Modelos Lineares Mistos Assimétricos

Víctor Hugo Lachos Dávila

Tese apresentada

ao

Instituto de Matemática e Estatística

da

Universidade de São Paulo

para

obtenção do grau

de

Doutor em Estatística

Área de Concentração: Estatística

Orientador: **Prof. Dr. Heleno Bolfarine**

Co-orientador: **Prof. Dr. Reinaldo Boris Arellano-Valle**

- São Paulo, Outubro de 2004 -

Modelos Lineares Mistos

Assimétricos

Este Exemplar corresponde à redação final da tese, devidamente corrigida e defendida por Víctor Hugo Lachos Dávila e aprovada pela comissão julgadora.

São Paulo, 14 de março de 2005.

Banca Examinadora:

- . Prof. Dr. Heleno Bolfarine (Presidente) IME-USP
- . Profa. Dra. Márcia D'Élia Branco IME-USP
- . Prof. Dr. Josemar Rodrigues UFSCar
- . Prof. Dr. Filidor Edilson Vilca Labra UNICAMP
- . Prof. Dr. Mário de Castro Andrade Filho ICMC - USP

*À minha esposa Rosângela, pela
compreensão, dedicação e apoio
nesta caminhada*

*Os grandes logros na vida são produtos
de pequenos esforços do dia a dia.*

Anônimo

Agradecimentos

- Ao Professor Heleno Bolfarine pela confiança, sugestões, apoio e pela excelente orientação durante a elaboração deste trabalho. Admiro sua capacidade de pesquisa, assim como, sua presteza na colaboração com os outros.
- Ao Professor Reinaldo B. Arellano-Valle pela co-orientação, seu aporte foi fundamental na obtenção dos resultados dos Capítulos 3 e 4.
- Ao Professor Filidor Vilca Labra pela sua amizade, seu aporte como orientador durante o mestrado foi importante para minha formação e inclinação pela pesquisa.
- Especiais agradecimentos a minha esposa, pelo apoio; as minhas filhas: Thuany e Stephany, pelo carinho e momentos de felicidade que sempre me concedem; a minha mãe **Rosa** pelo incentivo constante e por nunca medir esforços na tarefa de me-educar e também a meus irmãos, eternos companheiros: Roque e Alberto pelo apoio e preocupação constante.
- Aos membros da banca examinadora, Prof. Mário de Castro e Prof. Josemar Rodrigues pelas sugestões e revisão ortográfica deste trabalho.
- À CAPES pelo apoio financeiro e ao curso de pós-graduação do IME-USP, pelas condições de trabalho.
- Aos professores do Departamento de Estatística IME-USP, pelos ensinamentos concedidos.
- A todos os amigos da pós que dividiram estes três anos de estudo e a todos aqueles que de alguma forma, colaboraram na minha formação acadêmica e na execução da tese.
- E finalmente, a Deus pela capacidade e dádiva da vida.

Muito obrigado
Víctor Hugo

Resumo

Modelos lineares mistos tem sido freqüentemente usados na análise de dados onde as respostas são agrupadas, pelo fato de serem flexíveis para modelar a correlação entre e intra-indivíduos (ou grupos). A normalidade (simetria) dos efeitos e erros aleatórios é uma suposição rotineira em modelos lineares mistos, que pode ser não realista e obscurecer importantes características da variação entre e intra - indivíduos (ou grupos). Neste trabalho relaxamos a suposição de normalidade considerando que, tanto os erros como os efeitos aleatórios, seguem uma distribuição normal - assimétrica, que inclui a distribuição normal como caso especial e fornece flexibilidade em capturar uma ampla variedade de comportamentos não normais, por simplesmente adicionar um parâmetro que controla o grau de assimetria. A densidade marginal das quantidades observadas é encontrada e mostramos que tem forma fechada, de modo que inferências podem ser abordadas usando programas computacionais conhecidos (R, S-plus, Matlab) e técnicas de otimização padrão. Explorando propriedades estatísticas do modelo considerado implementamos o algoritmo EM que fornece algumas vantagens sobre a maximização direta da função log - verossimilhança. Apresentamos também, para esta distribuição normal - assimétrica multivariada, vários resultados relacionados com a teoria da distribuição das formas quadráticas, transformações lineares, densidade marginal e condicionamento. Em um segundo estagio do trabalho, usando uma segunda versão de distribuição normal - assimétrica multivariada, os modelos lineares mistos normal assimétricos bayesianos são definidos e procedimentos relacionados com o método Monte Carlo via cadeias de Markov (MCMC) são apresentados fazendo da inferência bayesiana uma alternativa viável para tais modelos. Em ambos os casos, resultados de estudo de simulação e aplicações a conjuntos de dados reais são fornecidos mostrando que os critérios de informação padrão, tais como AIC, BIC e HQ podem ser usados para detectar afastamentos da normalidade (simetria). Finalmente, apresentamos métodos para estimação em modelos lineares mistos com erros nas variáveis, baseados na função escore corrigido de Nakamura (1990), simulação - extrapolação (SIMEX) de Stefanski e Cook (1995) e máxima verossimilhança. Um estudo de simulação comparando os métodos SIMEX e escore corrigido é apresentado.

Abstract

Linear mixed models have been widely used in analysis of data where responses are clustered around some random effects, because they are more flexible to modelling the correlation within-subject, often present in this type of data. Normality (symmetric) of the random effects and the within-subject errors is a routine assumption for the linear mixed model, but it may be unrealistic, obscuring important features of among- and within-subjects variation. We relax this assumption by considering that the random effects and model errors follow a skew-normal distributions, which includes normality as a special case and provides flexibility in capturing a broad range of non-normal behavior. The marginal distribution for the observed quantity is derived which is expressed in closed form, so inference may be carried out using existing statistical software and standard optimization techniques. By exploring statistical properties of the model considered, we implement an EM-type algorithm which seem to provide some advantages over a direct maximization of the likelihood. In addition, we derived for this skew-normal distribution several result associated with the distributional theory of linear, marginal, conditional and quadratic forms. In a second stage, by introducing also a new class of skew-normal distributions, a new version of skew-normal linear mixed model is defined. Then, the MCMC scheme is derived which makes a bayesian inference a viable alternative for such models. In both cases, results of simulation studies and applications to real data sets are provided demonstrating that standard information criteria may be used to detect departures from normality. Finally, we present methods for estimation in linear mixed models with errors-in-variables, based in the corrected escore function of Nakamura (1990), simulation-extrapolation of Stefanski and Cook (1995) and maximum likelihood approach, a simulation study is also reported.

Conteúdo

1	Introdução	1
1.1	Motivação	1
1.2	Uma Breve Revisão de MLM	3
1.2.1	Algoritmo EM para EMV	4
1.2.2	Inferência Baseada nos EMV	7
1.2.3	Estimação sob Restrições de Igualdade no Parâmetro β	9
1.3	Objetivo do Trabalho	11
1.4	Organização do Trabalho	12
2	A Distribuição Normal-Assimétrica	14
2.1	Introdução	14
2.2	A Distribuição Normal-Assimétrica Univariada	16
2.3	A Distribuição Normal-Assimétrica Multivariada	20
2.3.1	Função Geratriz de Momentos	25
2.3.2	Momentos	31
2.3.3	Distribuição Marginal e Independência	33
2.3.4	Distribuições Condicionais	36
2.3.5	Distribuição Normal-Assimétrica Fundamental	39
2.4	Aplicação a Modelos de Regressão Multivariados	42
2.4.1	Estudo de Simulação	47
2.4.2	Aplicação	52
3	Modelos Lineares Mistos Normal-Assimétrico: Enfoque Clássico	55
3.1	Introdução	55
3.2	O Modelo com Interceptos Aleatórios Normal-Assimétrico	57
3.2.1	Algoritmo EM para encontrar os EMV no MIASN	61
3.3	O Modelo Linear Misto Normal-Assimétrico	63

3.3.1	Algoritmo EM quando $\lambda_{e1} = \dots = \lambda_{em} = 0$	66
3.3.2	Algoritmo EM quando $\lambda_b = 0$	68
3.4	Estudo de Simulação	70
3.5	Aplicação	73
4	Modelos Lineares Mistos Normal-Assimétrico: Enfoque Bayesiano	76
4.1	Introdução	76
4.2	O Modelo com Interceptos Aleatórios Normal-Assimétrico Bayesiano	78
4.2.1	Modelagem Bayesiana para o MIASN	79
4.2.2	O Esquema MCMC	80
4.2.3	Aplicação	84
4.3	O Modelo Linear Misto Normal-Assimétrico Bayesiano	89
4.3.1	Modelagem Bayesiana para o MLMSN	93
4.3.2	O Esquema MCMC	95
4.4	Estudo de Simulação	97
4.5	Aplicação	101
5	Modelo Linear Misto com Erros de Medida	108
5.1	Método de Simulação-Extrapolação	110
5.1.1	Procedimento de Estimação	110
5.1.2	Propriedades Assintóticas do Estimador	111
5.2	Método do Escore Corrigido	112
5.2.1	Procedimento de Estimação	112
5.2.2	Resultados Assintóticos	113
5.3	Aplicação do EC e SIMEX no MLMeM	114
5.3.1	Comparando os Métodos EC e SIMEX através de Simulação	118
5.4	Algoritmo EM no MLMeM Estrutural	122
5.5	O MLMeM Estrutural Normal-Assimétrico	125
6	Considerações Finais	129
6.1	Conclusões	129
6.2	Perspectivas para Pesquisas Futuras	130
A	Resultados de Álgebra Matricial	134

Capítulo 1

Introdução

1.1 Motivação

O modelo linear misto (MLM) é uma estrutura bastante usada para analisar dados de medidas repetidas, dados agrupados e dados longitudinais (entre outros), sendo de grande aplicabilidade em pesquisas de diversas áreas como agricultura, biologia, economia, geofísica e ciências sociais (veja, por exemplo, Diggle et al., 1994). A crescente popularidade destes modelos é explicada pela flexibilidade que eles oferecem para modelar a correlação entre e intra-unidades amostrais, freqüentemente presente em dados longitudinais (Laird & Ware, 1982) e dados agrupados (Henderson, 1984), pela capacidade de modelar dados balanceados e desbalanceados e pela existência de programas computacionais eficientes e confiáveis para ajustá-los (Wolfinger, Tobias & Sall, 1991). O MLM tem como casos particulares o modelo linear clássico, o modelo de componentes de variância e também os modelos hierárquicos (multiníveis).

Uma suposição rotineira, na versão clássica do modelo, é que tanto os erros como os efeitos aleatórios são normalmente distribuídos. Sob estas suposições, inferências para os parâmetros fixos e efeitos aleatórios do modelo podem ser tratadas usando uma ampla variedade de programas, tais como *SAS proc mixed* (Littell, Milliken, Stroup & Wolfinger, 1996) e *S-plus lme* (Pinheiro & Bates, 2000). Apesar deste modelo oferecer uma grande flexibilidade para modelar estes efeitos (Verbeke & Lesaffre, 1997), este sofre da mesma falta de robustez como outros modelos estatísticos baseados na distribuição gaussiana e pode ser demasiadamente restritivo ao fornecer uma representação adequada da estrutura que está presente em dados de medidas repetidas ou dados agrupados. Por exemplo,

densidade assimétrica estimada pode sugerir a exclusão de importantes covariáveis e revelar características críticas da heterogeneidade inerente entre sujeitos. Do ponto de vista prático, o método mais adotado para alcançar a normalidade é a transformação de variáveis que funciona bem em muitos casos. Embora esta metodologia possa fornecer resultados empíricos razoáveis, esta deve ser evitada se um modelo teórico mais conveniente for encontrado (veja Azzalini & Capitanio, 1999 ou Capítulo 2). Assim, considerável esforço tem sido direcionado para relaxar a suposição de normalidade e, conjuntamente, estimar a densidade dos efeitos aleatórios e parâmetros do modelo.

