *et al.* (2007), um vetor aleatório $\mathbf{Y}$ segue uma distribuição SSN p-variada com vetor de locação $\boldsymbol{\mu} \in \mathbb{R}^p$, matriz de dispersão $\boldsymbol{\Sigma}$ (definida positiva $p \times p$) e matriz de assimetria $(p \times k)$ $\boldsymbol{\Lambda}$ se sua pdf é

$$f(\mathbf{y}) = 2^k \phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Omega}) \Phi_k(\boldsymbol{\Lambda}^\top \boldsymbol{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\mu})|\boldsymbol{\Delta}), \quad \mathbf{y} \in \mathbb{R}^p, \tag{3.1}$$

onde $\boldsymbol{\Omega} = \boldsymbol{\Sigma} + \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top$, $\boldsymbol{\Delta} = (\mathbf{I}_k + \boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda})^{-1} = \mathbf{I}_k - \boldsymbol{\Lambda}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\Lambda}$ ($\mathbf{I}_k$ é uma matriz identidade $k \times k$), $\phi_p(.|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e $\Phi_p(.|\boldsymbol{\Sigma})$ denota a pdf da $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e cdf da $N_p(\mathbf{0}, \boldsymbol{\Sigma})$, respectivamente. Assim, consideramos $SSN_{p,k}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda})$ para indicar que $\mathbf{Y}$ tem densidade dada em (3.1) e $SSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda})$ quando $p = k$. Note que para $\boldsymbol{\Lambda} = \mathbf{0}$, a densidade (3.1) reduz para a usual densidade da $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Esta distribuição SSN multivariada generaliza a distribuição skew-normal de Sahu *et al.* (2003). De fato, para $k = p$ e $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \ldots, \lambda_p\}$, a densidade (3.1) reduz para a densidade da skew-normal de Sahu *et al.* (2003) discutida, por exemplo, em De La Cruz (2008), Jara *et al.* (2008) e Lin (2009). Neste caso, diag{.} denota a matriz diagonal obtida por extrair os elementos da diagonal principal de uma matriz quadrada ou da diagonalização de um dado vetor. Além disso, observe que o modelo definido em (3.1) é um caso especial da distribuição skew-normal fundamental introduzida na Definição 2.1 em Arellano-Valle &

Genton (2005). Quando $\boldsymbol{\mu} = \mathbf{0}$ e $\boldsymbol{\Omega} = \mathbf{I}_p$, temos a distribuição skew-normal fundamental canônica introduzida na Definição 2.2 em Arellano-Valle & Genton (2005). Por sua vez, se $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Omega} = \mathbf{I}_p$ e $\boldsymbol{\Lambda} = \boldsymbol{\delta} = \dfrac{\boldsymbol{\lambda}}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}}$, temos a distribuição skew-normal definida em (2.1). Neste contexto, propomos um estudo unificado da classe de distribuições SMSN.

Pela Proposição 1 de Arellano-Valle *et al.* (2007), a distribuição SSN (3.1) apresenta a seguinte representação estocástica

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}|\mathbf{T}_1| + \boldsymbol{\Sigma}^{1/2}\mathbf{T}_2, \tag{3.2}$$

onde $\mathbf{T}_1 \sim N_k(0, \mathbf{I}_k)$ é independente de $\mathbf{T}_2 \sim N_p(0, \mathbf{I}_p)$, tal que $\boldsymbol{\tau} = |\mathbf{T}_1|$ segue uma distribuição half-normal padrão k-dimensional, denotada por $HN_k(\mathbf{0}, \mathbf{I}_k)$. Note que a expressão (3.2) é uma ferramenta útil para gerar números aleatórios e para propósitos teóricos. De acordo com Arellano-Valle *et al.* (2007), o vetor de médias e a matriz de covariâncias de $\mathbf{Y} \sim SSN_{p,k}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda})$ são dados por

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\boldsymbol{\lambda} \text{ e } Var[\mathbf{Y}] = \boldsymbol{\Sigma} + (1 - \frac{2}{\pi})\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top, \tag{3.3}$$

onde $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)^\top$ denota os elementos da diagonal principal de $\boldsymbol{\Lambda}$. Note que se $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \ldots, \lambda_p\}$, a inclusão de assimetria em um modelo não afeta a estrutura de correlação do modelo original simétrico, i.e., alterações ocorrem nos valores das correlações, mas a estrutura permanece a mesma. Entretanto, isso não é verdade para as distribuições assimétricas de Azzalini & Capitanio (2003), uma vez que a introdução de assimetria altera a estrutura de correlação do modelo original.

A seguir propomos a generalização da distribuição skew-normal considerada em Arellano-Valle *et al.* (2007), denominamos por distribuições misturas de escala skew-normal estendidas (SSMSN). Um vetor aleatório p-dimensional com distribuição SSMSN

é construído a partir da mistura escala de uma variável aleatória skew-normal (SSN) e uma variável aleatória positiva, i.e.,

$$\mathbf{Y} = \boldsymbol{\mu} + \kappa^{1/2}(U)\mathbf{Z}, \tag{3.4}$$

onde $\mathbf{Z} \sim SSN_{p,k}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda})$, $\kappa(.)$ é uma função de pesos, U é variável aleatória de mistura positiva com cdf $H(u;\boldsymbol{\nu})$ e pdf $h(u;\boldsymbol{\nu})$, independente de $\mathbf{Z}$, onde $\boldsymbol{\nu}$ é um vetor ou um escalar relacionado à distribuição de U. Do lema B.1.1 dado no Apêndice B, temos que $\mathbf{Y}|U = u \sim SSN_{p,k}(\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma}, \kappa^{1/2}(u)\boldsymbol{\Lambda})$. Assim, as funções densidade e geradora de momentos (mgf) de $\mathbf{Y}$ são, respectivamente, dadas por

$$f(\mathbf{y}) = 2^k \int_0^\infty \phi_p(\mathbf{y}|\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Omega})\Phi_k(\kappa^{-1/2}(u)\boldsymbol{\Lambda}^\top\boldsymbol{\Omega}^{-1}(\mathbf{y}-\boldsymbol{\mu})|\boldsymbol{\Delta})dH(u;\boldsymbol{\nu}), \tag{3.5}$$

e

$$M_{\mathbf{Y}}(\mathbf{s}) = 2^k \int_0^\infty \exp\left(\mathbf{s}^\top\boldsymbol{\mu} + \frac{1}{2}\mathbf{s}^\top\kappa(u)\boldsymbol{\Omega}\mathbf{s}\right)\Phi_k(\kappa^{1/2}(u)\boldsymbol{\Lambda}^\top\mathbf{s})dH(u;\boldsymbol{\nu}), \tag{3.6}$$

onde $\mathbf{s}, \mathbf{y} \in \mathbb{R}^p$ e $\Phi_k(.)$ denota a cdf da normal multivariada padrão. Neste caso, denotamos por $\mathbf{Y} \sim SSMSN_{p,k}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}; H)$ quando $\mathbf{Y}$ tem pdf dada em (3.5). Note que quando $\boldsymbol{\Lambda} = \mathbf{0}$, a classe de distribuições SSMSN reduz para classe SMN, um caso particular das distribuições SMSN definidas na Seção 2.2.

Em seguida, obtemos o vetor de médias e matriz de covariâncias para o vetor aleatório SSMSN. A prova é simples e decorre de (3.3) e (3.4).

Proposição 3.2.1 *Suponha que* $\mathbf{Y} \sim SSMSN_{p,k}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}; H)$. *Então,*
a) Se $E[\kappa^{1/2}(U)] < \infty$, *temos que* $E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}E[\kappa^{1/2}(U)]\boldsymbol{\lambda}$;
b) Se $E[\kappa(U)] < \infty$, *temos que* $Var[\mathbf{Y}] = E[\kappa(U)]\left(\boldsymbol{\Omega} + \frac{2}{\pi}(\boldsymbol{\lambda}\boldsymbol{\lambda}^\top - \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top)\right) - \frac{2}{\pi}E^2[\kappa^{1/2}(U)]\boldsymbol{\lambda}\boldsymbol{\lambda}^\top$.

A classe de distribuições SSMSN tem como membros particulares as distribuições skew-t (SST), skew-slash (SSL) e skew-normal contaminada (SSCN). Essas distribuições têm caudas mais pesadas que a distribuição skew-normal (SSN) e dessa forma, podem ser consideradas para a obtenção de procedimentos robustos. Note que diferentemente da distribuição SSN, os componentes da classe SSMSN são em geral correlacionados quando a assimetria está presente e até mesmo se $\boldsymbol{\Omega}$ é uma matriz diagonal. Este fato tem uma explicação informal, uma vez que todos os componentes da classe de distribuições SSMSN compartilham do mesmo fator de mistura $\kappa^{1/2}(U)$.

Portanto, a seguir descrevemos algumas distribuições que pertencem à classe SSMSN e para cada elemento desta classe, calculamos o vetor de médias e a matriz de covariâncias. Para todos os casos, considere $\mathbf{A} = \boldsymbol{\Lambda}^\top \boldsymbol{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\mu})$.

- **Distribuição SST Multivariada**

  A distribuição skew-t estendida multivariada com ν graus de liberdade, denotada por $SST_{p,k}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}; \nu)$, é obtida de (3.5) considerando $\kappa(u) = 1/u$ e $U \sim Gamma(\nu/2, \nu/2)$, $\nu > 0$. Segue do Lema B.1.2 dado no Apêndice B que a pdf de $\mathbf{Y}$ é

  $$f(\mathbf{y}) = 2^k \, t_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Omega}; \nu) T_k \left(\sqrt{\frac{p+\nu}{d+\nu}} \mathbf{A} | \boldsymbol{\Delta}; \nu + p \right), \quad \mathbf{y} \in \mathbb{R}^p,$$

  onde $t_p(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$ e $T_p(\cdot|\boldsymbol{\Sigma}; \nu)$ denotam, respectivamente, as pdf e cdf da distribuição t-Student p-variada e $d = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ é a distância de Mahalanobis. Um caso particular da distribuição skew-t estendida é a distribuição skew-cauchy estendida quando $\nu = 1$. Além disso, quando $\nu \uparrow \infty$, obtemos a distribuição SSN como caso limite. Aplicações da distribuição SST com $p = k$ e

$\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \ldots, \lambda_p\}$ podem ser encontradas em De La Cruz (2008). Dos resultados da Proposição 3.2.1, temos que o vetor de médias e matriz de covariâncias de $\mathbf{Y}$ são, respectivamente,

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{\nu}{\pi}}\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\boldsymbol{\lambda}, \quad \nu > 1 \quad \text{e}$$

$$Var[\mathbf{Y}] = \frac{\nu}{\nu-2}\left(\boldsymbol{\Omega} + \frac{2}{\pi}(\boldsymbol{\lambda}\boldsymbol{\lambda}^\top - \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top)\right) - \frac{\nu}{\pi}\left(\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\right)^2 \boldsymbol{\lambda}\boldsymbol{\lambda}^\top, \quad \nu > 2.$$

- **Distribuição SSL Multivariada**

A distribuição skew-slash estendida multivariada, denotada por $SSL_{p,k}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}; \nu)$, foi obtida considerando $\kappa(u) = 1/u$ e $U \sim Beta(\nu, 1)$, $\nu > 0$. Segue de (3.5) que a pdf de $\mathbf{Y}$ é dada por

$$f(\mathbf{y}) = 2^k \nu \int_0^1 u^{\nu-1} \phi_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Omega}) \Phi_k(u^{1/2}\mathbf{A}|\boldsymbol{\Delta}) du, \quad \mathbf{y} \in \mathbb{R}^p.$$

A distribuição skew-slash estendida reduz para a distribuição SSN quando $\nu \uparrow \infty$. Da Proposição 3.2.1, obtemos

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\frac{2\nu}{2\nu-1}\boldsymbol{\lambda}, \quad \nu > 1/2 \quad \text{e}$$

$$Var[\mathbf{Y}] = \frac{\nu}{\nu-1}\left(\boldsymbol{\Omega} + \frac{2}{\pi}(\boldsymbol{\lambda}\boldsymbol{\lambda}^\top - \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top)\right) - \frac{2}{\pi}\left(\frac{2\nu}{2\nu-1}\right)^2 \boldsymbol{\lambda}\boldsymbol{\lambda}^\top, \quad \nu > 1.$$

- **Distribuição SSCN Multivariada**

A distribuição skew-normal contaminada estendida é obtida quando U é uma variável aleatória discreta assumindo dois estados, onde sua pdf dado o vetor de parâmetros $\boldsymbol{\nu} = (\nu_1, \nu_2)^\top$ é dada por

$$h(u; \boldsymbol{\nu}) = \nu_1 \mathbb{I}_{(u=\nu_2)} + (1-\nu_1)\mathbb{I}_{(u=1)}, \quad 0 < \nu_1 < 1, \, 0 < \nu_2 \leq 1, \tag{3.7}$$

Dessa forma, de (3.5), temos que

$$f(\mathbf{y}) = 2^k \left\{ \nu_1 \phi_p(\mathbf{y}|\boldsymbol{\mu}, \nu_2^{-1}\boldsymbol{\Omega}) \Phi_k(\nu_2^{1/2}\mathbf{A}|\boldsymbol{\Delta}) + (1-\nu_1)\phi_k(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Omega}) \Phi_p(\mathbf{A}|\boldsymbol{\Delta}) \right\},$$

e neste caso, especificamos por $SSCN_{p,k}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}; \nu_1, \nu_2)$, $0 < \nu_1 < 1$, $0 < \nu_2 \leq 1$. O parâmetro ν_1 pode ser interpretado como a proporção de outliers enquanto que ν_2 pode ser visto como o fator de escala. A distribuição skew-normal contaminada estendida reduz para a distribuição SSN quando $\nu_2 = 1$. Assim, da Proposição 3.2.1, temos que

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}} \left(\frac{\nu_1}{\sqrt{\nu_2}} + 1 - \nu_1 \right) \boldsymbol{\lambda} \text{ e}$$

$$Var[\mathbf{Y}] = \left(\frac{\nu_1}{\nu_2} + 1 - \nu_1 \right) \left(\boldsymbol{\Omega} + \frac{2}{\pi}(\boldsymbol{\lambda}\boldsymbol{\lambda}^\top - \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top) \right) - \frac{2}{\pi} \left(\frac{\nu_1}{\sqrt{\nu_2}} + 1 - \nu_1 \right)^2 \boldsymbol{\lambda}\boldsymbol{\lambda}^\top.$$

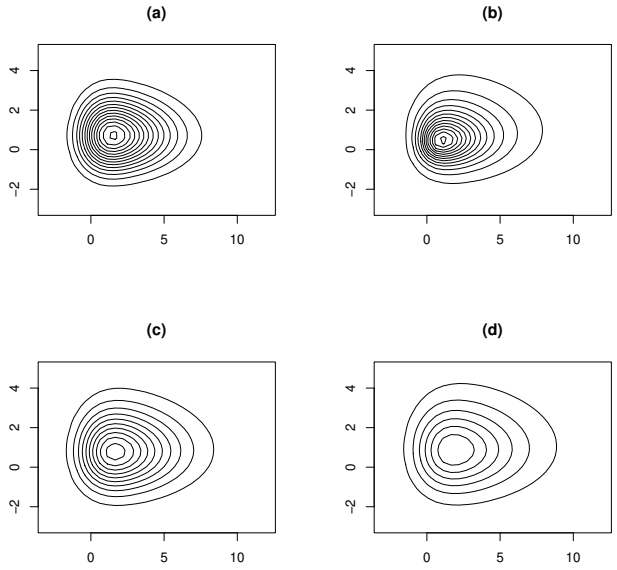


Figura 3.1: Contornos de alguns elementos da classe SSMSN bivarida padrão. (a) $SSN_{2,1}(\boldsymbol{\Lambda})$ (b) $SST_{2,1}(\boldsymbol{\Lambda}; 2)$ (c) $SSCN_{2,1}(\boldsymbol{\Lambda}; 0.5, 0.5)$ (d) $SSL_{2,1}(\boldsymbol{\Lambda}; 1)$, onde $\boldsymbol{\Lambda} = (3, 1)^\top$.

A Figura 3.1 mostra os contornos das densidades associadas com algumas distribuições SSMSN bivariadas padrão ($\boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \mathbf{I}_2$), por exemplo, $SSN_{2,1}(\boldsymbol{\Lambda})$, $SST_{2,1}(\boldsymbol{\Lambda}; 2)$, $SSL_{2,1}(\boldsymbol{\Lambda}; 1)$ e $SSCN_{2,1}(\boldsymbol{\Lambda}; 0.5, 0.5)$, com $\boldsymbol{\Lambda} = (3, 1)^\top$. Note que estes contornos não são elípticos e eles podem ser bastante assimétricos dependendo da escolha do parâmetro $\boldsymbol{\Lambda}$.

Em seguida, caracterizamos a distribuição de uma transformação linear arbitrária $\mathbf{b} + \mathbf{BY}$ através de sua mgf e da pdf quando $\mathbf{B}$ é uma matriz não singular. A prova segue diretamente de (3.6), tal que $M_{\mathbf{b}+\mathbf{BY}}(\mathbf{s}) = \exp\left(\mathbf{s}^\top \mathbf{b}\right) M_{\mathbf{Y}}(\mathbf{B}^\top \mathbf{s})$ e de (3.5), tal que $f_{\mathbf{b}+\mathbf{BY}}(\mathbf{v}) = |\det(\mathbf{B})|^{-1} \mathrm{f}_{\mathbf{Y}}(\mathbf{B}^{-1}(\mathbf{v} - \mathbf{b}))$.

Proposição 3.2.2 *Considere que* $\mathbf{Y} \sim SSMSN_{p,k}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}; H)$. *Então, para quaisquer vetor fixo* $\mathbf{b} \in \mathbb{R}^m$ *e matriz* $\mathbf{B} \in \mathbb{R}^{m \times p}$ *de posto completo, temos que*

$$\mathbf{V} = \mathbf{b} + \mathbf{BY} \sim SSMSN_{m,k}(\mathbf{b} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^\top, \boldsymbol{B\Lambda}; H).$$

Além disso, a matriz $\mathbf{B}$ é não singular quando $m = p$ e então,

$$\begin{aligned} f_{\mathbf{b}+\mathbf{BY}}(\mathbf{v}) &= 2^k |\det(\mathbf{B})|^{-1} \int_0^\infty \phi_{\mathrm{p}}(\mathbf{B}^{-1}(\mathbf{v} - \mathbf{b}) | \boldsymbol{\mu}, \kappa(\mathrm{u})\boldsymbol{\Omega}) \\ &\times \Phi_k(\kappa^{-1/2}(u)\boldsymbol{\Lambda}^\top \boldsymbol{\Omega}^{-1}(\mathbf{B}^{-1}(\mathbf{v} - \mathbf{b}) - \boldsymbol{\mu}) | \boldsymbol{\Delta}) dH(u; \boldsymbol{\nu}), \mathbf{v} \in \mathbf{R}^p. \end{aligned}$$

Da Proposição 3.2.2 com $\mathbf{b} = \mathbf{0}$, temos as seguintes propriedades adicionais das distribuições SSMSN.

Corolário 3.2.3 *Considere que* $\mathbf{Y} \sim SSMSN_{p,k}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}; H)$. *Assim,*

(i) $-\mathbf{Y} \sim SSMSN_{p,k}(-\boldsymbol{\mu}, \boldsymbol{\Sigma}, -\boldsymbol{\Lambda}; H)$;

(ii) $\boldsymbol{a}^\top \mathbf{Y} \sim SSMSN_{1,k}(\boldsymbol{a}^\top \boldsymbol{\mu}, \boldsymbol{a}^\top \boldsymbol{\Sigma} \boldsymbol{a}, \boldsymbol{a}^\top \boldsymbol{\Lambda}; H)$ *para qualquer vetor unitário* $\boldsymbol{a} \in \mathbb{R}^p$;

(iii) $\boldsymbol{B}\mathbf{Y} \sim SSMSN_{p,k}(\boldsymbol{B\mu}, \boldsymbol{B\Sigma B}^\top, \boldsymbol{B\Lambda}; H)$ *para qualquer matriz ortogonal* $\boldsymbol{B}$ $k \times k$.

Na Proposição 3.2.2, mostramos que o vetor aleatório SSMSN é invariante sob transformações lineares o que implica que a distribuição marginal de $\mathbf{Y}$ é ainda SSMSN.

Proposição 3.2.4 *Considere que* $\mathbf{Y} \sim SSMSN_{p,k}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}; H)$ *e que* $\mathbf{Y}$ *está particionado como* $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ *com dimensões* p_1 *e* p_2 *($p_1 + p_2 = p$), respectivamente. Adicionalmente, considere* $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top$ *e* $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_1^\top, \boldsymbol{\Lambda}_2^\top)^\top$ *sendo as correspondentes partições de* $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$ *e* $\boldsymbol{\Lambda}$. *Dessa forma, a densidade marginal de* $\mathbf{Y}_1$ *é* $SSMSN_{p_1,k}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\Lambda}_1; H)$.

Prova A prova é baseada na Proposição 3.2.2, considerando $\mathbf{B} = [\mathbf{I}_{p_1}, \mathbf{0}_{p_2}]$ e $\mathbf{b} = \mathbf{0}$.

A seguir denote por $\boldsymbol{\Lambda}$ a matriz $(\lambda_{ij}; i = 1, \ldots, p, j = 1, \ldots, k)$ e por $\boldsymbol{\Lambda}_i^\top = (\lambda_{i1}, \ldots, \lambda_{ik})$, $i = 1, \ldots, p$, as linhas da matriz $\boldsymbol{\Lambda}$. Além disso, considere o vetor $p \times 1$ com 1 na j-ésima posição, $j = 1, \ldots, k$, definido por $\mathbf{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)^\top$.

Proposição 3.2.5 *Considere que* $\mathbf{Y} = (Y_1, \ldots, Y_p)^\top \sim SSMSN_{p,k}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}; H)$. *Então,* $\mathbf{Y}_i \sim SSMSN_{1,k}(\mu_i, \sigma_{ii}, \boldsymbol{\Lambda}_i^\top; H), i = 1, \ldots, p$.

Prova Note primeiro que $\mathbf{e}_i^\top \mathbf{Y} = Y_i$, $\mathbf{e}_i^\top \mu = \mu_i$, $\mathbf{e}_i^\top \boldsymbol{\Lambda} = \boldsymbol{\Lambda}_i^\top$ e $\mathbf{e}_i^\top \boldsymbol{\Sigma} \mathbf{e}_i = \sigma_{ii}$, onde $||\mathbf{e}_i|| = 1$ e σ_{ii} são os elementos da diagonal principal de $\boldsymbol{\Sigma}, i = 1, \ldots, p$. Pela parte (ii) do Corolário 3.2.3 segue a prova desta proposição.

Proposição 3.2.6 *Sob a notação definida na Proposição 3.2.4, considerando* $\boldsymbol{\Omega}_{12} = \boldsymbol{\Omega}_{21} = \mathbf{0}$, *temos que os vetores aleatórios* $\mathbf{Y}_1$ *e* $\mathbf{Y}_2$ *são independentes dado* $U = u$, *sob as seguintes condições na matriz* $\boldsymbol{\Lambda}$*:*

(i) $\boldsymbol{\Lambda}_{12} = \boldsymbol{\Lambda}_{21} = 0$. *Neste caso, dado* $U = u$, $\mathbf{Y}_i \sim SSN_{p_i,k_i}(\boldsymbol{\mu}_i, \kappa(u)\boldsymbol{\Sigma}_{ii}, \kappa^{1/2}(u)\boldsymbol{\Lambda}_{ii})$, $i = 1, 2$; *ou*

(ii) $\boldsymbol{\Lambda}_{11} = \boldsymbol{\Lambda}_{22} = 0$. *Nesta caso, dado* $U = u$, $\mathbf{Y}_1 \sim SSN_{p_1,k_2}(\boldsymbol{\mu}_1, \kappa(u)\boldsymbol{\Sigma}_{11}, \kappa^{1/2}(u)\boldsymbol{\Lambda}_{12})$ *e* $\mathbf{Y}_2 \sim SSN_{p_2,k_1}(\boldsymbol{\mu}_2, \kappa(u)\boldsymbol{\Sigma}_{22}, \kappa^{1/2}(u)\boldsymbol{\Lambda}_{21})$.

Além disso, se $\boldsymbol{\mu}_2 = \mathbf{0}$, *temos que* $\mathbf{Y}_1$ *e* $\mathbf{Y}_2$ *são não correlacionados.*

Prova Dado $U = u$, $\mathbf{Y} \sim SSN_{p,k}(\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma}, \kappa^{1/2}(u)\boldsymbol{\Lambda})$ e segue da Proposição 2.7 em Arellano-Valle & Genton (2005) que $\mathbf{Y}_1$ e $\mathbf{Y}_2$ são independentes. Se $\boldsymbol{\mu}_2 = 0$, da Proposição 3.2.4, temos que $\mathbf{Y}_1 \sim SSMSN_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\Lambda}_1; H)$ e $\mathbf{Y}_2 \sim SSMN_{p_2}(\mathbf{0}, \boldsymbol{\Sigma}_{22}, \boldsymbol{\Lambda}_2; H)$. Dessa forma, $Cov[\mathbf{Y}_1, \mathbf{Y}_2] = E[\mathbf{Y}_1\mathbf{Y}_2^\top] = E[E[\mathbf{Y}_1\mathbf{Y}_2^\top|U]] = E[E[\mathbf{Y}_1|U]E[\mathbf{Y}_2^\top|U]] = \mathbf{0}$, concluindo a prova.

Um importante resultado que será útil na implementação do algoritmo EM para a classe SSMSN é dado a seguir. Lin (2009) propõe o algoritmo EM para estimar os parâmetros do modelo skew-t introduzido por Sahu *et al.* (2003), i.e., especificando $k = p$ e $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \ldots, \lambda_p\}$.

Proposição 3.2.7 *Considere que* $\mathbf{Y} \sim SSMSN_{p,k}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ *e* $U \sim H$ *é a variável de mistura. Assim,*

$$\begin{aligned}
\kappa_r &= E[\kappa^{-r}(U)|\mathbf{y}] = \frac{2^k f_0(\mathbf{y})}{f(\mathbf{y})} E[\kappa^{-r}(U_\mathbf{y})\Phi_k(\kappa^{-1/2}(U_\mathbf{y})A_*)], \\
\tau_r &= E[\kappa^{-r/2}(U)W_{\Phi_k}(\kappa^{-1/2}(U)A)|\mathbf{y}] = \frac{2^k f_0(\mathbf{y})}{f(\mathbf{y})} E[\kappa^{-r/2}(U_\mathbf{y})\phi_k(\kappa^{-1/2}(U_\mathbf{y})A_*)],
\end{aligned}$$

onde f_0 *é a pdf of* $\mathbf{Y}_0 \sim SMN_p(\boldsymbol{\mu}, \boldsymbol{\Omega}; H)$, $U_\mathbf{y} \overset{d}{=} U|\mathbf{Y}_0 = \mathbf{y}$, $W_{\Phi_k}(x) = \phi_k(x)/\Phi_k(x)$, $x \in \mathbb{R}$, *tal que* $\phi_k(.)$ *e* $\Phi_k(.)$ *denotam as pdf e cdf da normal multivariada padrão, respectivamente e* $A_* = \boldsymbol{\Delta}^{-1/2}\boldsymbol{\Lambda}^\top\boldsymbol{\Omega}^{-1}(\mathbf{Y} - \boldsymbol{\mu})$.

Prova A prova deste resultado é similar à prova da Proposição 2.2.1, com $g(u) = \kappa^{-r}(u)$ e $g(u) = \kappa^{-r/2}(u)W_{\Phi_k}(\kappa^{-1/2}(u)\mathrm{A}_*)$, respectivamente.

Há algumas outras variantes das distribuições assimétricas disponíveis na literatura. As distribuições desenvolvidas nesta seção, entretanto, são mais fáceis, no sentido algébrico, para trabalhar com análise bayesiana do que outras. A análise bayesiana em problemas de regressão sob distribuições com caudas pesadas tem recebido considerável atenção na literatura estatística recente. Na próxima seção, propomos o modelo de regressão sob a classe de distribuições misturas de escala skew-normal estendidas (SSMSN).

3.3 Modelo de Regressão com Assimetria

Nesta seção, propomos o modelo de regressão, onde a variável resposta segue uma distribuição SSMSN (SSMSN-RM), considerando um caso especial com $p = k$ e $\mathbf{\Lambda} = \text{diag}\{\boldsymbol{\lambda}\}$, mas não $\mathbf{\Sigma}$ diagonal, condição implicitamente suposta por Sahu *et al.* (2003). Neste caso, especificamos $\mathbf{\Lambda}\mathbf{\Lambda}^\top = \mathbf{\Lambda}^2$ e segue que para $\boldsymbol{\lambda} = \mathbf{0}$, temos a distribuição normal. Além disso, para valores positivos de $\boldsymbol{\lambda}$, temos distribuição assimétrica à direita e para valores negativos, obtemos distribuição assimétrica à esquerda.

De acordo com Sahu *et al.* (2003), o modelo de regressão sob a classe de distribuições SSMSN será definido como segue. Suponha que temos n observações independentes, $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, onde

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n,$$

em que $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ip})^\top$ é um vetor $p \times 1$ de respostas contínuas observadas para o i-ésimo indivíduo, $\mathbf{X}_i$ de dimensão $n \times p$ é a matriz de planejamento conhecida, $\boldsymbol{\beta}$ é um vetor $p \times 1$ dos coeficientes de regressão desconhecidos e $\boldsymbol{\epsilon}_i$ é o vetor $p \times 1$ de erros

independentes, tal que

$$\mathbf{Y}_i \sim SSMSN_p(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}; H), i = 1, \ldots, n.$$

A matriz $\boldsymbol{\Sigma}$ $(p \times p)$ definida positiva pode ser estruturada ou não. Neste caso, assumimos uma matriz não estruturada $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\gamma})$, onde $\boldsymbol{\gamma}$ denota os elementos da matriz triangular superior de $\boldsymbol{\Sigma}$.

Uma característica importante deste modelo é que ele pode ser formulado através de uma representação hierárquica, permitindo uma implementação fácil de aplicações estatísticas em programas bayesianos como WinBUGS. De (3.4) e da representação estocástica do vetor aleatório SSN dada em (3.2), temos que o SSMSN-RM pode ser escrito hierarquicamente como

$$\begin{aligned} \mathbf{Y}_i|\mathbf{T}_i = \mathbf{t}_i, U_i = u_i, &\overset{\text{ind}}{\sim} N_p(\mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\Lambda}\mathbf{t}_i, \kappa(u_i)\boldsymbol{\Sigma}), \\ \mathbf{T}_i|U_i = u_i &\overset{\text{ind}}{\sim} HN_p(0, \kappa(u_i)\mathbf{I}_p), \\ U_i &\overset{\text{iid}}{\sim} H(\boldsymbol{\nu}), \end{aligned}$$

$i = 1, \ldots, n$, todos independentes. Definindo $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$ como sendo a amostra observada, $\mathbf{u} = (u_1, \ldots, u_n)^\top$ e $\mathbf{t} = (\mathbf{t}_1, \ldots, \mathbf{t}_n)^\top$ segue que a função verossimilhança dos dados completos $(\mathbf{y}^\top, \mathbf{u}^\top, \mathbf{t}^\top)^\top$ (verossimilhança aumentada) é dada por

$$L(\boldsymbol{\theta}|\mathbf{y}, \mathbf{u}, \mathbf{t}) \propto \prod_{i=1}^{n} [\phi_p(\mathbf{y}_i|\mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\Lambda}\mathbf{t}_i, \kappa(u_i)\boldsymbol{\Sigma})\phi_p(\mathbf{t}_i|0, \kappa(u_i)\mathbf{I}_p)\mathbf{I}_{\{\mathbf{t}_i > \mathbf{0}\}}h(u_i; \boldsymbol{\nu})], \quad (3.8)$$

onde $\mathbf{I}_A$ é a função indicadora do conjunto A.

Assim, para a especificação bayesiana completa do modelo, precisamos assumir distribuições prioris para o vetor de parâmetros desconhecido $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\lambda}^\top, \boldsymbol{\nu}^\top)^\top$.

3.3.1 Distribuições Prioris e Posteriores Conjuntas

Nesta seção, descrevemos a escolha de prioris para os parâmetros do SSMSN-RM. Assumindo os elementos do vetor de parâmetros independentes, consideramos que a distribuição conjunta à priori para os parâmetros desconhecidos tem densidade dada por

$$\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\beta})\pi(\boldsymbol{\Sigma})\pi(\boldsymbol{\lambda})\pi(\boldsymbol{\nu}).$$

Com o intuito de garantir posteriores próprias, adotamos prioris próprias. Além disso, assim como Sahu *et al.* (2003) (veja página 11), consideramos uma distribuição priori normal multivariada para $\boldsymbol{\beta}$ com hiperparâmetros conhecidos $\boldsymbol{\beta}_0$ e $\mathbf{S}_{\boldsymbol{\beta}}$. A distribuição Wishart inversa ($IW_p(\mathbf{r}, \mathbf{T})$) foi usada como priori para a matriz de dispersão $\boldsymbol{\Sigma}$; veja Sahu *et al.* (2003), página 10, para mais detalhes sobre a distribuição Wishart inversa. A priori para $\boldsymbol{\lambda}$ é a distribuição normal multivariada com hiperparâmetros conhecidos $\boldsymbol{\lambda}_0$ e $\mathbf{S}_{\lambda}$. Finalmente, a distribuição priori de $\boldsymbol{\nu}$ com densidade $\pi(\boldsymbol{\nu})$ depende de uma particular distribuição SSMSN escolhida. Considerando a função de verossimilhança (3.8) e as prioris especificadas acima, temos que a distribuição posteriori aumentada de $\boldsymbol{\theta}$ é

$$\begin{aligned}\pi(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\nu}, \mathbf{u}, \mathbf{t}|\mathbf{y}) \quad \propto \quad & \prod_{i=1}^{n}[\phi_p(\mathbf{y}_i|\mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\Lambda}\mathbf{t}_i, \kappa(u_i)\boldsymbol{\Sigma}) \\ & \times \phi_p(\mathbf{t}_i|0, \kappa(u_i)\mathbf{I}_p)\mathbf{I}_{\{\mathbf{t}_i>\mathbf{0}\}}h(u_i;\boldsymbol{\nu})]\pi(\boldsymbol{\theta}). \qquad (3.9)\end{aligned}$$

Entretanto, é bastante complicado obter analiticamente as distribuições posteriores marginais das quantidades de interesse envolvidas na equação (3.9). Desse modo, para conduzir uma análise bayesiana, vamos considerar métodos MCMC, tal como o amostrador de Gibbs, pois as distribuições posteriores condicionais podem ser inferidas.

3.3.2 Esquema MCMC

Métodos MCMC são facilitados usando a distribuição posteriori aumentada proveniente do modelo. Dessa forma, a especificação completa do modelo é como segue

$$\begin{aligned}
\mathbf{Y}_i|\mathbf{t}_i, u_i, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda} &\sim N_p(\mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\Lambda}\mathbf{t}_i, \kappa(u_i)\boldsymbol{\Sigma}) && (3.10)\\
\mathbf{T}_i|u_i &\sim HN_p(0, \kappa(u_i)\mathbf{I}_p) && (3.11)\\
U_i|\boldsymbol{\nu} &\sim H(u_i;\boldsymbol{\nu}) && (3.12)\\
\boldsymbol{\beta} &\sim N_m(\boldsymbol{\beta}_0, \mathbf{S}_{\boldsymbol{\beta}}) && (3.13)\\
\boldsymbol{\Sigma} &\sim IW_p(\mathbf{r}, \mathbf{T}) && (3.14)\\
\boldsymbol{\lambda} &\sim N_p(\boldsymbol{\lambda}_0, \mathbf{S}_{\boldsymbol{\lambda}})\mathbf{I}_{\{\boldsymbol{\lambda}>\mathbf{0}\}} && (3.15)\\
\boldsymbol{\nu} &\sim \pi(\boldsymbol{\nu}). && (3.16)
\end{aligned}$$

Note que a representação (3.10)-(3.16) é importante, pois permite escrever facilmente o código BUGS. Além disso, a partir desta representação, as distribuições posteriores condicionais aumentadas requeridas para implementar o amostrador de Gibbs são simples de serem obtidas.

Observe que dado $U = u$, as distribuições condicionais posteriores aumentadas têm a mesma forma para qualquer elemento da classe SSMSN. Portanto, sob o modelo aumentado descrito em (3.10)–(3.16), as distribuições posteriores condicionais aumentadas são dadas por

$$\boldsymbol{\beta}|\mathbf{u}, \ldots \sim N_m(\mathbf{A}_\beta^{-1}\mathbf{a}_\beta,\ \mathbf{A}_\beta^{-1}),$$

onde $\mathbf{A}_\beta = \mathbf{S}_\beta^{-1} + \sum_{i=1}^n \kappa^{-1}(u_i)\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i$ e $\mathbf{a}_\beta = \mathbf{S}_\beta^{-1}\boldsymbol{\beta}_0 + \sum_{i=1}^n \kappa^{-1}(u_i)\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \boldsymbol{\Lambda}\mathbf{t}_i)$;

$$\boldsymbol{\Sigma}|\mathbf{u}, \ldots \sim IW_p(\mathbf{r} + n, (\mathbf{T}^{-1} + \sum_{i=1}^n \kappa^{-1}(u_i)(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \boldsymbol{\Lambda}\mathbf{t}_i)(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \boldsymbol{\Lambda}\mathbf{t}_i)^\top)^{-1});$$

$$\boldsymbol{\lambda}|\mathbf{u},\ldots \;\sim\; N_p(\mathbf{B}^{-1}\mathbf{b},\mathbf{B}^{-1})$$

onde $\mathbf{B}=\mathbf{S}_{\boldsymbol{\lambda}}^{-1}+\sum_{i=1}^{n}\kappa^{-1}(u_i)\text{diag}\{\mathbf{t}_i\}\boldsymbol{\Sigma}^{-1}\text{diag}\{\mathbf{t}_i\}$ e $\mathbf{b}=\mathbf{S}_{\boldsymbol{\lambda}}^{-1}\boldsymbol{\lambda}_0+\sum_{i=1}^{n}\kappa^{-1}(u_i)\text{diag}\{\mathbf{t}_i\}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})$;

$$\mathbf{T}_i|\mathbf{u},\ldots \;\sim\; N_p(\mathbf{A}_t^{-1}\mathbf{a}_{t_i},\kappa(u_i)\mathbf{A}_t^{-1})\mathbf{I}_{\{\mathbf{t}_i>\mathbf{0}\}},$$

onde $\mathbf{A}_t=\boldsymbol{\Lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\Lambda}+\mathbf{I}_p$ e $\mathbf{a}_{t_i}=\boldsymbol{\Lambda}^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})$, $i=1,\ldots,n$. Para completar a especificação do modelo, precisamos das distribuições posteriores condicionais aumentadas de U e do parâmetro $\boldsymbol{\nu}$. Para cada U_i, temos que

$$\begin{aligned}\pi(u_i|\boldsymbol{\theta},\mathbf{y},\mathbf{t}) \;\propto\;& \kappa^{-p}(u_i)h(u_i;\boldsymbol{\nu})\\ &\times\exp\left\{-\frac{1}{2}\kappa^{-1}(u_i)(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\boldsymbol{\Lambda}\mathbf{t}_i)^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\boldsymbol{\Lambda}\mathbf{t}_i)\right.\\ &\left.-\frac{1}{2}\kappa^{-1}(u_i)\mathbf{t}_i^\top\mathbf{t}_i\mathbf{I}_{\{\mathbf{t}_i>\mathbf{0}\}}\right\},\quad i=1,\ldots,n \end{aligned}\tag{3.17}$$

e para o parâmetro $\boldsymbol{\nu}$, temos que

$$\pi(\boldsymbol{\nu}|\boldsymbol{\theta}_1,\mathbf{y},\mathbf{u},\mathbf{t})\propto\pi(\boldsymbol{\nu})\prod_{i=1}^{n}h(u_i|\boldsymbol{\nu}),\tag{3.18}$$

onde $\boldsymbol{\theta}_1=(\boldsymbol{\beta}^\top,\boldsymbol{\gamma}^\top,\boldsymbol{\lambda}^\top)^\top$. As distribuições obtidas em (3.17) e (3.18) dependem da distribuição específica SSMSN adotada e, para cada caso SSMSN, da distribuição priori escolhida para $\boldsymbol{\nu}$; veja Apêndice B.2 para mais detalhes. Na próxima seção, apresentamos um exemplo prático dos resultados teóricos desenvolvidos para SSMSN-RM.

3.4 Aplicação: "Stack-Loss"

A seguir analisamos o conjunto de dados "Stack-Loss" apresentado por Brownlee (1960) no contexto gaussiano. Estes dados têm sido utilizados em ilustrações númericas por um grande número de programas estatísticos e autores, por exemplo, no programa

bayesiano WinBUGS (exemplos volume 1).

Os dados referem-se a $n = 21$ dias de observação de um processo químico, no qual a variável de interesse y = "stack-loss" está relacionada com outras três variáveis químicas, denominadas x_1 = "air flow", x_2 = "water temperature" e x_3 = "acid concentration", como descritos em detalhes em Dodge (1996).

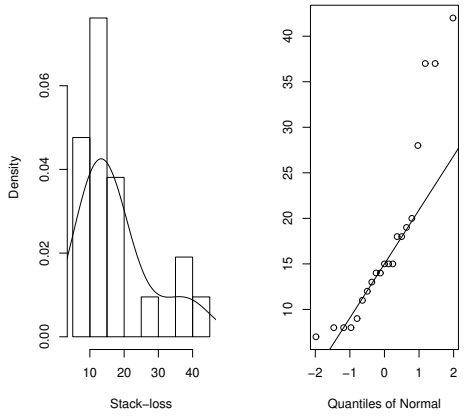


Figura 3.2: Histograma e gráfico Q-Q normal dos dados "stack-loss".

Com o intuito de verificar a existência de assimetria nos dados, a Figura 3.2 apresenta o histograma dos dados e mostra que há um padrão não normal, i.e., assimetria à direita. Além disso, o gráfico Q-Q normal claramente sugere o uso de distribuições com caudas pesadas.

De acordo com as observações acima, consideramos o modelo de regressão linear definido na Seção 3.3 com a variável explicativa $\mathbf{x}_i = (1, x_{1i}, x_{2i}, x_{3i})^\top$ e a variável resposta seguindo as distribuições SSN, SST, SSL ou SSCN provenientes da classe

SSMSN, apenas para comparações, i.e., $y_i \sim SSMSN_1(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2, \lambda; H), i = 1, \ldots, 21$, com $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top$.

Os modelos foram ajustados considerando as estratégias descritas na Seção 3.3.2, usando o programa WinBUGS e as seguintes especificações relativas às prioris (Sahu *et al.*, 2003) $\beta_j \sim N(0, 100)$, $j = 0, \ldots, 3$, $\lambda \sim N(0, 100)\mathbf{I}_{\{\lambda>0\}}$, tal que o truncamento da densidade considera que a distribuição dos dados é assimétrica à direita (positiva), $\sigma^2 \sim I\Gamma(0.1, 0.1)$, pois a especificação univariada da distribuição Wishart inversa é a distribuição gamma inversa ($I\Gamma$), $\nu \sim E(0.5)\mathbf{I}_{\{\nu>\mathbf{1}\}}$ (distribuição exponencial truncada) para a skew-t, $\nu_1 \sim Beta(2, 2)$ e $\nu_2 \sim U(0, 1)$ (distribuição uniforme) para a skew-normal contaminada e $\nu \sim Gamma(0.01, 0.001)\mathbf{I}_{\{\nu>\mathbf{1}\}}$ para a skew-slash. Note que as prioris acima são todas não informativas, no sentido que os hiperparâmetros foram escolhidos com a finalidade de considerarmos prioris com variâncias grandes. Assim, esperamos que os resultados sejam robustos com respeito às prioris especificadas.

Consideramos 50000 iterações, descartamos as primeiras 20000 ("burn-in") e com as iterações restantes realizamos inferência. Além disso, para evitar problemas de correlação nas cadeias geradas, o valor do "lag" considerado foi igual a 5. Os resultados dos ajustes são dados nas Tabelas 3.1 e 3.2. A Tabela 3.2 mostra o intervalo de credibilidade de 95% para os parâmetros dos modelos ajustados. Note que nenhum dos intervalos de credibilidade para λ dos modelos ajustados inclui o valor zero. Podemos concluir que há evidências de assimetria para todos os casos e então, a análise do conjunto de dados "Stack-Loss" no contexto das distribuições assimétricas é apropriada.

Com o intuito de comparar os modelos, calculamos o BIC e o critério de informação do desvio (DIC) como apresentado por Spiegelhalter *et al.* (2002). Os autores afirmam

que o valor do DIC como implementado no WinBUGS pode ser usado para comparar modelos complexos. Dessa forma, usando os valores BIC e DIC, obtemos que o modelo skew-t (SST) é o que melhor se ajusta aos dados. Em todos os modelos observamos convergência, porém por brevidade, apresentaremos apenas os históricos das cadeias de Markov geradas para todos os parâmetros sob o melhor modelo segundo os critérios de comparação; veja Figura 3.3.

Tabela 3.1: Dados "Stack-Loss". Média e desvio padrão (entre parênteses) posteriores para os quatro modelos ajustados.

Dist.	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\sigma^2}$	$\widehat{\lambda}$	$\widehat{\nu}$	$\widehat{\gamma}$	DIC	BIC
SSN	-16.1000	0.7216	1.2050	-0.4447	9.2260	3.8400	-	-	401.2000	-67.8897
	(8.9550)	(0.1684)	(0.4242)	(0.1459)	(6.8600)	(2.419)	-	-		
SST	-18.5100	0.7371	1.0160	-0.3758	6.1970	2.8820	4.4600	-	119.9890	-56.2716
	(8.9480)	(0.1877)	(0.4722)	(0.1567)	(5.0860)	(2.0750)	(2.1680)	-		
SSL	-16.6300	0.7148	1.2100	-0.4347	6.5350	3.4090	6.7250	-	129.7610	-67.0987
	(8.5730)	(0.1674)	(0.4425)	(0.1422)	(5.6050)	(2.173)	(6.922)	-		
SSCN	-16.6700	0.7195	1.1880	-0.4326	6.3780	3.3140	0.4779	0.5460	121.6420	-61.4074
	(8.4810)	(0.1677)	(0.4299)	(0.1462)	(5.4710)	(2.1210)	(0.2907)	(0.2142)		

Tabela 3.2: Dados "Stack-Loss". Intervalo de credibilidade de 95% para os parâmetros dos modelos ajustados.

	SSN		SST		SSL		SSCN	
Parâmetro	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%
β_0	-32.7500	2.5360	-35.0400	-0.3624	-31.9700	2.0090	-32.4000	0.3179
β_1	0.4045	1.0800	0.3656	1.0990	0.4077	1.0430	0.3968	1.0560
β_2	0.3234	2.009	0.1477	1.9650	0.3486	2.0730	0.3586	2.0400
β_3	-0.7509	-0.1695	-0.6902	-0.08591	-0.7184	-0.1613	-0.7290	-0.1411
σ^2	0.2953	25.580	0.2855	19.0400	0.2058	20.9700	0.2368	20.7900
λ	0.1557	8.7090	0.1082	7.7990	0.1409	7.6250	0.1497	7.6560
ν	-	-	2.1740	10.0100	1.0980	27.760	0.1497	7.6560
γ	-	-	-	-	-	-	0.1397	0.9244

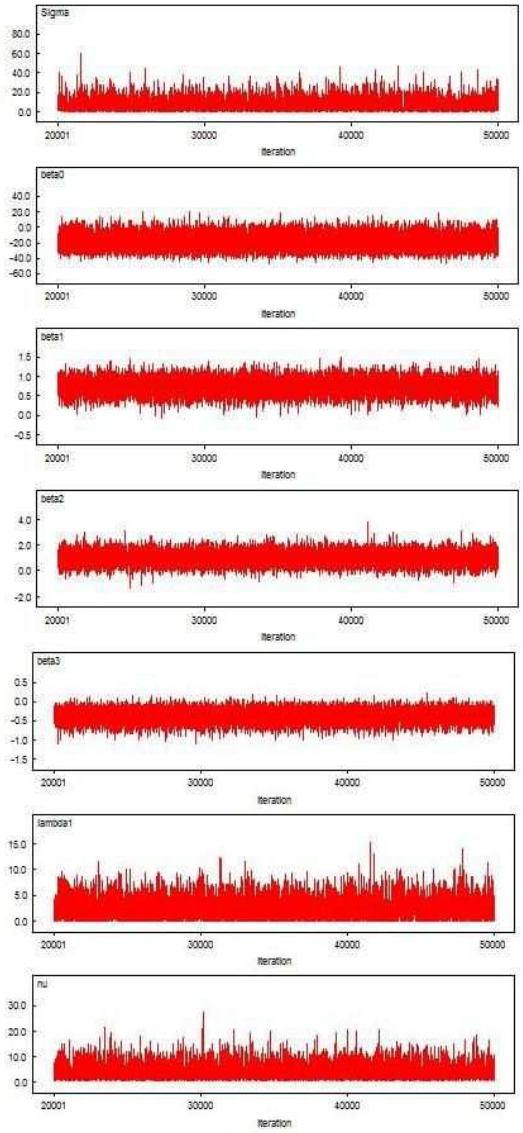


Figura 3.3: Dados "Stack-Loss". Histórico das cadeias de Markov geradas sob o modelo SST para todos os parâmetros.

Motivados por uma considerável pesquisa que tem sido feita para introduzir novas distribuições, a seguir apresentamos um estudo inicial com a finalidade de introduzir algumas generalizações das distribuições SMSN que podem ser úteis na modelagem de imagens, por exemplo.

3.5 Algumas Generalizações

Nesta seção, propomos algumas versões generalizadas das distribuições SMSN introduzidas no Capítulo 2 e destacamos algumas propriedades dessas extensões. Tais propriedades podem ser obtidas através de procedimentos similares aos utilizados para as distribuições SMSN. Além disso, algumas dessas propriedades são compartilhadas por diversas classes de distribuições multivariadas.

3.5.1 Extensões quando U tem Distribuição Multivariada

Considere que um vetor aleatório $\mathbf{Y}$ é dito ser decomposto independentemente se $\mathbf{Y} \stackrel{d}{=} \mathbf{V} \odot \mathbf{W}$, tal que $\mathbf{V}$ e $\mathbf{W}$ são independentes. $\mathbf{V} \odot \mathbf{W}$ especifica o produto Hadamard (Styan, 1973), i.e., $\mathbf{V} \odot \mathbf{W} = (V_1W_1, ..., V_pW_p)^\top$ quando $\mathbf{V}$ e $\mathbf{W}$ são vetores p-dimensionais e $V \odot \mathbf{W} = (VW_1, ..., VW_p) = V\mathbf{W}$ se V é um escalar, onde W_k's são os componentes de $\mathbf{W}$. Este conceito tem sido aplicado em procedimentos de testes não-paramétricos (Zhu & Neuhaus, 2000) e é uma propriedade compartilhada por diversas classes de distribuições multivariadas, por exemplo, as distribuições elípticas.

Além disso, este conceito será a base para a construção das generalizações das distribuições SMSN desenvolvidas nesta seção. Por exemplo, quando $V = \kappa^{1/2}(U)$ é um escalar e $\mathbf{W} = \mathbf{Z} \sim \mathrm{SN}_p(\mathbf{0}, \mathbf{\Sigma}, \boldsymbol{\lambda})$, temos que $\mathbf{Y} \sim SMSN_p(\mathbf{0}, \mathbf{\Sigma}, \boldsymbol{\lambda}; H)$. A generalização

deste caso é dada por

$$\mathbf{Y} = \boldsymbol{\mu} + \kappa^{1/2}(\mathbf{U}) \odot \mathbf{Z}, \tag{3.19}$$

onde $\mathbf{Z} \sim \mathrm{SN}_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ e $\mathbf{U}$ é um vetor aleatório p-dimensional de variáveis de misturas positivas com cdf conjunta $H_p(.;\boldsymbol{\nu})$ e pdf conjunta $h(.;\boldsymbol{\nu})$, tal que $\kappa(\mathbf{U}) = (\kappa(U_1), ..., \kappa(U_p))^\top$ é um vetor de funções pesos, independente de $\mathbf{Z}$, onde $\boldsymbol{\nu}$ é um escalar ou um vetor de parâmetros indexando a distribuição de $\mathbf{U}$. Considerando que o vetor $\kappa^{1/2}(\mathbf{U})$ denota $\kappa^{1/2}(\mathbf{U}) = (\kappa^{1/2}(U_1), ..., \kappa^{1/2}(U_p))^\top$, temos que dado $\mathbf{U} = \mathbf{u}$, $\mathbf{Y}|\mathbf{U} = \mathbf{u} \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}_\mathbf{u}, \boldsymbol{\lambda})$, onde $\boldsymbol{\mu}$ é o vetor de locação, $\boldsymbol{\Sigma}_u = \mathrm{diag}^{1/2}\{\kappa(\mathbf{u})\}\boldsymbol{\Sigma}\mathrm{diag}^{1/2}\{\kappa(\mathbf{u})\}$ é a matriz de dispersão e $\boldsymbol{\lambda}$ é o vetor de assimetria. Assim, a pdf de $\mathbf{Y}$ é dada por

$$f(\mathbf{y}) = 2\int_{\mathbb{R}^p_+} \phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}_u)\Phi(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}_u^{-1/2}(\mathbf{y} - \boldsymbol{\mu}))dH_p(\mathbf{u}), \tag{3.20}$$

onde $\phi_p(.|\boldsymbol{\mu}, \boldsymbol{\Sigma}_u)$ especifica a pdf $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}_u)$ e $\Phi(.)$ representa a cdf da distribuição normal padrão univariada. Neste caso, usamos a notação $\mathbf{Y} \sim GSMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H_p)$ quando $\mathbf{Y}$ tem a pdf dada em (3.20). Um caso particular dessa distribuição é a distribuição skew-normal quando H_p é degenerada e $\kappa(\mathbf{u}) = \mathbf{1}_p$. Se $\boldsymbol{\lambda} = \mathbf{0}$, $\boldsymbol{\mu} = \mathbf{0}$ e $\kappa^{1/2}(\mathbf{U}) = (U_1^{1/2}, ..., U_p^{1/2})^\top = \mathbf{U}^{1/2}$, temos que a distribuição GSMSN reduz para a distribuição proposta por Lyu & Simoncelli (2008) utilizada em aplicações de imagens fotográficas. Eles propuseram um modelo denominado "*Fields of Gaussian Scale Mixtures*" (FoGSM) e a distribuição proposta em (3.20) representa uma extensão deste modelo. Por exemplo, se $\mathbf{Y} \sim GSMSN_p(0, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H_p)$ e $\kappa(\mathbf{U}) = \mathbf{U}$, temos que a densidade de $\mathbf{Y}$ condicionada à $\mathbf{U} = \mathbf{u}$ é dada por

$$\begin{aligned} f(\mathbf{y}|\mathbf{u}) &= 2\phi_p(\mathbf{y}|\mathbf{0}, \boldsymbol{\Sigma}_u)\Phi(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}_u^{-1/2}\mathbf{y}) \\ &= \frac{2|\boldsymbol{\Sigma}|^{-1/2}}{\sqrt{(2\pi)^p \prod_{i=1}^p u_i}} \exp\left(-\frac{1}{2}\mathbf{y}^\top \boldsymbol{\Sigma}_u^{-1}\mathbf{y}\right) \Phi(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}_u^{-1/2}\mathbf{y}), \end{aligned} \tag{3.21}$$

onde $\boldsymbol{\Sigma}_u = \mathrm{diag}\{\sqrt{\mathbf{u}}\}\boldsymbol{\Sigma}\mathrm{diag}\{\sqrt{\mathbf{u}}\}$. Para $\boldsymbol{\lambda} = \mathbf{0}$, a densidade condicional dada em (3.21) reduz para a densidade condicional obtida por Lyu & Simoncelli (2008), onde

eles conside-ram que $\mathbf{U}$ tem pdf dada por

$$h(\mathbf{u}) = c_u \frac{|\mathbf{\Psi}|^{-1/2}}{\prod_{i=1}^{p} u_i} \exp\left(-\frac{1}{2}(\log \mathbf{u})^\top \mathbf{\Psi}^{-1}(\log \mathbf{u})\right),$$

uma extensão natural da distribuição log-normal univariada; veja Lyu & Simoncelli (2008), página 3, para mais detalhes desta distribuição. Neste caso, $\log \mathbf{u}$ denota o vetor $\log \mathbf{u} = (\log u_1, ..., \log u_n)^\top$.

Uma versão particular da distribuição GSMSN é obtida quando a dimensão de $\mathbf{U}$ é menor que p. Por exemplo, considere $\mathbf{U} = (U_1, U_2)^\top$ e $\mathbf{Z} \sim \mathrm{SN}_p(\mathbf{0}, \mathbf{\Sigma}, \boldsymbol{\lambda})$, onde $\mathbf{Z}$ está particionado como $\mathbf{Z} = (\mathbf{Z}_1^\top, \mathbf{Z}_2^\top)^\top$ de dimensões p_1 e p_2 $(p_1 + p_2 = p)$, respectivamente. Similarmente à (3.19), definimos

$$\mathbf{Y} = \boldsymbol{\mu} + \kappa^{1/2}(\mathbf{U}) \odot \mathbf{Z} = \boldsymbol{\mu} + \begin{pmatrix} \kappa^{1/2}(U_1)\mathbf{Z}_1 \\ \kappa^{1/2}(U_2)\mathbf{Z}_2 \end{pmatrix}.$$

Neste caso, usamos a notação $\mathbf{Y} \sim GSMSN_p(\boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{\lambda}; H_2)$, onde H_2 é a cdf de $\mathbf{U}$.

Além disso, considere que $\kappa(\mathbf{U})$ é uma função escalar do vetor aleatório $\mathbf{U}$. Dessa forma, de (3.19), temos que $\mathbf{Y}|\mathbf{U} = \mathbf{u} \sim SN_p(\boldsymbol{\mu}, \kappa(\mathbf{u})\mathbf{\Sigma}, \boldsymbol{\lambda})$. Esta abordagem pode ser útil para obter modelos mais flexíveis que as distribuições normal e skew-normal. Por exemplo,

a) Quando $\kappa(\mathbf{U}) = 1/(U_1U_2)$, $U_1 \sim U(0,1)$, $U_2 \sim Gamma(\nu/2, \nu/2)$, $\nu > 0$ e $\mathbf{Z} \sim SN_p(\mathbf{0}, \mathbf{\Sigma}, \mathbf{0})$, i.e., $\mathbf{Z} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$ todos independentes. Então,

$$\mathbf{Y} \overset{d}{=} \boldsymbol{\mu} + U_1^{-1/2}U_2^{-1/2}\mathbf{Z} \tag{3.22}$$

tem a distribuição slash-Student p-variada definida em Tan & Peng (2005). A prova segue diretamente do fato que

$$\mathbf{T} = U_2^{-1/2}\mathbf{Z} \sim t_p(\boldsymbol{\mu}, \mathbf{\Sigma}; \nu).$$

b) Se $\mathbf{Z} \sim \mathrm{SN}_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, temos que $\mathbf{Y}$ em (3.22) tem uma distribuição skew-slash-Student multivariada, como mostrado em Tan & Peng (2005). Obviamente, muitas outras distribuições podem ser construídas através da escolha apropriada de pdf para U_1 e U_2.

A seguir apresentamos o vetor de médias e a matriz de covariâncias para o vetor aleatório GSMSN. A prova decorre de (3.19), do vetor de médias e da matriz de covariâncias do vetor aleatório SMSN, momentos descritos na Proposição 2.2.7.

Proposição 3.5.1 *Considere que* $\mathbf{Y} \sim GSMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H_p)$, $E_{\kappa,1}(\mathbf{U}) = E[\kappa^{1/2}(\mathbf{U})]$ *e* $E_{\kappa,2}(\mathbf{U}) = E[\kappa^{1/2}(\mathbf{U})[\kappa^{1/2}(\mathbf{U})]^\top]$. *Então,*

a) Se $E_{\kappa,1}(\mathbf{U}) < \infty$, *temos que*

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}} \left(E_{\kappa,1}(\mathbf{U}) \odot (\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}) \right).$$

b) Se $E_{\kappa,2}(\mathbf{U}) < \infty$, *temos que*

$$Var[\mathbf{Y}] = \boldsymbol{\Sigma}_y = (E_{\kappa,2}(\mathbf{U}) \odot \boldsymbol{\Sigma}) - \frac{2}{\pi} \left([E_{\kappa,1}(\mathbf{U}) E_{\kappa,1}(\mathbf{U})^\top] \odot [\boldsymbol{\Delta}\boldsymbol{\Delta}^\top] \right),$$

onde $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$.

Sob a distribuição GSMSN quando $\kappa(\mathbf{U})$ é uma função escalar do vetor aleatório $\mathbf{U}$, o resultado desenvolvido na Proposição 2.2.2 relacionado com transformações lineares é ainda válido, consequentemente a distribuição marginal de $\mathbf{Y}$ ainda pertence à classe de distribuições GSMSN. Para o caso em que $\kappa(\mathbf{U})$ é um vetor, esta propriedade não é satisfeita.

É importante ressaltar que outras possíveis extensões e propriedades dessas extensões podem ser derivadas. Além disso, alguns aspectos inferenciais podem ser discutidos, incluindo o algoritmo EM para obtenção dos estimadores de máxima verossimilhança dos parâmetros de interesse. Trata-se de uma linha interessante de pesquisa.

3.6 Observações Finais

Os resultados obtidos neste capítulo estendem muitas propriedades das distribuições assimétricas. As extensões das distribuições misturas de escala skew-normal propostas neste capítulo são muito gerais, bastante flexíveis e amplamente aplicáveis. Neste sentido, propomos um estudo unificado da classe de distribuições SMSN. Essas distribuições em geral, podem ser usadas para a modelagem de fenômenos aleatórios assimétricos e com caudas mais pesadas que a distribuição normal. Por exemplo, essas distribuições podem ser úteis em problemas de regressão e calibração no contexto de assimetria. Nesta tese, estudamos os modelos lineares sob a classe de distribuições SMSN, especificamente o modelo de regressão linear, o modelo linear misto e o modelo de Grubbs, considerando as distribuições skew-normal, skew-t, skew-slash e skew-normal contaminada.

Capítulo 4

Modelo de Regressão Linear Misturas de Escala Skew-Normal

Um procedimento de estimação robusto e técnicas de diagnóstico de influência para o modelo de regressão linear sob a classe de distribuições misturas de escala skew-normal multivariada (SMSN) serão desenvolvidos neste capítulo. A maior virtude em considerar o modelo de regressão linear sob a classe de distribuições SMSN é que ele apresenta uma representação hierárquica, útil para estudos de inferência. Inspirados pelo algoritmo EM, desenvolvemos a análise de influência local no modelo de regressão linear misturas de escala skew-normal segundo a metodologia de Zhu & Lee (2001), a qual é invariante sob reparametrizações. Alguns esquemas de perturbação úteis serão discutidos. Com a finalidade de examinar a sensibilidade dessa classe de distribuições com respeito às observações aberrantes e/ou influentes, alguns estudos de simulação serão apresentados. Finalmente, um conjunto de dados reais será analisado, ilustrando

a utilidade da metodologia proposta.

4.1 Introdução

A suposição de normalidade sempre foi muito atrativa para o modelo de regressão linear com resposta contínua e, mesmo quando não era alcançada, procurava-se alguma transformação na resposta no sentido de obter-se pelo menos a simetria. Contudo, com o passar do tempo, verificou-se que as estimativas obtidas para os coeficientes dos modelos normais mostraram-se sensíveis às observações extremas, comumente chamadas de observações aberrantes, incentivando o desenvolvimento de metodologias robustas contra tais observações.

Alternativas à suposição de normalidade têm sido propostas na literatura. Lange *et al.* (1989) propõem o modelo t-Student, Lange & Sinsheimer (1993) propõem a classe de distribuições normal/independente e suas aplicações em modelos de regressão, Yamaguchi (2001) sugere usar a distribuição normal contaminada, Galea-Rojas *et al.* (2003) e Osorio *et al.* (2007) desenvolvem o modelo de regressão linear sob a classe de distribuições elípticas e mostram que ele apresenta um bom desempenho na presença de "outliers". Baseados nos trabalhos de Azzalini (1985) e Azzalini & Capitanio (1999), muitos autores têm considerado a distribuição skew-normal em diversas áreas, tais como economia, engenharia, biologia, entre outras. Nessa linha de pesquisa, Lachos *et al.* (2007a) propõem o modelo de regressão skew-normal multivariado e Azzalini & Genton (2008) desenvolvem o modelo de regressão skew-t multivariado. Neste capítulo, propomos uma classe flexível de modelos de regressão linear com as distribuições misturas de escala skew-normal (SMSN-RM). Esta classe rica de modelos pode naturalmente atribuir pesos diferentes para cada observação e consequentemente controlar a

influência da observação nas estimativas dos parâmetros.

A avaliação dos aspectos de robustez das estimativas dos parâmetros em modelos estatísticos tem sido uma importante preocupação de vários pesquisadores nas décadas recentes. A metodologia de deleção, que consiste em estudar o impacto sobre as estimativas dos parâmetros após a exclusão de uma observação individual, é provavelmente a técnica mais utilizada para detectar observações influentes (Cook & Weisberg, 1982). No entanto, a investigação da influência de pequenas perturbações no modelo ou nos dados sobre as estimativas dos parâmetros tem recebido crescente atenção nos últimos anos. Isto pode ser estudado realizando a análise de influência local (Cook, 1986), uma técnica estatística geral usada para avaliar a estabilidade dos resultados de estimação com respeito às suposições do modelo. Considerando o trabalho pioneiro de Cook (1986), vários autores desenvolveram trabalhos nessa área de pesquisa; veja, por exemplo, Lesaffre & Verbeke (1998), Galea-Rojas *et al.* (2005) e Osorio *et al.* (2007). Motivada por estes trabalhos, neste capítulo também discutimos análise de influência local no modelo de regressão linear sob a classe de distribuições misturas de escala skew-normal. Entretanto, uma vez que a função log-verossimilhança observada do SMSN-RM envolve integrais, a aplicação direta da metodologia de influência local de Cook (1986) pode ser difícil, pois esta abordagem envolve derivadas parciais de primeira e segunda ordem dessa função. Ao invés disso, discutimos análise de influência local no SMSN-RM baseada na metodologia de Zhu & Lee (2001). Portanto, neste capítulo, consideramos um estudo de estimação robusta e discutimos análise de influência local no modelo de regressão linear sob a classe de distribuições SMSN. Esta proposta estende alguns resultados provenientes dos trabalhos de Galea-Rojas *et al.* (2003) e Osorio *et al.* (2007), onde substituímos a suposição de simetria por uma classe flexível de distribuições.

O capítulo está organizado como segue. Na Seção 4.2, apresentamos o SMSN-RM, incluindo o algoritmo EM para estimação por máxima verossimilhança e a matriz informação observada, útil no cálculo dos desvios padrões dessas estimativas. Na Seção 4.3, desenvolvemos a metodologia de influência local pertinente ao SMSN-RM, sendo que quatro esquemas de perturbação foram considerados. Nas Seções 4.4 e 4.5, exemplos numéricos considerando dados reais e simulados são apresentados para ilustrar a metodologia proposta. Finalmente, algumas observações finais são dadas na Seção 4.6.

4.2 O Modelo Proposto

Nesta seção, consideramos o modelo de regressão linear, onde as observações seguem distribuições misturas de escala skew-normal (SMSN-RM). Em geral, o modelo de regressão linear normal (N-RM) é definido como

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n, \tag{4.1}$$

onde $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ip})^\top$ é um vetor $p \times 1$ de respostas contínuas observadas para o i-ésimo indivíduo, $\mathbf{X}_i$ de dimensão $n \times p$ é a matriz de planejamento conhecida, $\boldsymbol{\beta}$ é um vetor $p \times 1$ dos coeficientes de regressão desconhecidos e $\boldsymbol{\epsilon}_i$ é o vetor $p \times 1$ de erros independentes, com $\boldsymbol{\epsilon}_i \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$. A matriz $\boldsymbol{\Sigma}$ de dimensão $p \times p$ é positiva definida e pode ser não estruturada ou estruturada, por exemplo, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\alpha})$, onde $\boldsymbol{\alpha}$ denota os elementos da matriz triangular superior de $\boldsymbol{\Sigma}$, no caso não estruturado. De acordo com Sahu *et al.* (2003) e Lachos *et al.* (2007a), estendemos o N-RM definido acima considerando a relação linear em (4.1) com a seguinte suposição:

$$\mathbf{Y}_i \overset{\text{ind}}{\sim} SMSN_p(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H), \ i = 1, \ldots, n.$$

O modelo de regressão linear multivariado sob a classe de distribuições misturas de escala skew-normal (SMSN-RM) definido acima pode ser escrito de forma similar ao

modelo definido em (4.1), mas neste caso $\boldsymbol{\epsilon}_i \overset{\text{ind}}{\sim} SMSN_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$, $i = 1, \ldots, n$. Além disso, note que o modelo definido nesta seção é diferente do modelo especificado na Seção 3.3, pois consideramos parametrizações diferentes. Por exemplo, se consideramos a seguinte parametrização para a matriz de assimetria $\boldsymbol{\Lambda}$ da distribuição SSN, definida em (3.1), dada por $\boldsymbol{\Lambda} = \boldsymbol{\delta} = \dfrac{\boldsymbol{\lambda}}{\sqrt{1 + \boldsymbol{\lambda}^\top \boldsymbol{\lambda}}}$, temos a distribuição skew-normal definida em (2.1) e utilizada na construção da classe das distribuições SMSN.

4.2.1 Estimação por Máxima Verossimilhança

Nesta seção, apresentamos um algoritmo EM para estimação por máxima verossimilhança dos parâmetros do modelo de regressão linear sob a classe de distribuições misturas de escala skew-normal multivariada (SMSN-RM). Para explorar o algoritmo EM, apresentamos SMSN-RM no contexto de dados incompletos, usando os resultados desenvolvidos na Seção 2.2. Dessa forma, de (2.4), o modelo definido acima (Seção 4.2) pode ser reescrito hierarquicamente como

$$\mathbf{Y}_i | T_i = t_i, U_i = u_i, \overset{\text{ind}}{\sim} N_p(\mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\Delta} t_i, \kappa(u_i)\boldsymbol{\Gamma}), \tag{4.2}$$

$$T_i | U_i = u_i \overset{\text{ind}}{\sim} HN_1(0, \kappa(u_i)), \tag{4.3}$$

$$U_i \overset{\text{iid}}{\sim} H(\boldsymbol{\nu}), \tag{4.4}$$

$i = 1, \ldots, n$, todos independentes, onde $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$, $\boldsymbol{\Gamma} = \boldsymbol{\Sigma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top$ e $HN_1(0, \sigma^2)$ denota a distribuição half-$N(0, \sigma^2)$ (Johnson *et al.*, 1994). Com o intuito de evitar as dificuldades de estimar o parâmetro $\boldsymbol{\nu}$ da variável de mistura (Lucas, 1997), fixamos $\boldsymbol{\nu}$ previamente, como recomendado por Lange *et al.* (1989), Berkane *et al.* (1994) e Osorio *et al.* (2007). Considere $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, o vetor de respostas observáveis para as n unidades amostrais, $\mathbf{u} = (u_1, \ldots, u_n)^\top$ e $\mathbf{t} = (t_1, \ldots, t_n)^\top$. Então, sob a representação hierárquica (4.2)–(4.4), segue que a função log-verossimilhança completa associada com

$\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{u}^\top, \mathbf{t}^\top)^\top$ é

$$\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = c - \frac{n}{2}\log|\boldsymbol{\Gamma}| - \frac{1}{2}\sum_{i=1}^{n}\kappa^{-1}(u_i)(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \boldsymbol{\Delta}t_i)^\top\boldsymbol{\Gamma}^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \boldsymbol{\Delta}t_i),$$

onde c é uma constante que independe do vetor de parâmetros $\boldsymbol{\theta}$. Considere $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\boldsymbol{\beta}}^{(k)\top}, \widehat{\boldsymbol{\gamma}}^{(k)\top}, \widehat{\boldsymbol{\Delta}}^{(k)\top})^\top$ a estimativa de $\boldsymbol{\theta}$ na k-ésima iteração, onde $\boldsymbol{\gamma}$ denota o conjunto minimal de parâmetros, tal que $\boldsymbol{\Gamma}$ está bem definida. Entretanto, note que $\boldsymbol{\beta}$ é o nosso parâmetro de interesse principal.

Após manipulações algébricas, a esperança condicional da função log-verossimilhança completa é dada por $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$, onde

$$\begin{aligned} Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) &= c - \frac{n}{2}\log|\boldsymbol{\Gamma}| - \frac{1}{2}\sum_{i=1}^{n}\widehat{u}_i^{(k)}\boldsymbol{\epsilon}_i^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\epsilon}_i + \sum_{i=1}^{n}\widehat{ut}_i^{(k)}\boldsymbol{\epsilon}_i^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta} \\ &\quad -\frac{1}{2}\sum_{i=1}^{n}\widehat{ut^2}_i^{(k)}\boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta}, \end{aligned} \tag{4.5}$$

com $\boldsymbol{\epsilon}_i = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}$, $\widehat{u}_i = \mathrm{E}[\kappa^{-1}(U_i)|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i]$, $\widehat{ut}_i = \mathrm{E}[\kappa^{-1}(U_i)T_i|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i]$, $\widehat{ut_i^2} = \mathrm{E}[\kappa^{-1}(U_i)T_i^2|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i]$ e usando propriedades conhecidas da esperança condicional, obtemos (veja Apêndice C.1 para detalhes)

$$\widehat{ut}_i^{(k)} = \widehat{u}_i^{(k)}\widehat{\mu}_{Ti}^{(k)} + \widehat{M}_T^{(k)}\widehat{\eta}_{1i}^{(k)} \tag{4.6}$$

$$\widehat{ut^2}_i^{(k)} = \widehat{u}_i^{(k)}\widehat{\mu}_{T_i}^{2(k)} + \widehat{M}_T^{2(k)} + \widehat{M}_T^{(k)}\widehat{\mu}_{T_i}^{(k)}\widehat{\eta}_{1i}^{(k)}, \tag{4.7}$$

onde $\widehat{M}_T^2 = 1/(1 + \widehat{\boldsymbol{\Delta}}^\top\widehat{\boldsymbol{\Gamma}}^{-1}\widehat{\boldsymbol{\Delta}})$ e $\widehat{\mu}_{T_i} = \widehat{M}_T^2\widehat{\boldsymbol{\Delta}}^\top\widehat{\boldsymbol{\Gamma}}^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$, $i = 1, \ldots, n$, com todas essas quantidades avaliadas em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$. Uma vez que $\frac{\mu_{T_i}}{M_T} = \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$, as esperanças condicionais dadas em (4.6) - (4.7), especificamente $\widehat{u}_i = \widehat{u}_{1i}$ e $\widehat{\eta}_{1i}$, podem ser facilmente obtidas dos resultados apresentados na Seção 2.2 (veja Proposição 2.2.1). Deste modo, pelo menos para as distribuições skew-t e skew-normal contaminada pertencentes à classe SMSN, temos expressões fechadas para as quantidades $\widehat{u}_i$ e

$\widehat{\eta}_{1i}$. Já para a distribuição skew-slash, a integração de Monte Carlo pode ser empregada e denominamos o algoritmo de MC-EM; veja Wei & Tanner (1990). Portanto, temos o seguinte algoritmo EM:

Passo E: Dado $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, calcule $\widehat{ut^2}_i^{(k)}$, $\widehat{ut}_i^{(k)}$ e $\widehat{u}_i^{(k)}$, $i = 1, \ldots, n$, usando (4.6)-(4.7).

Passo M: Atualize $\widehat{\boldsymbol{\theta}}^{(k+1)}$ maximizando $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ sobre $\boldsymbol{\theta}$, o que leva às seguintes expressões fechadas (veja Apêndice C.1 para detalhes):

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}^{(k+1)} &= (\sum_{i=1}^{n} \widehat{u}_i^{(k)} \mathbf{X}_i^\top \boldsymbol{\Gamma}^{-1(k)} \mathbf{X}_i)^{-1} \sum_{i=1}^{n} \mathbf{X}_i^\top \boldsymbol{\Gamma}^{-1(k)} (\widehat{u}_i^{(k)} \mathbf{y}_i - \widehat{ut}_i^{(k)} \boldsymbol{\Delta}^{(k)}), \\
\widehat{\boldsymbol{\Gamma}}^{(k+1)} &= \frac{1}{n} \sum_{i=1}^{n} \Big[\widehat{u}_i^{(k)} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}^{(k)})(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}^{(k)})^\top - \widehat{ut}_i^{(k)} \boldsymbol{\Delta}^{(k)} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}^{(k)})^\top \\
&\quad - \widehat{ut}_i^{(k)} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}^{(k)}) \boldsymbol{\Delta}^{\top(k)} + \widehat{ut^2}_i^{(k)} \boldsymbol{\Delta}^{(k)} \boldsymbol{\Delta}^{(k)\top} \Big], \\
\widehat{\boldsymbol{\Delta}}^{(k+1)} &= \frac{\sum_{i=1}^{n} \widehat{ut}_i^{(k)} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}^{(k)})}{\sum_{i=1}^{n} \widehat{ut^2}_i^{(k)}} \\
\widehat{\mathbf{D}}^{(k+1)} &= \widehat{\boldsymbol{\Gamma}}^{(k+1)} + \widehat{\boldsymbol{\Delta}}^{(k+1)} \widehat{\boldsymbol{\Delta}}^{\top(k+1)} \\
\widehat{\boldsymbol{\lambda}}^{(k+1)} &= \widehat{\mathbf{D}}^{(k+1)-1/2} \widehat{\boldsymbol{\Delta}}^{(k+1)} / (1 - \widehat{\boldsymbol{\Delta}}^{\top(k+1)} \widehat{\mathbf{D}}^{(k+1)-1} \widehat{\boldsymbol{\Delta}}^{(k+1)})^{1/2}
\end{aligned}
$$

As iterações são repetidas até que uma regra de convergência adequada seja satisfeita. Valores iniciais são necessários para implementar este algoritmo. Eles são obtidos sob a suposição de normalidade e considerando $\widehat{\lambda}_j^{(0)} = 3\text{sign}(\widehat{\rho}_j)$, onde $\widehat{\rho}_j$ é o coeficiente de assimetria amostral para a variável j, $j = 1, \ldots, p$. Entretanto, 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 $\boldsymbol{\lambda} = \mathbf{0}$ (ou $\boldsymbol{\Delta} = \mathbf{0}$), as equações do passo M se reduzem às equações obtidas quando assumimos as distribuições SMN. Por outro lado, este algoritmo claramente generaliza os resultados encontrados em Lachos *et al.* (2007a) na Seção 2, considerando $\kappa(U_i) = 1$, $i = 1, \ldots, n$.

4.2.2 Matriz Informação Observada

Suponha que temos observações de n indivíduos independentes, $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, onde $\mathbf{Y}_i \sim SMSN_p(\boldsymbol{\mu}_i(\boldsymbol{\beta}), \boldsymbol{\Sigma}(\boldsymbol{\alpha}), \boldsymbol{\lambda}; H)$, $i = 1, \ldots, n$. Então, a função log-verossimilhança para $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top, \boldsymbol{\lambda}^\top)^\top \in \mathbb{R}^q$, dada a amostra observada $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, é dada por $\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta})$, onde $\ell_i(\boldsymbol{\theta}) = \log 2 - \frac{p}{2} log 2\pi - \frac{1}{2}\log|\boldsymbol{\Sigma}| + \log K_i$, com

$$K_i = K_i(\boldsymbol{\theta}) = \int_0^\infty \kappa^{-p/2}(u_i) \exp\{-\frac{1}{2}\kappa^{-1}(u_i) d_i\} \Phi(\kappa^{-1/2}(u_i) A_i) dH(u_i; \boldsymbol{\nu}),$$

$d_i = (\mathbf{y}_i - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i)$ e $A_i = \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i)$. Usando a seguinte notação $I_i^F(w) = \int_0^\infty \kappa^{-w}(u_i) \exp\{-\frac{1}{2}\kappa^{-1}(u_i) d_i\} F(\kappa^{-1/2}(u_i) A_i) dH(u_i; \boldsymbol{\nu})$, onde $F(.)$ é a função $\Phi(.)$ ou $\phi(.)$, podemos escrever $K_i(\boldsymbol{\theta}) = I_i^\Phi(\frac{p}{2})$, $i = 1, \ldots, n$. Dessa forma, a matriz de segundas derivadas com respeito a $\boldsymbol{\theta}$ é dada por

$$\mathbf{L} = \sum_{i=1}^n \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = -\frac{n}{2} \frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} - \sum_{i=1}^n \frac{1}{K_i^2} \frac{\partial K_i}{\partial \boldsymbol{\theta}} \frac{\partial K_i}{\partial \boldsymbol{\theta}^\top} + \sum_{i=1}^n \frac{1}{K_i} \frac{\partial^2 K_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top},$$

onde

$$\frac{\partial K_i}{\partial \boldsymbol{\theta}} = I_i^\phi(\frac{p+1}{2}) \frac{\partial A_i}{\partial \boldsymbol{\theta}} - \frac{1}{2} I_i^\Phi(\frac{p+2}{2}) \frac{\partial d_i}{\partial \boldsymbol{\theta}}$$

e

$$\begin{aligned} \frac{\partial^2 K_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} &= \frac{1}{4} I_i^\Phi(\frac{p+4}{2}) \frac{\partial d_i}{\partial \boldsymbol{\theta}} \frac{\partial d_i}{\partial \boldsymbol{\theta}^\top} - \frac{1}{2} I_i^\Phi(\frac{p+2}{2}) \frac{\partial^2 d_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \\ &\quad -\frac{1}{2} I_i^\phi(\frac{p+3}{2}) (\frac{\partial A_i}{\partial \boldsymbol{\theta}} \frac{\partial d_i}{\partial \boldsymbol{\theta}^\top} + \frac{\partial d_i}{\partial \boldsymbol{\theta}} \frac{\partial A_i}{\partial \boldsymbol{\theta}^\top}) - I_i^\phi(\frac{p+3}{2}) A_i \frac{\partial A_i}{\partial \boldsymbol{\theta}} \frac{\partial A_i}{\partial \boldsymbol{\theta}^\top} \\ &\quad +I_i^\phi(\frac{p+1}{2}) \frac{\partial^2 A_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}. \end{aligned}$$

Para os casos particulares skew-t, skew-slash e skew-normal contaminada, as integrais $I^\Phi(.)$ e $I^\phi(.)$ podem ser obtidas considerando os resultados apresentados na Seção 2.2.2 e as definições de d_i e A_i descritas nesta seção. As derivadas de $\log|\boldsymbol{\Sigma}|$, d_i e A_i são dadas no Apêndice C.2. Dessa forma, $\mathbf{J} = -\mathbf{L}$ denota a matriz de informação observada

para a log-verossimilhança marginal $\ell(\boldsymbol{\theta})$. Então, intervalos de confiança assintóticos e testes de hipóteses para o parâmetro $\boldsymbol{\theta} \in \mathbb{R}^q$ podem ser obtidos, assumindo que o estimador de máxima verossimilhança tem aproximadamente distribuição $N_q(\boldsymbol{\theta}, \mathbf{J}^{-1})$. Na prática, $\mathbf{J}$ é usualmente desconhecida e a substituímos por $\widehat{\mathbf{J}}$ que é a matriz $\mathbf{J}$ avaliada na estimativa de máxima verossimilhança de $\boldsymbol{\theta}$.

4.3 Influência Local

Nesta seção, discutimos dignóstico de influência com ênfase na metodologia de influência local proposta por Zhu & Lee (2001). A seguir vamos obter a curvatura normal para o SMSN-RM proposto, considerando a formulação hierárquica dada em (4.2)–(4.4). Os detalhes do cálculo da hessiana $\ddot{Q}_\theta(\boldsymbol{\theta}) = \dfrac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}$ e da matriz $\boldsymbol{\nabla}_{\boldsymbol{\omega}} = \dfrac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}$, para os esquemas de perturbação propostos, são dados nas próximas subseções. A metodologia de diferenciação foi baseada na técnica de diferenciação matricial proposta por Magnus & Neudecker (1988).