Uma alternativa interessante à técnica apresentada acima é ajustar MLM's baseados em famílias paramétricas de distribuições multivariadas que representem desvios locais da família normal de distribuições, isto é, gostaríamos de tratar distribuições que exibam um bom comportamento, similar à distribuição normal, e que possam estar arbitrariamente próximas à distribuição normal dependendo de parâmetros convenientes. Neste sentido, a família normal-assimétrica de distribuições que é analiticamente tratável, acomoda valores práticos de assimetria e curtose e estritamente inclui a distribuição normal, pode ser usada para modelar dados, fazer análises estatísticas e robustificar os métodos da teoria normal. Assim, muitos autores, como Azzalini & Capitanio (1999), Azzalini & Kotz (2002), Capitanio et al. (2002), Sahu & Dey (2001) e Sahu et al. (2003) estudaram modelos de regressão com distribuições assimétricas, mas não existe literatura no contexto de modelos lineares com efeitos mistos. Neste trabalho propomos um método que é particularmente interessante para MLM por assumir que tanto os erros como os efeitos aleatórios seguem uma distribuição normal-assimétrica. Este enfoque oferece a vantagem de fornecer estimadores e algoritmos mais eficientes, assim como uma interpretação prática dos parâmetros.

Por outro lado, variáveis independentes, ou covariáveis em modelos lineares são frequentemente medidas com erro (Fuller, 1987; Carroll et al., 1995). Assim, a combinação de efeitos aleatórios e erros de medida para modelos lineares é também um tópico de crescente interesse na literatura estatística, pois os estimadores de máxima verossimilhança (EMV) são geralmente inconsistentes. Por exemplo, Davidian & Giltinan (1995) em um de seus problemas abertos relacionados aos MLM's mencionam: (Capítulo 12, pag. 328.) "é freqüente em casos práticos que os valores das covariáveis em indivíduos sejam medidas com erros e estudos inferenciais nestes modelos ainda não estão desenvolvidos adequadamente". Erros de medida são problemas freqüentes em diversas áreas de pesquisa e existe

um consenso sobre a importância de incorporá-los na modelagem estatística. Nesta linha de pesquisa podemos citar Wang et al. (1998) e Lin & Carroll (1999), que investigam o vício dos parâmetros estimados e testes de componentes de variância em modelos lineares generalizados mistos com erros de medida usando o método de simulação-extrapolação (SIMEX). Trabalhos relacionados a métodos para dados longitudinais com erros de medida foram publicados por Palta & Lin (1996). Recentemente, Xu-Ping et al. (2002) apresentam um método unificado para estimação em MLM com erros de medida, baseado na função escore corrigida (EC) de Nakamura (1990). Neste trabalho dedicamos um capítulo completo ao estudo inferencial de MLM com erros de medida (MLMeM) estabelecendo um paralelo entre os métodos SIMEX e EC e apresentando outros resultados interessantes no contexto de MLMeM estruturais sob distribuição normal e normal-assimétrica. A seguir, apresentamos uma revisão inferencial de MLM com o intuito de introduzir o modelo e estabelecer parte da notação que será usada no decorrer desta tese.

1.2 Uma Breve Revisão de MLM

O modelo linear misto mais freqüentemente encontrado na literatura para respostas contínuas foi proposto por Laird & Ware (1982) e é expresso como

$$\mathbf{Y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{b}_j + \boldsymbol{\epsilon}_j, \quad j = 1, \dots, m, \quad (1.1)$$

onde $\mathbf{Y}_j$ de dimensão $(n_j \times 1)$ é um vetor de respostas dos m sujeitos ou indivíduos, $\mathbf{X}_j$ de dimensão $(n_j \times p)$ é a matriz planejamento correspondente ao efeito fixo, $\boldsymbol{\beta}$ de dimensão $(p \times 1)$ é o vetor dos coeficientes da regressão da média populacional, chamado de efeitos fixos, $\mathbf{Z}_j$ de dimensão $(n_j \times q)$ é a matriz planejamento correspondente ao vetor de efeitos aleatórios $\mathbf{b}_j$, de dimensão $(q \times 1)$, e $\boldsymbol{\epsilon}_j$ de dimensão $(n_j \times 1)$ é o vetor de erros aleatórios. Tipicamente assume-se que os efeitos aleatórios $\mathbf{b}_j$ e os componentes residuais $\boldsymbol{\epsilon}_j$ são independentes com

$$\mathbf{b}_j \stackrel{\text{iid}}{\sim} N_q(\mathbf{0}, \mathbf{D}) \text{ e } \boldsymbol{\epsilon}_j \stackrel{\text{ind.}}{\sim} N_{n_j}(\mathbf{0}, \boldsymbol{\psi}_j), \quad (1.2)$$

$j = 1, \dots, m$, onde $\mathbf{D} = \mathbf{D}(\boldsymbol{\gamma})$ e $\boldsymbol{\psi}_j = \boldsymbol{\psi}_j(\boldsymbol{\alpha})$ são matrizes de dispersão, usualmente associadas com a variabilidade entre e intra-unidades amostrais sendo dependentes de um número desconhecido e reduzido de parâmetros $\boldsymbol{\gamma}$ e $\boldsymbol{\alpha}$, respectivamente. Dependendo do contexto, várias suposições podem ser feitas com relação à matriz de covariâncias $\boldsymbol{\psi}_j$

que usualmente depende de j através de suas dimensões. Com dados agrupados supõe-se, tipicamente, que $\psi_j = \sigma_e^2 \mathbf{I}_{n_j}$, $j = 1, \dots, m$, refletindo a suposição que sub-unidades de um mesmo grupo têm erros permutáveis. Correlações permutáveis são comuns em aplicações de dados longitudinais e também outras formas tais como AR(1) (Box, Jenkins & Reinsel, 1994) e em banda (*banded*) onde ψ_j depende de um número pequeno de parâmetros. A seguir, assumiremos que $\psi_j = \sigma_e^2 \mathbf{I}_{n_j}$, $j = 1, \dots, m$, neste contexto o modelo (1.1)-(1.2) é chamado de modelo de independência condicional.

Note que sob as suposições consideradas acima, o modelo pode ser representado hierarquicamente como

$$\mathbf{Y}_j | \mathbf{b}_j \stackrel{\text{ind.}}{\sim} N_{n_j}(\mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \mathbf{b}_j, \sigma_e^2 \mathbf{I}_{n_j}), \quad (1.3)$$

$$\mathbf{b}_j \stackrel{\text{iid}}{\sim} N_q(\mathbf{0}, \mathbf{D}), \quad j = 1, \dots, m. \quad (1.4)$$

O principal interesse é fazer inferência sobre o vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T, \sigma_e^2)^T$, que fornece informação sobre a relação entre $\mathbf{Y}$ e $\mathbf{X}$ e também sobre o grau de dependência entre indivíduos de um mesmo grupo (ou medidas em um mesmo indivíduo). O MLM é tipicamente simples de analisar, seja através de expressões diretas (tal como ANOVA) no caso balanceado ou através das equações de Henderson na situação desbalanceada.

Os métodos estimação mais difundidos para o vetor de parâmetros $\boldsymbol{\theta}$ em MLM são máxima verossimilhança (EMV) e máxima verossimilhança restrita (EMVR, Lindstrom & Bates, 1988); na prática EMV e EMVR não diferem muito. EMVR dos componentes de variância-covariância são usualmente preferíveis aos EMV, a razão básica é que os EMVR, ao menos no caso balanceado, são não viciados, enquanto que os EMV são. Se os EMVR são sempre não viciados, não é conhecido. Aqui, restringimos nossa atenção a EMV usando o algoritmo EM, o qual é detalhado a seguir.

1.2.1 Algoritmo EM para EMV

Nesta seção consideramos a EMV do vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T, \sigma_e^2)^T$ no modelo normal linear de efeitos mistos definido em (1.1)-(1.2), usando o algoritmo EM (Dempster, Laird & Rubin, 1977). Este método é um algoritmo iterativo para encontrar os EMV em modelos com dados incompletos. Mais especificamente, $\mathbf{Y}_{obs}$ denota os dados observados e $\mathbf{Y}_{mis}$ denota os dados não observáveis, com dados completos $\mathbf{Y}_{com} = (\mathbf{Y}_{obs}, \mathbf{Y}_{mis})$, onde

$\mathbf{Y}_{obs}$ é acrescido de $\mathbf{Y}_{mis}$. Denotamos por $f(\mathbf{Y}_{com}|\boldsymbol{\theta})$ a função de verossimilhança dos dados completos, $\boldsymbol{\theta} \in \Theta$, por $\ell_c(\boldsymbol{\theta}|\mathbf{Y}_{com}) = \log(f(\mathbf{Y}_{com}|\boldsymbol{\theta}))$ a função log-verossimilhança e por $Q(\boldsymbol{\theta}|\boldsymbol{\theta}')$ a esperança da log-verossimilhança dos dados completos, ou seja,

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}') = E\{\ell_c(\boldsymbol{\theta}|\mathbf{Y}_{com})|\mathbf{Y}_{obs}, \boldsymbol{\theta}'\}.$$

Cada iteração do algoritmo EM consiste de dois passos, uma etapa E (de esperança) e outra M (de maximização). Na etapa E calculamos $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ como função de $\boldsymbol{\theta}$. Na etapa M maximizamos a função resultante com relação a $\boldsymbol{\theta}$, e obtemos assim uma nova estimativa $\boldsymbol{\theta}^{(t+1)}$. Cada iteração do algoritmo EM incrementa o logaritmo da função de verossimilhança observada $\ell(\boldsymbol{\theta}|\mathbf{Y}_{obs})$ de modo que $\ell(\boldsymbol{\theta}^{(t)}|\mathbf{Y}_{obs}) \leq \ell(\boldsymbol{\theta}^{(t+1)}|\mathbf{Y}_{obs})$ e o algoritmo tipicamente converge a um máximo local ou global da função de verossimilhança. Quando o passo M do algoritmo é dificultoso de implementar, é comum substituir este por uma seqüência de passos de maximização restrita (CM), cada um das quais maximiza $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ sob $\boldsymbol{\theta}$ com alguma função de $\boldsymbol{\theta}$ mantida fixa. Isto leva a uma extensão simples do algoritmo EM, chamado de algoritmo ECM (veja, Meng & Rubin, 1993). Em todo o trabalho ambos os algoritmos serão chamados indistintamente de algoritmo EM, salvo algumas exceções no Capítulo 5.