4.3.1 Matriz Hessiana

Com intuito de obter as medidas de diagnóstico de influência local para um particular esquema de perturbação, é necessário obter $\ddot{Q}_\theta(\boldsymbol{\theta})$, onde $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\Delta}^\top)^\top$ é o vetor de parâmetros. Assim, a matriz hessiana tem elementos dados por

$$
\begin{aligned}
\frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} &= -(\widehat{\boldsymbol{u}} \odot \boldsymbol{X})^\top(\boldsymbol{I}_n \otimes \boldsymbol{\Gamma}^{-1})\boldsymbol{X}, \quad \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\Delta}^\top} = -(\widehat{\boldsymbol{ut}} \odot \boldsymbol{X})^\top(\mathbf{1}_n \otimes \boldsymbol{\Gamma}^{-1}), \\
\frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\gamma_r} &= -(\widehat{\boldsymbol{u}} \odot \boldsymbol{X})^\top(\boldsymbol{I}_n \otimes \dot{\boldsymbol{\Gamma}}_r)\boldsymbol{\epsilon} + (\widehat{\boldsymbol{ut}} \odot \boldsymbol{X})^\top(\mathbf{1}_n \otimes (\dot{\boldsymbol{\Gamma}}_r\boldsymbol{\Delta})), \\
\frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\Delta}\partial\boldsymbol{\Delta}^\top} &= -\mathbf{1}_n^\top\widehat{\boldsymbol{ut}^2}\boldsymbol{\Gamma}^{-1}, \quad \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\Delta}\partial\gamma_r} = -\dot{\boldsymbol{\Gamma}}_r\left((\widehat{\boldsymbol{ut}}^\top \otimes \boldsymbol{I}_p)\boldsymbol{\epsilon} - \mathbf{1}_n^\top\widehat{\boldsymbol{ut}^2}\boldsymbol{\Delta}\right),
\end{aligned}
$$

$$\frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\gamma_r\partial\gamma_s} = \frac{n}{2}\mathrm{tr}(\boldsymbol{A}_{rs}) - \frac{1}{2}[(\widehat{\boldsymbol{u}}\odot\boldsymbol{\epsilon})^\top(\boldsymbol{I}_n\otimes\boldsymbol{B}_{rs})\boldsymbol{\epsilon} - 2(\widehat{\boldsymbol{ut}}\odot\boldsymbol{\epsilon})^\top(\mathbf{1}_n\otimes(\boldsymbol{B}_{rs}\boldsymbol{\Delta})) + \mathbf{1}_n^\top\widehat{\boldsymbol{ut}^2}\boldsymbol{\Delta}^\top\boldsymbol{B}_{rs}\boldsymbol{\Delta}],$$

onde $\dot{\boldsymbol{\Gamma}}_r = \boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(r)\boldsymbol{\Gamma}^{-1}$, $\boldsymbol{A}_{rs} = \boldsymbol{\Gamma}^{-1}[\dot{\boldsymbol{\Gamma}}(r)\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(s) - \ddot{\boldsymbol{\Gamma}}(r,s)]$ e $\boldsymbol{B}_{rs} = \boldsymbol{\Gamma}^{-1}[\dot{\boldsymbol{\Gamma}}(r)\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(s) + \dot{\boldsymbol{\Gamma}}(s)\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(r) - \ddot{\boldsymbol{\Gamma}}(r,s)]\boldsymbol{\Gamma}^{-1}$, com $\dot{\boldsymbol{\Gamma}}(r) = \partial\boldsymbol{\Gamma}/\partial\gamma_r$, $\ddot{\boldsymbol{\Gamma}}(r,s) = \partial^2\boldsymbol{\Gamma}/\partial\gamma_r\partial\gamma_s$, $r,s = 1,\ldots,p^*$ e $p^* = dim(\boldsymbol{\gamma})$. Além disso, $\boldsymbol{X} = (\boldsymbol{X}_1^\top,...,\boldsymbol{X}_n^\top)^\top$, $\widehat{\boldsymbol{u}} = (\widehat{u}_1,...,\widehat{u}_n)^\top$, $\widehat{\boldsymbol{ut}} = (\widehat{ut}_1,...,\widehat{ut}_n)^\top$, $\widehat{\boldsymbol{ut}^2} = (\widehat{ut^2}_1,...,\widehat{ut^2}_n)^\top$ e $\mathbf{I}_n$ é a matriz identidade $n\times n$. Considere também que $\otimes$ denota o produto Kronecker, $\odot$ é o produto Hadamard (Styan, 1973) e $\boldsymbol{\epsilon}$ denota $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^\top,...,\boldsymbol{\epsilon}_n^\top)^\top$, com $\boldsymbol{\epsilon}_i$ como em (4.5), $i = 1,...,n$. Note que quando $p = m$ e $\mathbf{X}_i = \mathbf{I}_m$, $i = 1,...,n$, a matriz hessiana (avaliada em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$) é bloco diagonal com blocos

$$\ddot{Q}_{11}(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Delta}}) = -\begin{pmatrix} \sum_{i=1}^n \widehat{u}_i & \sum_{i=1}^n \widehat{tu}_i \\ \sum_{i=1}^n \widehat{tu}_i & \sum_{i=1}^n \widehat{ut^2}_i \end{pmatrix}\otimes\boldsymbol{\Gamma}^{-1} \text{ e } \ddot{Q}_{22}(\widehat{\boldsymbol{\gamma}}) = \ddot{Q}_\gamma(\widehat{\boldsymbol{\theta}}),$$

onde $\ddot{Q}_\gamma(\widehat{\boldsymbol{\theta}})$ é a matriz de derivadas de segunda ordem de $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ com respeito a $\boldsymbol{\gamma}$.

4.3.2 Esquemas de Perturbação

Nesta seção, apresentamos a matriz $\boldsymbol{\nabla}_{\boldsymbol{\omega}_o}$ sob os seguintes esquemas de perturbação para o SMSN-RM: *ponderação de casos* para detectar observações com notável contribuição na função log-verossimilhança e que podem exercer alta influência no processo de estimação, *perturbação na variável resposta* para detectar observações com alta influência sobre seus próprios valores previstos, *perturbação na escala* realizada na matriz $\boldsymbol{\Sigma}$ para revelar indivíduos que são mais influentes na modelagem da estrutura de escala assim como indicar aqueles indivíduos influentes na estimação dos parâmetro $\boldsymbol{\alpha}$ e finalmente *perturbação na variável explicativa* para detectar possíveis maus condicionamentos entre algumas colunas da matriz de planejamento $\mathbf{X}$. Para cada esquema de perturbação,

temos a seguinte forma particionada

$$\boldsymbol{\nabla}_{\boldsymbol{\omega}_o} = (\boldsymbol{\nabla}_{\beta}^{\top}, \boldsymbol{\nabla}_{\Delta}^{\top}, \boldsymbol{\nabla}_{\gamma}^{\top})^{\top},$$

onde $\boldsymbol{\nabla}_{\beta} = \dfrac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^{\top}}|_{\boldsymbol{\omega}_o} \in \mathbb{R}^{m \times g}$, $\boldsymbol{\nabla}_{\Delta} = \dfrac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\Delta} \partial \boldsymbol{\omega}^{\top}}|_{\boldsymbol{\omega}_o} \in \mathbb{R}^{p \times g}$ e $\boldsymbol{\nabla}_{\gamma} = (\boldsymbol{\nabla}_{\gamma 1}, \ldots, \boldsymbol{\nabla}_{\gamma p^*})^{\top}$, com $\boldsymbol{\nabla}_{\gamma r} = \dfrac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial \gamma_r \partial \boldsymbol{\omega}^{\top}}|_{\boldsymbol{\omega}_o} \in \mathbb{R}^{1 \times g}$, $r = 1, \ldots, p^*$ e g, p, m sendo as dimensões do vetor de perturbação $\boldsymbol{\omega}$, da variável resposta $\mathbf{Y}_i$ e do parâmetro de regressão $\boldsymbol{\beta}$, respectivamente.

Ponderação de Casos

Primeiramente, considere atribuir uma ponderação arbitrária para o valor esperado da função log-verossimilhança dos dados completos (função-Q perturbada), a qual captura violações em direções gerais, representada por

$$Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \mathrm{E}[\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{y}_c)] = \sum_{i=1}^{n} \omega_i \mathrm{E}[\ell_i(\boldsymbol{\theta}|\mathbf{y}_c)] = \sum_{i=1}^{n} \omega_i Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}),$$

com $Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ definida em (4.5), $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^{\top}$ é um vetor de dimensão $n \times 1$, $0 \leq \omega_i \leq 1$ para $i = 1, \ldots, n$ e $\boldsymbol{\omega}_o = (1, \ldots, 1)^{\top}$. Note que para $\omega_i = 0$ e $\omega_j = 1, j \neq i$, a i-ésima unidade experimental é excluída da função log-verossimilhança dos dados completos. Além disso, é possível mostrar que a técnica de influência local para este esquema de perturbação é equivalente a uma medida de influência por eliminação de casos (Osorio, 2006). Para esse esquema de perturbação, encontramos

$$\begin{aligned}
\boldsymbol{\nabla}_{\beta} &= (\widehat{\boldsymbol{u}} \odot \boldsymbol{X})^{\top}(\boldsymbol{I}_n \otimes \boldsymbol{\Gamma}^{-1})D(\boldsymbol{\epsilon}) - (\widehat{\boldsymbol{ut}} \odot \boldsymbol{X})^{\top}(\boldsymbol{I}_n \otimes (\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta})), \\
\boldsymbol{\nabla}_{\Delta} &= (\widehat{\boldsymbol{ut}}^{\top} \otimes \boldsymbol{\Gamma}^{-1})D(\boldsymbol{\epsilon}) - (\widehat{\boldsymbol{ut}^2}^{\top} \otimes (\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta})), \\
\boldsymbol{\nabla}_{\gamma_r} &= -\frac{1}{2}\mathrm{tr}\{\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(r)\}\mathbf{1}_n^{\top} + \frac{1}{2}[(\widehat{\boldsymbol{u}} \odot \boldsymbol{\epsilon})^{\top}(\boldsymbol{I}_n \otimes \dot{\boldsymbol{\Gamma}}_r)D(\boldsymbol{\epsilon}) - 2(\widehat{\boldsymbol{ut}} \odot \boldsymbol{\epsilon})^{\top}(\boldsymbol{I}_n \otimes (\dot{\boldsymbol{\Gamma}}_r\boldsymbol{\Delta})) \\
&\quad + \widehat{\boldsymbol{ut}^2}^{\top}\boldsymbol{\Delta}^{\top}\dot{\boldsymbol{\Gamma}}_r\boldsymbol{\Delta})], \ r = 1, \ldots, p^*.
\end{aligned}$$

Perturbação na Matriz Escala

Para estudar os efeitos de violações nas suposições da matriz escala $\boldsymbol{\Sigma}$ dos efeitos aleatórios, consideramos a perturbação $\boldsymbol{\Delta}(\omega_i) = \omega_i^{-1/2}\boldsymbol{\Delta}$ e $\boldsymbol{\Gamma}(\omega_i) = \omega_i^{-1}\boldsymbol{\Gamma}$, o que corresponde considerar que a distribuição dos efeitos aleatórios é heterocedástica, isto é,

$$\mathbf{Y}_i \sim SMSN_p\left(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}/w_i, \boldsymbol{\lambda}; H\right), i = 1, \ldots, n.$$

Sob este esquema de perturbação, o modelo postulado (não perturbado) é obtido quando $\boldsymbol{\omega}_o = (1, \ldots, 1)^\top$. Além disso, a função-Q perturbada é como em (4.5) substituindo $\boldsymbol{\Delta}(\omega_i)$ e $\boldsymbol{\Gamma}(\omega_i)$ com $\boldsymbol{\Delta}$ e $\boldsymbol{\Gamma}$, respectivamente. Note que, quando $\boldsymbol{\lambda} = \mathbf{0}$, a perturbação acima se reduz à proposta por Osorio *et al.* (2007). A matriz $\nabla_{\boldsymbol{\omega}_o}$ tem elementos dados por

$$\begin{aligned}
\boldsymbol{\nabla}_\beta &= (\widehat{\boldsymbol{u}} \odot \boldsymbol{X})^\top(\boldsymbol{I}_n \otimes \boldsymbol{\Gamma}^{-1})D(\boldsymbol{\epsilon}) - \frac{1}{2}(\widehat{\boldsymbol{ut}} \odot \boldsymbol{X})^\top(\boldsymbol{I}_n \otimes (\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta})),\\
\boldsymbol{\nabla}_\Delta &= \frac{1}{2}(\widehat{\boldsymbol{ut}}^\top \otimes \boldsymbol{\Gamma}^{-1})D(\boldsymbol{\epsilon}),\\
\boldsymbol{\nabla}_{\gamma_r} &= \frac{1}{2}(\widehat{\boldsymbol{u}} \odot \boldsymbol{\epsilon})^\top(\boldsymbol{I}_n \otimes \dot{\boldsymbol{\Gamma}}_r)D(\boldsymbol{\epsilon}) - \frac{1}{2}(\widehat{\boldsymbol{ut}} \odot \boldsymbol{\epsilon})^\top(\boldsymbol{I}_n \otimes (\dot{\boldsymbol{\Gamma}}_r\boldsymbol{\Delta})),\ r = 1, \ldots, p^*.
\end{aligned}$$

Perturbação na Variável Resposta

A perturbação nas respostas $(\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$ é introduzida substituindo $\mathbf{y}_i$ por $\mathbf{y}_i\boldsymbol{\omega} = \mathbf{y}_i + \omega_i\mathbf{S}_y$, onde $\mathbf{S}_y = \mathrm{Diag}(\mathbf{V})$, com $\mathbf{V}$ a matriz de covariâncias amostrais de $\mathbf{y}$. Neste caso, Diag(.) denota o vetor formado pelos elementos da diagonal principal de uma matriz quadrada. A função-Q perturbada é como em (4.5) substituindo $\mathbf{y}_i\boldsymbol{\omega}$ com $\mathbf{y}_i$. Sob este esquema de perturbação, o vetor $\boldsymbol{\omega}_o$ é dado por $\boldsymbol{\omega}_o = \mathbf{0} \in \mathbb{R}^{\mathbf{n}}$ e os elementos de $\boldsymbol{\nabla}_{\boldsymbol{\omega}_o}$ são

$$\begin{aligned}
\boldsymbol{\nabla}_\beta &= (\widehat{\boldsymbol{u}} \odot \boldsymbol{X})^\top(\boldsymbol{I}_n \otimes (\boldsymbol{\Gamma}^{-1}\mathbf{S}_y)),\\
\boldsymbol{\nabla}_\Delta &= \widehat{\boldsymbol{ut}}^\top \otimes (\boldsymbol{\Gamma}^{-1}\mathbf{S}_y),\\
\boldsymbol{\nabla}_{\gamma_r} &= (\widehat{\boldsymbol{u}} \odot \boldsymbol{\epsilon})^\top(\boldsymbol{I}_n \otimes (\dot{\boldsymbol{\Gamma}}_r\mathbf{S}_y)) - (\widehat{\boldsymbol{ut}} \otimes (\mathbf{S}_y^\top\dot{\boldsymbol{\Gamma}}_r\boldsymbol{\Delta})),\ r = 1, \ldots, p^*,
\end{aligned}$$

onde $\boldsymbol{X}$ e $\boldsymbol{\epsilon}$ são definidos em (4.5).

Perturbação na Variável Explicativa

Sob este esquema de perturbação, consideramos a influência que a perturbação nas variáveis explicativas podem produzir nas estimativas dos parâmetros. Considere $\mathbf{X}_{i\boldsymbol{\omega}} = \mathbf{X}_i + \boldsymbol{W}_i\boldsymbol{S}_n$, onde $\boldsymbol{W}_i = (\omega_{kj}^{(i)})$ é uma matriz de perturbações de dimensão $p \times m$, $i = 1, \ldots, n$ e $\boldsymbol{S}_n = \text{diag}\{s_1, \ldots, s_m\}$, onde s_j, $j = 1, \ldots, m$, denotam os fatores de escala para as diferentes unidades de medições associadas com as colunas de $\boldsymbol{X}$. Neste caso, consideramos $\boldsymbol{\omega} = (\boldsymbol{\omega}_1^\top, \ldots, \boldsymbol{\omega}_n^\top)^\top$, onde $\boldsymbol{\omega}_i = (\boldsymbol{\omega}_{(i)1}^\top, \ldots, \boldsymbol{\omega}_{(i)m}^\top)^\top$ é um vetor de dimensão $pm \times 1$ e $\boldsymbol{\omega}_{(i)j}$ é a j-ésima coluna da matriz $\boldsymbol{W}_i$. Dessa forma, a função log-verossimilhança completa perturbada é dada por $\ell_c(\boldsymbol{\theta}|\boldsymbol{\omega})$, alterando $\mathbf{X}_{i\boldsymbol{\omega}}$ com $\mathbf{X}_i$ na log-verossimilhança completa do modelo postulado e neste caso, $\boldsymbol{\omega}_o = \mathbf{0}$. Sob este esquema de perturbação, segue que a matriz $(p + m + p^*) \times mnp$ $\boldsymbol{\nabla}_{\boldsymbol{\omega}_o}$ é dada por $\boldsymbol{\nabla}_{\boldsymbol{\omega}_o} = (\boldsymbol{\nabla}_{\beta}^\top\boldsymbol{\omega}_o, \boldsymbol{\nabla}_{\Delta}^\top\boldsymbol{\omega}_o, \boldsymbol{\nabla}_{\boldsymbol{\gamma}}^\top\boldsymbol{\omega}_o)^\top$, onde $\boldsymbol{\nabla}_{\boldsymbol{\tau}}\boldsymbol{\omega}_o = (\boldsymbol{\nabla}_{1\boldsymbol{\tau}}, ..., \boldsymbol{\nabla}_{n\boldsymbol{\tau}})$, $\boldsymbol{\tau} = \boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{\gamma}$, com

$$\begin{aligned}
\boldsymbol{\nabla}_{i\beta} &= \boldsymbol{S}_n \otimes (\widehat{u}_i\boldsymbol{\epsilon}_i - \widehat{ut}_i\boldsymbol{\Delta})^\top\boldsymbol{\Gamma}^{-1} - \widehat{u}_i(\widehat{\boldsymbol{\beta}}^\top\boldsymbol{S}_n \otimes \mathbf{X}_i^\top\boldsymbol{\Gamma}^{-1}),\\
\boldsymbol{\nabla}_{i\Delta} &= -\widehat{ut}_i[\boldsymbol{S}_n \otimes \boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta} + \boldsymbol{\beta}^\top\boldsymbol{S}_n \otimes \boldsymbol{\Gamma}^{-1}],\\
\boldsymbol{\nabla}_{i\gamma_k} &= -\widehat{\boldsymbol{S}}_n\widehat{\boldsymbol{\beta}} \otimes \dot{\boldsymbol{\Gamma}}_r(\widehat{u}_i\boldsymbol{\epsilon}_i - \widehat{ut}_i\boldsymbol{\Delta}), i = 1, ..., n.
\end{aligned}$$

Quando estamos interessados em perturbar uma variável explicativa específica, consideramos a seguinte matriz explicativa perturbada $\mathbf{X}_{i\boldsymbol{\omega}} = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{is}(\omega_i), \ldots, \mathbf{x}_{im})$, onde $\mathbf{x}_{is}(\omega_i) = \mathbf{x}_{is} + \omega_i\mathbf{1}_p$, $s = 1, \ldots, m$, $\mathbf{x}_{is}$ é a s-ésima coluna da matriz $\mathbf{X}_i$ e $\mathbf{1}_p$ é um vetor $p \times 1$ de uns. Assim, este caso pode cobrir situações onde x é medida com erro. Note que os elementos de $\boldsymbol{\nabla}_{\omega_0}$ são

$$\boldsymbol{\nabla}_{\beta} = \widehat{\boldsymbol{u}}^\top \odot [\boldsymbol{A}_s(\boldsymbol{I}_n \otimes \mathbf{1}_p^\top)\boldsymbol{\epsilon}] - \beta_s\widehat{\boldsymbol{u}}^\top \odot [\boldsymbol{X}^\top(\boldsymbol{I}_n \otimes (\boldsymbol{\Gamma}^{-1}\mathbf{1}_p))] - \widehat{\boldsymbol{ut}}^\top \odot [\mathbf{1}_p^\top \otimes (\boldsymbol{A}_s^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta})],$$

$$\begin{aligned}
\boldsymbol{\nabla}_{\Delta} &= -\beta_s[\widehat{\boldsymbol{ut}}^{\top} \odot (\mathbf{1}_n^{\top} \otimes \boldsymbol{\Gamma}^{-1})], \\
\boldsymbol{\nabla}_{\gamma_r} &= \beta_s \left[\widehat{\boldsymbol{u}}^{\top} \odot (\boldsymbol{\epsilon})^{\top}[\boldsymbol{I}_n \otimes (\dot{\boldsymbol{\Gamma}}_r \mathbf{1}_p)] - \widehat{\boldsymbol{ut}} \odot [\mathbf{1}_n^{\top} \otimes (\mathbf{1}_p^{\top} \boldsymbol{\Gamma}_r \boldsymbol{\Delta})]\right], \ r = 1, \ldots, p^*,
\end{aligned}$$

onde $\boldsymbol{A}_s = (\mathbf{0}_p, \ldots, \mathbf{1}_p, \ldots, \mathbf{0}_p)$ é uma matriz $p \times m$, com $\mathbf{1}_p$ na s-ésima coluna da matriz $\boldsymbol{A}_s$, $s = 1, \ldots, m$.

Nas próximas seções, um exemplo real e estudos de simulação são apresentados com o intuito de ilustrar o desempenho da metodologia desenvolvida neste capítulo.

4.4 Estudos de Simulação

Com o intuito de examinar o desempenho dos métodos propostos, apresentamos alguns estudos de simulação. O primeiro estudo considera a análise de influência de um único outlier no método de estimação de $\boldsymbol{\theta}$, procedimento discutido em Pinheiro *et al.* (2001) no contexto dos modelos lineares mistos sob a distribuição t-Student. O segundo estudo mostra a capacidade da metodologia desenvolvida em detectar dados atípicos e a sensibilidade das estimativas de máxima verossimihança às observações extremas sob distribuições alternativas à skew-normal, considerando os esquemas ponderação de casos, perturbação na variável resposta e perturbação na variável explicativa. Os dados simulados são obtidos facilmente usando a representação estocástica dada em (2.4) com $\kappa(U) = 1/U$. Por exemplo, a distribuição skew-t multivariada é obtida de (2.4) considerando $U \sim Gamma(\nu/2, \nu/2)$, $\nu > 0$.

4.4.1 Estudo I: Influência de um Único Outlier

A flexibilidade dos modelos SMSN pode ser estudada considerando a influência de uma única observação aberrante na estimativa de máxima verossimilhança de $\boldsymbol{\theta}$. Sem

perda de generalidade, simulamos um conjunto de dados proveniente do ST-RM, com tamanho amostral $n = 100$, $\mathbf{X}_i = \mathbf{I}_2$, $i = 1, \ldots, 100$, $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top = (15, 20)^\top$, $\boldsymbol{\Sigma} = \mathbf{I}_2$, $\boldsymbol{\lambda} = (6, -4)^\top$ e $\nu = 3$. Para esta amostra, ajustamos o SN-RM, o ST-RM, o SCN-RM e o SSL-RM. Para simplificar, estudamos a influência da variação de $\Im$ unidades em uma única observação y_{ik} na estimativa de máxima verossimilhança de $\boldsymbol{\theta}$, isto é, substituímos a observação y_{ik} pelo valor contaminado $y_{ik}(\Im) = y_{ik} + \Im$. Neste exemplo, contaminamos a segunda observação da unidade experimental 43 variando $\Im$ entre -4 e 4. Na Figura 4.1, apresentamos os resultados das mudanças relativas das estimativas de β_0 e β_1, para diferentes contaminações de $\Im$, sob o SN-RM, o ST-RM, o SCN-RM e o SSL-RM. Como esperado, as estimativas sob os modelos assimétricos com caudas pesadas (ST, SCN e SSL) são menos afetadas pelas variações de $\Im$ em relação ao SN-RM.

4.4.2 Estudo II: Esquemas de Perturbação

Com o intuito de avaliar a capacidade da metodologia para detectar dados atípicos e a sensibilidade das estimativas de máxima verossimilhança às observações aberrantes sob distribuições alternativas à skew-normal, conduzimos um pequeno estudo de simulação. A seguir vamos simular quatro conjunto de dados provenientes de SMSN-RM, onde $\mathbf{Y}_i \sim SMSN_2(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{I}_2, \boldsymbol{\lambda}; H), i = 1, \ldots, 200$, com $\boldsymbol{\beta} = (-1, 2)^\top$, $\mathbf{X}_i = \begin{pmatrix} 1 & x_{i1} \\ 1 & x_{i2} \end{pmatrix}$, $x_{ij} \sim U(0, 1)$, $j = 1, 2$, $\boldsymbol{\lambda} = (2, 10)^\top$ e $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \epsilon_{i2})^\top$, tal que $\nu = 3$ para a skew-t, $\nu = 2$ para a skew-slash e $\boldsymbol{\nu} = (0.3, 01)^\top$ para a distribuição skew-normal contaminada. Note que geramos dados provenientes de modelos assimétricos e com caudas pesadas.

Como sugerido por Ortega *et al.* (2003), após gerar x_{ij}, $i = 1, \ldots, 200$ e $j = 1, 2$, perturbamos o valor máximo da amostra, considerando $x_{max} \leftarrow x_{max} + \sqrt{(\mathbf{x}^\top\mathbf{x})}$, onde

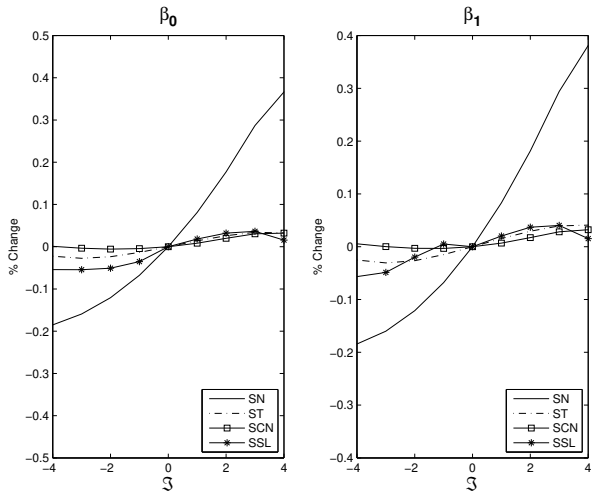


Figura 4.1: Dados simulados. Mudanças relativas nas estimativas de β_0 e β_1 sob o SN-RM, o ST-RM, o SCN-RM e o SSL-RM para diferentes contaminações de $\Im$ na observação 43. %change= $100 \times \left((\widehat{\theta}(\Im) - \widehat{\theta})/\widehat{\theta}\right)$, onde $\widehat{\theta}$ denota a estimativa original e $\widehat{\theta}(\Im)$ a estimativa para os dados contaminados.

$\mathbf{x} = (x_1, \ldots, x_{200})^\top$ e $x_i = (x_{i1}, x_{i2})^\top$. Este ponto atípico corresponde à unidade experimental 180. Para este tipo de perturbação nos dados, no contexto de influência local, consideramos ponderação de casos e perturbação na variável explicativa. Além disso, perturbamos a variável resposta da observação 180, tal que $y_{i2} \leftarrow y_{i2} + \sqrt{(\mathbf{y}^{*\top}\mathbf{y}^*)}$, onde $\mathbf{y}^* = (y_{1,2}, \ldots, y_{200,2})^\top$. Segundo a metodologia descrita na Seção 4.3.2, as Figuras 4.2 e 4.3 mostram os gráficos de $M(0)$ para as perturbações nas variáveis resposta e explicativa, respectivamente, considerando a marca de referência obtida segundo Lee & Xu (2004) com $c^* = 3$. Para estes esquemas de perturbação, observamos nas Figuras 4.2a e 4.3a, a influência da observação 180, justamente a unidade experimental perturbada. Isso confirma a alta sensibilidade das estimativas de máxima verossimilhança dos parâmetros na presença de dados atípicos quando o modelo skew-normal é considerado.

As Figuras 4.2-4.3 também mostram que as estimativas de máxima verossimilhança são mais flexíveis (menos sensíveis) na presença de dados atípicos quando distribuições como skew-t, skew-slash e skew-normal contaminada são usadas, uma vez que a observação perturbada não foi detectada. Para a perturbação ponderação de casos, notamos os mesmos resultados e então, não serão mostrados aqui por brevidade.

Na próxima seção, reanalisaremos o conjunto de dados "Stack–Loss", previamente analisado por Brownlee (1960) no contexto gaussiano. A técnica de influência local será considerada para avaliar a sensibilidade das estimativas de máxima verossimilhança dos parâmetros do modelo de regressão sob as distribuições skew-normal, skew-t, skew-slash e skew-normal contaminada, sob diversos esquemas de perturbação.

4.5 Aplicação: Conjunto de Dados "Stack-Loss"

Consideremos novamente o conjunto de dados "Stack-Loss", introduzido no Capítulo 3 para ilustrar uma análise bayesiana no modelo de regressão sob a classe alternativa de distribuições misturas de escala skew-normal (SSMSN). Nesta aplicação, reanalisamos este conjunto de dados, do ponto de vista clássico e no contexto de estimação e diagnóstico, considerando o modelo de regressão linear sob a classe de distribuições SMSN. Como já mencionado, o modelo de regressão linear que será considerado nesta aplicação difere do modelo especificado no Capítulo 3 e como esperado, os resultados serão diferentes.

É importante ressaltar que o conjunto de dados "Stack-Loss" apresentado por Brownlee (1960) representa uma referência clássica no contexto de regressão linear múltipla, de procedimentos robustos, de presença de "outliers" e de questões relacionadas. Dessa

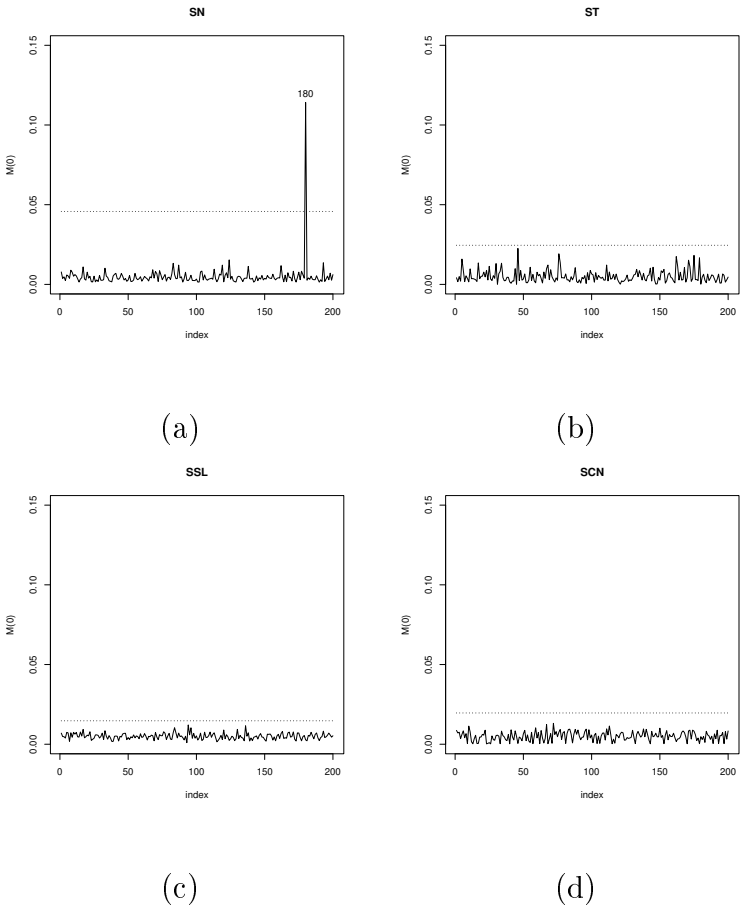


Figura 4.2: Dados simulados. Gráficos de $M(0)$ para a perturbação na variável resposta: (a) skew-normal; (b) skew-t; (c) skew-slash e (d) skew-normal contaminada. A linha horizontal delimita a marca de referência de Lee & Xu (2004) para M(0) com $c* = 3$.

forma, esses dados têm sido utilizados em ilustrações númericas por um grande número de programas estatísticos e autores, por exemplo, no programa bayesiano WinBUGS (exemplos volume 1), as observações 1, 3, 4 e 21 são detectadas como influentes, considerando regressão ridge e distribuições simétricas com caudas pesadas, tais como t-Student e exponencial potência.

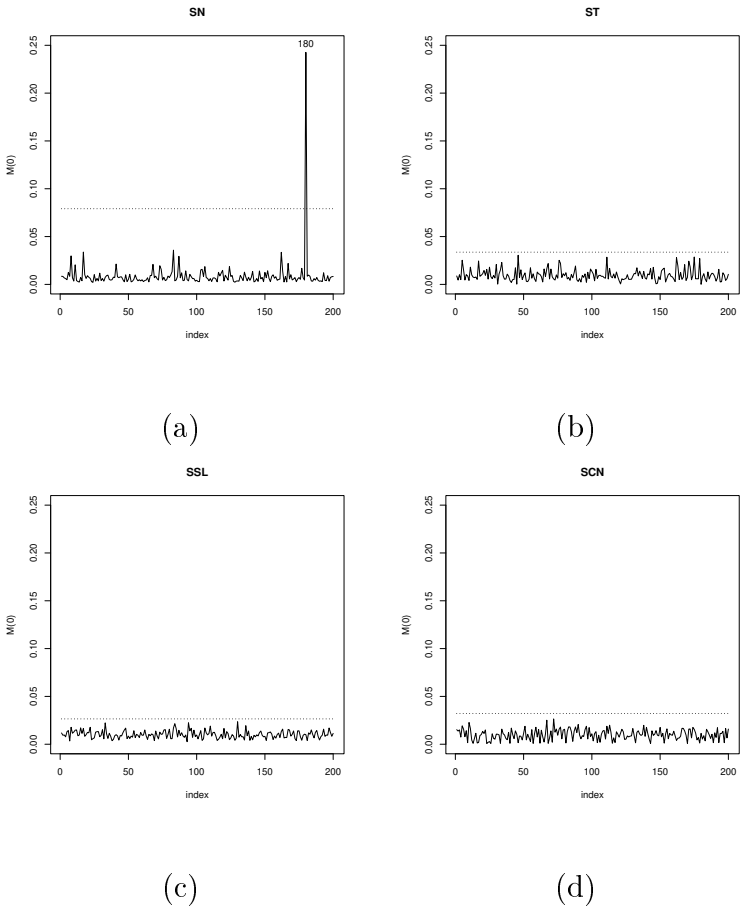


Figura 4.3: Dados simulados. Gráficos de $M(0)$ para perturbação na variável explicativa: (a) skew-normal; (b) skew-t; (c) skew-slash e (d) skew-normal contaminada. A linha horizontal delimita a marca de referência de Lee & Xu (2004) para M(0) com $c* = 3$.

Conforme descrito no Capítulo 3, com propósitos exploratórios, o histograma e o gráfico Q-Q normal dos dados foram construídos. A Figura 3.2 sugere um padrão assimétrico e o uso de distribuições com caudas pesadas. Assim, nesta seção, revisitamos o conjunto de dados "Stack-Loss" com o objetivo de fornecer inferências adicionais, usando as distribuições SMSN. Consideramos o modelo de regressão definido na Seção 4.2 com as variáveis explicativas $\mathbf{x}_i = (1, x_{1i}, x_{2i}, x_{3i})^\top$, sendo $x_1 =$"air flow", $x_2 =$ "wa-

ter temperature, x_3 = "acid concentration" e a variável resposta "stack-loss" seguindo as distribuições SN, ST, SSL ou SCN, provenientes da classe SMSN, apenas para comparações, i.e., $y_i \sim SMSN_1(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2, \lambda; H), i = 1, \ldots, 21$, com $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top$.

A Tabela 4.1 contém as estimativas de máxima verossimilhança para os parâmetros dos quatro modelos, i.e., SN-RM, ST-RM, SCN-RM e SSL-RM, acompanhadas dos seus respectivos desvios padrões obtidos via matriz informação observada. Como sugerido por Lange *et al.* (1989), a função log-verossimilhança perfilada foi usada na escolha do valor de $\boldsymbol{\nu}$. Encontramos que $\nu = 1.14$, $\nu = 1.12$ e $\boldsymbol{\nu} = (0.33, 0.03)^\top$ são apropriados para a ST, a SSL e a SCN, respectivamente. Pela Tabela 4.1, notamos que há poucas evidências de assimetria, entretanto Azzalini & Genton (2008) analisaram o mesmo conjunto de dados considerando um ST-RM. Além disso, os valores do AIC indicam que as distribuições SMSN com caudas pesadas (em particular, as distribuições ST e SCN) apresentam um ajuste melhor do que o modelo SN.

Tabela 4.1: Dados "Stack-Loss". Estimativas de máxima verossimilhança para os quatro modelos SMSN selecionados. Os valores SE, entre parênteses, são os desvios padrões assintóticos estimados.

Dist.	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\sigma^2}$	$\widehat{\lambda}$	$\ell(\widehat{\boldsymbol{\theta}})$	AIC
SN	-39.9533	0.7156	1.2953	-0.1521	8.5168	0.0144	-52.2878	58.2878
	(22.7490)	(0.1213)	(0.3311)	(0.1406)	(0.4434)	(8.6271)		
ST	-38.0525	0.8584	0.4760	-0.0794	0.9800	0.2835	-49.4816	56.4816
	(3.7302)	(0.0578)	(0.1451)	(0.0548)	(0.2237)	(0.6463)		
SCN	-36.4958	0.8375	0.4881	-0.0870	0.8141	0.3167	-47.3065	55.3065
	(3.8325)	(0.0558)	(0.1400)	(0.0514)	(0.1575)	(0.5580)		
SSL	-39.6103	0.8160	0.7439	-0.1032	2.0932	0.5584	-51.2626	58.2626
	(6.7244)	(0.1149)	(0.2664)	(0.0901)	(0.3450)	(0.9855)		

De acordo com os resultados obtidos na Proposição 2.2.8, identificamos as ob-

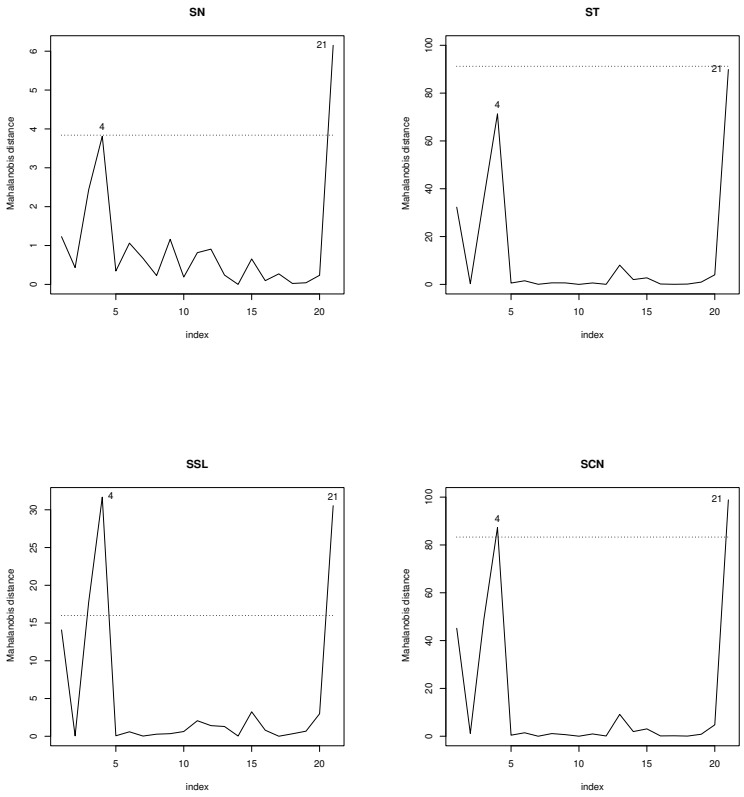


Figura 4.4: Dados "Stack-Loss". Gráficos da distância de Mahalanobis para os quatro modelos ajustados, considerando $\xi = 0.95$.

servações atípicas considerando a distância de Mahalanobis $d_i = (y_i - \mathbf{x}_i\boldsymbol{\beta})^2/\sigma^2,\ i = 1, \ldots, 21$; veja Osorio *et al.* (2007). A Figura 4.4 mostra tal distância para os quatro modelos ajustados. Nota-se que as observações 4 e 21 aparecem como "outliers" para os modelos SN, SSL e SCN. Entretanto, sob o ST-RM, essas duas observações aparecem de forma menos evidente.

Proveniente do algoritmo EM, os pesos estimados $(\hat{u}_i, i = 1, \ldots, 21)$ das observações

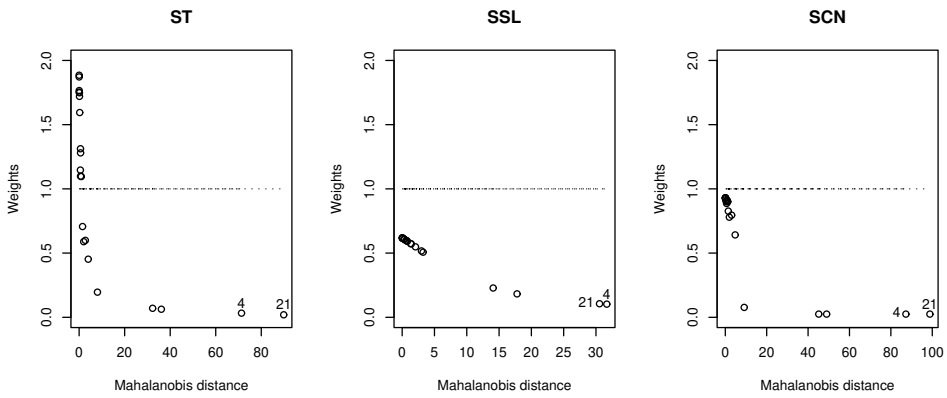


Figura 4.5: Dados "Stack-Loss". Valores de u_i estimados para os modelos ST, SSL e SCN.

4 e 21 são os menores para todos os modelos ajustados com distribuições de caudas pesadas (veja Figura 4.5), confirmando a flexibilidade das estimativas de máxima verossimilhança dos modelos SMSN com caudas pesadas contra observações atípicas. Para os casos normal e skew-normal $u_i = 1\ \forall i$ e estão representados por uma linha tracejada na Figura 4.5. Note que para as distribuições ST, SSL e SCN, o peso u_i é inversamente proporcional à distância de Mahalanobis. Dessa forma, o processo de estimação de $\boldsymbol{\theta}$ tende a atribuir menor peso para as observações anômalas no sentido da distância de Mahalanobis.

A seguir identificaremos as observações influentes provenientes do conjunto de dados "Stack-Loss" utilizando a quantidade $M(0)$, calculada através da curvatura conformal $B_{f_Q,\mathbf{u}}$, obtida considerando ponderação de casos e os esquemas de perturbação na escala e variável resposta. As Figuras 4.6, 4.7 e 4.8 apresentam os gráficos de $M(0)$ para os quatro modelos selecionados.