Seja $\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_m^T)^T$, $\mathbf{b} = (\mathbf{b}_1^T, \dots, \mathbf{b}_m^T)^T$. Então, a log-verossimilhança para os dados completos no modelo linear de efeitos mistos (1.1)-(1.2) é dada por

$$\begin{aligned} \ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{b}) = & \sum_{j=1}^m \left[-\frac{n_j}{2} \log \sigma_e^2 - \frac{1}{2\sigma_e^2} (\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta} - \mathbf{Z}_j\mathbf{b}_j)^T (\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta} - \mathbf{Z}_j\mathbf{b}_j) \right. \\ & \left. - \frac{1}{2} \log |\mathbf{D}| - \frac{1}{2} \mathbf{b}_j^T \mathbf{D}^{-1} \mathbf{b}_j \right], \end{aligned} \quad (1.5)$$

de modo que sendo $\hat{\mathbf{b}}_j = E(\mathbf{b}_j|\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}, \mathbf{Y} = \mathbf{y})$ e $\hat{\boldsymbol{\Omega}}_j = \text{Cov}(\mathbf{b}_j|\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}, \mathbf{Y} = \mathbf{y})$, temos das propriedades de cálculo de esperanças condicionais da distribuição normal multivariada (Muirhead, 1982) que

$$\begin{aligned} \hat{\boldsymbol{\Omega}}_j &= \hat{\mathbf{D}} - \hat{\mathbf{D}}\mathbf{Z}_j^T \left(\mathbf{Z}_j\hat{\mathbf{D}}\mathbf{Z}_j^T + \hat{\sigma}_e^2 \mathbf{I}_{n_j} \right)^{-1} \hat{\mathbf{Z}}_j\hat{\mathbf{D}} = \left(\hat{\mathbf{D}}^{-1} + \frac{1}{\hat{\sigma}_e^2} \mathbf{Z}_j^T \mathbf{Z}_j \right)^{-1} \\ \hat{\mathbf{b}}_j &= \hat{\mathbf{D}}\mathbf{Z}_j^T \left(\mathbf{Z}_j\hat{\mathbf{D}}\mathbf{Z}_j^T + \hat{\sigma}_e^2 \mathbf{I}_{n_j} \right)^{-1} (\mathbf{y}_j - \mathbf{X}_j\hat{\boldsymbol{\beta}}). \end{aligned} \quad (1.6)$$

Usando uma álgebra simples obtemos (vide, relação (A-1))

$$E(\ell_c(\theta|Y, b)|y, \hat{\theta}) = \sum_{j=1}^m \left[-\frac{n_j}{2} \log \sigma_e^2 - \frac{1}{2\sigma_e^2} (y_j - X_j \beta - Z_j \hat{b}_j)^T (y_j - X_j \beta - Z_j \hat{b}_j) - \frac{1}{2\sigma_e^2} \text{tr}(Z_j \hat{\Omega}_j Z_j^T) - \frac{1}{2} \log |D| - \frac{1}{2} \text{tr} \left(D^{-1} (\hat{b}_j \hat{b}_j^T + \hat{\Omega}_j) \right) \right]. \quad (1.7)$$

Assim, obtemos o seguinte algoritmo EM:

Etapa E : Dado $\theta = \hat{\theta}$, calcular $\hat{b}_j$ e $\hat{\Omega}_j$, para $j = 1, \dots, m$, usando (1.6).

Etapa M: Atualizar $\hat{\theta}$, maximizando (1.7) sobre θ , resultando em

$$\begin{aligned} \hat{\beta} &= \left(\sum_{j=1}^m X_j^T X_j \right)^{-1} \sum_{j=1}^m X_j^T (y_j - Z_j \hat{b}_j), \\ \hat{\sigma}_e^2 &= \frac{1}{N} \sum_{j=1}^m \left[(y_j - X_j \hat{\beta} - Z_j \hat{b}_j)^T (y_j - X_j \hat{\beta} - Z_j \hat{b}_j) + \text{tr}(\hat{\Omega}_j Z_j^T Z_j) \right], \\ \hat{D} &= \frac{1}{m} \sum_{j=1}^m (\hat{b}_j \hat{b}_j^T + \hat{\Omega}_j), \end{aligned} \quad (1.8)$$

onde $N = \sum_{j=1}^m n_j$. Como na maioria de procedimentos iterativos de maximização, valores iniciais dos parâmetros de interesse precisam ser escolhidos para o algoritmo EM definido previamente. Um método simples e usual para obter valores iniciais dos parâmetros fixos β e das componentes de variância σ_e^2 e D é ajustar separadamente modelos de regressão para cada grupo na amostra e calcular a média dos parâmetros individuais. Isto é, se $\hat{\beta}_j$ e $\hat{\sigma}_{e_j}^2$ denotam os parâmetros individuais estimados obtidos a partir do ajuste de uma regressão linear aos dados do j -ésimo sujeito, $j = 1, \dots, m$, os valores iniciais para o algoritmo EM são calculados como sendo

$$\hat{\beta}^o = \sum_{j=1}^m \hat{\beta}_j / m, \quad \hat{D}^o = \sum_{j=1}^m (\hat{\beta}_j - \hat{\beta}^o)(\hat{\beta}_j - \hat{\beta}^o)^T / (m-1), \quad \hat{\sigma}_e^{2o} = \sum_{j=1}^m \hat{\sigma}_{e_j}^2 / m. \quad (1.9)$$

O algoritmo EM descrito acima tende a ser robusto com relação aos valores iniciais, mas depende fortemente das características dos dados, e além disso é possível que a convergência aconteça num máximo local ou ponto de sela. É também recomendável que diferentes valores iniciais sejam usados a fim de garantir a estabilidade das estimativas resultantes.

1.2.2 Inferência Baseada nos EMV

Como discutido por Verbeke & Molenberghs (2000), a menos que os dados sejam analisados em uma estrutura bayesiana, inferências neste tipo de modelos são baseados na distribuição marginal para a resposta $\mathbf{Y}_j$. Assim, sob as hipóteses do modelo (1.1)-(1.2), não é difícil verificar que, marginalmente, $\mathbf{Y}_j$ tem distribuição

$$\mathbf{Y}_j \stackrel{\text{ind.}}{\sim} N_{n_j}(\mathbf{X}_j\boldsymbol{\beta}, \boldsymbol{\Sigma}_j), \quad j = 1, \dots, m, \quad (1.10)$$

em que $\boldsymbol{\Sigma}_j = \sigma_e^2 \mathbf{I}_{n_j} + \mathbf{Z}_j \mathbf{D} \mathbf{Z}_j^T = \sigma_e^2 \mathbf{I}_{n_j} + \sum_{k=1}^q \sum_{s=1}^q [\mathbf{D}]_{ks} \mathbf{Z}_j^k (\mathbf{Z}_j^s)^T$, com $\mathbf{Z}_j^k$ a k -ésima coluna de $\mathbf{Z}_j$ e $[\mathbf{D}]_{ks}$ o elemento na posição (k,s) da matriz de dispersão $\mathbf{D}$, $k, s = 1, \dots, q$. A partir de (1.10) podemos escrever a log-verossimilhança marginal como

$$\ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{j=1}^m \ell_j(\boldsymbol{\theta}|\mathbf{y}_j), \quad (1.11)$$

onde, desprezando a parte fixa (não dependente de $\boldsymbol{\theta}$),

$$\ell_j(\boldsymbol{\theta}|\mathbf{y}_j) = -\frac{1}{2} \log |\boldsymbol{\Sigma}_j| - \frac{1}{2} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})^T \boldsymbol{\Sigma}_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}), \quad j = 1, \dots, m.$$

Fazendo $p_1 = q(q+1)/2$, $\sigma_o^2 = \sigma_e^2$, $\mathbb{G}_j^o = \mathbf{I}_{n_j}$ e estruturando

$$\sigma_1^2 = [\mathbf{D}]_{11}, \sigma_2^2 = [\mathbf{D}]_{12}, \dots, \sigma_{p_1}^2 = [\mathbf{D}]_{qq},$$

$$\mathbb{G}_j^1 = \mathbf{Z}_j^1 (\mathbf{Z}_j^1)^T, \mathbb{G}_j^2 = \mathbf{Z}_j^1 (\mathbf{Z}_j^2)^T + \mathbf{Z}_j^2 (\mathbf{Z}_j^1)^T, \mathbb{G}_j^3 = \mathbf{Z}_j^1 (\mathbf{Z}_j^3)^T + \mathbf{Z}_j^3 (\mathbf{Z}_j^1)^T, \dots, \mathbb{G}_j^{p_1} = \mathbf{Z}_j^q (\mathbf{Z}_j^q)^T,$$

$j = 1, \dots, m$, podemos escrever

$$\boldsymbol{\Sigma}_j = \sum_{i=0}^{p_1} \sigma_i^2 \mathbb{G}_j^i. \quad (1.12)$$

Se $p_o = p_1 + p + 1$ e $\boldsymbol{\sigma}^2 = (\sigma_e^2, \sigma_1^2, \dots, \sigma_{p_1}^2)^T$, o espaço paramétrico Θ para o modelo (1.1)-(1.2) é dado por

$$\begin{aligned} \Theta = & \left[\boldsymbol{\theta} \in \mathbb{R}^{p_o} \mid \boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\sigma}^T)^T, \boldsymbol{\beta} \in \mathbb{R}^p; \sigma_e^2 > 0 \text{ e } (\sigma_1^2, \dots, \sigma_{p_1}^2) \in \mathbb{R}_+^{p_1}, \right. \\ & \left. \text{tal que } \mathbf{D} \text{ é semi-definida positiva} \right]. \end{aligned} \quad (1.13)$$

Note que o espaço paramétrico considerado é diferente do espaço paramétrico corres-

pondente ao modelo marginal, pois neste último é suficiente que Σ_j , $j = 1, \dots, m$, seja definida positiva.

Considerando propriedades de derivadas matriciais (veja, equações (A-2)-(A-7)), temos que a função escore é dada por

$$S(\theta) = S(\theta|X, y) = \frac{\partial \ell(\theta|y)}{\partial \theta} = \sum_{j=1}^m \frac{\partial \ell_j(\theta|y_j)}{\partial \theta}, \quad (1.14)$$

em que

$$\begin{aligned} \frac{\partial \ell_j(\theta|y_j)}{\partial \beta} &= X_j^T \Sigma_j^{-1} (y_j - X_j \beta), \\ \frac{\partial \ell_j(\theta|y_j)}{\partial \sigma_k^2} &= -\frac{1}{2} \text{tr}(\Sigma_j^{-1} G_j^k) + \frac{1}{2} (y_j - X_j \beta)^T \Sigma_j^{-1} G_j^k \Sigma_j^{-1} (y_j - X_j \beta), \end{aligned}$$

com $k = 0, \dots, p_1$.

Similarmente, a matriz de informação observada é dada por

$$I_F(\theta) = I_F(\theta|X, y) = -\ddot{L} = -\frac{\partial^2 \ell(\theta|y)}{\partial \theta \partial \theta^T} = -\sum_{j=1}^m \frac{\partial^2 \ell_j(\theta|y_j)}{\partial \theta \partial \theta^T},$$

onde

$$\begin{aligned} \frac{\partial^2 \ell_j(\theta|y_j)}{\partial \beta \partial \beta^T} &= -X_j^T \Sigma_j^{-1} X_j, \\ \frac{\partial^2 \ell_j(\theta|y_j)}{\partial \sigma_k^2 \partial \beta} &= -X_j^T \Sigma_j^{-1} G_j^k \Sigma_j^{-1} (y_j - X_j \beta), \\ \frac{\partial^2 \ell_j(\theta|y_j)}{\partial \sigma_k^2 \partial \sigma_s^2} &= \frac{1}{2} \text{tr}(\Sigma_j^{-1} G_j^k \Sigma_j^{-1} G_j^s) - (y_j - X_j \beta)^T \Sigma_j^{-1} G_j^k \Sigma_j^{-1} G_j^s \Sigma_j^{-1} (y_j - X_j \beta), \end{aligned} \quad (1.15)$$

com $k, s = 0, \dots, p_1$.

A matriz de informação esperada de Fisher, denotada por $J_F(\theta)$, é da forma

$$J_F(\theta) = J_F(\theta|X, y) = -\sum_{j=1}^m E \left(\frac{\partial^2 \ell_j(\theta|y_j)}{\partial \theta \partial \theta^T} \right) = \begin{pmatrix} J_{\beta\beta} & \mathbf{0} \\ \mathbf{0} & J_{\sigma\sigma} \end{pmatrix}, \quad (1.16)$$

em que $J_{\beta\beta} = \sum_{j=1}^m X_j^T \Sigma_j^{-1} X_j$ e $[J_{\sigma\sigma}]_{ks} = \frac{1}{2} \sum_{j=1}^m \text{tr}(\Sigma_j^{-1} G_j^k \Sigma_j^{-1} G_j^s)$, $k, s = 0, \dots, p_1$.

onde $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_k)$. Igualando a zero e usando propriedades de inversas matriciais (vide, equação (A-20); Nyquist, 1991), segue que

$$\begin{aligned}
\beta(\lambda) &= \left(\frac{1}{\sigma_e^2} \sum_{j=1}^m \mathbf{X}_j^T \mathbf{X}_j + \mathbf{C}^T \Lambda \mathbf{C} \right)^{-1} \left(\frac{1}{\sigma_e^2} \sum_{j=1}^m \mathbf{X}_j^T (\mathbf{y}_j - \mathbf{Z}_j \mathbf{b}) + \mathbf{C}^T \Lambda \mathbf{d} \right) \\
&= \left(\sum_{j=1}^m \mathbf{X}_j^T \mathbf{X}_j \right)^{-1} \sum_{j=1}^m \mathbf{X}_j^T (\mathbf{y}_j - \mathbf{Z}_j \mathbf{b}_j) + \sigma_e^2 \left(\sum_{j=1}^m \mathbf{X}_j^T \mathbf{X}_j \right)^{-1} \mathbf{C}^T [\Lambda^{-1} \\
&\quad + \sigma_e^2 \mathbf{C} \left(\sum_{j=1}^m \mathbf{X}_j^T \mathbf{X}_j \right)^{-1} \mathbf{C}^T]^{-1} [\mathbf{d} - \mathbf{C} \left(\sum_{j=1}^m \mathbf{X}_j^T \mathbf{X}_j \right)^{-1} \sum_{j=1}^m \mathbf{X}_j^T (\mathbf{y}_j - \mathbf{Z}_j \mathbf{b}_j)].
\end{aligned} \tag{1.17}$$

Finalmente, o estimador restrito para β é da forma

$$\begin{aligned}
\hat{\beta}_o &= \lim_{\lambda_1, \dots, \lambda_k \rightarrow \infty} \beta(\lambda) = \left(\sum_{j=1}^m \mathbf{X}_j^T \mathbf{X}_j \right)^{-1} \sum_{j=1}^m \mathbf{X}_j^T (\mathbf{y}_j - \mathbf{Z}_j \mathbf{b}_j) + \left(\sum_{j=1}^m \mathbf{X}_j^T \mathbf{X}_j \right)^{-1} \mathbf{C}^T \times \\
&\quad [\mathbf{C} \left(\sum_{j=1}^m \mathbf{X}_j^T \mathbf{X}_j \right)^{-1} \mathbf{C}^T]^{-1} [\mathbf{d} - \mathbf{C} \left(\sum_{j=1}^m \mathbf{X}_j^T \mathbf{X}_j \right)^{-1} \sum_{j=1}^m \mathbf{X}_j^T (\mathbf{y}_j - \mathbf{Z}_j \mathbf{b}_j)] \\
&= \hat{\delta} + \left(\sum_{j=1}^m \mathbf{X}_j^T \mathbf{X}_j \right)^{-1} \mathbf{C}^T [\mathbf{C} \left(\sum_{j=1}^m \mathbf{X}_j^T \mathbf{X}_j \right)^{-1} \mathbf{C}^T]^{-1} (\mathbf{d} - \mathbf{C} \hat{\delta}),
\end{aligned} \tag{1.18}$$

onde $\hat{\delta} = \left(\sum_{j=1}^m \mathbf{X}_j^T \mathbf{X}_j \right)^{-1} \sum_{j=1}^m \mathbf{X}_j^T (\mathbf{y}_j - \mathbf{Z}_j \mathbf{b}_j)$. É interessante notar que $\hat{\beta}_o$ é o estimador não restrito $\hat{\beta}$ como em (1.8) o qual foi acrescentado um termo de correção. O processo iterativo (algoritmo EM) para estimar os parâmetros do modelo sujeito à restrição $\mathbf{C}\beta - \mathbf{d} = \mathbf{0}$, denotado por $\hat{\theta}_o$, segue o mesmo procedimento dado na Seção 1.2.1, substituindo $\hat{\beta}$ por $\hat{\beta}_o$ na etapa M do algoritmo. O algoritmo proposto foi testado por simulação Monte Carlo, fornecendo boas aproximações dos parâmetros considerados. O programa implementado em Matlab encontra-se no Apêndice A.

Suponha agora que estamos interessados em testar a hipótese $H_o : \mathbf{C}\beta = \mathbf{d}$ contra a alternativa $H_1 : \mathbf{C}\beta \neq \mathbf{d}$. Os métodos usuais para testar estas hipóteses lineares são razão de verossimilhanças (RV), Wald (W) e escore (Es). Estas estatísticas são, respectivamente,

Note que, sob condições de regularidade adequadas, os EMV dos efeitos fixos β e os elementos de σ^2 da matriz de variâncias-covariâncias são assintoticamente independentes.

Intervalos de confiança e testes de hipóteses baseados nos EMV para o vetor de parâmetros θ podem ser obtidos usando a matriz de informação observada ou esperada. Por exemplo, dado que a matriz de informação de Fisher é bloco diagonal, intervalos de confiança e testes de hipóteses assintóticos para os efeitos fixos são obtidos assumindo que o $\hat{\beta}$ tem aproximadamente uma distribuição $N_p(\beta, (\mathbf{J}_{\beta\beta})^{-1})$. Na prática, $\mathbf{J}_{\beta\beta}$ é desconhecida e deve ser substituída pelo seu estimador consistente, $\hat{\mathbf{J}}_{\beta\beta}$. A seguir, discutiremos o processo iterativo para encontrar os EMV de θ sob condições de igualdade dos parâmetros de regressão β no MLM definido em (1.1)-(1.2).

1.2.3 Estimação sob Restrições de Igualdade no Parâmetro β

Suponha que estamos interessados em estimar o vetor de parâmetros β sob k restrições linearmente independentes $\mathbf{C}_i^T \beta - d_i = 0$, onde $\mathbf{C}_i, i = 1, \dots, k$, são vetores $p \times 1$ e $d_i, i = 1, \dots, k$, são escalares, ambos conhecidos e fixados. O problema aqui é maximizar a função log-verossimilhança $\ell_c(\theta|\mathbf{y}, \mathbf{b})$ sujeita à restrição linear $\mathbf{C}\beta - \mathbf{d} = \mathbf{0}$, onde $\mathbf{C} = (\mathbf{C}_1, \dots, \mathbf{C}_k)^T$ e $\mathbf{d} = (d_1, \dots, d_k)^T$. Similarmente a Nyquist (1991), que investiga esta classe de problema em modelos lineares generalizados, aplicamos a metodologia de funções penalizadas considerando a função penalizada quadrática

$$P(\theta, \lambda) = \ell_c(\theta|\mathbf{y}, \mathbf{b}) - \frac{1}{2} \sum_{i=1}^k \lambda_i (d_i - \mathbf{C}_i^T \beta)^2,$$

em que $\ell_c(\theta|\mathbf{y}, \mathbf{b})$ é como em (1.5). O procedimento consiste em encontrar a solução do problema $\max_{\theta} P(\theta, \lambda)$ para valores fixos e positivos de $\lambda_i, i = 1, \dots, k$, onde a solução de β denotada por $\beta(\lambda)$ produz o estimador restrito $\hat{\beta}_o$ obtido de

$$\hat{\beta}_o = \lim_{\lambda_1, \dots, \lambda_k \rightarrow \infty} \beta(\lambda).$$

Assim, diferenciando $P(\theta, \lambda)$ com relação a β temos

$$\frac{\partial P(\theta, \lambda)}{\partial \beta} = \frac{1}{\sigma_e^2} \sum_{j=1}^m \mathbf{X}_j^T (\mathbf{Y}_j - \mathbf{X}_j \beta - \mathbf{Z}_j \mathbf{b}_j) + \mathbf{C}^T \Lambda (\mathbf{d} - \mathbf{C} \beta),$$

dadas por

$$\begin{aligned}
\xi_{RV} &= 2\{\ell(\hat{\theta}|\mathbf{y}) - \ell(\hat{\theta}_o|\mathbf{y})\}, \\
\xi_W &= (\mathbf{C}\hat{\beta} - \mathbf{d})^T \{\mathbf{C}\hat{\mathbf{J}}_{\beta\beta}^{-1}\mathbf{C}^T\}^{-1}(\mathbf{C}\hat{\beta} - \mathbf{d}) \text{ e} \\
\xi_{Es} &= \{\mathbf{S}_\beta(\hat{\theta}_o) - \mathbf{S}_\beta(\hat{\theta})\}^T \widehat{\text{Var}}_o(\hat{\beta}) \{\mathbf{S}_\beta(\hat{\theta}_o) - \mathbf{S}_\beta(\hat{\theta})\} \\
&= \mathbf{S}_\beta(\hat{\theta}_o)^T \{\hat{\mathbf{J}}_{\beta\beta}^o\}^{-1} \mathbf{S}_\beta(\hat{\theta}_o),
\end{aligned}$$

onde $\mathbf{S}_\beta$ é a função escore correspondente a β dada em (1.14), $\mathbf{J}_{\beta\beta}$ dado em (1.16), onde $\hat{\mathbf{J}}_{\beta\beta}$ e $\hat{\mathbf{J}}_{\beta\beta}^o$ são avaliados em $\hat{\theta}$ e $\hat{\theta}_o$, respectivamente. Sob H_o e para n suficientemente grande, temos que ξ_{RV} , ξ_W e ξ_{Es} têm distribuição qui-quadrado central com k graus de liberdade.

1.3 Objetivo do Trabalho

O objetivo deste trabalho é o desenvolvimento de procedimentos de estimação na classe dos modelos lineares mistos normal-assimétrico, sendo estes procedimentos baseados em técnicas clássicas e bayesianas. Para a análise bayesiana será usado o programa WinBUGS14 (<http://www.mrc-bsu.cam.ac.uk/bugs/>), já que fornece uma estrutura amigável de modo que modelos complexos podem ser facilmente manipulados. Embora alguns resultados deste trabalho possam ser facilmente tratados usando técnicas padrão de maximização direta, a obtenção dos estimadores de máxima verossimilhança será via algoritmo EM (Dempster et al., 1977). Conjecturas relacionadas com os parâmetros do modelo podem ser facilmente testadas do ponto de vista clássico usando a estatística da razão de verossimilhanças apresentada na seção anterior. Podemos então relacionar os seguintes objetivos:

- i. Apresentar uma nova classe de distribuições normal-assimétrica multivariada que possua uma estrutura conveniente para fazer inferências em diversos modelos estatísticos e estudar, de forma detalhada, as propriedades destas distribuições.
- ii. Propor procedimentos de estimação no modelo linear misto normal-assimétrico via algoritmo EM e métodos Monte Carlo via Cadeias de Markov (MCMC). Neste último, basicamente o amostrador de Gibbs.
- iii. Estudar procedimentos de estimação em modelos lineares mistos com erros de medida sob as distribuições normal e normal-assimétrica.

1.4 Organização do Trabalho

Este trabalho contém 6 capítulos e 4 apêndices. No Capítulo 2 apresentamos a distribuição normal-assimétrica, com ênfase no caso multivariado. Estudamos basicamente as principais propriedades desta distribuição associadas com a distribuição de transformações lineares, marginalização, condicionamento e formas quadráticas. Uma representação estocástica para o vetor normal-assimétrico é considerada, podendo alternativamente ser usada para obter os momentos desta distribuição e para o desenvolvimento de estudos relacionados a aspectos inferenciais deste modelo. Adicionalmente, uma aplicação aos modelos de regressão multivariados é apresentada.