Observando essas figuras, a observação 21 parece ser a mais influente para as estima-

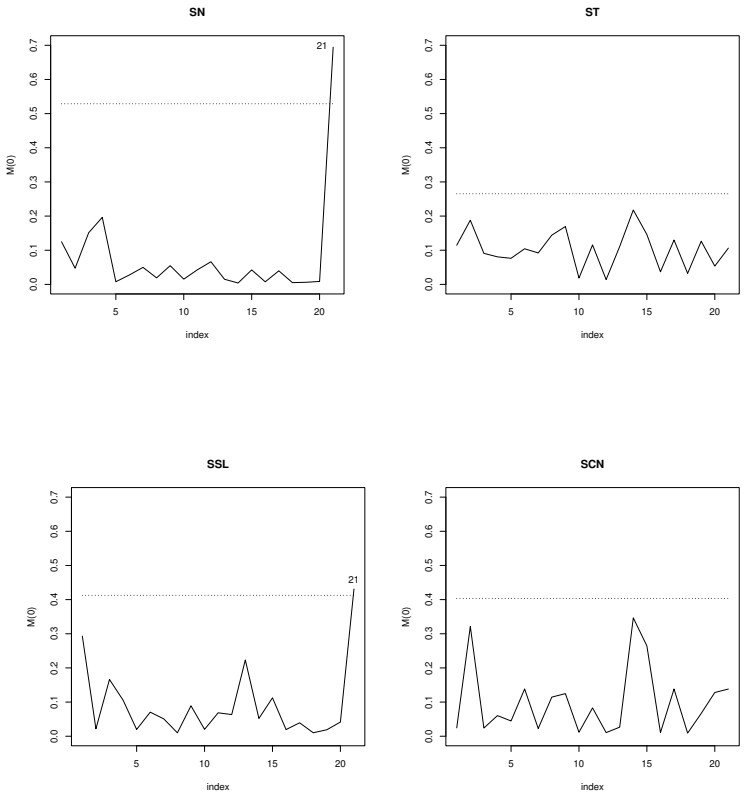


Figura 4.6: Dados "Stack-Loss". Gráficos de $M(0)$ sob ponderação de casos para os quatro modelos ajustados. A linha horizontal delimita a marca de referência de Lee & Xu (2004) para M(0) com $c* = 3$.

tivas de máxima verossimilhança do modelo SN sob ponderação de casos e o esquema de perturbação na escala. A mesma observação também parece ser a mais influente em $\widehat{\boldsymbol{\theta}}$ quando consideramos o modelo SSL e a perturbação ponderação de casos, enquanto que para os demais esquemas de perturbação nenhuma influência foi observada, como mostrado nas Figuras 4.7c e 4.8c. Além disso, note que as estimativas de máxima verossimilhança são bastante estáveis quando consideramos a perturbação na variável

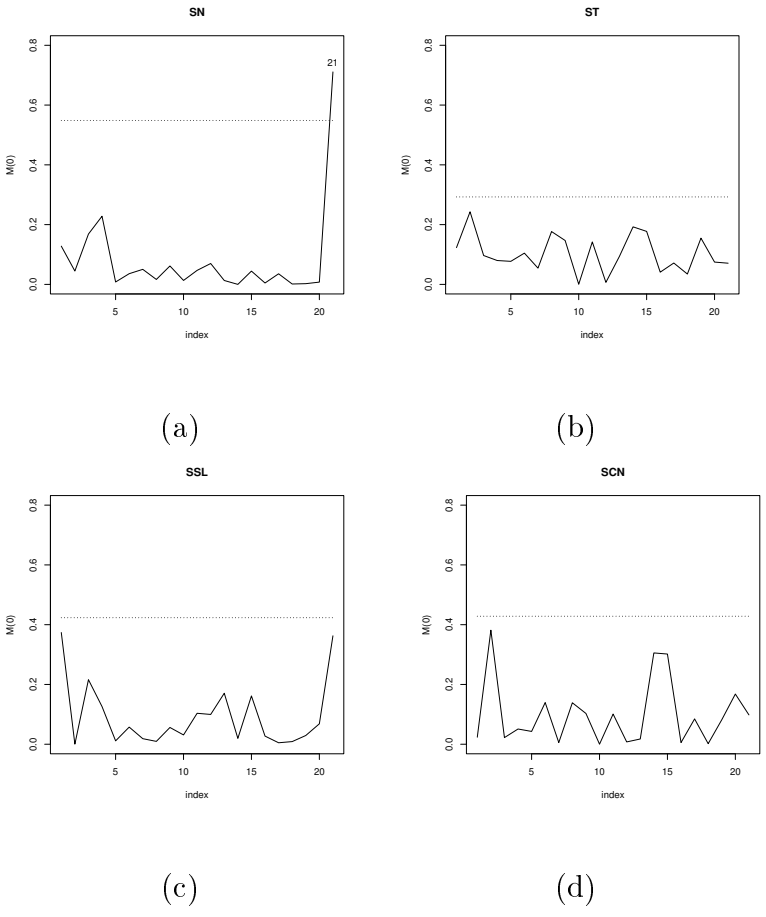


Figura 4.7: Dados "Stack-Loss". Gráficos de $M(0)$ sob a perturbação na escala para os quatro modelos ajustados. A linha horizontal delimita a marca de referência de Lee & Xu (2004) para M(0) com $c* = 3$.

resposta nos quatros modelos ajustados, como pode ser visto na Figura 4.8.

Como sugerido por Lee *et al.* (2006), usamos as quantidades TRC e MRC para revelar o impacto da observação influente detectada na estimativa de máxima verossi-

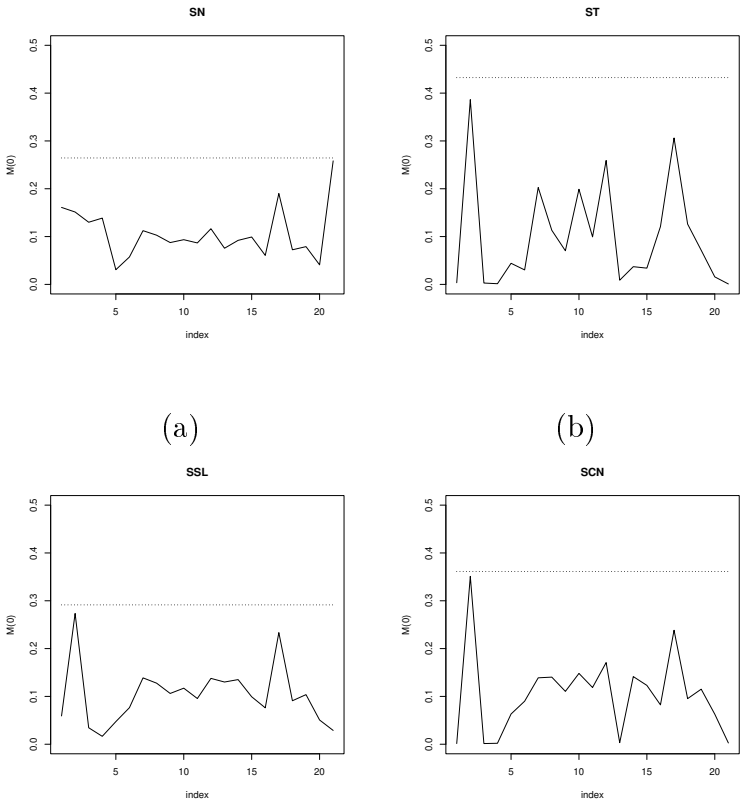


Figura 4.8: Dados "Stack-Loss". Gráficos de $M(0)$ sob perturbação na variável resposta. A linha horizontal delimita a marca de referência de Lee & Xu (2004) para M(0) com $c* = 3$.

milhança de $\boldsymbol{\theta}$. Elas são definidas como

$$TRC = \sum_{j=1}^{n_p} |\frac{\hat{\boldsymbol{\theta}}_j - \hat{\boldsymbol{\theta}}_{[i]j}}{\hat{\boldsymbol{\theta}}_j}| \quad \text{e} \quad MRC = \max_{j=1,\ldots,n_p} |\frac{\hat{\boldsymbol{\theta}}_j - \hat{\boldsymbol{\theta}}_{[i]j}}{\hat{\boldsymbol{\theta}}_j}|, \tag{4.8}$$

onde n_p é a dimensão de $\boldsymbol{\theta}$ e o subscrito $[i]$ indica que a estimativa de máxima verossimilhança de $\boldsymbol{\theta}$ foi obtida excluindo i-ésima observação, y_i. Na tabela 4.2, as comparações dessas medidas são dadas, considerando diferentes modelos, quando excluímos a observação 21. Note que a maior mudança ocorreu sob a distribuição skew-normal. Como

esperado, os resultados indicam que as estimativas de máxima verossimilhança são menos sensíveis à presença de dados atípicos quando usamos distribuições com caudas mais pesadas que a SN, por exemplo, a distribuição ST, fato também observado nos três esquemas de perturbação desenvolvidos nesta ilustração.

Tabela 4.2: Dados "Stack-Loss". Comparação das mudanças relativas nas estimativas de máxima verossimilhança em termos das quantidades TRC e MRC para os quatro modelos SMSN selecionados.

	TRC	MRC
SN	6.2714×10^4	6.2712×10^4
ST	2.4853	2.1442
SSL	3.9003	3.3005
SCN	2.2664	2.0840

A flexibilidade do modelo ST às observações atípicas pode ser estudada considerando a influência de uma única observação anômala na estimativa de máxima verossimilhança de $\boldsymbol{\theta}$. Nesta aplicação, propomos estudar a influência que a alteração de $\Im$ unidades em uma observação acarreta em $\widehat{\boldsymbol{\theta}}$ e em seu respectivo desvio padrão. Substituímos uma observação y_k pelo valor contaminado $y_k(\Im) = y_k + \Im$. Neste exemplo, contaminamos a observação 21 e variamos $\Im$ entre -10 e 10.

Na Figura 4.9, apresentamos os resultados das estimativas (primeira linha) e desvios padrões (segunda linha) de $\widehat{\beta}_1$, $\widehat{\beta}_2$ e $\widehat{\beta}_3$ para diferentes contaminações de $\Im$, sob os modelos de regressão SN e ST. Como esperado, as estimativas de máxima verossimilhança do modelo ST são menos afetadas por variações de $\Im$ do que as estimativas do modelo SN. Assim, sob o modelo SN, a observação atípica parece ter mais impacto no processo

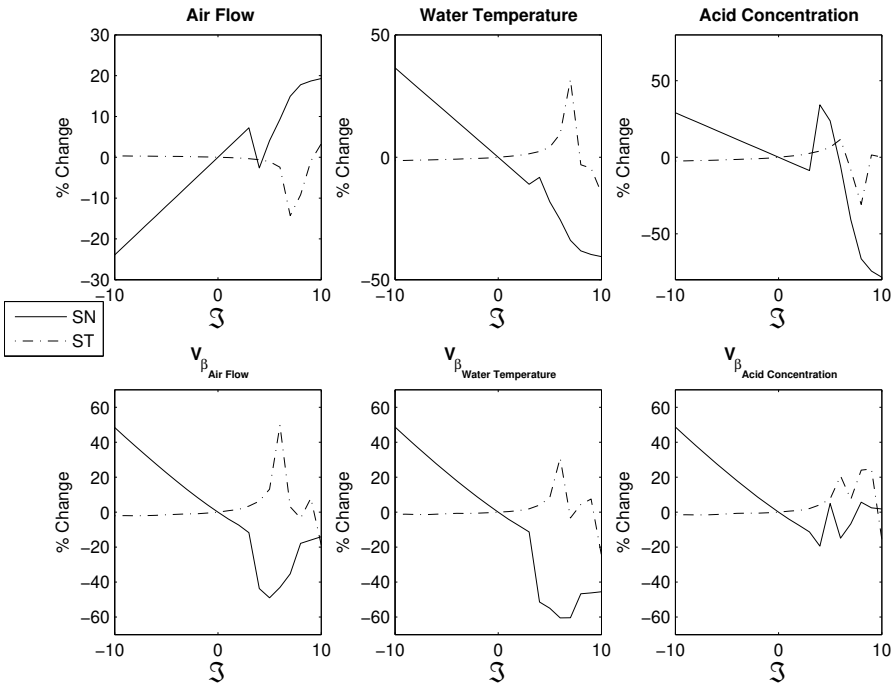


Figura 4.9: Dados "Stack-Loss". Mudanças relativas nas estimativas de máxima verossimilhança (primeira linha) e desvios padrões (segunda linha) de $\widehat{\boldsymbol{\theta}}$ para os modelos SN (linha contínua) e ST (linha tracejada), considerando diferentes contaminações de $\Im$ na observação 21.

de estimação e também no tamanho do intervalo de confiança.

4.6 Observações Finais

Neste capítulo, desenvolvemos uma nova classe de modelos de regressão usando as distribuições misturas de escala skew-normal, sendo a classe de distribuições misturas de escala normal um caso especial. Essa classe de modelos possui parâmetros adicionais que podem ser usados, simultaneamente, para controlar a assimetria e as caudas pe-

sadas das distribuições. Além disso, com um moderado trabalho computacional, os modelos de regressão SMSN com caudas pesadas podem fornecer procedimentos mais robustos às observações aberrantes, possíveis observações influentes, do que os obtidos quando usamos a distribuição skew-normal (normal). Expressões com formas fechadas foram obtidas para a hessiana $\ddot{\mathbf{Q}}$ e a matriz $\boldsymbol{\nabla}_{\boldsymbol{\omega}_o}$ sob quatro esquemas de perturbação. Através do estudo de diagnóstico de influência, notamos que os modelos SMSN com caudas mais pesadas que a distribuição skew-normal acomodam melhor as observações aberrantes e/ou influentes. Dessa forma, acreditamos que a metodologia desenvolvida para o SMSN-RM possa apre-sentar resultados satisfatórios em outros modelos multivariados, por exemplo, modelos lineares mistos.

Capítulo 5

Modelo Linear Misto Misturas de Escala Skew-Normal

Extensões de alguns procedimentos de diagnóstico para o modelo linear misto misturas de escala skew-normal serão discutidas. Esta classe fornece uma generalização útil do modelo linear misto normal (e skew-normal) uma vez que assume que os efeitos aleatórios e os erros aleatórios seguem conjuntamente uma distribuição SMSN multivariada. Inspirados pelo algoritmo EM, uma análise de influência local para os modelos lineares mistos, segundo a metodologia de Zhu & Lee (2001), será desenvolvida. Quatro modelos específicos de esquemas de perturbação serão discutidos. Finalmente, um conjunto de dados reais será analisado com o intuito de ilustrar a utilidade da metodologia proposta.

5.1 Introdução

Os modelos lineares mistos tornaram-se importantes ferramentas em análises estatísticas. Eles são comumente utilizados em análises de dados contínuos com medidas repetidas, sendo de grande aplicabilidade em pesquisas de diversas áreas como agricultura, biologia, economia, ciências sociais, entre outras. Apesar de sua ampla aplicabilidade estatística, um pressuposto possivelmente, uma suposição restritiva nos modelos lineares mistos é que os efeitos aleatórios e os erros aleatórios seguem conjuntamente uma distribuição normal multivariada. Assim, considerável atenção tem sido dada para relaxar a suposição de normalidade e consequentemente estimar os efeitos aleatórios e parâmetros do modelo.

Há uma considerável linha de pesquisa em modelos lineares mistos nessa direção. Verbeke & Lesaffre (1996) introduziram o modelo linear misto (LMM) heterogêneo, cujas distribuições dos efeitos aleatórios foram relaxadas para a mistura finita de normais. Pinheiro *et al.* (2001) propuseram o modelo linear misto t-Student multivariado e mostram que este novo modelo apresenta um bom desempenho na presença de "outliers". Zhang & Davidian (2001) propuseram um modelo linear misto, onde os efeitos aleatórios seguem uma distribuição semi-paramétrica (SNP). Rosa *et al.* (2003) adotaram uma abordagem bayesiana para realizar análises posteriores do LMM, usando a classe de distribuições normal/independente (NI) (Lange & Sinsheimer, 1993). Deste modo, o NI-LMM foi definido. Ghidey *et al.* (2004) desenvolveram um LMM com densidade suave para os efeitos aleatórios. Arellano-Valle *et al.* (2005) e Lin & Lee (2008) propuseram o modelo linear misto skew-normal (SN-LMM), baseados na distribuição skew-normal de Azzalini & Dalla-Valle (1996). Zhou & He (2008) propuseram o modelo linear misto skew-t (ST-LMM), onde um procedimento de estimação com 3 pas-

sos é proposto para obter os estimadores de máxima verossimilhança. Recentemente, Lachos *et al.* (2010) introduziram o modelo linear misto skew-normal/independente (SNI-LMM), onde os efeitos aleatórios e os erros aleatórios seguem conjuntamente uma distribuição skew-normal/independente multivariada (Lachos & Vilca, 2007). Eles também desenvolveram um algoritmo tipo-EM para estimar por máxima verossimilhança os parâmetros do modelo. Neste capítulo, considerando que a função de pesos positiva $\kappa(U)$, definida em (2.2), é uma função monótona, descrevemos o modelo linear misto sob a classe de distribuições misturas de escala skew-normal (SMSN-LMM), conforme discutido em Lachos *et al.* (2010). Esta classe de modelos é interessante, uma vez que o NI-LMM, o SN-LMM e o ST-LMM são casos particulares dessa classe. Dessa forma, a análise de influência local, desenvolvida neste capítulo, será baseada no algoritmo tipo-EM apresentado por Lachos *et al.* (2010).

Por outro lado, uma investigação de influência é um passo importante na análise de dados após a estimação dos parâmetros. Isso pode ser realizado através da análise de influência local, uma técnica estatística utilizada para avaliar a estabilidade dos resultados do processo de estimação com respeito às suposições do modelos (como heterocedasticidade ou covariáveis medidas com erros). Considerando o pioneiro trabalho de Cook (1986), esta área de pesquisa tem recebido considerável atenção na literatura estatística; veja, por exemplo, Lesaffre & Verbeke (1998), Zhu & Lee (2001), Fung *et al.* (2002), Lee & Xu (2004), Osorio *et al.* (2007), entre outros. Entretanto, como a função de log-verossimilhança observada do SMSN-LMM envolve integrais, a aplicação direta da abordagem de Cook (1986) para este modelo não é atrativa. Inspirados por Montenegro *et al.* (2009a) que estudou influência local no contexto SN-LMM, aplicamos a abordagem de influência local de Zhu & Lee (2001) para os SMSN-LMMs. Acreditamos que os resultados que serão desenvolvidos neste capítulo são um suplemento necessário

para o trabalho de Lachos *et al.* (2010).

O capítulo está organizado como segue. Na Seção 5.2, o SMSN-LMM é definido e um algoritmo tipo-EM para obter as estimativas de máxima verossimilhança é descrito, como discutidos em Lachos *et al.* (2010). Na Seção 5.3, desenvolvemos a metodologia de influência local pertinente ao SMSN-LMM, sendo que quatro esquemas de perturbação foram considerados. A metodologia será ilustrada na Seção 5.4, considerando o conjunto de dados "Framingham Cholesterol". Finalmente, algumas observações finais são dadas na Seção 5.5.

5.2 Descrição do Modelo

Nesta seção, consideramos o modelo linear misto sob a classe de distribuições misturas de escala skew-normal (SMSN-LMM). Detalhes dessa seção podem ser obtidos em Lachos *et al.* (2010). Em geral, o modelo linear misto normal (N-LMM) é definido como

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n\,,$$

onde $\mathbf{Y}_i$ é um vetor $n_i \times 1$ de respostas contínuas observadas para a i-ésima unidade amostral, $\mathbf{X}_i$ é a matriz $n_i \times p$ de desenho dos efeitos fixos, $\boldsymbol{\beta}$ é um vetor $p \times 1$ de efeitos fixos, $\mathbf{Z}_i$ é a matriz $n_i \times q$ de desenho dos efeitos aleatórios $\mathbf{b}_i$ (vetor de dimensão $q \times 1$) e $\boldsymbol{\epsilon}_i$ é o vetor $n_i \times 1$ de erros aleatórios. Assume-se que os efeitos aleatórios $\mathbf{b}_i$ e os erros aleatórios $\boldsymbol{\epsilon}_i$ são independentes com $\mathbf{b}_i \overset{\text{iid}}{\sim} N_q(\mathbf{0}, \mathbf{D})$ e $\boldsymbol{\epsilon}_i \overset{\text{ind}}{\sim} N_{n_i}(\mathbf{0}, \boldsymbol{\Sigma}_i)$. A matriz de covariâncias $(q \times q)$ $\mathbf{D}$ pode ser estruturada ou não. As matrizes de covariâncias $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\boldsymbol{\gamma})$ são tipicamente definidas para depender de i através de suas dimensões $n_i \times n_i$, $i = 1, \ldots, n$, sendo parametrizadas por um conjunto paramétrico $\boldsymbol{\gamma}$ fixo geralmente pequeno como, por exemplo, uma estrutura de covariância AR(1). Neste contexto

de modelos mistos, uma situação especial (e comum) é considerar $\boldsymbol{\Sigma}_i = \sigma_e^2 \mathbf{R}_i$, onde $\mathbf{R}_i$ é uma matriz conhecida de dimensã $n_i \times n_i$ e neste caso, $\boldsymbol{\gamma} = \sigma_e^2$.

De acordo com Lachos *et al.* (2010), o SMSN-LMM é definido considerando que

$$\mathbf{b}_i \overset{\text{iid}}{\sim} SMSN_q(\mathbf{0}, \mathbf{D}, \boldsymbol{\lambda}; H) \text{ e } \boldsymbol{\epsilon}_i \overset{\text{ind}}{\sim} SMN_{n_i}(\mathbf{0}, \sigma_e^2 \mathbf{R}_i; H), \ i = 1, \ldots, n.$$

De (2.4), podemos escrever o SMSN-LMM como segue

$$\mathbf{Y}_i | \mathbf{b}_i, U_i = u_i \overset{\text{ind}}{\sim} N_{n_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \kappa(u_i)\sigma_e^2\mathbf{R}_i), \tag{5.1}$$

$$\mathbf{b}_i | T_i = t_i, U_i = u_i \overset{\text{ind}}{\sim} N_q(\boldsymbol{\Delta} t_i, \kappa(u_i)\boldsymbol{\Gamma}), \tag{5.2}$$

$$T_i | U_i = u_i \overset{\text{ind}}{\sim} HN_1(0, \kappa(u_i)) \tag{5.3}$$

$$U_i \overset{\text{iid}}{\sim} H(\boldsymbol{\nu}), \tag{5.4}$$

$i = 1, \ldots, n$, todos independentes, $\boldsymbol{\Delta} = \mathbf{D}^{1/2}\boldsymbol{\delta}$ e $\boldsymbol{\Gamma} = \mathbf{D} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top$, com $\boldsymbol{\delta} = \boldsymbol{\lambda}/\sqrt{1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda}}$, $\mathbf{D}^{1/2}$ é a raiz quadrada de $\mathbf{D}$ contendo $q(q+1)/2$ elementos distintos, digamos $\boldsymbol{\alpha}$ e $HN_1(0, \sigma^2)$ é a distribuição half-$N_1(0, \sigma^2)$. Quando $U_i = 1$ $(i = 1 \ldots, n)$, o SMSN-LMM reduz para o SN-LMM, como definido em Arellano-Valle *et al.* (2005) e Lin & Lee (2008). Além disso, se $\boldsymbol{\lambda} = \mathbf{0}$, o SMSN-LMM reduz para o usual SMN-LMM, amplamente discutido na literatura; veja Pinheiro *et al.* (2001), Rosa *et al.* (2003) e Osorio *et al.* (2007).

É importante notar que o SMSN-LMM assume que os erros aleatórios $\boldsymbol{\epsilon}_i$ são simetricamente distribuídos, o parâmetro de assimetria $\boldsymbol{\lambda}$ incorpora a assimetria nos efeitos aleatórios $\mathbf{b}_i$ e consequentemente em $\mathbf{Y}_i$. Assim, a densidade marginal de $\mathbf{Y}_i$, na qual a inferência é tipicamente baseada, é dada por

$$f(\mathbf{y}_i|\boldsymbol{\theta}) = 2\int_0^\infty \phi_{n_i}(\mathbf{y}_i|\mathbf{X}_i\boldsymbol{\beta}, \kappa(u_i)\boldsymbol{\Psi}_i)\Phi\left(\kappa^{-1/2}(u_i)\bar{\boldsymbol{\lambda}}_i^\top\boldsymbol{\Psi}_i^{-1/2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right) dH(u_i; \boldsymbol{\nu}),$$

i.e., $\mathbf{Y}_i \overset{\text{ind}}{\sim} SMSN_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Psi}_i, \bar{\boldsymbol{\lambda}}_i; H)$, $i = 1, \ldots, n$, onde $\boldsymbol{\Psi}_i = \sigma_e^2\mathbf{R}_i + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top$, $\boldsymbol{\Lambda}_i = \sigma_e^2(\sigma_e^2\mathbf{D}^{-1} + \mathbf{Z}_i^\top\mathbf{R}_i^{-1}\mathbf{Z}_i)^{-1}$ e $\bar{\boldsymbol{\lambda}}_i = \dfrac{\boldsymbol{\Psi}_i^{-1/2}\mathbf{Z}_i\mathbf{D}\boldsymbol{\zeta}}{\sqrt{1+\boldsymbol{\zeta}^\top\boldsymbol{\Lambda}_i\boldsymbol{\zeta}}}$, com $\boldsymbol{\zeta} = \mathbf{D}^{-1/2}\boldsymbol{\lambda}$. Assim, a função log-verossimilhança para $\boldsymbol{\theta}$ baseada nos dados observados $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$ é dada por $\ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^n \log\left(f(\mathbf{y}_i|\boldsymbol{\theta})\right)$, onde $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma_e^2, \boldsymbol{\alpha}^\top, \boldsymbol{\lambda}^\top)^\top \in \mathbb{R}^h$, com $h = p+q(q+1)/2+1+q$. Como a função log-verossimilhança observada envolve expressões complexas, não é atrativo trabalhar diretamente com $\ell(\boldsymbol{\theta}|\mathbf{y})$ tanto no processo de estimação como na análise de influência local. Para o SMSN-LMM, um algoritmo tipo-EM foi desenvolvido por Lachos *et al.* (2010) para obter o estimador de máxima verossimilhança de $\boldsymbol{\theta}$. Neste processo de estimação, $\mathbf{b}$, $\mathbf{u}$ e $\mathbf{t}$ são considerados hipoteticamente como dados faltantes e o vetor de dados completos é dado por $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{b}^\top, \mathbf{u}^\top, \mathbf{t}^\top)^\top$, onde $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$ e $\mathbf{t} = (t_1, \ldots, t_n)^\top$. Assim, sob a representação hierárquica (5.1)–(5.4), o algoritmo tipo-EM é aplicado na função log-verossimilhança dos dados completos $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}|\mathbf{y}_c)$, dada por

$$\begin{aligned}\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) \;=\; \sum_{i=1}^n \Bigg[& -\frac{1}{2}\log|\sigma_e^2\mathbf{R}_i| - \frac{\kappa^{-1}(u_i)}{2\sigma_e^2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top\mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i) \\ & -\frac{1}{2}\log|\boldsymbol{\Gamma}| - \frac{\kappa^{-1}(u_i)}{2}(\mathbf{b}_i - \boldsymbol{\Delta}t_i)^\top\boldsymbol{\Gamma}^{-1}(\mathbf{b}_i - \boldsymbol{\Delta}t_i)\Bigg] + c,\end{aligned}$$

onde c é uma constante que independe do vetor de parâmetros $\boldsymbol{\theta}$. Para o valor atual de $\boldsymbol{\theta}$, o passo E do algoritmo tipo-EM requer a avaliação de $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \mathrm{E}[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}] = \sum_{i=1}^m Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$, onde o valor esperado é considerado com respeito à distribuição condicional conjunta de $\mathbf{b}$, $\mathbf{u}$ e $\mathbf{t}$ dado $\mathbf{y}$ e $\widehat{\boldsymbol{\theta}}$. Dessa forma, temos que

$$Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = Q_{1i}(\boldsymbol{\beta}, \sigma_e^2|\widehat{\boldsymbol{\theta}}) + Q_{2i}(\boldsymbol{\alpha}, \boldsymbol{\lambda}|\widehat{\boldsymbol{\theta}}),$$

onde

$$\begin{aligned} Q_{1i}(\boldsymbol{\beta}, \sigma_e^2|\widehat{\boldsymbol{\theta}}) &= -\frac{1}{2}\log|\sigma_e^2\mathbf{R}_i| - \frac{\widehat{u}_i}{2\sigma_e^2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top\mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) \\ &\quad + \frac{1}{\sigma_e^2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top\mathbf{R}_i^{-1}\mathbf{Z}_i\widehat{\mathbf{ub}}_i - \frac{1}{2\sigma_e^2}\mathrm{tr}\left(\mathbf{R}_i^{-1}\mathbf{Z}_i\widehat{\mathbf{ub^2}}_i\mathbf{Z}_i^\top\right), \qquad (5.5) \\ Q_{2i}(\boldsymbol{\alpha}, \boldsymbol{\lambda}|\widehat{\boldsymbol{\theta}}) &= -\frac{1}{2}\log|\boldsymbol{\Gamma}| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\widehat{\mathbf{ub^2}}_i\right) + \boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}\widehat{\mathbf{utb}}_i \\ &\quad - \frac{\widehat{ut^2}_i}{2}\boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta}, \qquad (5.6) \end{aligned}$$

com $\widehat{u}_i$, $\widehat{\mathbf{ub}}_i$, $\widehat{\mathbf{ub^2}}_i$, $\widehat{\mathbf{utb}}_i$ e $\widehat{ut^2}_i$ $i = 1, \ldots, n$, todas definidas em Lachos *et al.* (2010) para os casos skew-normal, skew-t, skew-slash e skew-normal contaminada. Neste trabalho, supomos que o parâmetro associado com a distribuição da variável de mistura U é conhecido; veja, por exemplo, Lange *et al.* (1989); Berkane *et al.* (1994) e Osorio *et al.* (2007). Neste caso, a função log-verossimilhança perfilada e o critério de informação de Schwarz podem ser usados para determinar o valor ótimo de $\boldsymbol{\nu}$.

O passo M requer a maximização de $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ com respeito a $\boldsymbol{\theta}$. A motivação para considerar o algoritmo tipo-EM, no processo de estimação dos parâmetros de interesse, é que ele pode ser utilizado eficientemente para obter expressões com forma fechada nos passos CM. Neste capítulo, como nosso interesse não é o método de estimação, sugerimos ao leitor interessado ver Lachos *et al.* (2010) para mais detalhes do algoritmo tipo-EM.

Lachos *et al.* (2010) também desenvolveu um estimador de Bayes empírico para os efeitos aleatórios $\mathbf{b}_i$ que pode ser expresso como

$$\mathbf{b}_i(\boldsymbol{\theta}) = \boldsymbol{\mu}_{b_i} + \frac{\eta_{-1i}}{\sqrt{1 + \boldsymbol{\zeta}^\top\boldsymbol{\Lambda}_i\boldsymbol{\zeta}}}\boldsymbol{\Lambda}_i\boldsymbol{\zeta},$$

onde $\boldsymbol{\mu}_{b_i} = \mathbf{D}\mathbf{Z}_i^\top(\sigma_e^2\mathbf{R}_i + \mathbf{Z}_i^\top\mathbf{D}\mathbf{Z}_i)^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) = (\mathbf{D}^{-1} + \frac{1}{\sigma_e^2}\mathbf{Z}_i^\top\mathbf{R}_i^{-1}\mathbf{Z}_i)^{-1}\frac{1}{\sigma_e^2}\mathbf{Z}_i^\top\mathbf{R}_i^{-1}$

$(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$, $\eta_{-1i} = E\left\{\kappa^{1/2}(U_i)\dfrac{\phi(\kappa^{-1/2}(U_i)A_{\mathbf{y}_i})}{\Phi(\kappa^{-1/2}(U_i)A_{\mathbf{y}_i})}|\mathbf{y}_i\right\}$ (veja Proposição 2.2.1) e $A_{\mathbf{y}_i} = \bar{\boldsymbol{\lambda}}_i^\top \boldsymbol{\Psi}_i^{-1/2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$, com $\boldsymbol{\Lambda}_i$ e $\boldsymbol{\zeta}$ definidos em (5.5). Como em Pinheiro *et al.* (2001), substituindo $\boldsymbol{\theta}$ e $\mathbf{b}_i$ por suas respectivas estimativas, obtemos a seguinte decomposição da distância de Mahalanobis $d_i(\boldsymbol{\theta}) = d_i = (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top \boldsymbol{\Psi}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$ para a classe de distribuições SMSN:

$$\begin{aligned} d_i(\widehat{\boldsymbol{\theta}}) &= (\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}})^\top(\widehat{\sigma_e^2}\mathbf{R}_i + \mathbf{Z}_i^\top\widehat{\mathbf{D}}\mathbf{Z}_i)^{-1}(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}) \\ &= \frac{1}{\widehat{\sigma_e^2}}\widehat{\mathbf{e}}_i^\top\mathbf{R}_i^{-1}\widehat{\mathbf{e}}_i + \widehat{\boldsymbol{\mu}}_{b_i}^\top\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{\mu}}_{b_i} = \widehat{d_{\mathbf{e}_i}} + \widehat{d_{\mathbf{b}_i}}, \end{aligned} \tag{5.7}$$

onde $\widehat{\mathbf{e}}_i = \mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}} - \mathbf{Z}_i\widehat{\boldsymbol{\mu}}_{b_i}$ são os resíduos estimados. Eq. (5.7) fornece uma forma simples de calcular a distância de Mahalanobis. As distâncias estimadas d_i, $d_{\mathbf{b}_i}$ e $d_{\mathbf{e}_i}$ são ferramentas de diagnóstico úteis para identificar observações atípicas, como discutido em Pinheiro *et al.* (2001). Dessa forma, a decomposição de d_i apresentada em (5.7) para a classe de distribuições SMSN representa uma extensão de alguns resultados propostos por Pinheiro *et al.* (2001) para o caso t-Student.

5.3 Influência Local

Nesta seção, discutimos dignóstico de influência com ênfase na metodologia de influência local proposta por Zhu & Lee (2001). A seguir vamos obter a curvatura normal para o SMSN-LMM considerando a formulação hierárquica dada em (5.1)–(5.4). Os detalhes do cálculo da hessiana $\ddot{Q}_\theta(\boldsymbol{\theta}) = \dfrac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}$ e da matriz $\boldsymbol{\nabla}_{\boldsymbol{\omega}} = \dfrac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}$ para diferentes esquemas de perturbação são dados abaixo.

5.3.1 Matriz Hessiana

Seja $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma_e^2, \boldsymbol{\alpha}^\top, \boldsymbol{\lambda}^\top)^\top$ o vetor de parâmetros de interesse. A reparametrização $\mathbf{D} = \mathbf{F}^2$ foi considerada para facilitar a obtenção das derivadas teóricas, onde $\mathbf{F}$ é a raiz quadrada de $\mathbf{D}$. Assim, a matriz hessiana é dada por $\ddot{Q}_\theta(\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^n \ddot{Q}_i(\boldsymbol{\theta})$ com

$$\ddot{Q}_i(\boldsymbol{\theta}) = -\frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \begin{pmatrix} \ddot{Q}_{1i}(\boldsymbol{\beta}, \sigma_e^2) & \mathbf{0} \\ \mathbf{0} & \ddot{Q}_{2i}(\boldsymbol{\alpha}, \boldsymbol{\lambda}) \end{pmatrix},$$

onde as matrizes

$$\ddot{Q}_{1i}(\boldsymbol{\beta}, \sigma_e^2) = -\frac{\partial^2 Q_{1i}(\boldsymbol{\beta}, \sigma_e^2|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}^\top} \text{ com } \boldsymbol{\tau} = (\boldsymbol{\beta}^\top, \sigma_e^2)^\top \text{ e}$$
$$\ddot{Q}_{2i}(\boldsymbol{\alpha}, \boldsymbol{\lambda}) = -\frac{\partial^2 Q_{2i}(\boldsymbol{\alpha}, \boldsymbol{\lambda}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\pi}\partial\boldsymbol{\pi}^\top} \text{ com } \boldsymbol{\pi} = (\boldsymbol{\alpha}^\top, \boldsymbol{\lambda}^\top)^\top,$$

têm seus elementos apresentados no Apêndice D.

5.3.2 Esquemas de Perturbação

Nesta seção, apresentamos a matriz $\boldsymbol{\nabla}_{\boldsymbol{\omega}_o}$ considerando os esquemas de perturbação abordados para o modelo de regressão linear no capítulo anterior. Para cada esquema de perturbação, temos a forma particionada

$$\boldsymbol{\nabla}_{\boldsymbol{\omega}_o} = (\boldsymbol{\nabla}_\beta^\top, \boldsymbol{\nabla}_{\sigma_e^2}^\top, \boldsymbol{\nabla}_\alpha^\top, \boldsymbol{\nabla}_\lambda^\top)^\top,$$

para $\boldsymbol{\nabla}_{\boldsymbol{\tau}} = (\boldsymbol{\nabla}_{1\boldsymbol{\tau}}^\top, \ldots, \boldsymbol{\nabla}_{g\boldsymbol{\tau}}^\top)^\top$, $\boldsymbol{\tau} = \boldsymbol{\beta}, \boldsymbol{\sigma}_e^2, \boldsymbol{\alpha}$ ou $\boldsymbol{\lambda}$ e $\boldsymbol{\nabla}_{l\boldsymbol{\tau}} = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\tau}\partial\omega_l}|\boldsymbol{\omega}_o, l = 1, \ldots, g$, onde $\boldsymbol{\nabla}_\beta = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^\top}|\boldsymbol{\omega}_o \in \mathbb{R}^{p\times g}$, $\boldsymbol{\nabla}_{\sigma_e^2} = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\sigma_e^2\partial\boldsymbol{\omega}^\top}|\boldsymbol{\omega}_o \in \mathbb{R}^{1\times g}$, $\boldsymbol{\nabla}_\alpha = (\boldsymbol{\nabla}_{\alpha_1}, \ldots, \boldsymbol{\nabla}_{\alpha_{p^*}})$, com $\boldsymbol{\nabla}_{\alpha_r} = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\alpha_r\partial\boldsymbol{\omega}^\top}|\boldsymbol{\omega}_o \in \mathbb{R}^{1\times g}$, $r = 1, \ldots, p^*$ e $\boldsymbol{\nabla}_\lambda = (\boldsymbol{\nabla}_{\lambda_1}, \ldots, \boldsymbol{\nabla}_{\lambda_q})$, com $\boldsymbol{\nabla}_{\lambda_k} = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\lambda_k\partial\boldsymbol{\omega}^\top}|\boldsymbol{\omega}_o \in \mathbb{R}^{1\times g}$, $k = 1, \ldots, q$, $p^* = dim(\boldsymbol{\alpha})$ e g sendo a dimensão do vetor de perturbação $\boldsymbol{\omega}$.

Ponderação de Casos

Considere a inclusão de pesos no valor esperado da função log-verossimilhança dos dados completos (função-Q perturbada) como

$$Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \omega_i \mathrm{E}[\ell_i(\boldsymbol{\theta}|\mathbf{y}_c)] = \sum_{i=1}^{n} w_i Q_{1i}(\boldsymbol{\beta}, \sigma_e^2|\widehat{\boldsymbol{\theta}}) + \sum_{i=1}^{n} w_i Q_{2i}(\boldsymbol{\alpha}, \boldsymbol{\lambda}|\widehat{\boldsymbol{\theta}}), \quad (5.8)$$

onde $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ é um vetor de dimensão $n \times 1$ com $0 \leq \omega_i \leq 1$ para $i = 1, \ldots, n$, $\boldsymbol{\omega}_o = (1, \ldots, 1)^\top$, $Q_{1i}(\boldsymbol{\beta}, \sigma_e^2|\widehat{\boldsymbol{\theta}})$ e $Q_{2i}(\boldsymbol{\alpha}, \boldsymbol{\lambda}|\widehat{\boldsymbol{\theta}})$ são como apresentadas em (5.5) e (5.6), respectivamente. Além disso, note que esta perturbação geral definida em (5.8) pode ser decomposta em duas sub-perturbações definidas por (i) $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \omega_i Q_{1i}(\boldsymbol{\beta}, \sigma_e^2|\widehat{\boldsymbol{\theta}}) + \sum_{i=1}^{n} Q_{2i}(\boldsymbol{\alpha}, \boldsymbol{\lambda}|\widehat{\boldsymbol{\theta}})$ and (ii) $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} Q_{1i}(\boldsymbol{\beta}, \sigma_e^2|\widehat{\boldsymbol{\theta}}) + \sum_{i=1}^{n} \omega_i Q_{2i}(\boldsymbol{\alpha}, \boldsymbol{\lambda}|\widehat{\boldsymbol{\theta}})$. Para este esquema de perturbação, encontramos

$$\begin{aligned}
\boldsymbol{\nabla}_{i\beta} &= \frac{1}{\sigma_e^2}\mathbf{X}_i^\top \mathbf{R}_i^{-1}\left(\widehat{u}_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) - \mathbf{Z}_i\widehat{\mathbf{ub}_i}\right), \\
\boldsymbol{\nabla}_{i\sigma_e^2} &= -\frac{n_i}{2\sigma_e^2} + \frac{1}{2\sigma_e^4}\mathrm{tr}\left(\mathbf{R}_i^{-1}\mathbf{M}_i\right), \\
\boldsymbol{\nabla}_{i\alpha_r} &= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\alpha_r)\right) + \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\alpha_r)\boldsymbol{\Gamma}^{-1}\mathbf{N_i}\right) \\
&\quad + \boldsymbol{\delta}^\top\dot{\mathbf{F}}(r)\boldsymbol{\Gamma}^{-1}(\widehat{\mathbf{utb}_i} - \widehat{ut^2}_i\boldsymbol{\Delta}), \\
\boldsymbol{\nabla}_{i\lambda_k} &= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\lambda_k)\right) + \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\lambda_k)\boldsymbol{\Gamma}^{-1}\mathbf{N_i}\right) \\
&\quad + \dot{\boldsymbol{\delta}}(k)\mathbf{F}\boldsymbol{\Gamma}^{-1}(\widehat{\mathbf{utb}_i} - \widehat{ut^2}_i\boldsymbol{\Delta}),
\end{aligned}$$

onde $\dot{\boldsymbol{\Gamma}}(\alpha_r) = \dot{\mathbf{F}}(r)\mathbf{F} + \mathbf{F}\dot{\mathbf{F}}(r) - \dot{\mathbf{F}}(r)\boldsymbol{\delta}\boldsymbol{\delta}^\top\mathbf{F} - \mathbf{F}\boldsymbol{\delta}\boldsymbol{\delta}^\top\dot{\mathbf{F}}(r)$, $\dot{\mathbf{F}}(r) = \dfrac{\partial \mathbf{F}}{\partial \alpha_r}$, $r = 1, \ldots, \dim(\boldsymbol{\alpha})$, $\dot{\boldsymbol{\Gamma}}(\lambda_k) = -\mathbf{F}\dot{\mathbf{G}}(k)\mathbf{F}$, $\dot{\mathbf{G}}(k) = \dfrac{\partial \boldsymbol{\delta}\boldsymbol{\delta}^\top}{\partial \lambda_k}$, $\dot{\boldsymbol{\delta}}(k) = \dfrac{\partial \boldsymbol{\delta}}{\partial \lambda_k}$, $k = 1, \ldots, q$, $\mathbf{M}_i = \widehat{u}_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top - 2\mathbf{Z}_i\widehat{\mathbf{ub}}_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top + \mathbf{Z}_i\widehat{\mathbf{ub}^2}_i\mathbf{Z}_i^\top$, $\mathbf{N_i} = \widehat{\mathbf{ub}_i^2} - \widehat{\mathbf{utb}}_i\boldsymbol{\Delta}^\top - \boldsymbol{\Delta}\widehat{\mathbf{utb}}_i^\top + \widehat{ut_i^2}\boldsymbol{\Delta}\boldsymbol{\Delta}^\top$ e $\mathbf{I}_{n_i}$ denota a matriz identidade $n_i \times n_i$, $i = 1, \ldots, n$. De Osorio (2006), temos a equivalência entre uma medida de influência obtida por eliminação de casos e a técnica de influência local para este esquema de perturbação.