No Capítulo 3 abordaremos os modelos lineares mistos normal-assimétrico (MLMSN) estendendo o MLM usual. Uma forma fechada da densidade marginal das quantidades envolvidas é obtida integrando sob o efeito aleatório, de modo que a log-verossimilhança marginal pode ser maximizada diretamente usando programas computacionais existentes como Ox, Matlab, R, S-Plus, entre outros. Notamos também que as expressões analíticas desta natureza não são dadas na literatura. Adicionalmente, o algoritmo EM é apresentado que parece apresentar algumas vantagens sobre a maximização direta, especificamente em termos de robustez com relação aos valores iniciais e forma fechada dos parâmetros do efeito fixos no processo iterativo de estimação. Realizamos um estudo de simulação onde mostramos que existe perda de eficiência na estimação de certos parâmetros quando a assimetria esta presente nos dados, levando a inferências menos precisas sobre os parâmetros de interesse. Finalmente uma aplicação a um conjunto de dados reais é apresentada, onde os resultados coincidem com os reportados por Zhang e Davidian (2001) usando técnicas semi-paramétricas.

No Capítulo 4 desenvolveremos métodos bayesianos de estimação em MLMSN usando o esquema MCMC e uma distribuição conveniente para este tipo de análise. Sob esta estrutura encontramos também a verossimilhança marginal das quantidades envolvidas e mostramos através de simulação que critérios de informação, tais como AIC (critério de informação de Akaike), BIC (critério de informação bayesiano) e HQ (critério de informação de Hannan-Quinn), podem ser usados para detectar afastamentos da normalidade e escolher o modelo que melhor represente os dados. Baseados nas estratégias inferenciais discutidas no estudo de simulação, uma análise bayesiana de um conjunto de dados de colesterol é apresentada.

No capítulo 5 trataremos o problema de estimação em modelos lineares mistos com erros nas variáveis. Após uma revisão bibliográfica dos métodos usuais de estimação, aplicamos os métodos a este tipo de modelos em um marco teórico. Em seguida, apresentamos um estudo de simulação, onde o objetivo é comparar a robustez dos métodos quanto ao aumento da variabilidade do erro de medição. Finalmente, desenvolvemos procedimentos de estimação, via máxima verossimilhança, no MLM estrutural com erros de medida sob as suposições de normalidade e distribuição normal-assimétrica em todas as variáveis aleatórias consideradas.

O sexto e último capítulo finaliza esta tese com algumas conclusões de nosso estudo, pesquisas que estão sendo desenvolvidas motivadas pelos resultados deste trabalho e algumas perspectivas para trabalhos futuros.

Capítulo 2

A Distribuição Normal-Assimétrica

2.1 Introdução

Existe uma tendência geral na literatura estatística para encontrar métodos mais flexíveis que representem características dos dados tão adequadas quanto possíveis reduzindo suposições não realistas. A motivação deve-se ao crescente poder computacional e, também, a inúmeros conjuntos de dados em distintos campos da ciência, que freqüentemente não satisfazem algumas suposições, tais como, independência e normalidade. Para modelar desvios destas suposições e particularmente da suposição de normalidade, que desempenha um papel importante na análise estatística, diferentes enfoques podem ser encontrados na literatura. Do ponto de vista prático, o método mais comum adotado para alcançar a normalidade (ou simetria) é transformação de variáveis. Embora tal método possa dar resultados empíricos razoáveis, deve ser evitado se um modelo mais conveniente pode ser encontrado. Azzalini & Capitanio (1999) apresentam algumas razões para evitar este procedimento:

- a) A transformação não fornece informação útil para entender o mecanismo de geração dos dados.
- b) A transformação de variáveis dificulta a interpretação, especialmente quando cada variável é transformada usando diferentes funções.
- c) A transformação para um conjunto de dados pode freqüentemente não ser aplicável para outros conjuntos de dados.
- d) Quando a suposição de homocedasticidade é necessária, algumas vezes a transformação requerida difere da transformação para alcançar a normalidade.

Um enfoque alternativo à transformação de variáveis, para modelagem de dados de um ponto de vista paramétrico, apropriado para o tratamento de observações contínuas multivariadas, consiste em construir classes paramétricas flexíveis de distribuições multivariadas que exibam assimetria e curtose diferentes da distribuição normal. A classe das distribuições elípticas é provavelmente o exemplo mais conhecido deste enfoque. Esta classe foi introduzida por Kelker (1970) e sistematicamente discutida por Cambanis et al. (1981). Para uma revisão desta classe, ver por exemplo, Fang et al. (1990) e Arellano-Valle (1994). A classe elíptica inclui um vasto conjunto de distribuições, como por exemplo, normal, normal composta, t de Student, exponencial potência, Pearson tipo II. A principal vantagem é que representa uma extensão natural do conceito de simetria no contexto multivariado. Embora modelos elípticos forneçam alternativas para o modelo normal, estes somente podem ser aplicados em situações práticas onde simetria parece razoável. Alternativamente, existem muitas outras classes paramétricas de distribuições que podem ser escolhidas, contudo esta escolha não é tão ampla como no caso univariado; muitas delas são estudadas no livro de Johnson & Kotz (1972). Um outro enfoque para uma extensão multivariada da teoria univariada é mediante o uso de copulas. Alguns inconvenientes deste método podem ser encontrados em Azzalini & Dalla-Valle (1996). Portanto, é de interesse prático estudar modelos estatísticos que sejam menos sensíveis do que a distribuição gaussiana a certos desvios das suposições consideradas, construindo famílias paramétricas de *distribuições assimétricas* que sejam analiticamente tratáveis, possam acomodar valores práticos de assimetria e curtose, e estritamente incluam a distribuição normal, evitando assim, a necessidade de transformações *ad hoc*.

Este capítulo está centrado no estudo da distribuição normal-assimétrica, com ênfase no caso multivariado. Nosso interesse é basicamente introduzir uma nova distribuição numa estrutura conveniente para fazer inferências em inúmeros modelos estatísticos e estudar as principais propriedades desta distribuição. O restante do capítulo está organizado como segue. Na Seção 2.2 apresentamos uma revisão bibliográfica da distribuição normal-assimétrica univariada de Azzalini (1985). Na Seção 2.3 introduzimos uma extensão multivariada da distribuição normal-assimétrica e apresentamos um estudo detalhado das propriedades desta distribuição. A Seção 2.4 está dedicada a uma aplicação prática da distribuição normal-assimétrica em modelos de regressão multivariados, os resultados e métodos propostos são exemplificados através de um estudo de simulação e análise de um conjunto de dados reais.

2.2 A Distribuição Normal-Assimétrica Univariada

Uma variável aleatória Z é dita ter uma distribuição normal-assimétrica com parâmetro de assimetria λ , denotada por $SN_1(\lambda)$, se sua função densidade é da forma

$$f_Z(z) = 2\phi_1(z) \Phi_1(\lambda z), \quad z \in \mathbb{R}, \quad (2.1)$$

onde $\phi_1(\cdot)$ e $\Phi_1(\cdot)$ denotam a função densidade de probabilidade (fdp) e a função distribuição acumulada (fda), respectivamente, da normal padrão ($N(0,1)$). A densidade em (2.1) é uma densidade válida para $\lambda \in \mathbb{R}$ e é simétrica se $\lambda = 0$ (em tal caso coincide com a densidade da normal padrão). Valores positivos (negativos) de λ correspondem a uma assimetria positiva (negativa) na densidade (2.1). No limite, quando λ tende a $+\infty$ ($-\infty$), Z tende a uma variável aleatória metade-normal (*half-normal*) positiva (negativa). A família de densidades (2.1) foi discutida formalmente por Azzalini (1985) propondo que esta densidade poderia fornecer uma família assimétrica razoável como alternativa ao modelo normal clássico. No entanto, a idéia de modelagem assimétrica por meio da construção de famílias matematicamente tratáveis, que inclui a distribuição normal, foi proposto inicialmente por outros autores num contexto bayesiano (ver por exemplo, O'Hagan & Leonard, 1976). Exemplos da densidade (2.1) para $\lambda = -3, 1, 3, 5$ são mostrados na Figura 2.1.

Como discutido por Azzalini & Dalla-Valle (1996), três cenários interessantes podem ser utilizados para gerar extensões do modelo normal-assimétrico univariado, que são de interesse em nosso estudo e conseqüentemente são motivo de consideração.

Proposição 2.1. *Sejam $Y_i \stackrel{\text{iid}}{\sim} N(0,1)$, $i = 1, 2$, e definamos a variável $Z = [Y_1 | \lambda Y_1 > Y_2]$. Então $Z \sim SN_1(\lambda)$.*

Demonstração. Note que,

$$\begin{aligned} P(Z \leq z) &= P(Y_1 \leq z | \lambda Y_1 > Y_2) = \frac{P(Y_1 \leq z, \lambda Y_1 > Y_2)}{P(\lambda Y_1 > Y_2)} \\ &= \frac{1}{P(\lambda Y_1 > Y_2)} \int_{-\infty}^z \int_{-\infty}^{\lambda y} \phi(y) \phi(w) dw dy \\ &= \frac{1}{P(\lambda Y_1 > Y_2)} \int_{-\infty}^z \phi(y) \Phi(\lambda y) dy. \end{aligned} \quad (2.2)$$

A prova segue notando que $P(\lambda Y_1 > Y_2) = P(\lambda Y_1 - Y_2 > 0) = \frac{1}{2}$ (pois $\lambda Y_1 - Y_2$ tem

Figura 2.1: Função densidade de probabilidade de (2.1) para $\lambda = -3, 0, 1, 3$ e 5

distribuição normal com média zero). □

Proposição 2.2. *Suponha que o vetor (Y_1, Y_2) tem distribuição normal bi-variada com marginais $N(0, 1)$ e correlação $\delta = \frac{\lambda}{(1 + \lambda^2)^{1/2}}$ e definamos a variável $Z = [Y_1 | Y_2 > 0]$. Então $Z \sim SN_1(\lambda)$.*

Demonstração. Sem perda de generalidade escrevemos $Y_2 = \delta Y_1 - (1 - \delta^2)^{1/2} W$, com Y_1 e W i.i.d. $N(0, 1)$. Conseqüentemente, o evento $[Y_2 > 0]$ é equivalente a $[\delta Y_1 - (1 - \delta^2)^{1/2} W > 0]$ ou $[\frac{\delta}{(1 - \delta^2)^{1/2}} Y_1 > W]$. A prova segue da Proposição 2.1. □

Observação 2.1. *Sob as condições da Proposição 2.2, alternativamente podemos representar a variável aleatória Z como*

$$Z = \begin{cases} Y_1, & \text{se } Y_2 > 0, \\ -Y_1, & \text{caso contrário.} \end{cases} \quad (2.3)$$

Este resultado fornece uma representação estocástica da variável aleatória Z , que pode ser

usada para gerar números aleatórios e para propósitos teóricos (veja Azzalini & Dalla-Valle, 1996).

Muitas propriedades da densidade em (2.1) podem ser facilmente obtidas de uma terceira representação de Z , incluída em Azzalini (1986) (veja também Henze, 1986). O resultado é dado a seguir.

Proposição 2.3. *Sejam $X_i \stackrel{\text{iid}}{\sim} N(0, 1)$, $i = 0, 1$, e*

$$Z \stackrel{d}{=} \delta |X_0| + (1 - \delta^2)^{1/2} X_1, \quad \delta = \frac{\lambda}{(1 + \lambda^2)^{1/2}}, \quad (2.4)$$

onde $X \stackrel{d}{=} Y$ significa que X e Y são identicamente distribuídos. Então $Z \sim SN_1(\lambda)$.