Perturbação na Matriz Escala dos Efeitos Aleatórios

Este esquema de perturbação é introduzido considerando que o efeito aleatório

$$\mathbf{b}_i \sim SMSN_q(\mathbf{0}, \frac{\mathbf{D}}{w_i}, \boldsymbol{\lambda}; H), i = 1, \ldots, n.$$

Sob este esquema de perturbação, o modelo postulado é obtido quando $\boldsymbol{\omega}_o = (1, \ldots, 1)^\top \in \mathbb{R}^n$. Além disso, a função-Q perturbada é dada por

$$Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} Q_{1i}(\boldsymbol{\beta}, \sigma_e^2|\widehat{\boldsymbol{\theta}}) + \sum_{i=1}^{n} Q_{2i}(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \omega_i|\widehat{\boldsymbol{\theta}}),$$

onde $Q_{2i}(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \omega_i|\widehat{\boldsymbol{\theta}})$ é como em (5.6) alterando $\boldsymbol{\Delta}(\omega_i) = \omega_i^{-1/2}\boldsymbol{\Delta}$ e $\boldsymbol{\Gamma}(\omega_i) = \omega_i^{-1}\boldsymbol{\Gamma}$ com $\boldsymbol{\Delta}$ e $\boldsymbol{\Gamma}$, respectivamente. A matriz $\boldsymbol{\nabla}_{\boldsymbol{\omega}_o}$ tem elementos dados por

$$\begin{aligned}
\boldsymbol{\nabla}_{i\beta} &= \mathbf{0}, \quad \boldsymbol{\nabla}_{i\sigma_e^2} = 0, \\
\boldsymbol{\nabla}_{i\alpha_r} &= \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\alpha_r)\boldsymbol{\Gamma}^{-1}\mathbf{C}_i\right) + \frac{1}{2}\boldsymbol{\delta}^\top\dot{\mathbf{F}}(r)\boldsymbol{\Gamma}^{-1}\widehat{\mathbf{utb}}_i, \\
\boldsymbol{\nabla}_{i\lambda_k} &= \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\lambda_k)\boldsymbol{\Gamma}^{-1}\mathbf{C}_i\right) + \frac{1}{2}\dot{\boldsymbol{\delta}}(k)\mathbf{F}\boldsymbol{\Gamma}^{-1}\widehat{\mathbf{utb}}_i,
\end{aligned}$$

onde $C_i = \widehat{\mathbf{ub}_i^2} - \widehat{\mathbf{utb}}_i\boldsymbol{\Delta}^\top, i = 1, \ldots, n$, $r = 1, \ldots, dim(\boldsymbol{\alpha})$ e $k = 1, \ldots, q$.

Perturbação na Variável Explicativa

Podemos investigar o mau condicionamento entre algumas colunas da matriz de planejamento $\mathbf{X}_i$, por exemplo, perturbando uma particular variável explicativa. Sob esta condição, temos a seguinte matriz de desenho perturbada $\mathbf{X}_i(\boldsymbol{\omega}) = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iu}(\omega_i), \ldots, \mathbf{x}_{ip})$, onde $\mathbf{x}_{iu}(\omega_i) = \mathbf{x}_{iu} + \omega_i\mathbf{1}_{n_i}, u = 1, \ldots, p$, $\mathbf{x}_{iu}$ é a u-ésima coluna da matriz $\mathbf{X}_i$ e $\mathbf{1}_{n_i}$ é um vetor $n_i \times 1$ de uns. Assim, a função-Q perturbada é como em (5.5) e (5.6) substituindo $\mathbf{X}_{i\boldsymbol{\omega}}$ com $\mathbf{X}_i$ e o modelo postulado é obtido com $\boldsymbol{\omega}_o = \mathbf{0} \in \mathbb{R}^n$. Sob este

esquema de perturbação $\boldsymbol{\nabla}_{\boldsymbol{\omega}_o}$ tem os seguintes elementos

$$\begin{aligned}
\boldsymbol{\nabla}_{i\beta} &= \frac{\widehat{u}_i}{\sigma_e^2}\mathbf{B}_i^\top\mathbf{R}_i^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right) - \frac{\widehat{u}_i}{\sigma_e^2}\mathbf{X}_i^\top\mathbf{R}_i^{-1}\mathbf{B}_i\boldsymbol{\beta} - \frac{1}{\sigma_e^2}\mathbf{B}_i^\top\mathbf{R}_i^{-1}\mathbf{Z}_i\widehat{\mathbf{ub}}_i,\\
\boldsymbol{\nabla}_{i\sigma_e^2} &= -\frac{1}{\sigma_e^4}\boldsymbol{\beta}^\top\mathbf{B}_i^\top\mathbf{R}_i^{-1}\left(\widehat{u}_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) - \mathbf{Z}_i\widehat{\mathbf{ub}}_i\right),\\
\boldsymbol{\nabla}_{i\alpha_r} &= 0, \quad \boldsymbol{\nabla}_{i\lambda_k} = 0,
\end{aligned}$$

onde $\mathbf{B}_i = \dfrac{\partial\mathbf{X}_i(\boldsymbol{\omega})}{\partial\omega_i} = (\mathbf{0}_{n_i},\ldots,\mathbf{0}_{n_i},\mathbf{1}_{n_i},\mathbf{0}_{n_i},\ldots,\mathbf{0}_{n_i})$ denota uma matriz $n_i \times p$, $i = 1,\ldots,n$.

Perturbação na Variável Resposta

Considere uma perturbação aditiva para o vetor de respostas observadas $(\mathbf{y}_1^\top,\ldots,\mathbf{y}_n^\top)^\top$, isto é, adicionamos uma pequena perturbação às respostas de cada unidade experimental, obtendo $\mathbf{y}_i(\boldsymbol{\omega}) = \mathbf{y}_i + \omega_i\mathbf{1}_{n_i}$, onde $\mathbf{1}_{n_i}$ é um vetor $n_i \times 1$ de uns, $i = 1,\ldots,n$. A função-Q perturbada é como em (5.5) e (5.6) alterando $\mathbf{y}_i(\boldsymbol{\omega})$ com $\mathbf{y}_i$. Sob este esquema de perturbação o vetor $\boldsymbol{\omega}_o$ é dado por $\boldsymbol{\omega}_o = \mathbf{0} \in \mathbb{R}^{\mathbf{n}}$ e $\boldsymbol{\nabla}_{\boldsymbol{\omega}_o}$ tem os seguintes elementos

$$\begin{aligned}
\boldsymbol{\nabla}_{i\beta} &= \frac{\widehat{u}_i}{\sigma_e^2}\mathbf{X}_i^\top\mathbf{R}_i^{-1}\mathbf{1}_{n_i},\\
\boldsymbol{\nabla}_{i\sigma_e^2} &= \frac{1}{\sigma_e^4}\left(\widehat{u}_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) - \mathbf{Z}_i\widehat{\mathbf{ub}}_i\right)^\top\mathbf{R}_i^{-1}\mathbf{1}_{n_i},\\
\boldsymbol{\nabla}_{i\alpha_r} &= 0, \quad \boldsymbol{\nabla}_{i\lambda_k} = 0,
\end{aligned}$$

$r = 1,\ldots,dim(\boldsymbol{\alpha})$, $k = 1,\ldots,q$ e $i = 1,\ldots,n$.

Na próxima seção, um exemplo real será apresentado com o intuito de ilustrar a metodologia desenvolvida.

5.4 Aplicação: Conjunto de Dados "Framingham Cholesterol"

Para propósitos ilustrativos, aplicamos as técnicas apresentadas para o conjunto de dados "Framingham Cholesterol". Estes dados foram originalmente relatados por Zhang & Davidian (2001) e já foram analisados por Arellano-Valle *et al.* (2005), Lachos *et al.* (2007a) e Lin & Lee (2008) sob a distribuição skew-normal. Lachos *et al.* (2010) também consideraram o conjunto de dados "Framingham Cholesterol" e notaram que as distribuições SMSN com caudas pesadas são mais adequadas aos dados do que a distribuição skew-normal. Dessa forma, revisitamos este conjunto de dados com o intuito de aplicar a metodologia de influência local de Zhu & Lee (2001). O conjunto de dados inclui os níveis de colesterol no tempo, idade e gênero de $n = 200$ indivíduos selecionados aleatoriamente. Como em Zhang & Davidian (2001), assumimos o modelo linear misto

$$Y_{ij} = \beta_o + \beta_1 \text{sex}_i + \beta_2 \text{age}_i + \beta_3 t_{ij} + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij} \, , \tag{5.9}$$

onde Y_{ij} é o nível de colesterol (mg/DL), dividido por 100, no j-ésimo tempo para o indivíduo i, t_{ij} é (tempo-5)/10, com o tempo medido em anos desde o início do estudo, age_i é a idade ao início do estudo e sex_i é o indicador do gênero (0 = feminino, 1 = masculino). O número de repetições j de cada indivíduo varia de 1 a 6, ou seja, trata-se de um conjunto de dados desbalanceado.

O exemplo ilustrativo está organizado como segue. Primeiro, ajustamos o SMSN-LMM para o conjunto de dados "Framingham Cholesterol". Apesar de não serem testes formais, como em Zhang & Davidian (2001), comparamos o SMSN-LMM e o SMN-LMM (especificamente, os modelos normal, t-Student, slash e normal contaminada) inspecionando alguns critérios de informação. Em seguida, identificamos as observações

influentes deste conjunto de dados considerando a quantidade $M(0)$, obtida através da curvatura conformal $B_{f_Q,\mathbf{u}}$ e os esquemas de perturbação descritos na Seção 5.3.2. A Tabela 5.1 mostra as estimativas de máxima verossimilhança para os parâmetros dos quatro modelos (SN-LMM, ST-LMM, SCN-LMM e SSL-LMM), acompanhadas de seus respectivos desvios padrões assintóticos estimados, calculados via a matriz informação observada. Como nosso interesse não é o processo de estimação dos parâmetros do modelo linear misto, sugerimos ao leitor interessado Lachos *et al.* (2010).

Tabela 5.1: Dados "Framingham Cholesterol". Estimativas de máxima verossimilhança para os quatro modelos SMSN selecionados. (d_{11}, d_{12}, d_{22}) são os elementos distintos da matriz $\mathbf{D}^{1/2}$. Os valores SE são os desvios padrões assintóticos estimados.

	SN–LMM		ST–LMM		SCN–LMM		SSL–LMM	
Parâmetro	Estimativa	SE	Estimativa	SE	Estimativa	SE	Estimativa	SE
β_o	1.3520	0.1502	1.3888	0.1311	1.4045	0.1396	1.4089	0.1433
β_1	-0.0488	0.0509	-0.0548	0.0447	-0.0461	0.0468	-0.0430	0.0482
β_2	0.0152	0.0033	0.0149	0.0029	0.0144	0.0030	0.0140	0.0031
β_3	0.3562	0.0667	0.3641	0.0611	0.4006	0.0630	0.3998	0.0638
σ_e^2	0.0430	0.0017	0.0325	0.0025	0.0264	0.0028	0.0228	0.0025
d_{11}	0.5261	0.0474	0.4417	0.0477	0.4079	0.0541	0.3918	0.0472
d_{12}	0.0018	0.0302	-0.0030	0.0305	-0.0246	0.0290	-0.0232	0.0277
d_{22}	0.2166	0.0330	0.2035	0.0370	0.2099	0.0386	0.1953	0.0353
λ_1	13.8050	4.2423	13.7822	4.4242	13.4875	4.6855	14.1171	4.7110
λ_2	-6.3654	4.3984	-8.0691	3.9867	-8.7607	4.0621	-8.4215	4.2099
ν	-	-	8.1799	-	0.2981	-	2.0898	-
γ	-	-	-	-	0.3345	-	-	-

De acordo com Lange *et al.* (1989), obtemos o valor de $\boldsymbol{\nu}$ maximizando a função verossimilhança perfilada como segue. Suponha $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, onde $\boldsymbol{\theta}_1$ é o parâmetro de interesse e $\boldsymbol{\theta}_2$ é o parâmetro nuisance, então a log-verossimilhança perfilada de $\boldsymbol{\theta}_1$ é $l_p(\boldsymbol{\theta}_1) = max_{\boldsymbol{\theta}_2} l(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. Usando este procedimento, encontramos $\nu = 8.1799$ para a

ST, temos $\nu = 2.0898$ para a SSL e obtemos $\nu = 0.2981$ e $\gamma = 0.3345$ para a SCN.

A Tabela 5.2 mostra o intervalo de confiança de 95% baseado na aproximação normal do estimador de máxima verossimilhança (CI: $\lambda_i \pm 1.96$ SE, $i = 1, 2$). Note que apenas o intervalo de confiança de λ_2 do SN-LMM ajustado inclui o valor zero. Entretanto, nossa experiência em estimar a variância assintótica do estimador de máxima verossimilhança de $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^\top$ indica que ele é tipicamente superestimado; veja Montenegro *et al.* (2009b) para uma discussão detalhada sobre este assunto.

Tabela 5.2: Dados "Framingham Cholesterol". Intervalo de confiança de 95% (CI, baseado na aproximação normal do estimador de máxima verossimilhança).

Modelos	CI	
	λ_1	λ_2
SN-LMM	[5.4901; 22.1199]	[-14.9863; 2.2555]
ST-LMM	[5.1108; 22.4536]	[-15.8824; -0.2558]
SCN-LMM	[4.3039; 22.6711]	[-16.7224; -0.7990]
SCN-LMM	[4.8835; 23.3507]	[-16.6729; -0.1701]

De acordo com Montenegro *et al.* (2009b), acreditamos que um intervalo de confiança mais preciso para λ_2 pode ser obtido usando a estatística da razão de verossimilhanças (LR) e a verossimilhança perfilada. Um intervalo de confiança de aproximadamente 95% (PCI) para λ_2, por exemplo, pode ser construído considerando $2\left\{l(\widehat{\boldsymbol{\theta}}) - l_p(\lambda_2)\right\} < \chi^2_{0.95,1}$, onde $l_p(\lambda_2)$ é a função log-verossimilhança perfilada. Dessa forma, o PCI de λ_2 para o SN-LMM é [-11;-5], veja a Figura 5.1. Note que o PCI para o parâmetro λ_2 proporciona fortes evidências de que ele é menor do que zero, diferente das evidências fornecidas pelo intervalo de confiança simétrico CI (Tabela 5.2). Estes resultados favorecem a evidência de que os modelos assimétricos são mais apropriados para o conjunto

de dados "Framingham Cholesterol".

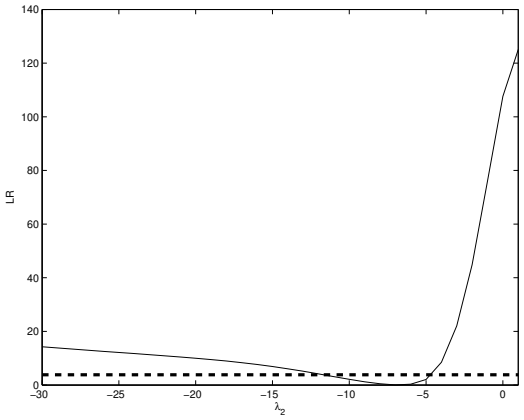


Figura 5.1: Dados "Framingham Cholesterol". Razão de verossimilhanças (LR) baseada na verossimilhança perfilada. A linha no gráfico é a linha limite de 95% ($\chi^2_{0.95,1}$).

Em seguida, comparamos o SMSN-LMM e o SMN-LMM inspecionando os critérios de informação descritos na Seção 1.3. Comparando os modelos simétricos e assimétricos através dos valores dos critérios de informação dados na Tabela 5.3, notamos que os modelos assimétricos SMSN-LMM apresentam um melhor ajuste do que os modelos simétricos SMN-LMM. Além disso, as distribuições ST, SCN e SSL apresentam um melhor ajuste do que o SN-LMM.

Ademais, assim como Montenegro *et al.* (2009b), usamos a estatística da razão de verossimilhanças para testar a hipótese nula $H_0 : \boldsymbol{\lambda} = \mathbf{0}$, i.e., para testar se o modelo simétrico é suficiente. A implementação desta estatística é computacionalmente simples, apesar de requerer as estimativas de máxima verossimilhança de $\boldsymbol{\theta}$ sob a hipótese nula (modelo simétrico) e sob o modelo irrestrito (modelo assimétrico). Note que quando $\boldsymbol{\lambda} = \mathbf{0}$, as equações dos passos CM, provenientes do algoritmo tipo-EM, se reduzem às equações quando assumimos as distribuições SMN; veja Lachos *et al.* (2010). Este

Tabela 5.3: Dados "Framingham Cholesterol". Alguns critérios de informação.

Modelos	$l(\hat{\boldsymbol{\theta}})$	AIC	BIC	HQ
Normal	-236.9451	244.9451	258.1384	250.2842
t-Student	-206.2968	215.2968	230.1392	221.3033
Slash	-207.2307	216.2307	231.0731	222.2372
Normal Contaminada	-201.7945	211.7945	228.2861	218.4684
SN	-152.0090	162.0090	178.5006	168.6829
ST	-127.4155	138.4155	156.5562	145.7568
SSL	-130.2786	141.2786	159.4193	148.6199
SCN	-125.9170	137.9170	157.7069	145.9257

procedimento de testes é importante, pois uma vez que há suspeita de que $\boldsymbol{\lambda}$ seja muito próximo de zero, a matriz informação observada $\mathbf{J}$ pode ser singular, fato provado apenas para modelos mais simples; veja, por exemplo, DiCiccio & Monti (2004). Podemos observar na Tabela 5.4 que há evidências suficientes nos dados para rejeitar H_0 para todos os casos e por isso as distribuições SMSN são mais apropriadas para este exemplo.

Tabela 5.4: Dados "Framingham Cholesterol". Resultados da estatística da razão de verossimilhanças para testar a hipótese de interesse, H_0.

Modelos	Valor	P-valor
Normal/SN-LMM	169.8722	0.0000
t-Student/ST-LMM	157.7626	0.0000
Slash/SSL-LMM	153.9042	0.0000
Normal Contaminada/SCN-LMM	151.7550	0.0000

Para exemplificar nossa proposta metodológica no contexto de SMSN-LMM, conduzimos a seguir análise de diagnóstico de influência. Neste estudo, assumimos apenas os modelos skew-normal, skew-t, skew-slash e skew-normal contaminada para propósitos comparativos. Com o intuito de detectar observações atípicas, primeiro usamos a distância de Mahalanobis $d_i = (\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}})^\top \boldsymbol{\Psi}^{-1}(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}})$, $i = 1, \ldots, 200$; veja, por exemplo, Pinheiro *et al.* (2001) e Osorio *et al.* (2007). Como trata-se de um

conjunto de dados desbalanceado, a Figura 5.2 mostra tal distância para os quatro modelos ajustados, mas sem apresentar uma marca de referência que facilita a identificação de "outliers" (veja Proposição 2.2.8). Pode-se notar que as observações $8, 26, 69, 74, 90, 111, 122, 138, 146, 160, 162, 174, 175$ e 187 destacam-se como "outliers".

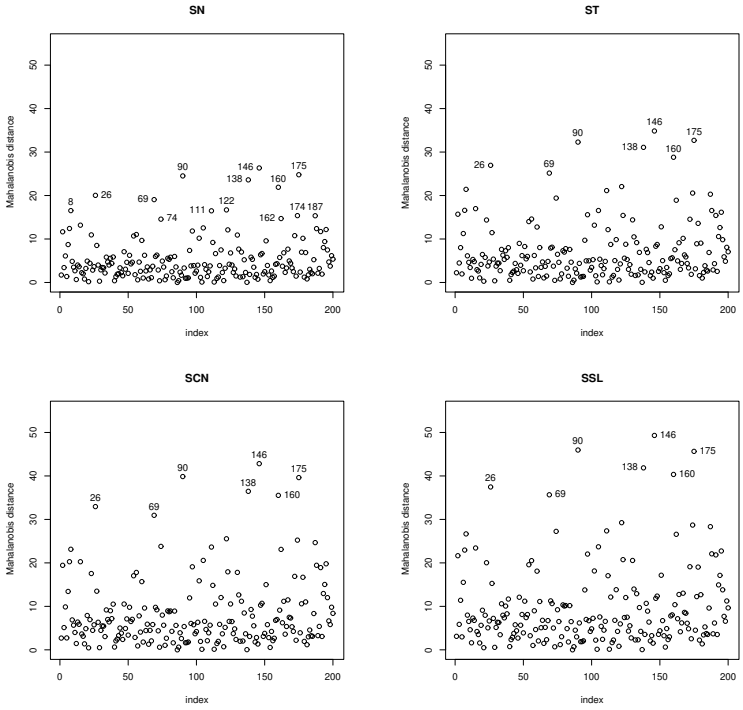


Figura 5.2: Dados "Framingham Cholesterol". Gráficos da distância de Mahalanobis para os quatro modelos ajustados.

As distâncias estimadas $d_{\mathbf{e}_i}$ (Error) e $d_{\mathbf{b}_i}$ (Random Effect–R.E.), obtidas considerando (5.7), são estatísticas de diagnóstico úteis para identificar observações anômalas. A Figura 5.3 apresenta tais distâncias para o SN-LMM. As observações $69, 90, 138, 145$ e 175 possuem valores grandes de $d_{\mathbf{e}_i}$, sugerindo que são **e**-"outliers". Além disso, as observações $8, 26$ e 160 apresentam valores grandes de $d_{\mathbf{b}_i}$ sugerindo que são **b**-"outliers".

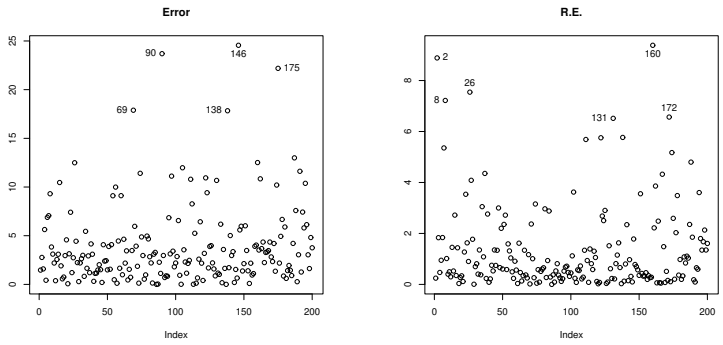


Figura 5.3: Dados "Framingham Cholesterol". Decomposição da distância de Mahalanobis: $d_{\mathbf{e}i}$ (error) e $d_{\mathbf{b}_i}$ (R.E.) estimadas pelo ajuste SN.

O gráfico de $d_{\mathbf{b}_i}$ também fornece evidências de que as observações $2, 131$ e 172 são **b**-"outliers", fato que não pode ser concluído pela Figura 5.2. Para as distribuições SMSN com caudas pesadas, observamos os mesmos resultados e portanto não serão mostrados aqui por brevidade.

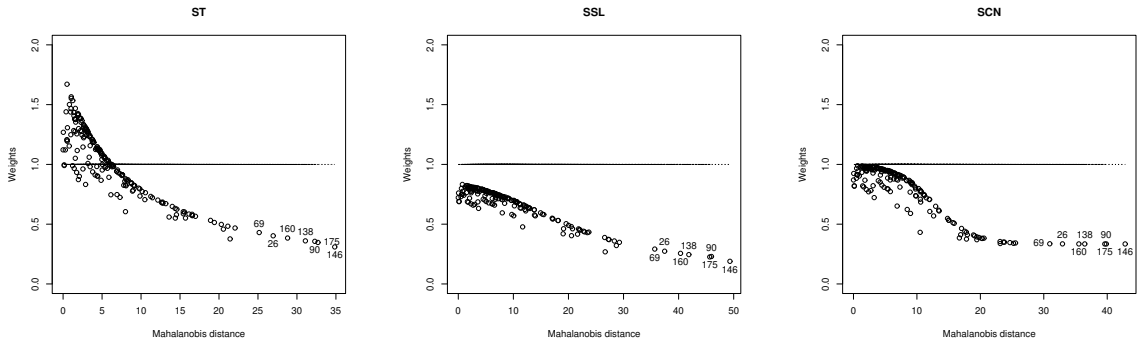


Figura 5.4: Dados "Framingham Cholesterol". Valores estimados de u_i para os modelos ST, SSL e SCN.

Quando usamos as distribuições com caudas mais pesadas que a skew-normal, o

algoritmo EM permite acomodar as observações discrepantes atribuindo menor peso no processo de estimação. Os pesos para a distribuição skew-normal ($u_i = 1, i = 1, \ldots, 200$) estão indicados na Figura 5.4 como uma linha contínua. Note por esta figura que nas distribuições ST, SSL e SCN, u_i é inversamente proporcional à distância de Mahalanobis. Assim, esta rica classe de distribuições pode naturalmente atribuir pesos diferentes para cada observação e consequentemente controla a influência da observação nas estimativas dos parâmetros. Estes resultados concordam com as considerações apresentadas por Osorio *et al.* (2007), no contexto simétrico.

Em seguida, conduzimos um estudo de influência local baseado na quantidade $M(0)$ com interesse em $\boldsymbol{\theta}$. Em análises preliminares, consideramos $c^* = 3$ para obter as marcas de referência. Entretanto, muitas observações foram identificadas como sendo influentes para esta escolha de c^*. Dessa forma, aqui nós apresentamos a análise de influência local considerando $c^* = 5$ para calcular a marca de referência.

Perturbação ponderação de casos: Na Figura 5.5 nota-se que sob os quatro modelos ajustados, a observação 39 é identificada como influente. Como esperado, a influência de tal observação é reduzida quando consideramos distribuições com caudas mais pesadas que a skew-normal.

Perturbação na matriz escala $\mathbf{D}$*:* Na Figura 5.6 nota-se que sob o SN-LMM, o ST-LMM e o SSL-LMM, as observações $2, 8, 26, 131$ e 172 (detectadas como **b**-"outliers" na Figura 5.3) são identificadas como influentes. Além disso, a observação 167 é identificada como influente sob o SCN-LMM.

Perturbação na variável explicativa: As estimativas de máxima verossimilhança são

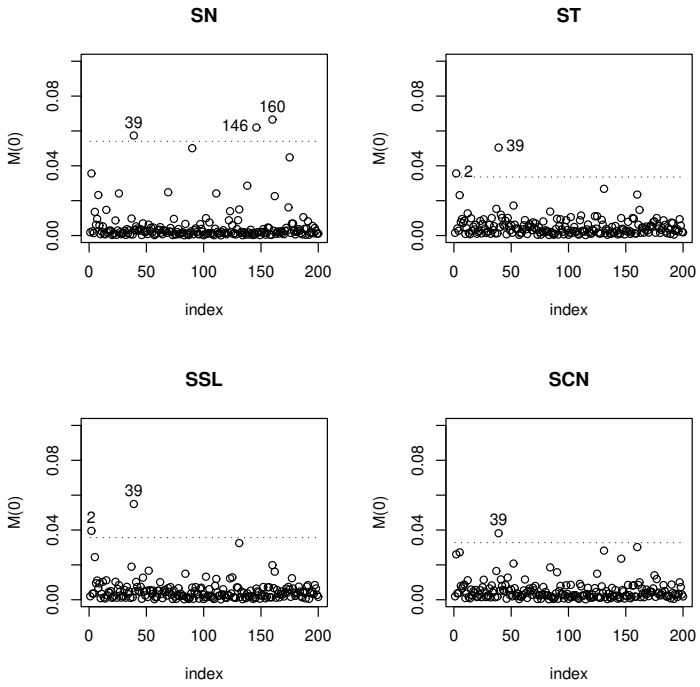


Figura 5.5: Dados "Framingham Cholesterol". Gráficos de M(0) sob a perturbação ponderação de casos para os quatro modelos ajustados. A linha horizontal delimita a marca de referência de Lee & Xu (2004) para M(0) com $c* = 5$.

estáveis sob a perturbação na variável explicativa para os quatro modelos considerados, como pode ser visto na Figura 5.7.

Perturbação na variável resposta: Da Figura 5.8 nota-se que apenas sob o modelo skew-normal, a observação 90 (detectada como "outlier" na Figura 5.2) é identificada como influente.

É importante notar que como esperado, a influência das observações é reduzida quando consideramos ditribuições com caudas mais pesadas que a skew-normal. Para

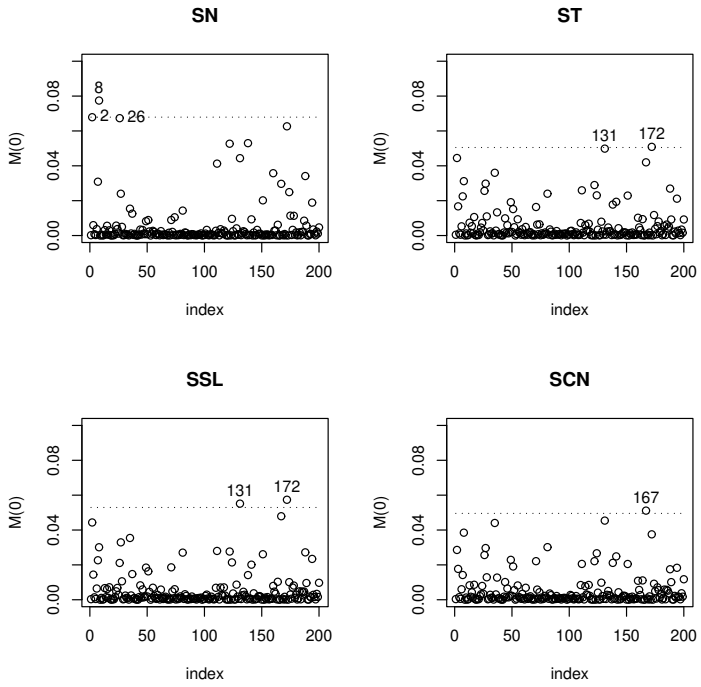


Figura 5.6: Dados "Framingham Cholesterol". Gráficos de M(0) sob a perturbação na matriz escala $\mathbf{D}$ para os quatro modelos ajustados. A linha horizontal delimita a marca de referência de Lee & Xu (2004) para M(0) com $c* = 5$.

este conjunto de dados, a distribuição skew-normal contaminada acomoda ligeiramente melhor as observações influentes.

Como sugerido por Lee *et al.* (2006), usamos as quantidades TRC e MRC para demonstrar o impacto das observações influentes detectadas na estimativa de máxima verossimilhança de $\boldsymbol{\beta}$, por exemplo. Elas são definidas por

$$TRC = \sum_{j=1}^{n_p} \mid \frac{\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\beta}}_{[i]j}}{\hat{\boldsymbol{\beta}}_j} \mid \quad \text{e} \quad MRC = max_{j=1,\ldots,n_p} \mid \frac{\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\beta}}_{[i]j}}{\hat{\boldsymbol{\beta}}_j} \mid,$$

onde n_p é a dimensão de $\boldsymbol{\beta}$ e o subscrito $[i]$ indica a estimativa de máxima verossimi-

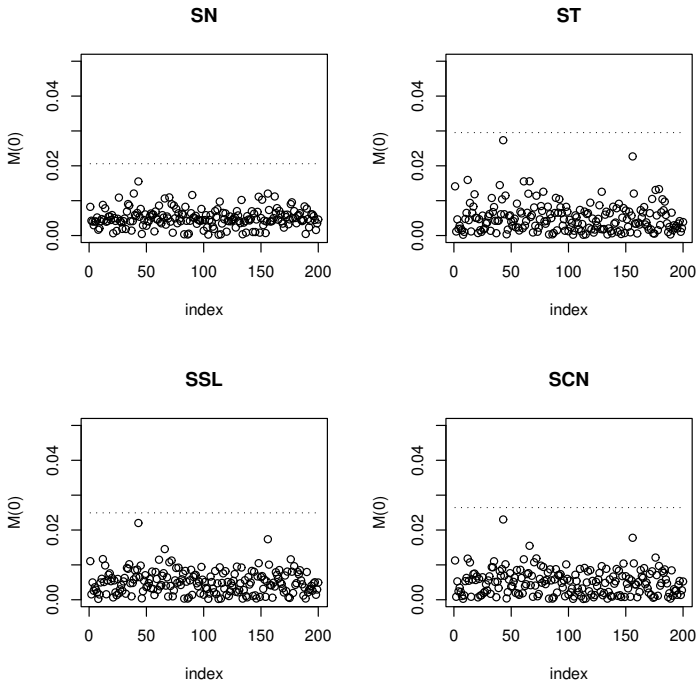


Figura 5.7: Dados "Framingham Cholesterol". Gráficos de M(0) sob a perturbação nas variáveis explicativas para os quatro modelos ajustados. A linha horizontal delimita a marca de referência de Lee & Xu (2004) para M(0) com $c* = 5$.

lhança de $\boldsymbol{\beta}$ com a i-ésima observação, y_i, excluída. A comparação dessas medidas, calculadas para os diferentes modelos quando a observação 90 é excluída, está apresentada na Tabela 5.5.

É interessante notar que as maiores modificações ocorrem sob a distribuição skew-normal. Como esperado, os resultados mostram que as estimativas de máxima verossimilhança são menos sensíveis à presença de dados atípicos e/ou influentes quando distribuições com caudas mais pesadas que a skew-normal são usadas, fato também observado nos quatro esquemas de perturbação considerados nesta ilustração.

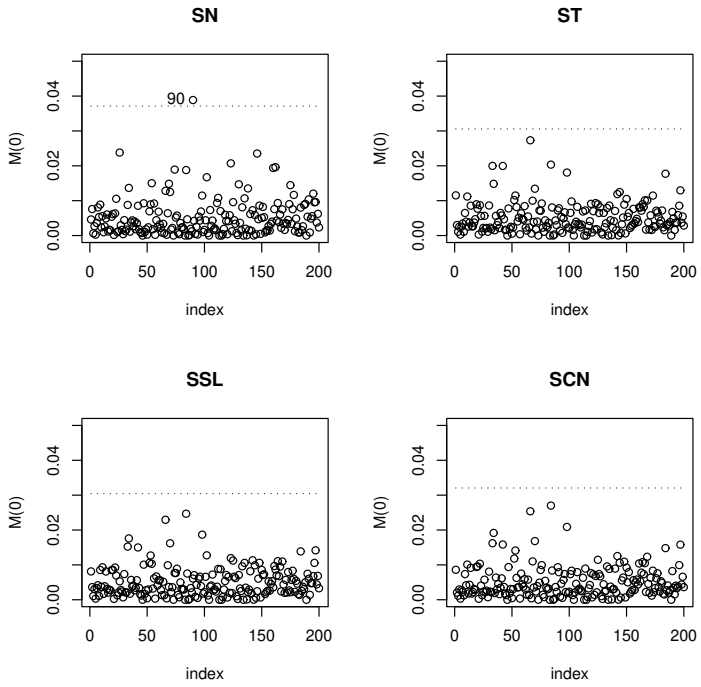


Figura 5.8: Dados "Framingham Cholesterol". Gráficos de M(0) sob a perturbação na variável resposta para os quatro modelos ajustados. A linha horizontal delimita a marca de referência de Lee & Xu (2004) para M(0) com $c* = 5$.

Tabela 5.5: Dados "Framingham Cholesterol". Comparação das mudanças relativas nas estimativas de máxima verossimilhança em termos de TRC e MRC para os quatro modelos SMSN selecionados.

Observação	Distribuição	TRC	MRC
90	SN	5.0506	4.0769
	ST	1.3903	1.2667
	SSL	2.0862	1.2126
	SCN	1.7453	0.6908

A flexibilidade dos modelos ST e SCN às observações atípicas pode ser estudada através da influência de um único "outlier" sobre as estimativas de máxima verossimilhança de $\boldsymbol{\theta}$, como discutido em Pinheiro *et al.* (2001). Por simplicidade, vamos avaliar quanto as estimativas de $\boldsymbol{\theta}$ são influenciadas pela modificação de $\Im$ unidades na observação y_{ik}. Dessa forma, substituímos a observação y_{ik} por $y_{ik}(\Im) = y_{ik} + \Im$ e anotamos a mudança relativa nas estimativas considerando $\left((\widehat{\theta}(\Im) - \widehat{\theta})/\widehat{\theta}\right)$, onde $\widehat{\theta}$ denota a estimativa original e $\widehat{\theta}(\Im)$ a estimativa para os dados contaminados. Neste exemplo, contaminamos a quinta observação do indivíduo 26 e variamos $\Im$ entre -1 e 1 por incrementos de 0.1. Na Figura 5.9, apresentamos os resultados das mudanças relativas na estimativa de $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top$, para diferentes contaminações de $\Im$, sob o SN-LMM, o ST-LMM e o SCN-LMM. Como esperado, os modelos ST e SCN são menos afetados pelas variações de $\Im$ do que o modelo SN.

5.5 Observações Finais

Neste capítulo, apresentamos estratégias para avaliar diagnóstico de influência em modelos lineares mistos sob a classe de distribuições SMSN. Diferentes esquemas de perturbação foram investigados para este propósito. Expressões explícitas foram obtidas para a hessiana $\ddot{\mathbf{Q}}$ e para a matriz $\boldsymbol{\nabla}_{\boldsymbol{\omega}_o}$ sob quatro esquemas de perturbação apropriados. Os resultados obtidos generalizam os encontrados por Lesaffre & Verbeke (1998) e Montenegro *et al.* (2009a) para uma ampla classe de modelos lineares mistos com estrutura longitudinal que inclui assimetria e caudas pesadas. Dessa forma, acreditamos que a metodologia desenvolvida possa apresentar resultados satisfatórios em outros modelos multivariados, por exemplo, no modelo de Grubbs. Tal modelo pode ser visto como um caso particular do modelo linear com efeitos mistos proposto por Verbeke &

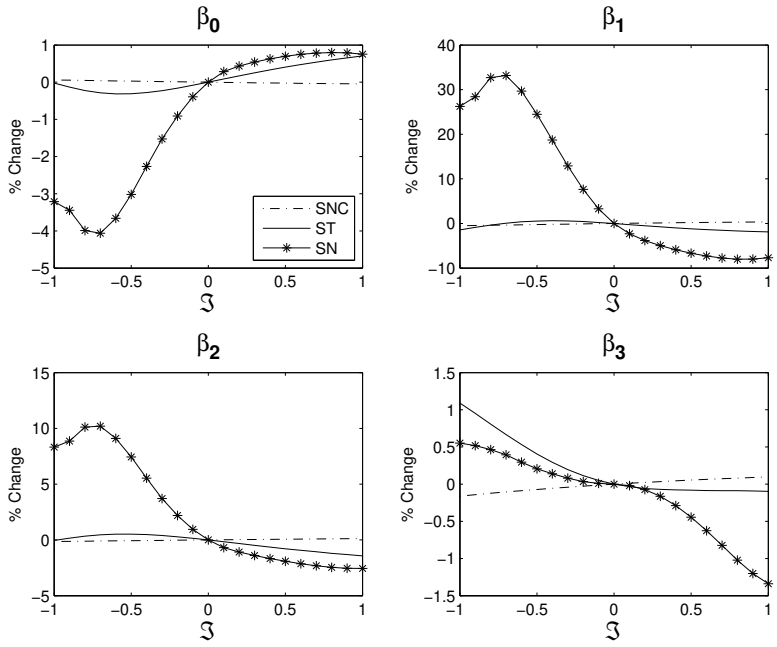


Figura 5.9: Dados "Framingham Cholesterol". Mudanças relativas nas estimativas de máxima verossimilhança de $\beta_0, \beta_1, \beta_2$ e β_3 dos ajustes SN-LMM, ST-LMM e SCN-LMM para diferentes contaminações de $\Im$ na quinta observação do indivíduo 26. %change= $100 \times \left((\widehat{\theta}(\Im) - \widehat{\theta})/\widehat{\theta}\right)$, onde $\widehat{\theta}$ denota a estimativa original e $\widehat{\theta}(\Im)$ a estimativa para os dados contaminados.

Molenberghs (2001) (veja página 118) e é definido como

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij},$$

em que μ representa a média geral, α_i são os efeitos aleatórios e ϵ_{ij} são os erros de medição, onde α_i e ϵ_{ij} são independentes e seguem distribuição normal com média zero e variâncias constantes σ_c^2 e σ^2, respectivamente. Contudo, desenvolvemos estimação e diagnóstico de influência para o modelo de Grubbs considerando-o como um caso particular do modelo de Barnett (1969). Os resultados estão resumidos no Capítulo 6.

Capítulo 6

Modelo de Grubbs Misturas de Escala Skew-Normal

Grubbs (1983) propôs um modelo, denominado modelo de Grubbs, para comparar diversos instrumentos de medição e usualmente, é comum assumir que os termos aleatórios seguem distribuição normal (ou simétrica). Neste capítulo, discutimos a extensão deste modelo sob a classe de distribuições misturas de escala skew-normal, onde as distribuições skew-normal, skew-t, skew-slash e skew-normal contaminada são membros particulares dessa família. Portanto, essas distribuições fornecem uma generalização do modelo de Grubbs simétrico introduzido por Osorio *et al.* (2009) e do modelo assimétrico skew-normal discutido por Montenegro *et al.* (2009b) e Montenegro *et al.* (2009c). Nesta classe flexível de modelos, discutimos um algoritmo tipo-EM para estimação dos parâmetros e o método de influência local (Zhu & Lee, 2001) para avaliar a sensibilidade das estimativas dos parâmetros sob esquemas de perturbações usuais.

Os resultados e métodos desenvolvidos neste capítulo serão ilustrados através de um exemplo numérico.

6.1 Introdução

O modelo de Grubbs é tipicamente usado para comparar a qualidade de vários instrumentos de medição, utilizados para medir a mesma quantidade de interesse desconhecida em um grupo comum de unidades experimentais. O problema da comparação de instrumentos de medição cujas características diferem em custo, velocidade e outros fatores, tais como eficiência, tem recebido uma crescente atenção em áreas como engenharia, medici-na, agricultura, entre outras; veja por exemplo, Barnett (1969), Grubbs (1983) e Fuller (1987). Zeller (2006) propõe o modelo de Grubbs em grupos e Lachos *et al.* (2007b) estuda o modelo de Grubbs, ambos sob a suposição de normalidade, evidenciando o fato bem conhecido da falta de robustez das estimativas de mínimos quadrados contra observações extremas. Para superar esta deficiência, recentemente, Osorio *et al.* (2009) propuseram o modelo de Grubbs sob as distribuições misturas de escala normal (modelo SMN-G). Nessa linha de pesquisa, veja Andrews & Mallows (1974), Lange & Sinsheimer (1993) e Osorio (2006).

No contexto assimétrico, Montenegro *et al.* (2009b) propuseram o modelo de Grubbs skew-normal (modelo SN-G) e mostraram as vantagens de utilizar distribuições assimétricas nesta subclasse dos modelos com erros nas variáveis. Através do modelo SN-G, observamos uma grande flexibilidade de introduzir assimetria, porém também notamos que a sua flexibilidade contra observações extremas pode ser seriamente afetada por observações na cauda da distribuição. Neste capítulo, estendemos os modelos SMN-G e SN-G assumindo que a observações seguem uma distribuição na classe SMSN.

Esta extensão resulta em uma classe flexível de modelos com erros nas variáveis e foi motivada pelo fato de que muitos conjuntos de dados considerados na literatura apresentam um significante comportamento não normal, tais como assimetria e caudas pesadas. Este é o caso do conjunto de dados estudado por Barnett (1969) que considerou uma transformação nos dados com o intuito de obter uma melhor aproximação da distribuição normal (ou, pelo menos, uma distribuição simétrica).

Estudos de influência local nos modelos com erros nas variáveis sob normalidade foram considerados por Zeller (2006), Lachos *et al.* (2007b) e De Castro *et al.* (2007), por exemplo. Osorio *et al.* (2009) desenvolveram análise de influência local, baseada na abordagem de Zhu & Lee (2001), para o modelo SMN-G. Além disso, estudos envolvendo a distribuição assimétrica skew-normal foram apresentados, por exemplo, por Lachos *et al.* (2008), Montenegro *et al.* (2009a), Montenegro *et al.* (2009b) e Montenegro *et al.* (2009c). Dessa forma, inspirados por Montenegro *et al.* (2009b) que estudou análise de influência local no contexto do modelo SN-G, neste capítulo, aplicamos a abordagem de influência local de Zhu & Lee (2001) no modelo SMSN-G.

O capítulo está organizado como segue. Na Seção 6.2, discutimos a formulação do modelo bem como alguns aspectos inferenciais, incluindo o algoritmo tipo-EM e o cálculo da matriz informação observada. Na Seção 6.3, a metodologia de influência local para o modelo SMSN-G, considerando três esquemas de perturbação, é apresentada. Os resultados desenvolvidos serão ilustrados na Seção 6.4 usando o famoso conjunto de dados de Barnett (1969). Finalmente, algumas observações finais são dadas na Seção 6.5.

6.2 O Modelo Proposto

Suponha que temos a nossa disposição $p \geq 2$ instrumentos para medir uma característica de interesse x em um grupo de n unidades experimentais. Considere x_i o valor da covariável não observado (verdadeiro) para i-ésima unidade experimental e Y_{ij} a resposta obtida com o instrumento j na unidade i, $i = 1, \ldots, n$ e $j = 1, \ldots, p$. Relacionando estas variáveis, temos o seguinte modelo proposto por Grubbs (1983):

$$\mathbf{Y}_i = \mathbf{a} + \mathbf{1}_p x_i + \boldsymbol{\epsilon}_i, \tag{6.1}$$

$i = 1, \ldots, n$, onde $\mathbf{a} = (0, \boldsymbol{\alpha}^\top)^\top = (0, \alpha_2, \ldots, \alpha_p)^\top$ e $\mathbf{1}_p = (1, \ldots, 1)^\top$ são vetores de dimensão $p \times 1$, sendo que α_i representa o vício aditivo do instrumento i. Neste caso, supomos que o primeiro instrumento é a referência ($\alpha_1 = 0$) e este será comparado com os restantes $p-1$ instrumentos. Além disso, considere que $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ip})^\top$ e $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{ip})^\top$ (o vetor de erros) são vetores aleatórios independentes de dimensão $p \times 1$, onde a suposição clássica é que $\boldsymbol{\epsilon}_i \overset{\text{iid}}{\sim} N_p(0, D(\boldsymbol{\phi}))$ e $x_i \overset{\text{iid}}{\sim} N_1(\mu_x, \phi_x)$, tal que $D(\boldsymbol{\phi}) = \text{diag}\{\boldsymbol{\phi}\}$ denota a matriz diagonal com $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^\top$. Neste capítulo, fornecemos uma generalização do modelo de Grubbs, assumindo que

$$\mathbf{r}_i = \begin{pmatrix} x_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \overset{\text{iid}}{\sim} SMSN_{p+1}\left(\begin{pmatrix} \mu_x \\ \mathbf{0} \end{pmatrix}, D(\phi_x, \boldsymbol{\phi}), \begin{pmatrix} \lambda_x \\ \mathbf{0} \end{pmatrix}; H \right), \tag{6.2}$$

$i = 1, \ldots, n$, onde $D(\phi_x, \boldsymbol{\phi}) = \text{diag}\{\phi_x, \phi_1, \ldots, \phi_p\}$. O modelo definido em (6.1)-(6.2) será chamado de modelo SMSN-G estrutural. De (2.4), esta formulação implica que

$$\mathbf{r}_i | U_i = u_i \overset{\text{ind}}{\sim} SN_{p+1}\left(\begin{pmatrix} \mu_x \\ \mathbf{0} \end{pmatrix}, \kappa(u_i) D(\phi_x, \boldsymbol{\phi}), \begin{pmatrix} \lambda_x \\ \mathbf{0} \end{pmatrix} \right), \tag{6.3}$$

$$U_i \overset{\text{iid}}{\sim} H(\boldsymbol{\nu}), i = 1, \ldots, n. \tag{6.4}$$

Além disso, de Ghosh *et al.* (2009), segue que

$$\boldsymbol{\epsilon}_i \overset{\text{iid}}{\sim} SMN_p(\mathbf{0}, D(\boldsymbol{\phi}); H) \text{ e } x_i \overset{\text{iid}}{\sim} SMSN(\mu_x, \phi_x, \lambda_x; H).$$

Uma vez que para cada $i = 1, \ldots, n$, $\boldsymbol{\epsilon}_i$ e x_i são indexados pelo mesmo fator de mistura de escala U_i, eles não são independentes. A independência corresponde ao caso que H é degenerada, com $\kappa(U_i) = 1, \forall i$ então, o modelo SMSN-G reduz ao modelo SN-G definido em Montenegro *et al.* (2009b). Entretanto, como afirmado em (6.3)-(6.4), condicionado à U_i, $\boldsymbol{\epsilon}_i$ e x_i são independentes para cada $i = 1, \ldots, n$, o que implica que $\boldsymbol{\epsilon}_i$ e x_i são não correlacionados, uma vez que $Cov(\boldsymbol{\epsilon}_i x_i) = E[\boldsymbol{\epsilon}_i x_i | U_i] = 0$.

Note que os erros $\boldsymbol{\epsilon}_i$ correspondem aos erros de medição então, espera-se que sejam simetricamente distribuídos. O parâmetro de assimetria λ_x incorpora a assimetria na variável latente x_i e consequentemente nas quantidades $\mathbf{Y}_i$, as quais têm marginalmente distribuição SMSN multivariada. Se $\lambda_x = 0$ então, o modelo assimétrico reduz para o modelo SMN-G definido por Osorio *et al.* (2009). Como em Lange *et al.* (1989), Berkane *et al.* (1994) e Osorio *et al.* (2009), assumimos que $\boldsymbol{\nu}$ é conhecido. Assim, para o modelo SMSN-G proposto, o vetor de parâmetros é dado por $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top, \mu_x, \phi_x, \lambda_x)^\top = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, onde $\boldsymbol{\theta}_1 = (\boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top)^\top$ e $\boldsymbol{\theta}_2 = (\mu_x, \phi_x, \lambda_x)^\top$. Ressaltamos que na prática, o parâmetro $\boldsymbol{\theta}_1$ é, em geral, nosso interesse principal, uma vez que está associado ao vício aditivo e a precisão dos instrumentos de medição. Inferência clássica para o vetor de parâmetros neste modelo é baseada na distribuição marginal de $\mathbf{Y}_i$ que está apresentada na proposição seguinte.

Proposição 6.2.1 *Sob o modelo SMSN-G estrutural definido em (6.1)-(6.2), a distribuição marginal de* $\mathbf{Y}_i$ *é dada por*

$$f(\mathbf{y}_i|\boldsymbol{\theta}) = 2\int_0^\infty \phi_p(\mathbf{y}_i|\boldsymbol{\mu}, \kappa(u_i)\boldsymbol{\Sigma})\Phi(\kappa^{-1/2}(u_i)\bar{\boldsymbol{\lambda}}_x^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}))dH(u_i;\boldsymbol{\nu}), \qquad (6.5)$$

onde

$$\boldsymbol{\mu} = \mathbf{a} + \mathbf{1}_p\mu_x, \;\; \boldsymbol{\Sigma} = \phi_x\mathbf{1}_p\mathbf{1}_p^\top + D(\boldsymbol{\phi}) \;\; e \;\; \bar{\boldsymbol{\lambda}}_x = \frac{\lambda_x\phi_x\boldsymbol{\Sigma}^{-1/2}\mathbf{1}_p}{\sqrt{\phi_x + \lambda_x^2\Lambda_x}},$$

com $\Lambda_x = \phi_x/c$ *e* $c = 1 + \phi_x\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi})\mathbf{1}_p$.

Prova: Veja Proposição 1 em Montenegro *et al.* (2009b) e substitua a matriz escala $\boldsymbol{\Sigma}$ por $\kappa(u)\boldsymbol{\Sigma}$.

De (6.5), é simples observar que $\mathbf{Y}_i$ são independentes e marginalmente distribuídos com $\mathbf{Y}_i \overset{\text{iid}}{\sim} SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \bar{\boldsymbol{\lambda}}_x; H)$, onde $\bar{\boldsymbol{\lambda}}_x = \dfrac{\lambda_x \phi_x \boldsymbol{\Sigma}^{-1/2}\mathbf{1}_p}{\sqrt{\phi_x + \lambda_x^2 \Lambda_x}}$ e $\boldsymbol{\Sigma}^{-1/2}$ é a raiz quadrada de $\boldsymbol{\Sigma}^{-1}$, sendo $\boldsymbol{\Sigma}$ positiva definida. Entretanto, note que o cálculo da raiz quadrada pode ser facilmente realizado usando o Matlab e não é necessário desenvolver uma metodologia neste trabalho.

Neste capítulo, para obter a estimativa de máxima verossimilhança de $\boldsymbol{\theta}$, consideramos o algoritmo tipo-EM e adicionalmente, derivamos a matriz de covariâncias assintótica dessas estimativas através da matriz informação observada. Os resultados são dados a seguir.

6.2.1 Algoritmo EM

Nesta seção, demonstraremos como utilizar o algoritmo tipo-EM para obter os estimadores de máxima verossimilhança dos parâmetros do modelo SMSN-G. Uma característica importante deste modelo é que ele pode ser formulado através de uma representação hierárquica, útil para derivações teóricas e para a implementação do algoritmo EM. De (2.4) segue que

$$\mathbf{Y}_i \mid x_i, U_i = u_i \overset{\text{ind}}{\sim} N_p(\mathbf{a} + \mathbf{1}_p x_i, \kappa(u_i) D(\boldsymbol{\phi})), \tag{6.6}$$

$$x_i \mid T_i = t_i, U_i = u_i \overset{\text{ind}}{\sim} N(\mu_x + \tau_x t_i, \kappa(u_i)\nu_x^2), \tag{6.7}$$

$$T_i | U_i = u_i \overset{\text{ind}}{\sim} HN(0, \kappa(u_i)), \tag{6.8}$$

$$U_i \overset{\text{iid}}{\sim} H(\boldsymbol{\nu}), \tag{6.9}$$

$i = 1, \ldots, n$, todos independentes, onde $HN(0, \sigma^2)$ denota a distribuição half-$N(0, \sigma^2)$, $\nu_x^2 = \phi_x(1 - \delta_x^2)$ e $\tau_x = \phi_x^{1/2}\delta_x$, com $\delta_x = \lambda_x/(1 + \lambda_x^2)^{1/2}$.

Neste processo de estimação, $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$ é o vetor de respostas observáveis para n unidades amostrais, $\mathbf{x} = (x_1, \ldots, x_n)^\top$, $\mathbf{u} = (u_1, \ldots, u_n)^\top$, $\mathbf{t} = (t_1, \ldots, t_n)^\top$ e $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\boldsymbol{\theta}}_1^{(k)\top}, \widehat{\boldsymbol{\theta}}_2^{(k)\top})^\top$ denota a estimativa de $\boldsymbol{\theta}$ na k-ésima iteração, onde $\widehat{\boldsymbol{\theta}}_1^{(k)} = (\widehat{\boldsymbol{\alpha}}^{(k)\top}, \widehat{\boldsymbol{\phi}}^{(k)\top})^\top$ e $\widehat{\boldsymbol{\theta}}_2^{(k)} = (\widehat{\mu}_x^{(k)}, \widehat{\phi}_x^{(k)}, \widehat{\lambda}_x^{(k)})^\top$. De (6.6)-(6.9), temos que a função log-verossimilhança completa associada com $\mathbf{y}_c = (\mathbf{y}, \mathbf{x}, \mathbf{t}, \mathbf{u})$ é dada por

$$\begin{aligned}\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) &= -\frac{n}{2}\log(|D(\boldsymbol{\phi})|) - \frac{1}{2}\sum_{i=1}^{n}\kappa^{-1}(u_i)(\mathbf{y}_i - \mathbf{a} - \mathbf{1}_p x_i)^\top D^{-1}(\boldsymbol{\phi})(\mathbf{y}_i - \mathbf{a} - \mathbf{1}_p x_i) \\ &\quad -\frac{n}{2}\log(\nu_x^2) - \frac{1}{2\nu_x^2}\sum_{i=1}^{n}\kappa^{-1}(u_i)(x_i - \mu_x - \tau_x t_i)^2 + c,\end{aligned}$$

onde c é uma constante que independe do vetor de parâmetros $\boldsymbol{\theta}$. Dada a estimativa atual $\widehat{\boldsymbol{\theta}}^{(k)}$, o passo E calcula $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = E[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}] = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$, onde $Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = Q_{1i}(\boldsymbol{\theta}_1|\widehat{\boldsymbol{\theta}}^{(k)}) + Q_{2i}(\boldsymbol{\theta}_2|\widehat{\boldsymbol{\theta}}^{(k)})$, com

$$\begin{aligned}Q_{1i}(\boldsymbol{\theta}_1|\widehat{\boldsymbol{\theta}}^{(k)}) &= -\frac{1}{2}\log(|D(\boldsymbol{\phi})|) - \frac{1}{2}\Big[\widehat{u}_i^{(k)}(\mathbf{y}_i - \mathbf{a})^\top D^{-1}(\boldsymbol{\phi})(\mathbf{y}_i - \mathbf{a}) \\ &\quad -2\widehat{ux}_i^{(k)}(\mathbf{y}_i - \mathbf{a})^\top D^{-1}(\boldsymbol{\phi})\mathbf{1}_p + \widehat{ux^2}_i^{(k)}\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi})\mathbf{1}_p\Big] \qquad (6.10) \\ Q_{2i}(\boldsymbol{\theta}_2|\widehat{\boldsymbol{\theta}}^{(k)}) &= -\frac{1}{2}\log(\nu_x^2) - \frac{1}{2\nu_x^2}\Big[\widehat{ux^2}_i^{(k)} + \mu_x^2\widehat{u}_i^{(k)} + \tau_x^2\widehat{ut^2}_i^{(k)} \\ &\quad -2\widehat{ux}_i^{(k)}\mu_x - 2\widehat{utx}_i^{(k)}\tau_x + 2\mu_x\tau_x\widehat{ut}_i^{(k)}\Big], \qquad (6.11)\end{aligned}$$

e requer o cálculo das seguintes expressões $\widehat{u}_i^{(k)} = E[\kappa^{-1}(U_i)|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i]$, $\widehat{ut}_i^{(k)} = E[\kappa^{-1}(U_i)t_i|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i]$, $\widehat{ut^2}_i^{(k)} = E[\kappa^{-1}(U_i)t_i^2|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i]$, $\widehat{ux}_i^{(k)} = E[\kappa^{-1}(U_i)x_i|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i]$, $\widehat{ux^2}_i^{(k)} = E[\kappa^{-1}(U_i)x_i^2|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i]$ e $\widehat{utx}_i^{(k)} = E[\kappa^{-1}(U_i)t_i x_i|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i]$, as quais podem ser avaliadas por

$$\begin{aligned}\widehat{ut}_i^{(k)} &= \widehat{u}_i^{(k)}\widehat{\mu}_{T_i}^{(k)} + \widehat{M}_T^{(k)}\widehat{\eta}_{1i}^{(k)}, \quad \widehat{ut^2}_i^{(k)} = \widehat{u}_i^{(k)}\widehat{\mu}_{T_i}^{2(k)} + \widehat{M}_T^{2(k)} + \widehat{M}_T^{(k)}\widehat{\mu}_{T_i}^{(k)}\widehat{\eta}_{1i}^{(k)}, \\ \widehat{utx}_i^{(k)} &= \widehat{r}_i^{(k)}\ \widehat{ut}_i^{(k)} + \widehat{s}^{(k)}\ \widehat{ut^2}_i^{(k)}, \quad \widehat{ux}_i^{(k)} = \widehat{r}_i^{(k)}\ \widehat{u}_i^{(k)} + \widehat{s}^{(k)}\ \widehat{ut}_i^{(k)}, \\ \widehat{ux^2}_i &= \widehat{T}_x^{2(k)} + \widehat{r}_i^{2(k)}\ \widehat{u}_i^{(k)} + 2\widehat{r}_i^{(k)}\widehat{s}^{(k)}\ \widehat{ut}_i^{(k)} + \widehat{s}^{2(k)}\ \widehat{ut^2}_i^{(k)},\end{aligned}$$

onde $\widehat{\eta}_{1i}^{(k)} = E[\kappa^{-1/2}(U_i)W_\Phi(\frac{\kappa^{-1/2}(U_i)\widehat{\mu}_{T_i}}{\widehat{M}_T})|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i]$, com $W_\Phi(\gamma) = \phi(\gamma)/\Phi(\gamma)$, $\gamma \in \mathbb{R}$, $\widehat{M}_T^2 = \widehat{c}_1/\widehat{c}$, $\widehat{\mu}_{T_i} = \frac{\widehat{\tau}_x}{c}\widehat{a}_i$, $\widehat{T}_x^2 = \widehat{\nu}_x{}^2/\widehat{c}_1$, $\widehat{r}_i = \widehat{\mu}_x + \widehat{T}_x^2\widehat{a}_i$, $\widehat{s} = \widehat{\tau}_x/\widehat{c}_1$ e $c_1 = 1 + \nu_x^2\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi})\mathbf{1}_p$, todas essas quantidades avaliadas em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$. Em cada passo, as esperanças condicionais $\widehat{u}_i = \widehat{u}_{1i}$ e $\widehat{\eta}_{1i}$ podem ser facilmente calculadas através dos resultados desenvolvidos na Seção 2.2 (veja Proposição 2.2.1).

Os passos CM condicionalmente maximizam $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ com respeito a $\boldsymbol{\theta}$, obtendo uma nova estimativa $\widehat{\boldsymbol{\theta}}^{(k+1)}$ e estão descritos abaixo.

Passo CM-1: Atualizar $\widehat{\boldsymbol{\alpha}}^{(k+1)}$ como

$$\widehat{\boldsymbol{\alpha}}^{(k+1)} = \overline{\mathbf{z}}_u^{(k)} - \overline{x}_u^{(k)}\mathbf{1}_{p-1}.$$

Passo CM-2: Fixar $\widehat{\boldsymbol{\alpha}}^{(k+1)}$ e atualizar $\widehat{\phi}_j^{(k+1)}$ como

$$\widehat{\phi}_j^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{u}_i^{(k)}(y_{ij} - \widehat{\alpha}_j^{(k+1)})^2 - 2\widehat{ux}_i^{(k)}(y_{ij} - \widehat{\alpha}_j^{(k+1)}) + \widehat{ux^2}_i^{(k)}\right).$$

Passo CM-3: Atualizar $\widehat{\tau}_x^{(k+1)}$ como

$$\widehat{\tau}_x^{(k+1)} = \frac{\sum_{i=1}^{n}(\widehat{utx}_i^{(k)} - \overline{x}_u^{(k)}\widehat{ut}_i^{(k)})}{\sum_{i=1}^{n}(\widehat{ut^2}_i^{(k)} - \overline{t}_u^{(k)}\widehat{ut}_i^{(k)})}].$$

Passo CM-4: Fixar $\widehat{\tau}_x^{(k+1)}$ e atualizar $\widehat{\mu}_x^{(k+1)}$ como

$$\widehat{\mu}_x^{(k+1)} = \overline{x}_u^{(k)} - \widehat{\tau}_x^{(k+1)}\overline{t}_u^{(k)}.$$

Passo CM-5: Fixar $\widehat{\tau}_x^{(k+1)}$ e $\widehat{\mu}_x^{(k+1)}$, atualizar $\widehat{\nu}_x^{2(k+1)}$ como

$$\widehat{\nu}_x^{2(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\widehat{R}_i^{(k,k+1)}.$$

Passo CM-6: Fixar $\widehat{\tau}_x^{(k+1)}$ e $\widehat{\nu}_x^{2(k+1)}$, atualizar $\widehat{\lambda}_x^{(k+1)}$ como

$$\widehat{\lambda}_x^{(k+1)} = \widehat{\tau}_x^{(k+1)}/\widehat{\nu}_x^{(k+1)}.$$

Passo CM-7: Fixar $\widehat{\tau}_x^{(k+1)}$ e $\widehat{\nu}_x^{2(k+1)}$, atualizar $\widehat{\phi}_x^{(k+1)}$ como

$$\widehat{\phi}_x^{(k+1)} = \widehat{\tau}_x^{2(k+1)} + \widehat{\nu}_x^{2(k+1)},$$

onde

$$\overline{\mathbf{z}}_u^{(k)} = \frac{\sum_{i=1}^n \widehat{u}_i^{(k)}\mathbf{z}_i}{\sum_{i=1}^n \widehat{u}_i^{(k)}}, \mathbf{z}_i = (y_{i2}, \ldots, y_{ip})^\top, \overline{x}_u^{(k)} = \frac{\sum_{i=1}^n \widehat{ux}_i^{(k)}}{\sum_{i=1}^n \widehat{u}_i^{(k)}}, \overline{t}_u^{(k)} = \frac{\sum_{i=1}^n \widehat{ut}_i^{(k)}}{\sum_{i=1}^n \widehat{u}_i^{(k)}},$$

$$\begin{aligned}\widehat{R}_i^{(k,k+1)} &= \widehat{ux^2}_i^{(k)} + \left[\widehat{\mu}_x^{(k+1)}\right]^2 \widehat{u}_i^{(k)} + \left[\widehat{\tau}_x^{(k+1)}\right]^2 \widehat{ut^2}_i^{(k)} \\ &- 2[\widehat{ux}_i^{(k)}\widehat{\mu}_x^{(k+1)} + \widehat{utx}_i^{(k)}\widehat{\tau}_x^{(k+1)} - \widehat{\mu}_x^{(k+1)}\widehat{\tau}_x^{(k+1)}\widehat{ut}_i^{(k)}] \quad \text{e} \quad j = 1, \ldots, p.\end{aligned}$$

As iterações do algoritmo tipo-EM devem ser repetidas até que uma regra de convergência seja satisfeita. Uma crítica muitas vezes expressa é que o procedimento do algoritmo tipo-EM tende a ficar preso em modas locais. Uma forma conveniente para contornar tais limitações é trabalhar com as iterações do tipo-EM com uma ampla variedade de valores iniciais que sejam representativos do espaço paramétrico. Se existem várias modas, podemos encontrar a moda global comparando os valores da log-verossimilhança. Note que quando $\kappa(U_i) = 1$, $i = 1, \ldots, n$ (uma variável aleatória degenerada), as equações dos passos CM reduzem às equações obtidas por Montenegro *et al.* (2009b) sob a distribuição skew-normal. Além disso, quando $\lambda_x = 0$ (ou $\tau_x = 0$), as equações dos passos CM são reduzidas para as equações obtidas por Osorio *et al.* (2009).

6.2.2 Matriz Informação Observada

Da Proposição 6.2.1, segue que a função log-verossimilhança para $\boldsymbol{\theta}$ dada a amostra observada $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$ é dada por

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}), \tag{6.12}$$

onde $\ell_i(\boldsymbol{\theta}) = \log 2 - \frac{p}{2} log 2\pi - \frac{1}{2}\log|\boldsymbol{\Sigma}| + \log K_i$, com

$$K_i = K_i(\boldsymbol{\theta}) = \int_0^\infty \kappa^{-p/2}(u)\exp\{-\frac{1}{2}\kappa^{-1}(u)d_i\}\Phi(\kappa^{-1/2}(u)A_i)dH(u;\boldsymbol{\nu}),$$

onde $d_i = (\mathbf{y}_i - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})$ e $A_i = \bar{\boldsymbol{\lambda}}_x^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}) = A_x a_i$, com $A_x = \frac{\lambda_x\Lambda_x}{\sqrt{\phi_x + \lambda_x^2\Lambda_x}}$ e $a_i = (\mathbf{y}_i - \boldsymbol{\mu})^\top D^{-1}(\boldsymbol{\phi})\mathbf{1}_p$. Dessa forma, a matriz de segundas derivadas com respeito a $\boldsymbol{\theta}$ é dada por

$$\mathbf{L} = \sum_{i=1}^{n}\frac{\partial^2\ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = -\frac{n}{2}\frac{\partial^2\log|\boldsymbol{\Sigma}|}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} - \sum_{i=1}^{n}\frac{1}{K_i^2}\frac{\partial K_i}{\partial\boldsymbol{\theta}}\frac{\partial K_i}{\partial\boldsymbol{\theta}^\top} + \sum_{i=1}^{n}\frac{1}{K_i}\frac{\partial^2 K_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top},$$

onde

$$\frac{\partial K_i}{\partial\boldsymbol{\theta}} = I_i^{\phi}(\frac{p+1}{2})\frac{\partial A_i}{\partial\boldsymbol{\theta}} - \frac{1}{2}I_i^{\Phi}(\frac{p+2}{2})\frac{\partial d_i}{\partial\boldsymbol{\theta}},$$

e

$$\begin{aligned}\frac{\partial^2 K_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} &= \frac{1}{4}I_i^{\Phi}(\frac{p+4}{2})\frac{\partial d_i}{\partial\boldsymbol{\theta}}\frac{\partial d_i}{\partial\boldsymbol{\theta}^\top} - \frac{1}{2}I_i^{\Phi}(\frac{p+2}{2})\frac{\partial^2 d_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\\ &- \frac{1}{2}I_i^{\phi}(\frac{p+3}{2})(\frac{\partial A_i}{\partial\boldsymbol{\theta}}\frac{\partial d_i}{\partial\boldsymbol{\theta}^\top} + \frac{\partial d_i}{\partial\boldsymbol{\theta}}\frac{\partial A_i}{\partial\boldsymbol{\theta}^\top}) - I_i^{\phi}(\frac{p+3}{2})A_i\frac{\partial A_i}{\partial\boldsymbol{\theta}}\frac{\partial A_i}{\partial\boldsymbol{\theta}^\top}\\ &+ I_i^{\Phi}(\frac{p+1}{2})\frac{\partial^2 A_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top},\end{aligned}$$

com $I_i^F(w) = \int_0^\infty \kappa^{-w}(u)\exp\{-\frac{1}{2}\kappa^{-1}(u)d_i\}F(\kappa^{-1/2}(u)A_i)dH(u)$, $w > 0$ onde $F(.)$ é a função $\Phi(.)$ ou $\phi(.)$. Note que podemos escrever $K_i = I_i^{\Phi}(\frac{p}{2})$. Expressões explícitas de $I_i^{\Phi}(w)$ e $I_i^{\phi}(w)$ para as distribuições skew-t, skew-slash e skew-normal contaminada

podem ser obtidas na Secão 2.2.2, considerando as definições de d_i e A_i descritas nesta seção. As derivadas de $\log \boldsymbol{\Sigma}$, d_i e A_i envolvem algumas manipulações algébricas e podem ser encontradas em Montenegro *et al.* (2009b). Dessa forma, $\mathbf{J} = -\mathbf{L}$ denota a matriz de informação observada para a log-verossimilhança marginal $\ell(\boldsymbol{\theta})$. Na prática, $\mathbf{J}$ é usualmente desconhecida e a substituímos por $\widehat{\mathbf{J}}$ que é a matriz $\mathbf{J}$ avaliada na estimativa de máxima verossimilhança de $\boldsymbol{\theta}$.

6.2.3 Estimação da Variável Latente

Nesta seção, consideramos técnicas bayesianas para obter as estimativas das quantidades x_i, denominadas de estimativas de Bayes empíricas de x_i. Uma vez que as variáveis latentes no modelo (6.1)-(6.2) são assumidas como aleatórias, é natural estimá-las usando técnicas bayesianas; veja por exemplo, Verbeke & Molenberghs (2001). Na prática, muitas vezes construímos histogramas e o gráfico Q-Q normal para as estimativas de Bayes empíricas de x_i com propósitos exploratórios, tais como verificar desvios da suposição de normalidade e assimetria (positiva ou negativa) presente nos dados.

Assim, a distribuição condicional de x_i dado $(\mathbf{Y}_i, U_i) = (\mathbf{y}_i, u_i)$ pertence à família de distribuições skew-nomal estendida (EST) proposta por Azzalini & Capitanio (1999) e sua pdf é dada por

$$f(x_i|\mathbf{y}_i, u_i, \boldsymbol{\theta}) = \frac{1}{\Phi(\kappa^{-1/2}(u_i)\mathrm{A}_i)}\, \phi(x_i|\mu_x + \Lambda_x a_i, \kappa(u_i)\Lambda_x)\Phi(\kappa^{-1/2}(u_i)\phi_x^{-1/2}\lambda_x(x_i - \mu_x)),$$

onde a_i e A_i são dados em (6.12), λ_x e Λ_x estão definidos na Proposição 6.2.1. Assim, segue do Lema 2 em Lachos *et al.* (2010) que

$$\mathrm{E}[x_i|\mathbf{Y}_i = \mathbf{y}_i, U_i = u_i, \boldsymbol{\theta}] = \mu_x + \Lambda_x a_i + \kappa^{1/2}(u_i) W_\Phi(\kappa^{-1/2}(u_i)\mathrm{A}_i) A_x.$$

O estimador de x_i que minimiza o erro quadrático médio (MSE), obtido através da média condicional x_i dado $\mathbf{Y}_i = \mathbf{y}_i$, pode ser expresso como

$$\begin{aligned} \widehat{x}_i(\boldsymbol{\theta}) &= \mathrm{E}[x_i|\mathbf{Y}_i = \mathbf{y}_i, \boldsymbol{\theta}] = \mathrm{E}[\mathrm{E}[x_i|U_i, \mathbf{Y}_i = \mathbf{y}_i, \boldsymbol{\theta}]|\mathbf{Y}_i = \mathbf{y}_i, \boldsymbol{\theta}], \\ &= \mu_x + \Lambda_x a_i + A_x \eta_{-1i}, \end{aligned} \tag{6.13}$$

onde $\eta_{-1i} = \mathrm{E}[\kappa^{1/2}(U_i) W_\Phi(\kappa^{-1/2}(U_i)\mathrm{A}_i)|\mathbf{y}_i]$. Expressões explícitas de η_{-1i} para as distribuições skew-t, skew-slash e skew-normal contaminada podem ser obtidas da Proposição 2.2.1. Na prática, o estimativa de Bayes empírica de x_i, $\widehat{x}_i$, pode ser obtida, avaliando (6.13) em $\widehat{\boldsymbol{\theta}}$.

O modelo SMSN-G dado em (6.1)-(6.2) considera duas fontes de variação que podem gerar "outliers" na componente do erro (e-"outliers") bem como na componente da variável latente (x-"outliers"); veja, por exemplo, Pinheiro *et al.* (2001) e Osorio *et al.* (2009). Então, substituindo $\boldsymbol{\theta}$ em $x_i(\boldsymbol{\theta})$ por suas estimativas atuais, estendemos os resultados desenvolvidos por Pinheiro *et al.* (2001) e Osorio *et al.* (2009) para os casos t-Student e normal, respectivamente através da seguinte decomposição da distância de Mahalanobis $d_i(\widehat{\boldsymbol{\theta}}) = (\mathbf{y}_i - \widehat{\boldsymbol{\mu}})^\top \widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}_i - \widehat{\boldsymbol{\mu}})$ dada por

$$d_i(\widehat{\boldsymbol{\theta}}) = \widehat{\mathbf{e}}_i^\top D^{-1}(\widehat{\boldsymbol{\phi}})\widehat{\mathbf{e}}_i + \frac{1}{\widehat{\phi}_x}\widehat{\mu}_{xi}^2 = \widehat{d}_{\mathbf{e}i} + \widehat{d}_{xi}, \tag{6.14}$$

onde $\widehat{\mathbf{e}}_i = \mathbf{y}_i - \widehat{\boldsymbol{\mu}} - \mathbf{1}_p\widehat{\mu}_{xi}$ é o resíduo estimado, com $\widehat{\mu}_{xi} = \widehat{\Lambda}_x\widehat{a}_i$.

Uma investigação de influência é um passo importante na análise de dados após a estimação das quantidades de interesse. Dessa forma, um estudo para avaliar a sensibilidade dos resultados obtidos pode ser realizado através da análise de influência local.

6.3 Influência Local

Nesta seção, derivamos a curvatura normal para o modelo de Grubbs misturas de escala skew-normal. Analogamente ao desenvolvimento apresentado na Seção 1.4.1, obtemos a curvatura normal usando a função-Q (Zhu & Lee, 2001) como medida de influência. Os detalhes do cálculos da hessiana $\ddot{Q}_\theta(\widehat{\boldsymbol{\theta}}) = \dfrac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}$ e da matriz $\boldsymbol{\nabla}_{\boldsymbol{\omega}} = \dfrac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}$ são dados abaixo.

6.3.1 Matriz Hessiana

Seja $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ o vetor de parâmetros de interesse, onde $\boldsymbol{\theta}_1 = (\boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top)^\top$ e $\boldsymbol{\theta}_2 = (\mu_x, \phi_x, \lambda_x)^\top$. Para o modelo definido em (6.1)-(6.2), temos que a matriz hessiana é dada por $\ddot{Q}_\theta(\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^n \ddot{Q}_i(\boldsymbol{\theta})$ com

$$\ddot{Q}_i(\boldsymbol{\theta}) \;=\; -\frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \begin{pmatrix} \ddot{Q}_{1,i}(\boldsymbol{\theta}_1) & \mathbf{0} \\ \mathbf{0} & \ddot{Q}_{2,i}(\boldsymbol{\theta}_2) \end{pmatrix},$$

onde $\ddot{Q}_{1,i}(\boldsymbol{\theta}_1) = -\partial^2 Q_{1i}(\boldsymbol{\theta}_1|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1^\top$ e $\ddot{Q}_{2,i}(\boldsymbol{\theta}_2) = -\partial^2 Q_{2i}(\boldsymbol{\theta}_2|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^\top$ têm os seguintes elementos (Magnus & Neudecker, 1988)

$$\begin{aligned}
\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1,\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^\top} &= -\widehat{u}_i \mathbb{I}_{(p)} D^{-1}(\boldsymbol{\phi}) \mathbb{I}_{(p)}^\top, \\
\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1,\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\phi}\partial\boldsymbol{\alpha}^\top} &= \left(-\widehat{u}_i D(\mathbf{y}_i - \mathbf{a}) + \widehat{ux}_i \mathbf{I}_p\right) D^{-2}(\boldsymbol{\phi}) \mathbb{I}_{(p)}^\top, \\
\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1,\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^\top} &= \left[\frac{1}{2} D(\boldsymbol{\phi}) - \widehat{u}_i D^2(\mathbf{y}_i - \mathbf{a}) + 2\widehat{ux}_i D(\mathbf{y}_i - \mathbf{a}) - \widehat{ux^2}_i \mathbf{I}_p\right] D^{-3}(\boldsymbol{\phi}), \\
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2,\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\mu_x\partial\mu_x} &= -\frac{1}{\nu_x^2}\widehat{u}_i,
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2, \boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\mu_x\partial\phi_x} &= \frac{\lambda_x(1+\lambda_x^2)^{1/2}}{2\phi_x^{3/2}}\widehat{ut}_i + \frac{(1+\lambda_x^2)}{\phi_x^2}B_{1i},\\
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2, \boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\mu_x\partial\lambda_x} &= -\frac{2\lambda_x}{\phi_x}B_{1i} - \frac{(1+2\lambda_x^2)}{\phi_x^{1/2}(1+\lambda_x^2)^{1/2}}\widehat{ut}_i,\\
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2, \boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\phi_x\partial\phi_x} &= \frac{1}{2\phi_x^2} - \frac{3\lambda_x(1+\lambda_x^2)^{1/2}}{4\phi_x^{5/2}}B_{2i} - \frac{(1+\lambda_x^2)}{\phi_x^3}B_{3i},\\
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2, \boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\lambda_x\partial\phi_x} &= \frac{(1+2\lambda_x^2)}{2\phi_x^{3/2}(1+\lambda_x^2)^{1/2}}B_{2i} + \frac{\lambda_x}{\phi_x^2}B_{3i},\\
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2, \boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\lambda_x\partial\lambda_x} &= \frac{(1-\lambda_x^2)}{(1+\lambda_x^2)^2} - \frac{1}{\phi_x}B_{3i} - \widehat{ut^2}_i - \frac{\lambda_x(3+2\lambda_x^2)}{\phi_x^{1/2}(1+\lambda_x^2)^{3/2}}B_{2i},
\end{aligned}
$$

onde $\mathbb{I}_{(p)} = [\mathbf{0}_q, \mathbf{I}_q]$, tal que $\mathbf{0}_q = (0, \ldots, 0)^\top$ é um vetor de dimensão $q \times 1$ e $\mathbf{I}_q$ é matriz identidade $q \times q$, $B_{1i} = \widehat{u}_i\mu_x - \widehat{ux}_i$, $B_{2i} = \widehat{ut}_i\mu_x - \widehat{utx}_i$ e $B_{3i} = \widehat{ux^2}_i + \widehat{u}_i\mu_x^2 - 2\widehat{ux}_i\mu_x$.

6.3.2 Esquemas de Perturbação

Para cada um dos esquemas de perturbação, obtemos a matriz $\boldsymbol{\nabla}_{\boldsymbol{\omega}_o}$. Antes de obtermos as matrizes para avaliar influência local para os esquemas de perturbação prospostos, notamos que a matriz $\ddot{Q}_\theta(\widehat{\boldsymbol{\theta}})$ é bloco diagonal com blocos $\ddot{Q}_1(\widehat{\boldsymbol{\theta}}_1) = \ddot{Q}_1$ e $\ddot{Q}_2(\widehat{\boldsymbol{\theta}}_2) = \ddot{Q}_2$. Assim, temos que para qualquer vetor unitário $\mathbf{d}$

$$C_{f_Q,\mathbf{d}} = C_{f_Q,\mathbf{d}}(\widehat{\boldsymbol{\theta}}_1) + C_{f_Q,\mathbf{d}}(\widehat{\boldsymbol{\theta}}_2),$$

onde $C_{f_Q,\mathbf{d}}(\widehat{\boldsymbol{\theta}}_1) = 2\mathbf{d}^\top\boldsymbol{\nabla}_1^\top\boldsymbol{\omega}_o(-\ddot{Q}_1)^{-1}\Delta_1\boldsymbol{\omega}_o\mathbf{d}$ e $C_{f_Q,\mathbf{d}}(\widehat{\boldsymbol{\theta}}_2) = 2\mathbf{d}^\top\boldsymbol{\nabla}_2^\top\boldsymbol{\omega}_o(-\ddot{Q}_2)^{-1}\Delta_2\boldsymbol{\omega}_o\mathbf{d}$, com $\boldsymbol{\nabla}_1\boldsymbol{\omega}_o$ e $\boldsymbol{\nabla}_2\boldsymbol{\omega}_o$ definidas abaixo para cada esquema de perturbação. Nesta seção, obtemos resultados similares aos encontrados em Lachos *et al.* (2007b), mas neste caso $\boldsymbol{\theta}_1 = (\boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top)^\top$. Os esquemas de perturbação propostos para o modelo SMSN-G também foram considerados por Lachos *et al.* (2007b) e Osorio *et al.* (2009)

Ponderação de Casos

Considere a inclusão de pesos para cada unidade experimental no valor esperado da função log-verossimilhança dos dados completos (função-Q perturbada) como

$$Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n}\omega_i \mathrm{E}[\ell_i(\boldsymbol{\theta}|\mathbf{y}_c)] = \sum_{i=1}^{n} Q_{1i}(\boldsymbol{\theta}_1,\omega_i|\widehat{\boldsymbol{\theta}}) + \sum_{i=1}^{n} Q_{2i}(\boldsymbol{\theta}_2,\omega_i|\widehat{\boldsymbol{\theta}}),$$

onde $\boldsymbol{\omega} = (\omega_1,\ldots,\omega_n)^\top$ é um vetor de dimensão $n\times 1$, $\boldsymbol{\omega}_o = \mathbf{1}_n$, $Q_{1i}(\boldsymbol{\theta}_1,\omega_i|\widehat{\boldsymbol{\theta}}) = w_i Q_{1i}(\boldsymbol{\theta}_1|\widehat{\boldsymbol{\theta}})$ e $Q_{2i}(\boldsymbol{\theta}_2,\omega_i|\widehat{\boldsymbol{\theta}}) = w_i Q_{2i}(\boldsymbol{\theta}_2|\widehat{\boldsymbol{\theta}})$, tal que $Q_{1i}(\boldsymbol{\theta}_1|\widehat{\boldsymbol{\theta}})$ e $Q_{2i}(\boldsymbol{\theta}_2|\widehat{\boldsymbol{\theta}})$ estão definidas em (6.10) e (6.11), respectivamente. Alternativamente, podemos pensar em duas outras sub-perturbações definidas por *i)* $Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} Q_{1i}(\boldsymbol{\theta}_1,\omega_i|\widehat{\boldsymbol{\theta}}) + \sum_{i=1}^{n} Q_{2i}(\boldsymbol{\theta}_2|\widehat{\boldsymbol{\theta}})$ e *ii)* $Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} Q_{1i}(\boldsymbol{\theta}_1|\widehat{\boldsymbol{\theta}}) + \sum_{i=1}^{n} Q_{2i}(\boldsymbol{\theta}_2,\omega_i|\widehat{\boldsymbol{\theta}})$ que serão avaliadas baseadas em $C_{f_Q,\mathbf{d}}(\widehat{\boldsymbol{\theta}}_1)$ e $C_{f_Q,\mathbf{d}}(\widehat{\boldsymbol{\theta}}_2)$, respectivamente. Sob este esquema de perturbação, as matrizes $\boldsymbol{\nabla}_1\boldsymbol{\omega}_o$ e $\boldsymbol{\nabla}_2\boldsymbol{\omega}_o$ são dadas por $\boldsymbol{\nabla}_1\boldsymbol{\omega}_o = (\boldsymbol{\nabla}_1\boldsymbol{\omega}_1,...,\boldsymbol{\nabla}_1\boldsymbol{\omega}_n)^\top$ e $\boldsymbol{\nabla}_2\boldsymbol{\omega}_o = (\boldsymbol{\nabla}_2\boldsymbol{\omega}_1,...,\boldsymbol{\nabla}_2\boldsymbol{\omega}_n)^\top$, respectivamente, onde os elementos de $\boldsymbol{\nabla}_1\boldsymbol{\omega}_i = \partial^2 Q_{1i}(\boldsymbol{\theta}_1,\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}_1\partial\omega_i$ e $\boldsymbol{\nabla}_2\boldsymbol{\omega}_i = \partial^2 Q_{2i}(\boldsymbol{\theta}_2,\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}_2\partial\omega_i$ são

$$\begin{aligned}
\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1,\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\alpha}\partial\omega_i} &= \widehat{u}_i\mathbb{I}_{(p)}D^{-1}(\boldsymbol{\phi})(\mathbf{y}_i-\mathbf{a}) - \widehat{ux}_i\mathbb{I}_{(p)}D^{-1}(\boldsymbol{\phi})\mathbf{1}_p,\\
\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1,\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\phi}\partial\omega_i} &= \frac{1}{2}\big[-D(\boldsymbol{\phi}) + \widehat{u}_i D^2(\mathbf{y}_i-\mathbf{a}) - 2\widehat{ux}_i D(\mathbf{y}_i-\mathbf{a}) + \widehat{ux^2}_i\mathbf{I}_p\big]D^{-2}(\boldsymbol{\phi})\mathbf{1}_p,\\
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2,\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\mu_x\partial\omega_i} &= -\frac{1}{\nu_x^2}(\widehat{u}_i\mu_x - \widehat{ux}_i + \widehat{ut}_i\tau_x),\\
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2,\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\phi_x\partial\omega_i} &= -\frac{1}{2\phi_x} + \frac{\lambda_x(1+\lambda_x^2)^{1/2}}{2\phi_x^{3/2}}B_{2i} + \frac{1+\lambda_x^2}{2\phi_x^2}B_{3i},\\
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2,\boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\lambda_x\partial\omega_i} &= \frac{\lambda_x}{(1+\lambda_x^2)} - \frac{\lambda_x}{\phi_x}B_{3i} - \lambda_x\widehat{ut^2}_i - \frac{(1+2\lambda_x^2)}{\phi_x^{1/2}(1+\lambda_x^2)^{1/2}}B_{2i},
\end{aligned}$$

onde $\mathbb{I}_{(p)} = [\mathbf{0}_q,\mathbf{I}_q]$, tal que $\mathbf{0}_q = (0,\ldots,0)^\top$ é um vetor de dimensão $q\times 1$ e $\mathbf{I}_q$ é matriz identidade $q\times q$, $B_{1i} = \widehat{u}_i\mu_x - \widehat{ux}_i$, $B_{2i} = \widehat{ut}_i\mu_x - \widehat{utx}_i$ e $B_{3i} = \widehat{ux^2}_i + \widehat{u}_i\mu_x^2 - 2\widehat{ux}_i\mu_x$.