Demonstração. Veja prova da Proposição 2.4 e considere $n = 1$. □

Em resumo, os resultados das Proposições 2.1 e 2.2 são essencialmente equivalentes, enquanto que o resultado da Proposição 2.3 parece ser diferente e de fato o mais interessante por permitir um estudo detalhado desta família de distribuições, assim como facilitar extensões para casos multivariados. Uma quarta derivação da distribuição normal-assimétrica que envolve máximos e dependência também merece atenção. Para isso, suponha que o vetor (Y_1, Y_2) tem distribuição normal bivariada com marginais $N(0, 1)$ e correlação δ . Definamos a variável $Z = [\max(Y_1, Y_2)]$. Pode-se mostrar (veja Loperfido, 2002) que a variável $Z \sim SN_1(\lambda)$, com $\lambda = \sqrt{(1 - \delta)/(1 + \delta)}$.

As propriedades dadas a seguir para uma variável aleatória Z , com distribuição $SN_1(\lambda)$, são facilmente derivadas usando a representação estocástica em (2.4). As provas serão omitidas por ser de nosso interesse estudar as propriedades da distribuição normal-assimétrica multivariada, em que a densidade (2.1) é um caso particular. Entre as principais propriedades consideramos:

$$Z^2 \sim \chi_1^2, \quad E[Z] = \sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{1 + \lambda^2}}, \quad \text{Var}[Z] = 1 - \frac{2\lambda^2}{\pi(1 + \lambda^2)}.$$

Segue que os momentos de ordem par de Z , coincidem com os momentos correspondentes da distribuição normal padrão e os momentos de ordem ímpar são dados por

$$E[Z^{2k+1}] = \sqrt{\frac{2}{\pi}} \lambda (1 + \lambda^2)^{-(k+\frac{1}{2})} 2^{-k} [(2k+1)!] \sum_{j=0}^k \frac{j! (2\lambda)^{2j}}{(2j+1)!(k-j)!}.$$

Usando a relação $E[Z^3] = (2 + \frac{1}{1 + \lambda^2})E[Z]$, segue que os índices de assimetria e curtose são dados por

$$\gamma = \frac{1}{2}(4 - \pi)\left(\frac{E^2[Z]}{\text{Var}[Z]}\right)^{3/2}; \quad \kappa = 2(\pi - 3)\left(\frac{E^2[Z]}{\text{Var}[Z]}\right)^2,$$

implicando que $-0,9953 < \gamma < 0,9953$ e $3,0000 < \kappa < 3,8692$.

Por ser de interesse prático, é natural agora estender o modelo (2.1) introduzindo parâmetros de locação e escala. Neste caso dizemos que $Y \sim SN_1(\mu, \sigma^2, \lambda)$ se

$$f_Y(y) = \frac{2}{\sigma} \phi_1\left(\frac{y - \mu}{\sigma}\right) \Phi_1\left(\lambda \frac{y - \mu}{\sigma}\right). \quad (2.5)$$

Isto é, uma v.a. Y tem densidade em (2.5), se sua representação é da forma $Y \stackrel{d}{=} \mu + \sigma Z$, onde $Z \sim SN_1(\lambda)$.

A distribuição normal-assimétrica compartilha com a distribuição normal boas propriedades e tem a vantagem de ser conveniente para análises de dados com distribuições empíricas assimétricas e unimodais (veja Figura 2.1). Apesar das boas propriedades, a família normal-assimétrica tem alguns problemas inferenciais. De fato, a estimação do parâmetro de assimetria pode ser problemática. Por exemplo, considerando o modelo (2.1), se todas as observações são positivas (negativas), a verossimilhança é monótona crescente (decrecente), acarretando em um máximo infinito (menos infinito) para o parâmetro λ (Liseo, 1990). Mais ainda, isto acontece com probabilidade positiva, embora decresça rapidamente com o aumento do tamanho da amostra. No caso do modelo (2.5) o problema todavia persiste, mas é complicado identificar amostras que levam a uma estimativa de λ tendendo a infinito. Evidências empíricas mostram que os estimadores de máxima verossimilhança dos parâmetros de locação e escala sempre existem, sendo confinado os problemas inferenciais somente ao parâmetro de assimetria.

Extensões univariadas da distribuição normal-assimétrica tem sido propostas por muitos autores. Arnold & Beaver (2000) consideram extensões do modelo (2.1) deriváveis de (2.4) por permitir que X_o e X_1 sejam va's i.i.d. com densidade comum não-normal ψ . De um modo geral podemos considerar que X_o e X_1 em (2.4) têm densidades ψ_1 e ψ_2 . Variantes Cauchy-assimétrica e Laplace-assimétrica obtidas desta forma são discutidas em Arnold

& Beaver (2000).

Distribuições normal-assimétrica multivariadas são consideradas em Azzalini & Dalla-Valle (1996), Azzalini & Capitanio (1999), Branco & Dey (2001), Sahu et al. (2003), entre outros. Genton et al. (2001) derivam os momentos do vetor aleatório com distribuição normal-assimétrica multivariada e suas formas quadráticas. Arellano-Valle, Del Pino & San Martin (2002) mostram que muitas das propriedades da distribuição normal-assimétrica multivariada são satisfeitas para uma classe geral de distribuições assimétricas obtidas de uma classe simétrica, definidas em termos de independência, condições nos sinais e valores absolutos, e apresentam uma fórmula geral para obter densidades assimétricas. Destes resultados, Arellano-Valle & Genton (2003) introduzem a classe *fundamental de distribuições assimétricas*, fornecendo um estudo unificado para a obtenção de distribuições multivariadas assimétricas partindo de uma classe simétrica multivariada (veja também Arellano-Valle & Del Pino, 2003). Na seção seguinte consideramos a extensão multivariada da distribuição normal-assimétrica definida em (2.5) e que será usada em grande parte deste trabalho. Derivamos para esta distribuição resultados diversos associados com a distribuição de transformações lineares, marginalização, condicionamento e formas quadráticas. Uma representação estocástica para o vetor normal-assimétrico multivariado é também considerada, sendo uma generalização da representação estocástica (2.4), podendo ser usada alternativamente para obter os momentos desta distribuição, e como veremos, também para o desenvolvimento de estudos relacionados a aspectos inferenciais deste modelo. Enfatizamos também que a distribuição introduzida a seguir difere daquelas encontradas na literatura, e portanto, muitas das propriedades derivadas na seção seguinte são aportes deste trabalho ao desenvolvimento dos tópicos de distribuições assimétricas.

2.3 A Distribuição Normal-Assimétrica Multivariada

Como discutido em Arellano-Valle & Genton (2003) e Arellano-Valle & Del Pino (2003), existem muitas definições de distribuições normal-assimétrica multivariada. Nesta seção consideramos uma definição unificada das definições encontradas em Azzalini & Dalla-Valle (1996), Azzalini & Capitanio (1999) e Arellano-Valle & Genton (2003). A seguir introduzimos uma notação importante que será usada no decorrer deste trabalho.

Denotamos por $\phi_n(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e $\Phi_n(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ a fdp e a fda, respectivamente, da distribuição

$N_n(\boldsymbol{\mu}, \Sigma)$ avaliada em $\mathbf{x}$. Quando $\boldsymbol{\mu} = \mathbf{0}$ e $\Sigma = \mathbf{I}_n$ (a matriz identidade $n \times n$), denotaremos estas funções como $\phi_n(\mathbf{x})$ e $\Phi_n(\mathbf{x})$.

Definição 2.1. *Um vetor aleatório n -dimensional $\mathbf{Y}$ segue uma distribuição normal-assimétrica com vetor de locação $\boldsymbol{\mu} \in \mathbb{R}^n$, matriz de dispersão Σ (uma matriz $n \times n$ definida positiva) e vetor de assimetria $\boldsymbol{\lambda} \in \mathbb{R}^n$, se sua fdp é dada por*

$$f_{\mathbf{Y}}(\mathbf{y}) = 2\phi_n(\mathbf{y}|\boldsymbol{\mu}, \Sigma)\Phi_1(\boldsymbol{\lambda}^T \Sigma^{-1/2}(\mathbf{y} - \boldsymbol{\mu})), \quad \mathbf{y} \in \mathbb{R}^n, \quad (2.6)$$

denotada por $\mathbf{Y} \sim SN_n(\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda})$. Quando $\boldsymbol{\mu} = \mathbf{0}$ e $\Sigma = \mathbf{I}_n$ temos

$$f_{\mathbf{Y}}(\mathbf{y}) = 2\phi_n(\mathbf{y})\Phi_1(\boldsymbol{\lambda}^T \mathbf{y}), \quad \mathbf{y} \in \mathbb{R}^n, \quad (2.7)$$

e neste caso denotamos por $\mathbf{Y} \sim SN_n(\boldsymbol{\lambda})$.

Observação 2.2. *Uma vez que a condição $\Phi_1(-w) = 1 - \Phi_1(w)$, para todo $w \in \mathbb{R}$, é suficiente para garantir que (2.6) é uma fdp (veja Genton & Loperfido, 2002), podemos usar então diferentes reparametrizações para representar o parâmetro de assimetria $\boldsymbol{\lambda}$ como, por exemplo,*

$$\boldsymbol{\lambda} = \frac{\boldsymbol{\Delta}^{-1/2} \boldsymbol{\delta}}{\sqrt{1 - \boldsymbol{\delta}^T \boldsymbol{\Delta}^{-1} \boldsymbol{\delta}}}, \quad (2.8)$$

para algum $\boldsymbol{\delta} \in \mathbb{R}^n$ e matriz definida positiva $\boldsymbol{\Delta}$ de dimensão $n \times n$ tal que $\sqrt{\boldsymbol{\delta}^T \boldsymbol{\Delta}^{-1} \boldsymbol{\delta}} < 1$. Dois casos podem ser considerados $\boldsymbol{\Delta} = \Sigma$, (matriz de correlações) que é a reparametrização usada por Azzalini & Dalla-Valle (1996), e $\boldsymbol{\Delta} = \mathbf{I}_n$, que é um caso especial da classe definida por Arellano-Valle & Genton (2003). De um modo mais geral, podemos substituir em (2.6) a parte assimétrica (ou função de assimetria; veja Arellano Valle et al., 2002) $\Phi_1(\cdot)$ por uma função arbitrária $Q_m(\cdot)$ em $[0, 1]$, que depende de $\mathbf{y}$ através de uma função ímpar $w : \mathbb{R}^n \rightarrow \mathbb{R}^m$ (ou função anti-simétrica; veja Arellano-Valle & Del Pino, 2003) e em tal caso $Q_m(w(-\mathbf{y})) = Q_m(-w(\mathbf{y}))$, $\mathbf{y} \in \mathbb{R}^n$. Assim, a distribuição normal-assimétrica em (2.6) pode ser estendida considerando

$$f_{\mathbf{Y}}(\mathbf{y}) = 2^m \phi_n(\mathbf{y}|\boldsymbol{\mu}, \Sigma) Q_m(w(\mathbf{y})), \quad \mathbf{y} \in \mathbb{R}^n. \quad (2.9)$$

Como notado na observação acima, a densidade em (2.6) é diferente da definição encontrada em Azzalini & Dalla-Valle (1996) e Azzalini & Capitanio (1999), pois estes autores consideram $\boldsymbol{\omega} = \text{diag}(\sigma_{11}^2, \dots, \sigma_{nn}^2)$ (σ_{ii}^2 : i -ésimo elemento da diagonal principal de Σ) no lugar de Σ na parte assimétrica de (2.6).