Perturbação das Observações

Consideramos perturbações aditivas nas medições obtidas através dos instrumentos usados no estudo. A perturbação das observações $(\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$ é introduzida substituindo $\mathbf{y}_i$ por $\mathbf{y}_i(\boldsymbol{\omega})$, onde $\boldsymbol{\omega}$ é o vetor de perturbação. Abordamos os seguintes casos de interesse:

(1) Perturbação simultânea das medições fornecidas pelos p instrumentos: $\mathbf{y}_i(\boldsymbol{\omega}) = \mathbf{y}_i + \boldsymbol{\omega}_i$, onde $\boldsymbol{\omega} = (\boldsymbol{\omega}_1^\top, ..., \boldsymbol{\omega}_n^\top)^\top$ é um vetor de perturbação $np \times 1$ e

(2) Perturbação das medições fornecidas por um instrumento particular: $\mathbf{y}_i(\boldsymbol{\omega}) = \mathbf{y}_i + \omega_i \mathbf{c}_m$, onde $\mathbf{c}_m$ denota o vetor $p \times 1$ de zeros com 1 na m-ésima posição e $\boldsymbol{\omega} = (w_1, ..., w_n)^\top$ é um vetor de perturbação $n \times 1$.

Em ambos os casos, $\boldsymbol{\omega}_o = \mathbf{0}$ e a função-Q perturbada é dada por

$$Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} Q_{1i}(\boldsymbol{\theta}_1, \boldsymbol{\omega}_i|\widehat{\boldsymbol{\theta}}) + \sum_{i=1}^{n} Q_{2i}(\boldsymbol{\theta}_2|\widehat{\boldsymbol{\theta}}),$$

onde $Q_{1i}(\boldsymbol{\theta}_1, \boldsymbol{\omega}_i|\widehat{\boldsymbol{\theta}})$ está definida em (6.10) alterando $\mathbf{y}_i(\boldsymbol{\omega})$ com $\mathbf{y}_i$. Sob este esquema de perturbação, o procedimento de influência local é baseado na $C_{f_Q,\mathbf{d}}(\widehat{\boldsymbol{\theta}}_1)$, onde $\boldsymbol{\nabla}_1\boldsymbol{\omega}_o = (\boldsymbol{\nabla}_1\boldsymbol{\omega}_1, ..., \boldsymbol{\nabla}_1\boldsymbol{\omega}_n)^\top$, com $\boldsymbol{\nabla}_1\boldsymbol{\omega}_i = \partial^2 Q_{1i}(\boldsymbol{\theta}_1, \boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}_1\partial\boldsymbol{\omega}_i{}^\top$, cujos elementos são

$$\begin{aligned}
\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1, \boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\omega}_i{}^\top} &= \widehat{u}_i \mathbb{I}_{(p)} D^{-1}(\boldsymbol{\phi})\mathbf{p}_m, \\
\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1, \boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\phi}\partial\boldsymbol{\omega}_i{}^\top} &= \left(\widehat{u}_i D(\mathbf{y}_i - \mathbf{a}) - \widehat{ux}_i \mathbf{I}_p\right) D^{-2}(\boldsymbol{\phi})\mathbf{p}_m,
\end{aligned}$$

onde $\mathbb{I}_{(p)} = [\mathbf{0}_q, \mathbf{I}_q]$, tal que $\mathbf{0}_q = (0, \ldots, 0)^\top$ é um vetor $q \times 1$ e $\mathbf{I}_q$ é a matriz identidade $q \times q$, $\mathbf{p}_m = 1$ sob o esquema de perturbação dado no item (1) e $\mathbf{p}_m = \mathbf{c}_m$ sob o esquema de perturbação dado no item (2).

Perturbação Vício Multiplicativo

Considere o modelo perturbado $\mathbf{y}_i = \mathbf{a} + \mathbf{b}_\omega x_i + \boldsymbol{\epsilon}_i$, $i = 1, \ldots, n$, onde $\mathbf{b}_\omega = (1, \boldsymbol{\omega}^\top)^\top$, com $\boldsymbol{\omega} \in \mathbb{R}^{p-1}$ e o modelo postulado é obtido considerando $\boldsymbol{\omega}_o = \mathbf{1}_{p-1}$. Note que sob a distribuição normal este modelo foi originalmente proposto em Barnett (1969) e $\mathbf{b}_\omega$ está associado com o vício multiplicativo dos instrumentos de medição. A função-Q perturbada é dada por

$$Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} Q_{1i}(\boldsymbol{\theta}_1, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) + \sum_{i=1}^{n} Q_{2i}(\boldsymbol{\theta}_2|\widehat{\boldsymbol{\theta}}),$$

onde $Q_{1i}(\boldsymbol{\theta}_1, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = -\frac{1}{2}\log(|D(\boldsymbol{\phi})|) - \frac{1}{2}\big[\widehat{u}_i^{(k)}(\mathbf{y}_i - \mathbf{a})^\top D^{-1}(\boldsymbol{\phi})(\mathbf{y}_i - \mathbf{a}) - 2\widehat{ux}_i^{(k)}(\mathbf{y}_i - \mathbf{a})^\top D^{-1}(\boldsymbol{\phi})\mathbf{b}_\omega + \widehat{ux^2}_i^{(k)}\mathbf{b}_\omega^\top D^{-1}(\boldsymbol{\phi})\mathbf{b}_\omega\big]$. Sob este esquema de perturbação, temos $\boldsymbol{\nabla}_1\boldsymbol{\omega}_o = \sum_{i=1}^{n} \partial^2 Q_{1i}(\boldsymbol{\theta}_1, \boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}_1\partial\boldsymbol{\omega}^\top$, cujos elementos são

$$\begin{aligned}
\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1, \boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\omega}^\top} &= -\widehat{ux}_i \mathbb{I}_{(p)} D^{-1}(\boldsymbol{\phi}) \mathbb{I}_{(p)}^\top, \\
\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1, \boldsymbol{\omega}_o|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\phi}\partial\boldsymbol{\omega}^\top} &= \left(-\widehat{ux}_i D(\mathbf{y}_i - \mathbf{a}) + \widehat{ux^2}_i \mathbf{I}_p\right) D^{-2}(\boldsymbol{\phi}) \mathbb{I}_{(p)}^\top,
\end{aligned}$$

onde $\mathbb{I}_{(p)} = [\mathbf{0}_q, \mathbf{I}_q]$, tal que $\mathbf{0}_q = (0, \ldots, 0)^\top$ é um vetor $q \times 1$ e $\mathbf{I}_q$ é a matriz identidade $q \times q$.

Em seguida, com o intuito de ilustrar a metodologia desenvolvida, apresentamos uma aplicação em um conjunto de dados reais.

6.4 Aplicação: Conjunto de Dados Barnett

Nesta aplicação, consideramos o conjunto de dados apresentado por Barnett (1969) em que é avaliada a qualidade de dois métodos, um de tipo padrão e outro novo, utilizados por operadores experientes e não experientes para medir a capacidade vital pulmonar

num grupo de $n = 72$ pacientes. Dessa forma, como em Montenegro *et al.* (2009c), as seguintes combinações método/operador serão avaliadas ($j = 1, \ldots, 4$):

- Método padrão e operador experiente.
- Método padrão e operador não experiente.
- Método novo e operador experiente.
- Método novo e operador não experiente.

Nesta aplicação, o termo instrumento refere-se às combinações método/operador. Este conjunto de dados já foi analisado sob o modelo N-G por Lachos *et al.* (2007b) e sob o modelo SN-G por Montenegro *et al.* (2009b) e Montenegro *et al.* (2009c), nos contextos de estimação e influência local. Conforme descrito em Montenegro *et al.* (2009c), consideramos as medições divididas por 100 para melhorar a estabilidade numérica.

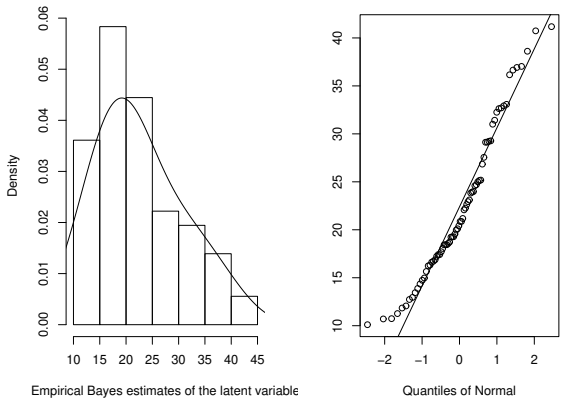


Figura 6.1: Dados Barnett: Histograma e o gráfico Q-Q normal das estimativas de Bayes empíricas de x_i sob normalidade.

Com o intuito de verificar a existência de assimetria e caudas pesadas na resposta latente (x_i), consideramos, inicialmente, um modelo N-G para os dados e calculamos as estimativas de Bayes empíricas de x_i. Com propósitos exploratórios, a Figura 6.1 mostra o histograma e o gráfico Q-Q normal para as estimativas de Bayes empíricas de x_i. Observa-se que a distribuição da variável latente apresenta assimetria positiva e assim, o modelo gaussiano pode não fornecer um bom ajuste. Além disso, o histograma das medições observadas para o instrumento de referência, não mostrado aqui por brevidade, claramente suporta o uso de distribuições assimétricas com caudas pesadas; veja Montenegro *et al.* (2009b). Dessa forma, o ajuste do modelo SMSN-G para esses dados parece adequado.

De acordo com as observações acima, consideramos uma distribuição SMSN para x_i e uma distribuição SMN para $\boldsymbol{\epsilon}_i$, conforme o modelo definido em (6.1) e (6.2). Nesta aplicação, consideramos as distribuições skew-normal (SN), skew-t (ST), skew-normal contaminada (SCN) e skew-slash (SSL) pertencentes à classe SMSN para propósitos comparativos.

A Tabela 6.1 apresenta as estimativas de máxima verossimilhança e seus respectivos desvios padrões assintóticos estimados, calculados através da matriz informação observada, para os modelos N-G e SMSN-G. Entretanto, nossa experiência em estimar a variância assintótica do estimador de máxima verossimilhança de $\boldsymbol{\lambda}$ indica que ele é tipicamente superestimado; veja Montenegro *et al.* (2009b), por exemplo, para uma discussão detalhada sobre este assunto. Além disso, de acordo com Lange *et al.* (1989), escolhemos o valor de $\boldsymbol{\nu}$ maximizando a função verossimilhança perfilada. Usando este procedimento, encontramos $\nu = 9.6560$ para a ST, temos $\nu = 2.4915$ para a SSL e obtemos $\nu = 0.6260$ e $\gamma = 0.2992$ para a SCN.

Em seguida, comparamos os modelos N-G e SMSN-G, inspecionando o critério de informação AIC, por exemplo. Este indica que os modelos assimétricos com caudas pesadas apresentam um melhor ajuste, sugerindo dessa forma, a violação da suposição de simetria dos dados. Adicionalmente, substituindo as estimativas de máxima verossimilhança de $\boldsymbol{\theta}$ na distância de Mahalanobis d_i, construímos os envelopes simulados, apresentados na Figura 6.2; veja Montenegro *et al.* (2009b) para mais detalhes. Pela Figura 6.2, temos fortes evidências de que as distribuições assimétricas com caudas pesadas são mais apropriadas para este conjunto de dados do que as distribuições normal e skew-normal. Por brevidade, o envelope do modelo SSL-G ajustado é similar ao do modelo ST-G e não será mostrado aqui.

Tabela 6.1: Dados Barnett. Estimativas de máxima verossimilhança para os modelos ajustados. Os valores SE, entre parênteses, são os desvios padrões assintóticos estimados.

	Normal	SN	ST	SSL	SCN
Parâmetro	Estimativa (SE)	Estimativa (SE)	Estimativa (SE)	Estimativa (SE)	Estimativa (SE)
α_1	-0.7042 (0.2984)	-0.7042 (0.2961)	-0.6735 (0.2690)	-0.6992 (0.2758)	-0.6231 (0.2651)
α_2	-0.9750 (0.3610)	-0.9750 (0.3649)	-1.0365 (0.3484)	-1.0801 (0.3547)	-0.8835 (0.3366)
α_3	-1.4389 (0.3657)	-1.4389 (0.3693)	-1.3660 (0.3549)	-1.4435 (0.3620)	-1.1897 (0.3382)
ϕ_1	4.9979 (0.9947)	5.0612 (0.9888)	3.7388 (0.8320)	2.8879 (0.6497)	1.9789 (0.4097)
ϕ_2	1.4129 (0.5698)	1.2515 (0.5641)	0.8924 (0.4565)	0.6800 (0.3542)	0.4919 (0.2323)
ϕ_3	4.3831 (0.9742)	4.5266 (0.9997)	3.7256 (0.8868)	2.8951 (0.6731)	1.7813 (0.4339)
ϕ_4	4.6330 (1.0321)	4.7578 (1.0646)	4.0136 (0.9739)	3.1342 (0.7423)	1.8736 (0.4594)
μ_x	22.4611 (0.9711)	12.1534 (1.3941)	12.2555 (1.5584)	12.2459 (1.5601)	12.2579 (1.6967)
ϕ_x	62.9065 (10.6442)	168.9810 (39.9995)	141.9690 (39.7241)	108.1170 (29.3612)	73.3300 (22.3333)
λ_x	- -	5.6843 (3.8465)	5.2148 (3.9439)	5.3509 (4.0655)	5.0883 (4.1955)
$l(\widehat{\boldsymbol{\theta}})$	-747.7896	-742.8945	-740.2310	-741.4170	-738.6340
AIC	1513.6000	1505.8000	1502.5000	1504.8000	1501.3000

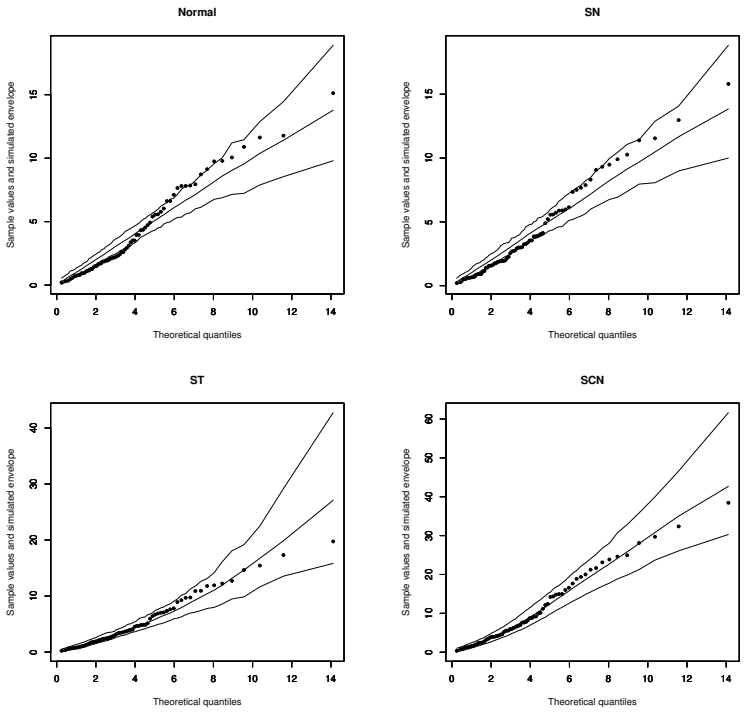


Figura 6.2: Dados Barnett. Envelopes simulados.

Com a finalidade de identificar as observações aberrantes, a distância de Mahalanobis foi considerada; veja os resultados descritos na Proposição 2.2.8 para mais detalhes. A Figura 6.3 ilustra tal distância para os modelos ajustados, sendo que o gráfico da distância de Mahalanobis para o modelo SSL-G é similar ao do modelo ST-G. Podemos observar pelos gráficos que as observações $4, 25, 54$ e 67 são detectadas como "outliers" para os modelos ST, SCN e SSL. Adicionalmente, as observações 1 e 72 aparecem como "outliers" para o caso SN e as observações $1, 23$ e 62 para o caso normal. Proveniente do algoritmo tipo-EM, os pesos estimados $(\widehat{u}_i, i = 1, \ldots, 72)$ para essas observações são menores quando consideramos os modelos SMSN-G com caudas pesadas (veja a Figura 6.4), confirmando a flexibilidade das estimativas de máxima verossimilhança desses modelos contra observações atípicas. Para os modelos normal e

skew-normal $u_i = 1\ \forall i$ e estão mostrados como uma linha tracejada na Figura 6.4.

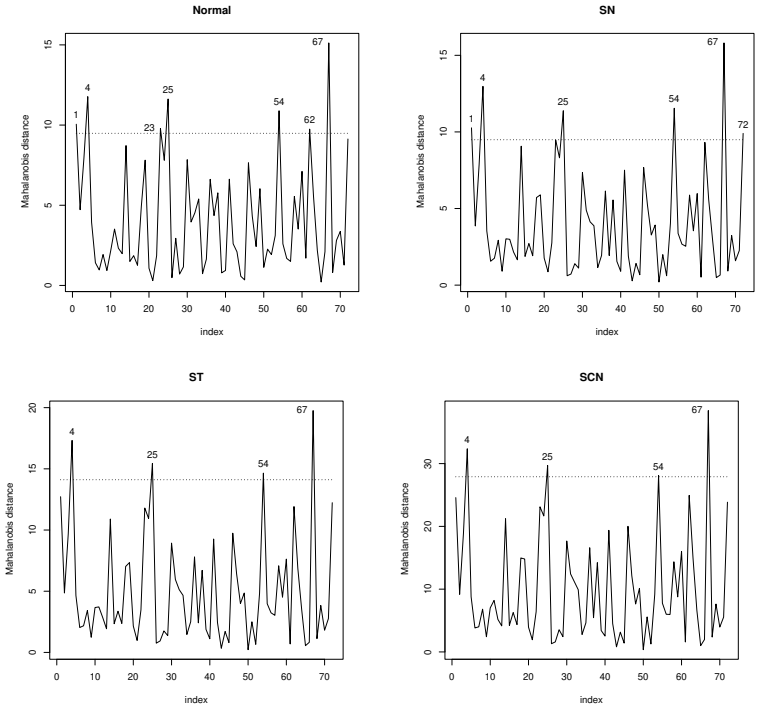


Figura 6.3: Dados Barnett. Gráficos da distância de Mahalanobis para quatro modelos ajustados, considerando $\xi = 0.95$.

As distâncias estimadas $d_{\mathbf{e}i}$ (Error) e d_{xi} (Latent), definidas em (6.14), estão apresentadas na Figura 6.5 para o modelo SN-G. As observações $4, 25$ e 67 apresentam apenas valores grandes da distância $d_{\mathbf{e}i}$, sugerindo que são **e**-"outliers". Além disso, as observações $1, 23$ e 54 possuem valores grandes da distância d_{xi}, sugerindo que são x-"outliers". O gráfico da distância d_{xi} destaca também as observações $3, 14, 30$ e 58 como x-"outliers". Note que esta conclusão não é possível a partir da Figura 6.3. Para os modelos N-G, ST-G, SSL-G e SCN-G, observa-se os mesmos resultados.

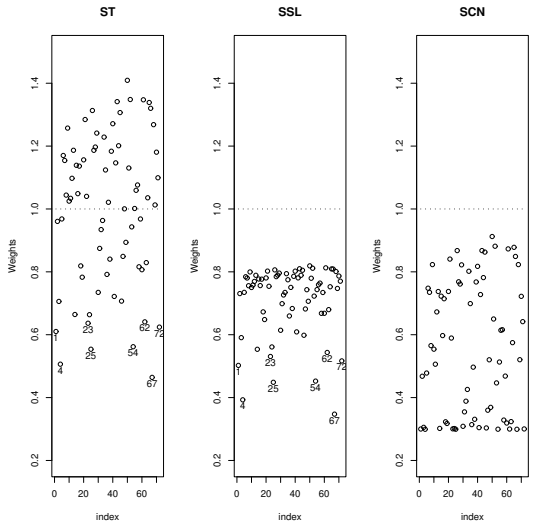


Figura 6.4: Valores de u_i estimados para os modelos ST, SSL e SCN.

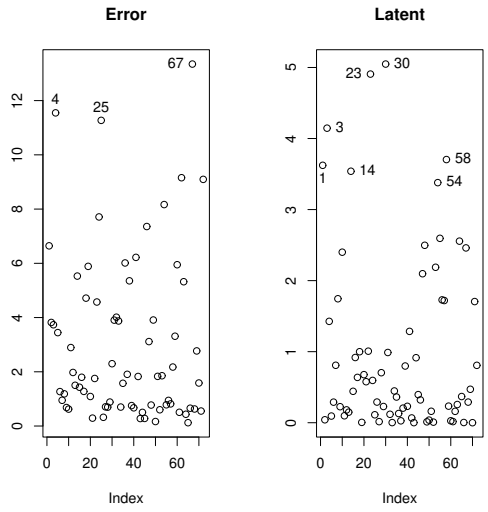


Figura 6.5: Dados Barnett. Decomposição da distância de Mahalanobis: $d_{\mathbf{e}i}$ (error) e d^2_{xi} (Latent) estimadas pelo ajuste SN.

A seguir identificaremos as observações influentes do conjunto de dados de Barnett utilizando a quantidade $M(0)$, calculada através da curvatura conformal $B_{f_Q,\mathbf{u}}$, obtida considerando ponderação de casos e os esquemas de perturbação nas observações e vício multiplicativo. As Figuras 6.6, 6.7 e 6.8 apresentam os gráficos de $M(0)$ para os quatro modelos selecionados. Na Figura 6.6, nota-se que sob o modelo SN-G, as observações 4 e 67 (detectadas como "outliers" na Figura 6.3) são identificadas como influentes. Como esperado, sob a perturbação ponderação de casos, a influência de tais observações é reduzida quando consideramos distribuições com caudas mais pesadas que a skew-normal, em particular as distribuições skew-t e skew-slash.

Na Figura 6.7, sob a perturbação simultânea das observações fornecidas pelos instrumentos, alguma influência é apreciada sob o modelo SN-G quando as medições do item 14 são perturbadas. Diferentemente, o item 18 é detectado como influente sob os modelos ST-G e SSL-G. Note que sob este esquema de perturbação, o modelo SCN-G parece acomodar melhor as observações influentes.

Na Figura 6.8, apresentamos os gráficos de influência local para a perturbação vício multiplicativo. Uma vez que os valores de $M(0)$ são bastante diferentes, podemos concluir que a suposição de igualdade dos vícios multiplicativos não é plausível e parece mais apropriado modificar o modelo incorporando o vício multiplicativo dos instrumentos. Dessa forma, um modelo de calibração (Barnett, 1969) pode ser mais apropriado para modelar este conjunto de dados; veja Lachos *et al.* (2009) para mais detalhes sobre o modelo de calibração sob a classe de distribuições SMSN. Esta conclusão concorda com os resultados dos testes de hipóteses, considerados em Montenegro *et al.* (2009b).

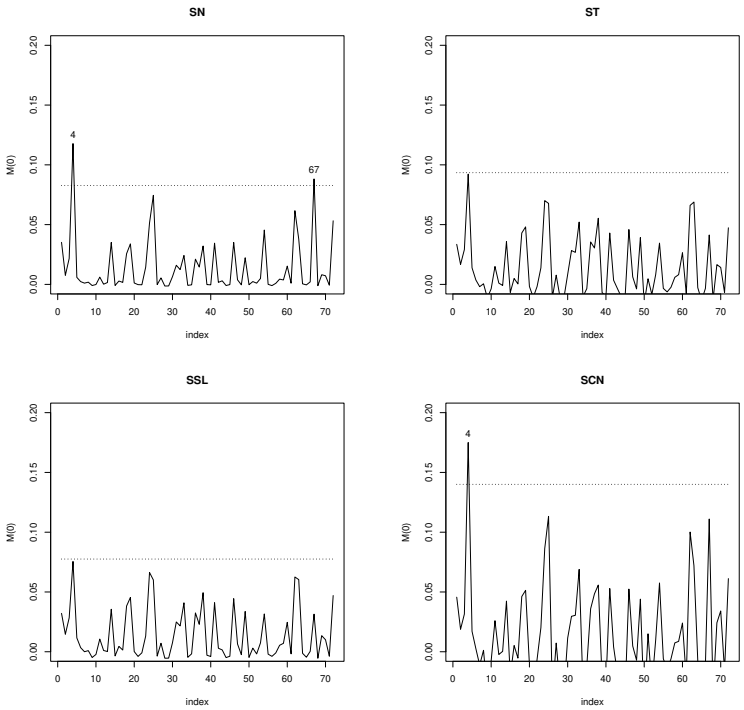


Figura 6.6: Dados Barnett. Gráficos de $M(0)$ sob ponderação de casos para os quatro modelos ajustados. A linha horizontal delimita a marca de referência de Lee & Xu (2004) para M(0) com $c* = 3$.

Como sugerido por Lee *et al.* (2006), usamos as quantidades TRC e MRC, definidas em (4.8), para revelar o impacto das observações influentes detectadas nas estimativas de máxima verossimilhança de $\boldsymbol{\theta}$. A Tabela 6.2 apresenta estas medidas, considerando diferentes modelos, quando excluímos as observações 4 e 67. Note que a maior mudança ocorreu sob a distribuição skew-normal. Como esperado, os resultados indicam que as estimativas de máxima verossimilhança são menos sensíveis à presença de dados atípicos quando usamos distribuições com caudas mais pesadas que a SN, fato também observado nos três esquemas de perturbação desenvolvidos nesta ilustração.

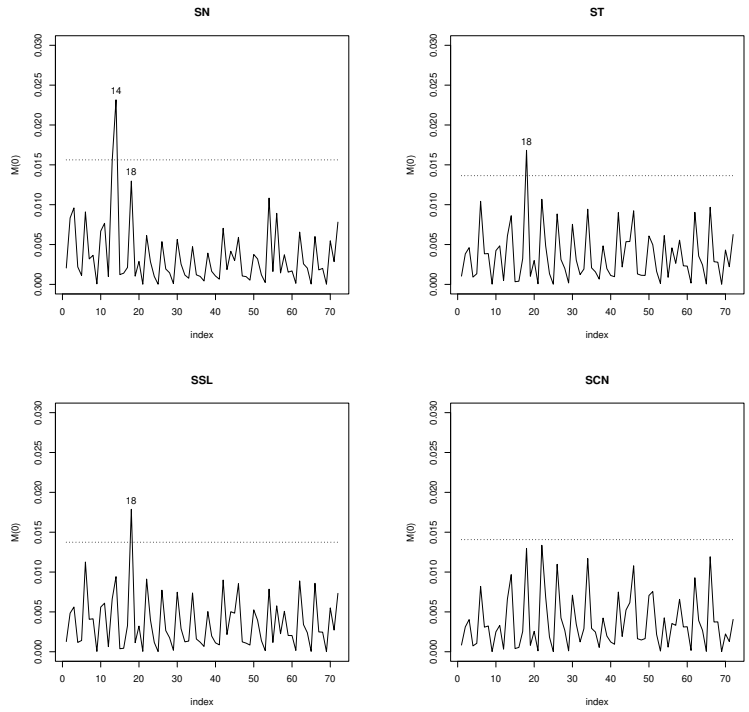


Figura 6.7: Dados Barnett. Gráficos de M(0) sob a perturbação simultânea das medições fornecidas pelos instrumentos para os quatro modelos ajustados. A linha horizontal delimita a marca de referência de Lee & Xu (2004) para M(0) com $c* = 3$.

Tabela 6.2: Dados Barnett. Comparação das mudanças relativas nas estimativas de máxima verossimilhança em termos de TRC e MRC para os quatro modelos SMSN selecionados.

Observação	Distribuição	TRC	MRC
4 e 67	SN	1.1860	0.2396
	ST	0.9130	0.1519
	SSL	0.9765	0.1578
	SCN	1.1314	0.2241

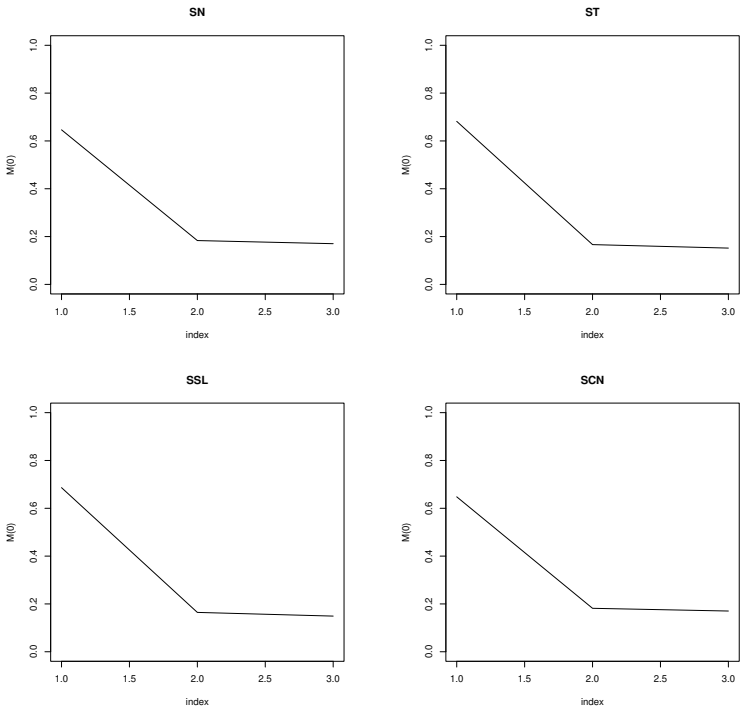


Figura 6.8: Dados Barnett. Gráficos de M(0) sob a perturbação vício multiplicativo para os quatro modelos ajustados.

6.5 Observações Finais

Neste capítulo, propomos uma metodologia robusta para os modelos SMSN-G que considera assimetria e/ou caudas pesadas, permitindo analisar dados em uma ampla variedade de situações. Para esta classe de modelos flexíveis, apresentamos um algoritmo tipo-EM para estimação por máxima verossimilhança e um estudo de influência local para avaliar a sensibilidade dos resultados no processo de estimação dos parâmetros. Expressões explícitas para a matriz $\boldsymbol{\nabla}$ foram obtidas sob diversos esquemas de perturbação. Ressaltamos, entretanto, que outras esquemas de perturbação podem ser desenvolvidos de forma análoga. Os resultados obtidos neste capítulo concordam com

os apresentados por Montenegro *et al.* (2009c) e Osorio *et al.* (2009).

Capítulo 7

Considerações Finais

7.1 Conclusões

Em resumo, nesta tese discutimos vários aspectos envolvendo modelos lineares sob a classe de distribuições misturas de escala skew-normal. Um dos aspectos abordados foi o desenvolvimento de técnicas de estimação nos modelos lineares SMSN, especificamente no modelo de regressão linear, no modelo linear misto e no modelo de Grubbs. Para estimação via máxima verossimilhança, a densidade marginal das quantidades envolvidas foi obtida de modo que os estimadores podem ser obtidos diretamente através de alguma técnica de maximização, usando programas estatísticos, tais como R e o Matlab, por exemplo. Alternativamente, o algoritmo EM foi proposto, explorando as propriedades estatísticas discutidas para a classe de distribuições misturas de escala skew-normal. Uma das características dessa classe rica de modelos é que as distribuições SMSN podem naturalmente atribuir pesos diferentes para cada observação e

consequentemente controlar a influência da observação no processo de estimação, por exemplo. Dessa forma, um procedimento de estimação robusto às observações aberrantes e/ou influentes foi considerado neste trabalho. Adicionalmente, mostramos que critérios de informação usuais, tais como, AIC, BIC e HQ podem ser usados para detectar afastamentos da normalidade e selecionar a distribuição dentro da classe SMSN que melhor se ajuste aos dados. Com o intuito de desenvolver procedimentos bayesianos de estimação, desenvolvemos uma classe alternativa de distribuições SMSN que generaliza as distribuições assimétricas introduzidas por Sahu *et al.* (2003) e Arellano-Valle *et al.* (2007), por exemplo, amplamente aplicadas em análises bayesianas. Neste sentido, propomos algumas extensões unificadas da classe de distribuições SMSN e que fazem da inferência bayesiana uma alternativa viável para tratar os modelos lineares. Em particular, propomos o modelo de regressão sob a classe alternativa de distribuições SMSN e desenvolvemos um procedimento bayesiano de estimação para tratar tal modelo. De uma forma geral, propusemos estratégias inferenciais necessárias para abordar análises de possíveis dados com características assimétricas, inclusive no contexto de dados parcialmente observados ("missing values"). Neste contexto, propomos um algoritmo simples para estimação dos parâmetros e imputação de dados ausentes com estrutura SMSN, pois consideramos diretamente o algoritmo EM desenvolvido para dados completamente observados.

Outro aspecto abordado foi o desenvolvimento de métodos de diagnóstico nos modelos lineares SMSN, especificamente no modelo de regressão linear, no modelo linear misto e no modelo de Grubbs. De acordo com Pinheiro *et al.* (2001) e Osorio *et al.* (2007), a distância de Mahalanobis foi considerada para identificar possíveis observações atípicas que podem prejudicar o ajuste dos modelos bem como as inferências estatísticas. Neste contexto, estendemos alguns resultados propostos por Pinheiro *et al.* (2001) para

o caso t-Student, desenvolvendo para o modelo linear misto e para o modelo de Grubbs no contexto assimétrico a decomposição da distância de Mahalanobis em duas partes, uma relacionada com o componente erro e a outra com o componente aleatório do modelo. Decomposição que fornece uma forma simples de calcular a distância de Mahalanobis bem como estatísticas de diagnóstico úteis na identificação de observações extremas para o componente erro e/ou componente aleatório do modelo, fato que não pode ser concluído simplesmente através da distância de Mahalanobis. Além disso, a metodologia de influência local proposta por Zhu & Lee (2001) pertinente aos modelos lineares SMSN foi desenvolvida, considerando alguns esquemas de perturbação de interesse. Temos notado que, para algumas aplicações, os modelos lineares SMSN com caudas mais pesadas que a distribuição skew-normal (normal) tendem a acomodar melhor as observações aberrantes e/ou influentes.

Os resultados obtidos foram aplicados em conjuntos de dados reais ou simulados. Foram utilizados os programas estatísticos R, Matlab e WinBUGS para as programações dos procedimentos de estimação e diagnóstico dos modelos estudados. Disponibilizamos alguns códigos em que as análises descritas neste trabalho podem ser realizadas. Informações rela-tivas a tais códigos estão disponíveis no seguinte endereço http:// www.ime.unicamp.br/∼hlachos/ListaPub.html.

Para o leitor interessado em aplicar a metodologia desenvolvida nesta tese com o objetivo de analisar um conjunto de dados, recomendamos como passo inicial uma análise exploratória dos dados. Na prática, muitas vezes usamos histogramas e gráficos Q-Q normal dos dados (por exemplo, no modelo de regressão linear) ou das estimativas de Bayes empíricas dos efeitos aleatórios sob normalidade (por exemplo, no modelo linear misto) com propósitos exploratórios, tais como verificar desvios da suposição de nor-

malidade e assimetria (positiva ou negativa) presente nos dados. De acordo com as observações acima, ajustamos um modelo para o conjunto de dados, tais que as estimativas de máxima verossimilhança dos parâmetros de interesse são obtidas via algoritmo EM e seus respectivos desvios padrões assintóticos estimados são calculados através da matriz informação observada. Para avaliar o ajuste do modelo proposto, sugerimos a construção de envelopes simulados e a inspeção de alguns critérios de informação, tais como AIC, BIC e HQ, por exemplo. Supondo que já assumimos um determinado modelo como correto, uma investigação de influência é um passo importante na análise de dados após a estimação dos parâmetros. Isso pode ser realizado através da análise de diagnóstico de influência.

Concluindo, esta tese é um esforço inicial para apresentar alguns tópicos nesta área de pesquisa e divulgar a utilidade da mesma.

7.2 Perspectivas Futuras

Várias linhas de pesquisa podem ser ainda consideradas, tais como:

- Estudar as distribuições gaussiana inversa skew-normal com a finalidade de introduzir mais propriedades e discutir alguns aspectos inferenciais, incluindo o algoritmo EM para a obtenção dos estimadores de máxima verossimilhança dos parâmetros de interesse, bem como a questão de identificabilidade.

- Propor resíduos para os modelos estatísticos lineares sob a classe de distribuições SMSN.

- Outras medidas de diagnóstico para os modelos lineares SMSN podem ser derivadas da metodologia proposta por Zhu & Lee (2001), tais como distância de Cook, afas-

tamento pela verossimilhança e leverage generalizado, todas baseadas na função de verossimilhança completa.

- Estender os resultados de estimação e diagnóstico encontrados para os modelos não-lineares sob a classe de distribuições misturas de escala skew-normal.

Apêndice A

Resultados Adicionais do Capítulo 2

A.1 Lemas da Seção 2.2

Lema A.1.1 *Considere que* $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$, $U \sim H$ *é a variável de mistura e* $g : \mathbb{R} \longrightarrow \mathbb{R}$ *é uma função mensurável. Então,*

$$E[g(U)|\mathbf{Y} = \mathbf{y}] \;\; = \;\; (2f_0(\mathbf{y})/f(\mathbf{y})) \int_0^\infty g(u)\Phi(\kappa^{-1/2}(u)A)h_0(u|\mathbf{y}_0)du,$$

onde $A = \boldsymbol{\lambda}^\top \mathbf{y}_0$, *com* $\mathbf{y}_0 = \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$, f_0 *é a pdf de* $\mathbf{Y}_0 \sim SMN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$ *e* $U_{\mathbf{y}} \stackrel{d}{=} U|\mathbf{Y}_0 = \mathbf{y}$.

Prova Para o caso em que U é (absolutamente) contínua, considere $h(u; \boldsymbol{\nu})$ sendo a pdf de U, $h(u|\mathbf{y}) = f(\mathbf{y}|u)h(u; \boldsymbol{\nu})/f(\mathbf{y})$ sendo a pdf condicional de U dado $\mathbf{Y} = \mathbf{y}$ e $h_0(u|\mathbf{y}_0)$ sendo a pdf condicional de U dado $\mathbf{Y}_0 = \mathbf{y}$, onde $\mathbf{Y}_0|U = u \sim N_p(\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma})$. Desde que $h(u|\mathbf{y}) = f(\mathbf{y}|u)h(u; \boldsymbol{\nu})/f(\mathbf{y})$ e $h_0(u|\mathbf{y}) = \phi_p(\mathbf{y}; \boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma})h(u; \boldsymbol{\nu})/f_0(\mathbf{y})$, onde f_0 é

a pdf de $\mathbf{Y}_0$, para qualquer função integrável $g(u)$, temos que

$$\begin{aligned}
\mathrm{E}[g(U)|\mathbf{Y}=\mathbf{y}] &= (1/f(\mathbf{y}))\int_0^\infty g(u)f(\mathbf{y}|u)h(u;\boldsymbol{\nu})du \\
&= (2/f(\mathbf{y}))\int_0^\infty g(u)\phi_p(\mathbf{y};\boldsymbol{\mu},\kappa(u)\boldsymbol{\Sigma})h(u;\boldsymbol{\nu})\Phi(\kappa^{-1/2}(u)A)du \\
&= (2/f(\mathbf{y}))\int_0^\infty g(u)f_0(\mathbf{y}|u)h(u;\boldsymbol{\nu})\Phi(\kappa^{-1/2}(u)A)du \\
&= (2f_0(\mathbf{y})/f(\mathbf{y}))\int_0^\infty g(u)\Phi(\kappa^{-1/2}(u)A)h_0(u|\mathbf{y}_0)du.
\end{aligned}$$

A prova quando U é discreta é análoga se substituímos $\int$ por $\sum$.

Lema A.1.2 *Se* $\mathbf{Y}\sim SN_p(\boldsymbol{\lambda})$, *para quaisquer vetor fixo p-dimensional* $\mathbf{b}$ *e matriz* $\mathbf{B}$ $(p\times p)$, *temos que*

$$E[\mathbf{Y}^\top\mathbf{B}\mathbf{Y}\mathbf{b}^\top\mathbf{Y}] = -\sqrt{\frac{2}{\pi}}[(\boldsymbol{\delta}^\top\mathbf{B}\boldsymbol{\delta}+tr(\mathbf{B}))\mathbf{b}^\top\boldsymbol{\delta}+2\boldsymbol{\delta}^\top\mathbf{B}\mathbf{b}],$$

onde $\boldsymbol{\delta}=\boldsymbol{\lambda}/\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}$.

Prova A prova segue usando a representação estocástica do vetor aleatório skew-normal dada em (2.4) e os momentos $E[|X_0|]$ e $E[|X_0|^3]$, onde $X_0\sim N(0,1)$ (Arellano-Valle *et al.*, 2005).

A.2 Esboço da Prova da Proposição 2.2.5

Considere que $\mathbf{Y}|U=u\sim SN_p(\boldsymbol{\mu},\kappa(u)\boldsymbol{\Sigma},\boldsymbol{\lambda})$, $\mathbf{Y}_1|U=u\sim SN_{p_1}(\boldsymbol{\mu}_1,\kappa(u)\boldsymbol{\Sigma}_{11},\boldsymbol{\Sigma}_{11}^{1/2}\widetilde{\boldsymbol{\upsilon}})$ e que a densidade $f(\mathbf{y}_2|\mathbf{y}_1,u)=f(\mathbf{y}|u)/f(\mathbf{y}_1|u)$ pode ser escrita como

$$f(\mathbf{y}_2|\mathbf{y}_1,u)=\phi_{(p-p_1)}(\mathbf{y}_2|\boldsymbol{\mu}_{2.1},\kappa(u)\boldsymbol{\Sigma}_{22.1})\frac{\Phi(\kappa^{-1/2}(u)\boldsymbol{\upsilon}^\top(\mathbf{y}-\boldsymbol{\mu}))}{\Phi(\kappa^{-1/2}(u)\widetilde{\boldsymbol{\upsilon}}^\top(\mathbf{y}_1-\boldsymbol{\mu}_1))}.$$

A prova segue diretamente da função geradora de momentos baseada na densidade condicional acima. Os detalhes dos cálculos são similares aos dados em Lin & Lee (2008).

Apêndice B

Resultados Adicionais do Capítulo 3

B.1 Lemas da Seção 3.2

Lema B.1.1 *Se* $\mathbf{Z} \sim SSN_{p,k}(\boldsymbol{\Delta}_*)$, *para quaisquer vetor fixo* $\boldsymbol{b} \in \mathbf{R}^m$ *e matriz* $\boldsymbol{B} \in \mathbf{R}^{m\times p}$ *de posto completo, temos que*

$$M_{\boldsymbol{B}\mathbf{Z}+\boldsymbol{b}}(\boldsymbol{s}) = 2^k \exp\left(\boldsymbol{s}\boldsymbol{b} + \frac{1}{2}\boldsymbol{s}^\top \boldsymbol{B}\boldsymbol{B}^\top \boldsymbol{s}\right) \Phi_k(\boldsymbol{\Delta}_*^\top \boldsymbol{B}^\top \boldsymbol{s}), \boldsymbol{s} \in \mathbf{R}^p.$$

Prova A prova segue diretamente da Proposição 2.4 em Arellano-Valle & Genton (2005), com $\mathbf{b} = \boldsymbol{\mu}$, $\mathbf{B} = \kappa^{1/2}(U)\boldsymbol{\Omega}^{1/2}$, $\boldsymbol{\Omega} = \boldsymbol{\Sigma} + \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top$ e $\boldsymbol{\Delta}_* = \boldsymbol{\Omega}^{-1/2}\boldsymbol{\Lambda}$.

O lema seguite é uma extensão dos resultados obtidos por Azzalini & Capitanio (2003) no contexto univariado, útil para avaliar a pdf da distribuição SST.

Lema B.1.2 *Se* $U \sim Gamma(\alpha, \beta)$, *para quaisquer vetor* $\mathbf{B} \in \mathbb{R}^p$ *e matriz definida positiva* $(p \times p)$ $\mathbf{\Sigma}$, *temos que*

$$E\{\Phi_p(\mathbf{B}\sqrt{U}|\mathbf{\Sigma})\} = T_p(\sqrt{\frac{\alpha}{\beta}}\mathbf{B}|\mathbf{\Sigma}; 2\alpha),$$

onde $T_p(\cdot|\mathbf{\Sigma};\nu)$ *especifica a cdf de uma distribuição t-Student p-variada.*

Prova Considere que $\mathbf{V} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$. Então, temos que

$$\begin{aligned} E\{\Phi_p(\mathbf{B}\sqrt{U}|\mathbf{\Sigma})\} &= E_U\{P(\mathbf{V} \leq \mathbf{B}\sqrt{u}|U=u)\} \\ &= E_U\{P(\frac{\mathbf{V}}{(u\beta/\alpha)^{1/2}} \leq \sqrt{\frac{\alpha}{\beta}}\mathbf{B}|U=u)\} \\ &= P(\mathbf{T} \leq \sqrt{\frac{\alpha}{\beta}}\mathbf{B}), \end{aligned}$$

onde claramente, $\mathbf{T} = \dfrac{\mathbf{V}}{(U\beta/\alpha)^{1/2}}$ tem uma distribuição t-Student multivariada, o que conclui a prova.

B.2 Distribuições Posteriores Condicionais para os Casos SMSN

- *Distribuição Skew-t*

Considere que ν é um parâmetro escalar e que quando $\nu = 1$, obtemos a distribuição skew-cauchy. Neste caso, adotamos a distribuição priori exponencial truncada para ν, denotada por $E(\frac{\varrho}{2})\mathbb{I}_{(2,\infty)}$. A distribuição posteriori condicional definida em (3.17) é

$$U_i|\boldsymbol{\theta}, \mathbf{t}, \mathbf{y} \sim Gamma(p + \nu/2; \nu/2 + C_i/2),$$

onde

$$C_i = (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{\Lambda}\mathbf{t}_i)^\top \mathbf{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{\Lambda}\mathbf{t}_i) + \mathbf{t}_i^\top \mathbf{t}_i \mathbf{I}_{\{\mathbf{t}_i > \mathbf{0}\}}. \tag{B.1}$$

A distribuição posteriori condicional completa de ν é

$$\begin{aligned}\pi(\nu|\boldsymbol{\theta}_1,\mathbf{u},\mathbf{t},\mathbf{y}) &\propto (2^{\nu/2}\Gamma(\nu/2))^{-n}\nu^{n\nu/2}\times \\ &\quad \exp\left(-\frac{\nu}{2}\left[\sum_{i=1}^{n}(u_i-\log u_i)+\varrho\right]\right)\mathbb{I}_{(2,\infty)}. \end{aligned} \tag{B.2}$$

Note que $\pi(\nu|\boldsymbol{\theta}_1,\mathbf{u},\mathbf{t},\mathbf{y}) \propto \pi_1(\nu)\times Gamma(\frac{n\nu}{2}+1,\frac{1}{2}\sum_{i=1}^{n}(u_i-\log u_i)+\varrho)\mathbb{I}_{(2,\infty)}$, onde $\pi_1(\nu)=(2^{\nu/2}\Gamma(\nu/2))^{-n}$ e $\boldsymbol{\theta}_1=(\boldsymbol{\beta}^\top,\boldsymbol{\gamma}^\top,\boldsymbol{\lambda}^\top)^\top$. Observe que a distribuição dada em (B.2) não tem forma fechada, mas o método Metropolis-Hastings pode ser utilizado.

- *Distribuição Skew-Slash*

Considere que ν é um parâmetro escalar. A distribuição $Gamma(a,b)\mathbb{I}_{\{\nu>1\}}$ com valores positivos e pequenos para a e b ($b \ll a$) pode ser adotada como distribuição priori para ν. Isso é conveniente para trabalharmos com famílias conjugadas. Neste caso, a distribuição posteriori condicional completa para cada U_i é

$$U_i|\boldsymbol{\theta},\mathbf{t},\mathbf{y} \sim Gamma(p+\nu/2; C_i/2)\mathbb{I}_{\{0<u_i<1\}},$$

em que C_i está definido em (B.1).

Adicionalmente, a distribuição posteriori condicional completa de ν é

$$\begin{aligned}\pi(\nu|\boldsymbol{\theta}_1,\mathbf{u},\mathbf{t},\mathbf{y}) &\propto \nu^{n+a-1}\times \\ &\quad \exp\left(-\nu\left[b-\sum_{i=1}^{n}\log u_i\right]\right)\mathbb{I}_{\{\nu>1\}}, \end{aligned} \tag{B.3}$$

isto é, $\nu|\boldsymbol{\theta}_1,\mathbf{u},\mathbf{t},\mathbf{y} \sim Gamma(n+a, b-\sum_{i=1}^{n}\log u_i)\mathbb{I}_{\{\nu>1\}}$, onde $\boldsymbol{\theta}_1=(\boldsymbol{\beta}^\top,\boldsymbol{\gamma}^\top,\boldsymbol{\lambda}^\top)^\top$.

• *Distribuição Skew-Normal Contaminada*

Os dois possíveis estados para os pesos u_i são ν_2 ou 1, com $\boldsymbol{\nu} = (\nu_1, \nu_2)^\top$. A distribuição $U(0,1)$ é usada como priori para ν_2 e uma distribuição $Beta(a,b)$ é adotada como priori para ν_1 com o intuito de trabalharmos com famílias conjugadas. A distribuição posteriori condicional completa para cada U_i é proporcional à

$$\begin{cases} \nu_1 \nu_2^p \exp\{-\frac{1}{2}\nu_2 C_i\}, & \text{if } u_i = \nu_2 \\ (1-\nu_1)\exp\{-\frac{1}{2}C_i\}, & \text{if } u_i = 1, \end{cases}, \tag{B.4}$$

em que C_i está definido em (B.1) e a distribuição posteriori condicional completa de ν_1 é

$$\pi(\nu_1|\boldsymbol{\theta}_2, \mathbf{u}, \mathbf{t}, \mathbf{y}) \propto \nu_1^{(a + \frac{n - \sum_{i=1}^n u_i}{1-\nu_2} - 1)} \times (1-\nu_1)^{(b + \frac{\sum_{i=1}^n u_i - n\nu_2}{1-\nu_2} - 1)},$$

isto é,

$$\nu_1|\boldsymbol{\theta}_2, \mathbf{u}, \mathbf{t}, \mathbf{y} \sim Beta\left(a + \frac{n - \sum_{i=1}^n u_i}{1-\nu_2}; b + \frac{\sum_{i=1}^n u_i - n\nu_2}{1-\nu_2}\right),$$

onde $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\lambda}^\top, \nu_2)^\top$.

A distribuição posteriori condicional completa de ν_2 é

$$\pi(\nu_2|\boldsymbol{\theta}_3, \mathbf{u}, \mathbf{t}, \mathbf{y}) \propto \nu_1^{(\frac{n - \sum_{i=1}^n u_i}{1-\nu_2})} \times (1-\nu_1)^{(\frac{\sum_{i=1}^n u_i - n\nu_2}{1-\nu_2})},$$

onde $\boldsymbol{\theta}_3 = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\lambda}^\top, \nu_1)^\top$. Um interessante método Metropolis–Hastings para ν_2 é descrito em Rosa *et al.* (2003).

Apêndice C

Resultados Adicionais do Capítulo 4

C.1 Etapas do Algoritmo EM

Esta seção descreve alguns detalhes dos passos E e M do algoritmo EM, utilizado para estimar os parâmetros do modelo de regressão linear sob a classe de distribuições misturas de escala skew-normal (SMSN-RM).

• **Passo E:**

A notação usada é a mesma da Seção 4.2.1. De (4.2)-(4.4), temos que

$$\begin{aligned} f(\mathbf{y}_i, t_i, u_i|\boldsymbol{\theta}) &= f(\mathbf{y}_i|t_i, u_i, \boldsymbol{\theta}) f(t_i|u_i, \boldsymbol{\theta}) f(u_i|\boldsymbol{\theta}) \\ &= f(\mathbf{y}_i|\boldsymbol{\theta}) f(u_i|\mathbf{y}_i, \boldsymbol{\theta}) f(t_i|u_i, \mathbf{y}_i, \boldsymbol{\theta}) \end{aligned}$$

e considerando sucessivamente o Lema 3 em Arellano-Valle *et al.* (2005), podemos ver

que

$$\begin{aligned} T_i|U_i = u_i, \mathbf{Y}_i = \mathbf{y}_i, \boldsymbol{\theta} &\sim N_1(u_i^{1/2}\mu_{T_i}, M_T^2)\mathbb{I}_{(0,\infty)}(t_i), \\ \mathbf{Y}_i|\boldsymbol{\theta} &\sim SMSN_p(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H),\ i = 1, \ldots, n \end{aligned}$$

e da Seção 2.2, os valores esperados condicionais de $U_i|\mathbf{Y}_i = \mathbf{y}_i, \boldsymbol{\theta}$ são conhecidos para as distribuições particulares da classe SMSN consideradas. Então, todos os valores esperados condicionais necessários podem ser obtidos percebendo que

$$\mathrm{E}\{U_i^r T_i^s|\mathbf{Y}_i, \boldsymbol{\theta}\} = \mathrm{E}\{U_i^r \mathrm{E}\{T_i^s|U_i, \mathbf{Y}_i\}|\mathbf{Y}_i, \boldsymbol{\theta}\},\ i = 1, \ldots, n.$$

• **Passo M:**

Para estimar o parâmetro $\boldsymbol{\beta}$, temos

$$\frac{\partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})}{\partial \boldsymbol{\beta}} = -\frac{1}{2}\sum_{i=1}^{n} \widehat{u}_i^{(k)}(-2\mathbf{X}_i^\top)\widehat{\boldsymbol{\Gamma}}^{-1(k)}(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k+1)}) + \sum_{i=1}^{n} \widehat{ut}_i^{(k)}(-\mathbf{X}_i^\top)\widehat{\boldsymbol{\Gamma}}^{-1(k)}\widehat{\boldsymbol{\Delta}}^{(k)} = 0.$$

Para estimar o parâmetro $\boldsymbol{\Delta}$, temos

$$\frac{\partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})}{\partial \boldsymbol{\Delta}}\widehat{\boldsymbol{\Gamma}}^{(k)} = \sum_{i=1}^{n} \widehat{ut}_i^{(k)}(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)})\widehat{\boldsymbol{\Gamma}}^{-1(k)}\widehat{\boldsymbol{\Gamma}}^{(k)} - \frac{1}{2}\sum_{i=1}^{n} \widehat{ut^2}_i^{(k)}(2\widehat{\boldsymbol{\Delta}}^{(k+1)})\widehat{\boldsymbol{\Gamma}}^{-1(k)}\widehat{\boldsymbol{\Gamma}}^{(k)} = 0.$$

Estimamos a matriz $\boldsymbol{\Gamma}$ através de sua inversa $\boldsymbol{\Gamma}^{-1}$, lembrando que a solução deve ser simétrica,

$$\begin{aligned} \frac{\partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})}{\partial \boldsymbol{\Gamma}^{-1}} &= \frac{1}{\partial \boldsymbol{\Gamma}^{-1}}\left[-\frac{n}{2}\log|\boldsymbol{\Gamma}| - \frac{1}{2}\sum_{i=1}^{n} tr(\boldsymbol{\Gamma}^{-1}\widehat{\mathrm{A}}_i^{(k)})\right] \\ &= -\frac{n}{2}\left[-2\widehat{\boldsymbol{\Gamma}}^{(k+1)} + diag(\widehat{\boldsymbol{\Gamma}}^{(k+1)})\right] - \frac{1}{2}\sum_{i=1}^{n}\left[(\widehat{A}_i^{(k)} + \widehat{A}_i^{(k)\top}) - diag(\widehat{A}_i^k)\right] = 0, \end{aligned}$$

onde $\widehat{A}_i^{(k)} = \widehat{u}_i^{(k)}\widehat{\boldsymbol{\epsilon}}_i^{(k)}\widehat{\boldsymbol{\epsilon}}_i^{(k)\top} - 2\widehat{ut}_i^{(k)}\widehat{\boldsymbol{\epsilon}}_i^{(k)}\widehat{\boldsymbol{\Delta}}^{(k)\top} + \widehat{ut^2}_i^{(k)}\widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\Delta}}^{\top(k)}$.

C.2 Matriz Informação Observada

Nesta seção, apresentamos os elementos necessários para a obtenção da matriz informação observada, útil para calcular os desvios padrões das estimativas de máxima verossimilhança dos parâmetros do modelo de regressão linear SMSN. Sem perda de generalidade, reparametrizamos $\boldsymbol{\Sigma} = \mathbf{D}^2(\boldsymbol{\alpha})$, onde $\boldsymbol{\alpha} = Vech(\mathbf{D})$, para facilitar a obtenção da matriz informação observada e as derivadas de primeira e segunda ordem de $\log|\mathbf{D}|$, A_i e d_i são obtidas. A notação usada é a mesma da Seção 4.2.2 e para o vetor p-dimensional $\boldsymbol{\rho} = (\rho_1 \ldots, \rho_p)^\top$, usaremos a notação $\dot{\boldsymbol{\Upsilon}}_r = \partial\boldsymbol{\Upsilon}(\boldsymbol{\rho})/\partial\rho_r$, com $r = 1, 2 \ldots, p$. Dessa forma,

- $\mathbf{D}$

$$\frac{\partial^2 \log|\mathbf{D}|}{\partial\alpha_k\partial\alpha_r} = -\mathrm{tr}(\mathbf{D}^{-1}\dot{\mathbf{D}}_s\mathbf{D}^{-1}\dot{\mathbf{D}}_k),$$

- A_i

$$\begin{aligned}
\frac{\partial A_i}{\partial\boldsymbol{\beta}} &= -\mathbf{X}_i^\top\mathbf{D}^{-1}\boldsymbol{\lambda}, \quad \frac{\partial A_i}{\partial\alpha_k} = -\boldsymbol{\lambda}^\top\mathbf{D}^{-1}\dot{\mathbf{D}}_k\mathbf{D}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i), \quad \frac{\partial A_i}{\partial\boldsymbol{\lambda}} = \mathbf{D}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i),\\
\frac{\partial^2 A_i}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} &= \mathbf{0}, \quad \frac{\partial^2 A_i}{\partial\boldsymbol{\beta}\partial\alpha_k} = -\mathbf{X}_i^\top\mathbf{D}^{-1}\dot{\mathbf{D}}_k\mathbf{D}^{-1}\boldsymbol{\lambda}, \quad \frac{\partial^2 A_i}{\partial\boldsymbol{\beta}\partial\boldsymbol{\lambda}^\top} = -\mathbf{X}_i^\top\mathbf{D}^{-1},\\
\frac{\partial^2 A_i}{\partial\alpha_k\partial\alpha_r} &= \boldsymbol{\lambda}^\top\mathbf{D}^{-1}(\dot{\mathbf{D}}_s\mathbf{D}^{-1}\dot{\mathbf{D}}_k + \dot{\mathbf{D}}_k\mathbf{D}^{-1}\dot{\mathbf{D}}_s)\mathbf{D}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i),\\
\frac{\partial^2 A_i}{\partial\alpha_k\partial\boldsymbol{\lambda}^\top} &= -(\mathbf{y}_j - \boldsymbol{\mu}_i)^\top\mathbf{D}^{-1}\dot{\mathbf{D}}_k\mathbf{D}^{-1}, \quad \frac{\partial^2 A_i}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}^\top} = \mathbf{0}
\end{aligned}$$

- d_i

$$\begin{aligned}
\frac{\partial d_i}{\partial \boldsymbol{\beta}} &= -2\mathbf{X}_i^\top \mathbf{D}^{-2}(\mathbf{y}_i - \boldsymbol{\mu}_i), \quad \frac{\partial d_i}{\partial \alpha_k} = -(\mathbf{y}_i - \boldsymbol{\mu}_i)^\top \mathbf{D}^{-1}(\dot{\mathbf{D}}_k \mathbf{D}^{-1} + \mathbf{D}^{-1}\dot{\mathbf{D}}_k)\mathbf{D}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i), \\
\frac{\partial d_i}{\partial \boldsymbol{\lambda}} &= \mathbf{0}, \frac{\partial^2 d_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}_k^\top} = 2\mathbf{X}_i^\top \mathbf{D}^{-2}\mathbf{X}_i, \quad \frac{\partial^2 d_i}{\partial \boldsymbol{\beta} \partial \alpha_k} = 2\mathbf{X}_i^\top \mathbf{D}^{-1}(\dot{\mathbf{D}}_k \mathbf{D}^{-1} + \mathbf{D}^{-1}\dot{\mathbf{D}}_k)\mathbf{D}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i), \\
\frac{\partial^2 d_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\lambda}^\top} &= \mathbf{0}, \quad \frac{\partial^2 d_i}{\partial \alpha_k \partial \boldsymbol{\lambda}^\top} = \mathbf{0}, \quad \frac{\partial^2 d_i}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^\top} = \mathbf{0}, \\
\frac{\partial^2 d_i}{\partial \alpha_k \partial \alpha_s} &= (\mathbf{y}_i - \boldsymbol{\mu})^\top \mathbf{D}^{-1}(\dot{\mathbf{D}}_s \mathbf{D}^{-1}\dot{\mathbf{D}}_k \mathbf{D}^{-1} + \dot{\mathbf{D}}_k \mathbf{D}^{-1}\dot{\mathbf{D}}_s \mathbf{D}^{-1} + \dot{\mathbf{D}}_k \mathbf{D}^{-2}\dot{\mathbf{D}}_s + \dot{\mathbf{D}}_s \mathbf{D}^{-2}\dot{\mathbf{D}}_k \\
&\quad \mathbf{D}^{-1}\dot{\mathbf{D}}_s \mathbf{D}^{-1}\dot{\mathbf{D}}_k + \mathbf{D}^{-1}\dot{\mathbf{D}}_k \mathbf{D}^{-1}\dot{\mathbf{D}}_s)\mathbf{D}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i).
\end{aligned}$$

Apêndice D

Resultados Adicionais do Capítulo 5

Neste apêndice, descrevemos os elementos da matriz hessiana $\ddot{Q}_\theta(\widehat{\boldsymbol{\theta}})$, usando a metodologia de diferenciação matricial descrita em Magnus & Neudecker (1988). A matriz hessiana é útil para obter as medidas de diagnóstico de influência local para o modelo linear misto misturas de escala skew-normal (SMSN-LMM). A seguir, os elementos de $\ddot{Q}_\theta(\widehat{\boldsymbol{\theta}})$ são dados por

$$\begin{aligned}
\frac{\partial^2 Q_{1i}(\boldsymbol{\beta}, \sigma_e^2|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} &= -\frac{\widehat{u}_i}{\sigma_e^2}\mathbf{X}_i^\top\mathbf{R}_i^{-1}\mathbf{X}_i, \\
\frac{\partial^2 Q_{1i}(\boldsymbol{\beta}, \sigma_e^2|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\sigma_e^2} &= -\frac{1}{\sigma_e^4}\mathbf{X}_i^\top\mathbf{R}_i^{-1}\left(-\widehat{u}_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) - \mathbf{Z}_i\widehat{\mathbf{ub}_i}\right), \\
\frac{\partial^2 Q_{1i}(\boldsymbol{\beta}, \boldsymbol{\gamma}|\widehat{\boldsymbol{\theta}})}{\partial\sigma_e^2\partial\sigma_e^2} &= \frac{n_i}{2\sigma_e^4} - \frac{1}{\sigma_e^6}\mathrm{tr}\left(\mathbf{M}_i\mathbf{R}_i^{-1}\right),
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 Q_{2i}(\boldsymbol{\alpha},\boldsymbol{\lambda}|\widehat{\boldsymbol{\theta}})}{\partial\lambda_k\partial\alpha_r} &= \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\lambda_k)\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\alpha_r)(\mathbf{I}_q-\boldsymbol{\Gamma}^{-1}\mathbf{N}_i)\right)\\
&\quad+\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\left(\dot{\mathbf{F}}(r)\dot{\mathbf{G}}(k)\mathbf{F}+\mathbf{F}\dot{\mathbf{G}}(k)\dot{\mathbf{F}}(r)\right)(\mathbf{I}_q-\boldsymbol{\Gamma}^{-1}\mathbf{N}_i)\right)\\
&\quad-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\alpha_r)\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\lambda_k)\boldsymbol{\Gamma}^{-1}\mathbf{N}_i\right)\\
&\quad-\dot{\boldsymbol{\delta}}(k)\mathbf{F}\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\alpha_r)\boldsymbol{\Gamma}^{-1}(\widehat{\mathbf{utb}}_i-\widehat{ut^2}_i\boldsymbol{\Delta})\\
&\quad+\left(\dot{\boldsymbol{\delta}}(k)\dot{\mathbf{F}}(r)\boldsymbol{\Gamma}^{-1}-\boldsymbol{\delta}^\top\dot{\mathbf{F}}(r)\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\lambda_k)\boldsymbol{\Gamma}^{-1}\right)(\widehat{\mathbf{utb}}_i-\widehat{ut^2}_i\boldsymbol{\Delta})\\
&\quad-\widehat{ut^2}_i\boldsymbol{\delta}^\top\dot{\mathbf{F}}(r)\boldsymbol{\Gamma}^{-1}\mathbf{F}\dot{\boldsymbol{\delta}}(k)^\top,\\
\frac{\partial^2 Q_{2i}(\boldsymbol{\alpha},\boldsymbol{\lambda}|\widehat{\boldsymbol{\theta}})}{\partial\alpha_s\partial\alpha_r} &= \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\alpha_s)\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\alpha_r)(\mathbf{I}_q-\boldsymbol{\Gamma}^{-1}\mathbf{N}_i)\right)\\
&\quad-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\left(\dot{\mathbf{F}}(r)\dot{\mathbf{F}}(s)+\dot{\mathbf{F}}(s)\dot{\mathbf{F}}(r)\right)(\mathbf{I}_q-\boldsymbol{\Gamma}^{-1}\mathbf{N}_i)\right)\\
&\quad+\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\left(\dot{\mathbf{F}}(r)\boldsymbol{\delta}\boldsymbol{\delta}^\top\dot{\mathbf{F}}(s)+\dot{\mathbf{F}}(s)\boldsymbol{\delta}\boldsymbol{\delta}^\top\dot{\mathbf{F}}(r)\right)(\mathbf{I}_q-\boldsymbol{\Gamma}^{-1}\mathbf{N}_i)\right)\\
&\quad-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\alpha_r)\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\alpha_s)\boldsymbol{\Gamma}^{-1}\mathbf{N}_i\right)\\
&\quad-\boldsymbol{\delta}^\top\left(\dot{\boldsymbol{\Gamma}}(\alpha_s)\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\alpha_r)+\dot{\mathbf{F}}(r)\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\alpha_s)\right)\boldsymbol{\Gamma}^{-1}(\widehat{\mathbf{utb}}_i-\widehat{ut^2}_i\boldsymbol{\Delta})\\
&\quad-\widehat{ut^2}_i\boldsymbol{\delta}^\top\dot{\mathbf{F}}(r)\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\alpha_s)\boldsymbol{\delta},\\
\frac{\partial^2 Q_{2i}(\boldsymbol{\alpha},\boldsymbol{\lambda}|\widehat{\boldsymbol{\theta}})}{\partial\lambda_l\partial\lambda_k} &= \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\lambda_l)\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\lambda_k)(\mathbf{I}_q-\boldsymbol{\Gamma}^{-1}\mathbf{N}_i)\right)\\
&\quad+\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\mathbf{F}\ddot{\mathbf{G}}(l,k)\mathbf{F}(\mathbf{I}_q-\boldsymbol{\Gamma}^{-1}\mathbf{N}_i)\right)\\
&\quad-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\lambda_l)\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(\lambda_k)\boldsymbol{\Gamma}^{-1}\mathbf{N}_i\right)\\
&\quad-\dot{\boldsymbol{\delta}}(l)\mathbf{F}\boldsymbol{\Gamma}^{1}\dot{\boldsymbol{\Gamma}}(\lambda_k)\boldsymbol{\Gamma}^{-1}(\widehat{\mathbf{utb}}_i-\widehat{ut^2}_i\boldsymbol{\Delta})\\
&\quad+\left(\ddot{\mathbf{G}}(l,k)\mathbf{F}-\dot{\boldsymbol{\delta}}(k)\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Gamma}}(l)\right)\boldsymbol{\Gamma}^{-1}(\widehat{\mathbf{utb}}_i-\widehat{ut^2}_i\boldsymbol{\Delta})-\widehat{ut^2}_i\dot{\boldsymbol{\delta}}(k)\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\delta}}(l),
\end{aligned}
$$

onde $\mathbf{M}_i=\widehat{u}_i(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})^\top-2\mathbf{Z}_i\widehat{\mathbf{ub}}_i(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})^\top+\mathbf{Z}_i\widehat{\mathbf{ub}^2}_i\mathbf{Z}_i^\top$, $\mathbf{N_i}=\widehat{\mathbf{ub}_i^2}-\widehat{\mathbf{utb}}_i\boldsymbol{\Delta}^\top-\boldsymbol{\Delta}\widehat{\mathbf{utb}}_i^\top+\widehat{ut_i^2}\boldsymbol{\Delta}\boldsymbol{\Delta}^\top$, $\dot{\boldsymbol{\Gamma}}(\lambda_k)=-\mathbf{F}\dot{\mathbf{G}}(k)\mathbf{F}$, $\dot{\boldsymbol{\Gamma}}(\alpha_r)=\dot{\mathbf{F}}(r)\mathbf{F}+\mathbf{F}\dot{\mathbf{F}}(r)-\dot{\mathbf{F}}(r)\boldsymbol{\delta}\boldsymbol{\delta}^\top\mathbf{F}-\mathbf{F}\boldsymbol{\delta}\boldsymbol{\delta}^\top\dot{\mathbf{F}}(r)$, $\dot{\mathbf{F}}(r)=\dfrac{\partial\mathbf{F}}{\partial\alpha_r}$, $\dot{\boldsymbol{\delta}}(k)=\dfrac{\partial\boldsymbol{\delta}}{\partial\lambda_k}$, $\dot{\mathbf{G}}(k)=\dfrac{\partial\boldsymbol{\delta}\boldsymbol{\delta}^\top}{\partial\lambda_k}$, $\ddot{\mathbf{G}}(l,k)=\dfrac{\partial^2\boldsymbol{\delta}\boldsymbol{\delta}^\top}{\partial\lambda_l\partial\lambda_k}$ e $\mathbf{I}_q$ denota a matriz identidade $q\times q$, $i=1,\ldots,n$, $r,s=1,\ldots,\dim(\boldsymbol{\alpha})$ e $l,k=1,\ldots,q$.

Referências Bibliográficas

Abramowitz, M. & Stegun, I. A. (1972). *Handbook of mathematical functons with formulas, graphis and mathematical tables*. New York, Dover.

Anderson, J. (2001). On the normal inverse gaussian stochastic volatility model. *Journal of Business & Economic Statistics*, **19**(1), 44–54.

Andrews, D. F. & Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B*, **36**, 99–102.

Arellano-Valle, R. B. & Genton, M. G. (2005). On fundamental skew distributions. *Journal of Multivariate Analysis*, **96**, 93–116.

Arellano-Valle, R. B., Bolfarine, H. & Lachos, V. H. (2005). Skew-normal linear mixed models. *Journal of Data Science*, **3**, 415–438.

Arellano-Valle, R. B., Bolfarine, H. & Lachos, V. H. (2007). Bayesian inference for skew-normal linear mixed models. *Journal of Applied Statistics*, **34**, 663–682.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171–178.

Azzalini, A. & Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society*, **61**, 579–602.

Azzalini, A. & Capitanio, A. (2003). Distributions generated and perturbation of symmetry with emphasis on the multivariate skew-t distribution. *Journal of the Royal Statistical Society, Series B*, **61**, 367–389.

Azzalini, A. & Dalla-Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**(4), 715–726.

Azzalini, A. & Genton, M. G. (2008). Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review*, **76**, 106–129.

Barndorff-Nielsen, O. E. (1997). Normal inverse gaussian distributions and stochasctic volatility modelling. *Scandinavian Journal of Statistics*, **24**, 1–14.

Barnett, V. D. (1969). Simultaneous pairwise linear structural relationships. *Biometrics*, **25**, 129–142.

Beale, E. M. L. & Little, R. J. A. (1975). Missing values in multivariate analysis. *Journal of the Royal Statistical Society, Series B*, **37**, 129–146.

Berkane, M., Kano, Y. & Bentler, P. M. (1994). Pseudo maximum likelihood estimation in elliptical theory: effects of misspecification. *Computational Statistics and Data Analysis*, **18**, 255–267.

Branco, M. D. & Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, **79**, 99–113.

Brownlee, K. A. (1960). *Statistical theory and methodology in science and engineering*. John Wiley, New York.

Cambanis, S., Fotopoulos, S. B. & He, L. (2000). On the conditional variance for scale mixtures of normal distributions. *Journal of Multivariate Analysis*, **74**, 163–92.

Chang, Y. P., Hung, M. C., Liu, H. & Jan, J. F. (2005). Testing symmetry of a NIG distribution. *Communication in Statistics–Simulation and Computation*, **34**, 851–862.

Chatterjee, S. & Hadi, A. S. (1988). *Sensivity analysis in linear regression*. John Wiley, New York.

Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society, Series B*, **48**, 133–169.

Cook, R. D. & Weisberg, S. (1982). *Residuals and influence in regression*. Chapman & Hall/CRC, Boca Raton, FL.

De Castro, M., Galea, M. & Bolfarine, H. (2007). Local influence assessment in heteroscedastic measurement error models. *Computational Statistics and Data Analysis*, **52**, 1132–1142.

De La Cruz, R. (2008). Bayesian non-linear regression models with skew-elliptical errors: applications to the classification of longitudinal profiles. *Computational Statistics and Data Analysis*, **53**(2), 436–449.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1–22.

DiCiccio, T. J. & Monti, A. C. (2004). Inferential aspects of the skew exponential power distribution. *Journal of the American Statistical Association*, **99**, 439–450.

Dodge, Y. (1996). *The guinea pig of multiple regression. In Robust Statistics, Data Analysis, and Computer Intensive Methods: In Honor of Peter Huber's 60th Birthday, Lecture Notes in Statistics* 109. New York, Springer-Verlag.

Embrechts, P. (1983). A property of the generalized inverse gaussian distribution with some applications. *Journal of Applied Probability*, **20**(3), 537–544.

Fang, K. T., Kotz, S. & Ng, K. W. (1990). *Simetric multivariate and related distribuitions*. Chapman & Hall, London.

Ferreira, C. S. (2008). *Inferência e diagnóstico em modelos assimétricos*. Tese de doutorado, Departamento de Estatística, IME-USP.

Fotopoulos, S. B., Jandhyala, V. K. & Chen, K. H. (2007). Non-linear properties of conditional returns under scale mixtures. *Computational Statistics and Data Analysis*, **51**, 3041–3056.

Fuller, W. A. (1987). *Measurement error models*. John Wiley, New York.

Fung, W. K., Wei, Z. Y. & He, X. (2002). Influence diagnostics and outlier tests for semiparametric mixed models. *Journal of the Royal Statistical Society, Series B*, **64**, 565–579.

Galea-Rojas, M., Paula, G. A. & Uribe-Opazo, M. (2003). On influence diagnostic in univariate elliptical linear regression models. *Statistical Papers*, **44**, 23–45.

Galea-Rojas, M., Paula, G. A. & Cysneiros, F. J. A. (2005). On diagnostics in symmetrical nonlinear models. *Statistics and Probability Letters*, **73**, 459–467.

Genton, M. G. (2004). *Skew-elliptical distributions and their applications: a journey beyond normality*. Chapman & Hall, London.

Genton, M. G., He, L. & Liu, X. (2001). Moments of skew-normal random vectors and their quadratic forms. *Statistics and Probability Letters*, **51**, 319–325.

Ghidey, W., Lesaffre, E. & Eilers, P. (2004). Smooth random effects distribution in a linear mixed model. *Biometrics*, **60**, 945–953.

Ghosh, P., Bayes, C. R. & Lachos, V. H. (2009). A robust bayesian approach to null intercept measurement error model with application to dental data. *Computational Statistics and Data Analysis*, **53**, 1066–1079.

Grubbs, F. E. (1983). Grubbs estimator. *Encyclopedia of Statistical Sciences*, **3**, 542–548.

Hill, M. A. & Dixon, W. J. (1982). Robustness in real life: a study of clinical laboratory data. *Biometrics*, **38**. 377-396.

Jara, A., Quintana, F. & Martín, E. S. (2008). Linear mixed models with skew-elliptical distributions: a bayesian approach. *Computational Statistics and Data Analysis*, **52**. 5033-5045.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994). *Continuous univariate distributions*, volume 1. John Wiley, New York.

Jorgensen, B. (1982). *Statistical properties of the generalized inverse gaussian distribtion*. Springer Verlag, Heidelberg.

Karlis, D. (2002). An EM type algoruthm for maximum likelihood estimation of the normal-inverse gaussian distribution. *Statistics and Probability Letters*, **57**, 43–52.

Kim, H. M. (2008). A note on scale mixtures of skew normal distribution. *Statistics and Probability Letters*, **78**. 1694-1701.

Lachos, V. H. (2004). *Modelos lineares mistos assimétricos*. Tese de doutorado, Departamento de Estatística, IME-USP.

Lachos, V. H. & Vilca, L. F. (2007). Skew-normal/ independent distributions, with applications. RT-IMECC 02, IMECC-UNICAMP.

Lachos, V. H., Bolfarine, H., Vilca, L. F. & Galea-Rojas, M. (2005). Estimation and influence diagnostic for structural comparative calibration models under the skew-normal distribution. RT-MAE 12, IME-USP.

Lachos, V. H., Bolfarine, H., Arellano-Valle, R. B. & Montenegro, L. C. (2007a). Likelihood based inference for multivariate skew-normal regression models. *Communications in Statistics–Theory and Methods*, **36**, 1769–1786.

Lachos, V. H., Vilca, L. F. & Galea, M. (2007b). Influence diagnostics for the grubbs' model. *Statistical Papers*, **48**, 419–436.

Lachos, V. H., Montenegro, L. C. & Bolfarine, H. (2008). Inference and influence diagnostics for skew-normal null intercept measurement errors models. *Journal of Statistical Computation and Simulation*, **78**. 395-419.

Lachos, V. H., Vilca, L. F., Bolfarine, H. & Ghosh, P. (2009). Robust multivariate measurement error models with scale mixtures of skew-normal distributions. *Statistics*, **doi: 10.1080/02331880903236926**.

Lachos, V. H., Ghosh, P. & Arellano-Valle, R. B. (2010). Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica*, **20**. SS-08-045.

Lange, K. L. & Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics*, **2**, 175–198.

Lange, K. L., Little, R. & Taylor, J. (1989). Robust statistical modeling using t distribution. *Journal of the American Statistical Association*, **84**, 881–896.

Lee, S. Y. & Xu, L. (2004). Influence analysis of nonlinear mixed-effects models. *Computational Statistics and Data Analysis*, **45**, 321–341.

Lee, S. Y., Lu, B. & Song, X. Y. (2006). Assessing local influence for nonlinear structural equation models with ignorable missing data. *Computational Statistics and Data Analysis*, **50**, 1356–1377.

Lesaffre, E. & Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, **54**, 570–582.

Lin, T. C. & Lin, T. (2009). Supervised learning of multivariate skew-normal mixture models with missing information. *Computational Statistics*, **doi:10.1007/s00180-009-0169-5**.

Lin, T. I. (2009). Robust mixture modeling using multivariate skew-t distributions. *Statistics and Computing*, **doi: 10.1007/s11222-009-9128-9**.

Lin, T. I. & Lee, J. C. (2008). Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. *Statistics in Medicine*, **27**, 1490–1507.

Lin, T. I., Lee, J. C. & Ho, H. J. (2006). On fast supervised learning for normal mixture models with missing information. *Pattern Recogn*, **39**, 1177–1187.

Lin, T. I., Ho, H. J. & Chen, C. L. (2009a). Analysis of multivariate skew-normal models with incomplete data. *Journal of Multivariate Analysis*, **doi: 10.1016/j.jmva.2009.07.005**.

Lin, T. I., Ho, H. J. & Shen, P. S. (2009b). Computationally efficient learning of multivariate t mixture models with missing information. *Computational Statistics*, **24**, 375–392.

Little, R. J. A. & Rubin, D. B. (1987). *Statistical Analysis With Missing Data*. John Wiley & Sons, New York.

Liu, C. (1999). Efficient ML estimation of the multivariate normal distribution from incomplete data. *Journal of Multivariate Analysis*, **69**, 206–217.

Lucas, A. (1997). Robustness of the student t based m-estimator. *Communications in Statistics–Theory and Methods*, **26**, 1165–1182.

Lyu, S. & Simoncelli, P. (2008). Modelling multiscale subbands of photographic images with fields of gaussian scale mixtures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **31**(4), 693–706.

Madan, D. B., Carr, P. P. & Chang, E. C. (1998). The variance-gamma process and option pricing. *European Finance Review*, **2**, 79–105.

Magnus, J. R. & Neudecker, H. (1988). *Differential calculus with applications in statistics and econometrics*. John Wiley & Sons, New York.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **657**, 519–530.

Meng, X. & Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika*, **81**, 633–648.

Montenegro, L. C., Bolfarine, H. & Lachos, V. (2009a). Local influence analysis of skew-normal linear mixed models. *Communication in Statistics–Theory and Methods*, **38**, 484–496.

Montenegro, L. C., Lachos, V. H. & Bolfarine, H. (2009b). Inference for a skew extension of the grubbs model. *Statistical Papers*, **doi: 10.1007/s00362-008-0157-9**.

Montenegro, L. C., Bolfarine, H. & Lachos, V. H. (2009c). Influence diagnostics for a skew extension of the grubbs measurement error model. *Communication in Statistics–Simulation and Computation*, **38**(4), 667–681.

Nielsen, S. F. (1997). Inference and missing data: asymptotic result. *Scandinavian Journal of Statistics*, **24**, 261–274.

O'hagan, A. & Leornard, T. (1976). Bayes estimation subject to uncertainty about constraints. *Biometrika*, **63**, 201–203.

Ortega, E. M., Bolfarine, H. & Paula, G. A. (2003). Influence diagnostics in generalized log-gamma regression models. *Computational Statistics and Data Analysis*, **42**, 165–186.

Osorio, F. (2006). *Diagnóstico de influência em modelos elípticos com efeitos mistos*. Tese de doutorado, Departamento de Estatística, IME-USP.

Osorio, F., Paula, G. A. & Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics and Data Analysis*, **51**, 4354–4368.

Osorio, F., Paula, G. A. & Galea, M. (2009). On estimation and influence diagnostics for the grubbs' model under heavy-tailed distributions. *Computational Statistics and Data Analysis*, **53**, 1249–1263.

Pinheiro, J. C. & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Springer, New York.

Pinheiro, J. C., Liu, C. H. & Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using a multivariate t-distribution. *Journal of Computational and Graphical Statistics*, **10**, 249–276.

Rosa, G. J. M., Padovani, C. R. & Gianola, D. (2003). Robust linear mixed models with normal/independent distributions and bayesian MCMC implementation. *Biometrical Journal*, **45**, 573–590.

Rosa, G. J. M., Gianola, D. & Padovani, C. R. (2004). Bayesian longitudinal data analysis with mixed models and thick-tailed distributions using MCMC. *Journal of Applied Statistics*, **31**(7), 855–873.

Sahu, S. K., Dey, D. K. & Branco, M. D. (2003). A new class of multivariate distributions with applications to bayesian regression models. *The Canadian Journal of Statistics*, **31**, 129–150.

Salberg, A. B., Swami, A., Oigard, T. A. & Hanssen, A. (2001). The normal inverse gaussian distribution as a model for MUI. *Asilomar Conference on Signals, Systems and Computers*, **2**, 1484–1488.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & Der, L. V. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical, Series B*, **64**, 583–639.

Styan, G. P. H. (1973). Hadamard products and multivariate statistical analysis. *Linear Algebra and its Applications*, **6**, 217–240.

Tan, F. & Peng, H. (2005). The slash and the skew-slash student t distributions. *Elsevier Science*. Preprint.

Teixeira, V. M. M. (2006). *Distribuições hiperbólicas generalizadas: aplicações ao mercado português*. Tese de mestrado, Departamento de Matemática e Engenharias, Universidade da Madeira.

Thabane, L. & Haq, M. S. (1999). Prediction from a model using a generalized inverse gaussian prior. *Statistical Papers*, **40**, 175–184.

Verbeke, G. & Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, **91**, 217–221.

Verbeke, G. & Molenberghs, G. (2001). *Linear mixed models for longitudinal data*. Springer, New York.

Wang, J. & Genton, M. G. (2006). The multivariate skew-slash distribution. *Journal of Statistical Planning and Inference*, **136**, 209–220.

Wang, J., Boyer, J. & Genton, M. (2004). A skew-symmetric representation of multivariate distributions. *Statistica Sinica*, **14**, 1259–1270.

Wei, G. C. G. & Tanner, M. A. (1990). A monte carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, **85**, 699–704.

Yamaguchi, K. (2001). Analysis of repeated measurements with multivariate t or contaminated multivariate normal errors. *Bulletin of the Computational Statistics of Japan*, **3**, 1–18.

Zeller, C. B. (2006). *Modelo de grubbs em grupos*. Tese de mestrado, Departamento de Estatística, IMECC-UNICAMP.

Zellner, A. (1976). Bayesian and non-bayesian analysis of the regression model with multivariate student-t error term. *Journal of the American Statistical Association*, **71**, 400–405.

Zhang, D. & Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*, **57**, 795–802.

Zhou, T. & He, X. (2008). Three-step estimation in linear mixed models with skew-t distributions. *Journal of Statistical Planning and Inference*, **138**, 1542–1555.

Zhu, H. & Lee, S. (2001). Local influence for incomplete-data models. *Journal of the Royal Statistical Society, Series B*, **63**, 111–126.

Zhu, H. & Lee, S. (2003). Local influence for generalized linear mixed models. *The Canadian Journal of Statistics*, **31**, 293–309.

Zhu, L. X. & Neuhaus, G. (2000). Nonparametric monte carlo tests for multivariate distribution. *Biometrika*, **87**(4), 919–928.

Zio, M., Guarnera, U. & Luzi, O. (2007). Imputation through finite gaussian mixture models. *Computational Statistics and Data Analysis*, **51**, 5305–5316.