O seguinte lema será muito usado no decorrer deste trabalho; por exemplo, pode ser usado para mostrar que (2.6) é uma fdp em $\mathbb{R}^n$.

Lema 2.1. *Seja $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Então, para algum vetor fixo $\mathbf{a}$ de dimensão k e alguma matriz fixa $\mathbf{B}$ de dimensão $k \times n$, temos*

$$E[\Phi_k(\mathbf{a} + \mathbf{B}\mathbf{Y}|\boldsymbol{\eta}, \boldsymbol{\Omega})] = \Phi_k(\mathbf{a}|\boldsymbol{\eta} - \mathbf{B}\boldsymbol{\mu}, \boldsymbol{\Omega} + \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T), \quad (2.10)$$

$$E[\phi_k(\mathbf{a} + \mathbf{B}\mathbf{Y}|\boldsymbol{\eta}, \boldsymbol{\Omega})] = \phi_k(\mathbf{a}|\boldsymbol{\eta} - \mathbf{B}\boldsymbol{\mu}, \boldsymbol{\Omega} + \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T). \quad (2.11)$$

Demonstração. A prova de (2.10) segue notando que

$$E[\Phi_k(\mathbf{a} + \mathbf{B}\mathbf{Y}|\boldsymbol{\eta}, \boldsymbol{\Omega})] = E[P(\mathbf{U} \leq \mathbf{a}|\mathbf{Y})] = P(\mathbf{U} \leq \mathbf{a}),$$

onde $\mathbf{U}|\mathbf{Y} = \mathbf{y} \sim N_k(\boldsymbol{\eta} - \mathbf{B}\mathbf{y}, \boldsymbol{\Omega})$, tal que $\mathbf{U} \sim N_k(\boldsymbol{\eta} - \mathbf{B}\boldsymbol{\mu}, \boldsymbol{\Omega} + \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T)$. Similarmente a prova de (2.11), segue do fato que

$$E[\phi_k(\mathbf{a} + \mathbf{B}\mathbf{Y}|\boldsymbol{\eta}, \boldsymbol{\Omega})] = \int_{\mathbb{R}^n} \phi_k(\mathbf{a}|\boldsymbol{\eta} - \mathbf{B}\mathbf{z}, \boldsymbol{\Omega})\phi_n(\mathbf{z}|\boldsymbol{\mu}, \boldsymbol{\Sigma})d\mathbf{z} = \int_{\mathbb{R}^n} f_{\mathbf{U}|\mathbf{Z}}(\mathbf{u})f_{\mathbf{Z}}(\mathbf{z})d\mathbf{z} = f_{\mathbf{U}}(\mathbf{u}).$$

□

Note que o Lema 2.1 garante em particular que (2.6) é uma função de densidade em $\mathbb{R}^n$, pois

$$\begin{aligned} & \int_{\mathbb{R}^n} \phi_n(\mathbf{z}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1(\boldsymbol{\lambda}^T\boldsymbol{\Sigma}^{-1/2}(\mathbf{z} - \boldsymbol{\mu}))d\mathbf{z} = E[\Phi_1(\boldsymbol{\lambda}^T\boldsymbol{\Sigma}^{-1/2}(\mathbf{z} - \boldsymbol{\mu}))] \\ &= \Phi_1(-\boldsymbol{\lambda}^T\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu} | -\boldsymbol{\lambda}^T\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}, 1 + \boldsymbol{\lambda}^T\boldsymbol{\lambda}) = \Phi_1(0) = 2^{-1}. \end{aligned}$$

A seguir apresentamos a representação estocástica de um vetor aleatório normal-assimétrico padronizado. A prova é baseada no seguinte lema.

Lema 2.2. *Seja $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e $\mathbf{X} \sim N_q(\boldsymbol{\eta}, \boldsymbol{\Omega})$. Então,*

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

onde $\boldsymbol{\Lambda} = (\boldsymbol{\Omega}^{-1} + \mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}$.

Demonstração. Fazendo $\mathbf{z} = \mathbf{y} - \boldsymbol{\mu} - \mathbf{A}\boldsymbol{\eta}$ e $\mathbf{w} = \mathbf{x} - \boldsymbol{\eta}$, temos após algumas manipulações algébricas que

$$\begin{aligned} (\mathbf{z} - \mathbf{A}\mathbf{w})^T \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \mathbf{A}\mathbf{w}) + \mathbf{w}^T \boldsymbol{\Omega}^{-1} \mathbf{w} = \\ \mathbf{z}^T (\boldsymbol{\Sigma} + \mathbf{A}\boldsymbol{\Omega}\mathbf{A}^T)^{-1} \mathbf{z} + (\mathbf{w} - \boldsymbol{\Lambda}\mathbf{A}^T \boldsymbol{\Sigma}^{-1} \mathbf{z})^T \boldsymbol{\Lambda}^{-1} (\mathbf{w} - \boldsymbol{\Lambda}\mathbf{A}^T \boldsymbol{\Sigma}^{-1} \mathbf{z}). \end{aligned}$$

A prova segue por se notar também que $|\boldsymbol{\Sigma} + \mathbf{A}\boldsymbol{\Omega}\mathbf{A}^T| |\boldsymbol{\Lambda}| = |\boldsymbol{\Sigma}| |\boldsymbol{\Omega}|$, com $|\mathbf{A}| = \det(\mathbf{A})$. $\square$

Proposição 2.4. *Seja $\mathbf{Z} \sim SN_n(\boldsymbol{\lambda})$. Então,*

$$\mathbf{Z} \stackrel{d}{=} \delta |X_0| + (\mathbf{I}_n - \delta \delta^T)^{1/2} \mathbf{X}_1, \quad \text{onde } \delta = \frac{\boldsymbol{\lambda}}{(1 + \boldsymbol{\lambda}^T \boldsymbol{\lambda})^{1/2}}, \quad (2.12)$$

$X_0 \sim N(0, 1)$, $\mathbf{X}_1 \sim N_n(\mathbf{0}, \mathbf{I}_n)$ e são independentes.

Demonstração. Seja $\mathbf{U} = \delta |X_0| + (\mathbf{I}_n - \delta \delta^T)^{1/2} \mathbf{X}_1$. Como $\mathbf{U} || X_0| = t \sim N_n(\delta t, \mathbf{I}_n - \delta \delta^T)$, onde $|X_0| \sim HN(0, 1)$ (a distribuição metade-normal padronizada). Então, pelo Lema 2.2, temos que

$$\begin{aligned} f_{\mathbf{U}}(\mathbf{z}) &= \int_0^\infty \phi_n(\mathbf{z} | \delta t, \mathbf{I}_n - \delta \delta^T) 2\phi(t) dt \\ &= \int_0^\infty \phi_n(\mathbf{z} | \mathbf{0}, \mathbf{I}_n) 2\phi(t | \delta^T \mathbf{z}, 1 - \delta^T \delta) dt \\ &= 2\phi_n(\mathbf{z}) \Phi_1 \left(\frac{\delta^T \mathbf{z}}{\sqrt{1 - \delta^T \delta}} \right), \end{aligned} \quad (2.13)$$

i.e., $\mathbf{U} \stackrel{d}{=} \mathbf{Z} \sim SN_n(\boldsymbol{\lambda})$, com $\boldsymbol{\lambda} = \frac{\delta}{\sqrt{1 - \delta^T \delta}}$, que conclui a prova. $\square$

Observação 2.3. *Da Proposição 2.4, podemos alternativamente denotar um vetor aleatório $SN_n(\boldsymbol{\lambda})$ definido em (2.7), como $SN_n(\boldsymbol{\delta})$ com densidade dada em (2.13).*

Observação 2.4. *Note que em conexão com a Proposição 2.2 do caso univariado, não é difícil ver que para $X_o \sim N(0, 1)$, $\mathbf{X}_1 \sim N_n(\mathbf{0}, \mathbf{I}_n)$ e $\text{Cov}(X_o, \mathbf{X}_1) = \boldsymbol{\delta}$, o vetor aleatório $\mathbf{Z} = [\mathbf{X}_1 | X_o > 0]$ tem distribuição $SN_n(\boldsymbol{\lambda})$. Assim, paralelamente a (2.3) e alternativamente a (2.12), podemos usar o seguinte mecanismo para gerar vetores aleatórios $SN_n(\boldsymbol{\lambda})$:*

$$\mathbf{Z} = \begin{cases} \mathbf{X}_1, & \text{se } X_o > 0, \\ -\mathbf{X}_1, & \text{caso contrário.} \end{cases} \quad (2.14)$$

Figura 2.2: Contornos da distribuição normal-assimétrica bivariada em (2.6), com $\mu = (0, 0)^T$, $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ e $\lambda = (\lambda_1, \lambda_2)^T$ para diferentes valores de λ_1, λ_2 e ρ .

(a) Para $\lambda = (0, 3)^T$ e $\rho = 0, 0, 5$ e $0, 9$, respectivamente.

(b) Para $\lambda = (2, 3)^T$ e $\rho = 0, 0, 5$ e $0, 9$, respectivamente.

(c) Para $\lambda = (-2, 2)^T$ e $\rho = 0, 0, 5$ e $0, 9$, respectivamente.

Wang et al. (2004) e Azzalini & Capitanio (1999) discutem uma representação estocástica condicional em (2.14) em um contexto mais geral de distribuições assimétricas.

É importante notar que a representação estocástica em (2.4) é um caso especial da representação estocástica dada em (2.12), considerando $n = 1$. Este fato não necessariamente acontece com a definição proposta por Azzalini & Dalla-Valle (1996). Sendo assim, muitas propriedades da distribuição normal-assimétrica multivariada, definida aqui, podem ser derivadas para o caso univariado e, mais ainda, as propriedades mostradas a seguir para um vetor padronizado normal-assimétrico podem ser facilmente estendidas para o correspondente modelo de locação-escala devido ao fato apresentado na seguinte

proposição.

Proposição 2.5. *Seja $\mathbf{Z} \sim SN_n(\boldsymbol{\lambda})$, e considere a transformação linear $\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{Z}$, onde $\boldsymbol{\Sigma}$ é definida positiva. Então,*

$$\mathbf{Y} \sim SN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}).$$

Demonstração. Visto que $\boldsymbol{\Sigma}$ é definida positiva, a prova segue usando o fato que $f_{\mathbf{Y}}(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-1/2} f_{\mathbf{Z}}(\boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}))$. $\square$

O seguinte resultado mostra que a fda de um vector aleatório $SN_n(\boldsymbol{\lambda})$ tem uma forma simples.

Proposição 2.6. *Se $\mathbf{Z} \sim SN_n(\boldsymbol{\lambda})$ então, sua fda é dada por*

$$F_{\mathbf{Z}}(\mathbf{z}) = 2\Phi_{n+1}((\mathbf{z}^T, 0)^T | \mathbf{0}, \boldsymbol{\Omega}), \text{ onde } \boldsymbol{\Omega} = \begin{pmatrix} \mathbf{I}_n & -\boldsymbol{\lambda} \\ -\boldsymbol{\lambda}^T & 1 + \boldsymbol{\lambda}^T \boldsymbol{\lambda} \end{pmatrix}.$$

Demonstração. Por (2.7) temos

$$\begin{aligned} F_{\mathbf{Z}}(\mathbf{z}) &= P(\mathbf{Z} \leq \mathbf{z}) = 2 \int_{\mathbf{u} \leq \mathbf{0}} \phi_n(\mathbf{u} + \mathbf{z}) \Phi_1(\boldsymbol{\lambda}^T(\mathbf{u} + \mathbf{z})) d\mathbf{u} \\ &= 2 \int_{\mathbf{u} \leq \mathbf{0}} \int_{v \leq 0} \phi_n(\mathbf{u} + \mathbf{z}) \phi_1(v + \boldsymbol{\lambda}^T(\mathbf{u} + \mathbf{z})) dv d\mathbf{u} = 2P(\mathbf{U} \leq \mathbf{0}, V \leq 0), \end{aligned}$$

onde $V | \mathbf{U} = \mathbf{u} \sim N_1(-\boldsymbol{\lambda}^T(\mathbf{u} + \mathbf{z}), 1)$ e $\mathbf{U} \sim N_n(-\mathbf{z}, \mathbf{I}_n)$. Assim, a prova segue do fato de que

$$\begin{pmatrix} \mathbf{U} \\ V \end{pmatrix} \sim N_{n+1} \left(\begin{pmatrix} -\mathbf{z} \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{I}_n & -\boldsymbol{\lambda} \\ -\boldsymbol{\lambda}^T & 1 + \boldsymbol{\lambda}^T \boldsymbol{\lambda} \end{pmatrix} \right).$$

$\square$

Na Figura 2.2 apresentamos alguns contornos da densidade em (2.6) para diferentes valores dos parâmetros. Note que estes contornos não são elípticos, como era de esperar.

2.3.1 Função Geratriz de Momentos

No próximo resultado, derivamos a função geratriz de momentos (fgm) da distribuição normal-assimétrica padronizada. Propriedades adicionais desta distribuição serão obtidas usando esta função.

Proposição 2.7. *Seja $\mathbf{Z} \sim SN_n(\boldsymbol{\lambda})$. Então, a fgm é dada por*

$$M_{\mathbf{Z}}(\mathbf{s}) = 2e^{\frac{1}{2}\mathbf{s}^T\mathbf{s}}\Phi_1(\boldsymbol{\delta}^T\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^n, \quad \boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{(1 + \boldsymbol{\lambda}^T\boldsymbol{\lambda})^{1/2}}. \quad (2.15)$$

Demonstração. Note que $e^{\mathbf{s}^T\mathbf{z}}\phi_n(\mathbf{z}) = e^{\frac{1}{2}\mathbf{s}^T\mathbf{s}}\phi_n(\mathbf{z} - \mathbf{s})$. Logo, de (2.7) temos

$$\begin{aligned} M_{\mathbf{Z}}(\mathbf{s}) &= E(e^{\mathbf{s}^T\mathbf{Z}}) = 2 \int_{\mathbb{R}^n} e^{\frac{1}{2}\mathbf{s}^T\mathbf{s}}\phi_n(\mathbf{u} - \mathbf{s})\Phi_1(\boldsymbol{\lambda}\mathbf{u})d\mathbf{u} \\ &= 2e^{\frac{1}{2}\mathbf{s}^T\mathbf{s}} \int_{\mathbb{R}^n} \phi_n(\mathbf{y})\Phi_1(\boldsymbol{\lambda}^T(\mathbf{y} + \mathbf{s}))d\mathbf{y} = 2e^{\frac{1}{2}\mathbf{s}^T\mathbf{s}} \int_{\mathbb{R}^n} \int_{\mathbf{u} \leq \boldsymbol{\lambda}^T\mathbf{s}} \phi_n(\mathbf{y})\phi_1(\mathbf{u} + \boldsymbol{\lambda}^T\mathbf{y})d\mathbf{y}d\mathbf{u} \\ &= 2e^{\frac{1}{2}\mathbf{s}^T\mathbf{s}} \int_{\mathbf{u} \leq \boldsymbol{\lambda}^T\mathbf{s}} E[\phi_1(\mathbf{u} + \boldsymbol{\lambda}^T\mathbf{Y})]d\mathbf{u}, \quad \text{com } \mathbf{Y} \sim N_n(\mathbf{0}, \mathbf{I}_n). \end{aligned}$$

O resultado segue usando o Lema 2.1. □

Como consequência imediata da Proposição 2.7, temos o seguinte corolário. A prova segue do fato de que $\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{Z}$ e $M_{\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{Z}}(\mathbf{s}) = e^{\mathbf{s}^T\boldsymbol{\mu}}M_{\mathbf{Z}}(\boldsymbol{\Sigma}^{1/2}\mathbf{s})$.

Corolário 2.1. *Seja $\mathbf{Y} \sim SN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. Então, a fgm é dada por*

$$M_{\mathbf{Y}}(\mathbf{s}) = 2e^{\mathbf{s}^T\boldsymbol{\mu} + \frac{1}{2}\mathbf{s}^T\boldsymbol{\Sigma}\mathbf{s}}\Phi_1(\boldsymbol{\delta}^T\boldsymbol{\Sigma}^{1/2}\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^n, \quad (2.16)$$

De modo similar ao corolário acima podemos caracterizar a distribuição de uma transformação linear arbitrária na forma $\boldsymbol{\Delta}\mathbf{Z} + \mathbf{u}$. A prova segue diretamente da Proposição 2.7 e será omitida.

Proposição 2.8. *Seja $\mathbf{Z} \sim SN_n(\boldsymbol{\lambda})$. Então, para uma matriz $\boldsymbol{\Delta}$ de dimensão $k \times n$ e um vetor $\mathbf{u}$ de dimensão k temos que*

$$M_{\boldsymbol{\Delta}\mathbf{Z} + \mathbf{u}}(\mathbf{s}) = 2e^{\mathbf{s}^T\mathbf{u} + \frac{1}{2}\mathbf{s}^T\boldsymbol{\Delta}\boldsymbol{\Delta}^T\mathbf{s}}\Phi_1(\boldsymbol{\delta}^T\boldsymbol{\Delta}^T\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^k. \quad (2.17)$$

De (2.17), com $\mathbf{u} = \mathbf{0}$, temos as seguintes propriedades adicionais da distribuição normal-assimétrica padronizada.

Corolário 2.2. *Se $\mathbf{Z} \sim SN_n(\boldsymbol{\lambda})$, então*

- i) $-\mathbf{Z} \sim SN_n(-\boldsymbol{\lambda})$,
- ii) $\mathbf{a}^T\mathbf{Z} \sim SN_1(\mathbf{a}^T\boldsymbol{\lambda})$, onde $\mathbf{a}$ é um vetor unitário no $\mathbb{R}^n$,
- iii) $\mathbf{A}\mathbf{Z} \sim SN_n(\mathbf{A}\boldsymbol{\lambda})$, onde $\mathbf{A}$ é uma matriz ortogonal $n \times n$.

O seguinte resultado está relacionado com a transformação linear de um vetor aleatório normal-assimétrico. Como veremos é fechada sob transformação linear usando uma matriz de posto completo.

Proposição 2.9. *Seja $\mathbf{Y} \sim SN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ e $\mathbf{C}$ uma matriz $n \times k$ de posto completo. Então,*

$$\mathbf{C}^T \mathbf{Y} \sim SN_k(\mathbf{C}^T \boldsymbol{\mu}, \mathbf{C}^T \boldsymbol{\Sigma} \mathbf{C}, \boldsymbol{\lambda}^*),$$

onde $\boldsymbol{\lambda}^* = \boldsymbol{\delta}_*/(1 - \boldsymbol{\delta}_*^T \boldsymbol{\delta}_*)^{1/2}$, com $\boldsymbol{\delta}_* = (\mathbf{C}^T \boldsymbol{\Sigma} \mathbf{C})^{-1/2} \mathbf{C}^T \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta}$.

Demonstração. De (2.16), temos para $\mathbf{s} \in \mathbb{R}^k$ que

$$\begin{aligned} M_{\mathbf{C}^T \mathbf{Y}}(\mathbf{s}) &= M_{\mathbf{Y}}(\mathbf{C}\mathbf{s}) = 2e^{\mathbf{s}^T \mathbf{C}^T \boldsymbol{\mu} + \frac{1}{2} \mathbf{s}^T \mathbf{C}^T \boldsymbol{\Sigma} \mathbf{C} \mathbf{s}} \Phi_1(\boldsymbol{\delta}^T \boldsymbol{\Sigma}^{1/2} \mathbf{C} \mathbf{s}) \\ &= 2e^{\mathbf{s}^T \mathbf{C}^T \boldsymbol{\mu} + \frac{1}{2} \mathbf{s}^T \mathbf{C}^T \boldsymbol{\Sigma} \mathbf{C} \mathbf{s}} \Phi_1(\boldsymbol{\delta}^T \boldsymbol{\Sigma}^{1/2} \mathbf{C} (\mathbf{C}^T \boldsymbol{\Sigma} \mathbf{C})^{-1/2} (\mathbf{C}^T \boldsymbol{\Sigma} \mathbf{C})^{1/2} \mathbf{s}). \end{aligned}$$

Novamente, em analogia a (2.16), o resultado segue. $\square$

Note na Proposição 2.9 que fazendo $\mathbf{C} = -\mathbf{I}_n$, então $-\mathbf{Y} \sim SN_n(-\boldsymbol{\mu}, \boldsymbol{\Sigma}, -\boldsymbol{\lambda})$. O seguinte resultado é uma consequência direta da Proposição 2.9, Sua importância é clara pois fornece uma caracterização para qualquer vetor aleatório normal-assimétrico na mesma forma que a distribuição normal multivariada.

Corolário 2.3. $\mathbf{Y} \sim SN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ se, e somente se, $\mathbf{a}^T \mathbf{Y} \sim SN_1(\mu_a, \Sigma_a, \delta_a)$ para todo $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{a} \neq 0$, onde $\mu_a = \mathbf{a}^T \boldsymbol{\mu}$, $\Sigma_a = \mathbf{a}^T \boldsymbol{\Sigma} \mathbf{a}$ e $\delta_a = (\mathbf{a}^T \boldsymbol{\Sigma} \mathbf{a})^{-1/2} \mathbf{a}^T \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta}$.

Demonstração. A prova da implicação é direta da Proposição 2.9 com $k = 1$. Para mostrar a condição necessária, observe que se $\mathbf{a}^T \mathbf{Y} \sim SN_1(\mu_a, \Sigma_a, \delta_a)$ para todo $\mathbf{a} \neq 0$, de (2.16) com $n = 1$, temos

$$\begin{aligned} M_{\mathbf{a}^T \mathbf{Y}}(s) = M_{\mathbf{Y}}(\mathbf{a}s) &= 2e^{s \mathbf{a}^T \boldsymbol{\mu} + \frac{1}{2} s^2 \mathbf{a}^T \boldsymbol{\Sigma} \mathbf{a}} \Phi_1(\delta_a (\mathbf{a}^T \boldsymbol{\Sigma} \mathbf{a})^{1/2} s) \\ &= 2e^{s \mathbf{a}^T \boldsymbol{\mu} + \frac{1}{2} s^2 \mathbf{a}^T \boldsymbol{\Sigma} \mathbf{a}} \Phi_1(\boldsymbol{\delta}^T \boldsymbol{\Sigma}^{1/2} \mathbf{a} s), \quad s \in \mathbb{R}. \end{aligned}$$

Tomando $s = 1$, a expressão acima se reduz a

$$M_{\mathbf{Y}}(\mathbf{a}) = 2e^{\mathbf{a}^T \boldsymbol{\mu} + \frac{1}{2} \mathbf{a}^T \boldsymbol{\Sigma} \mathbf{a}} \Phi_1(\boldsymbol{\delta}^T \boldsymbol{\Sigma}^{1/2} \mathbf{a}),$$

e, como $\mathbf{a}$ é arbitrário, por (2.16) segue que $\mathbf{Y} \sim SN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. $\square$

A seguir caracterizamos a distribuição das formas quadráticas de um vetor aleatório normal-assimétrico através de sua fgm.

Proposição 2.10. Sejam $\mathbf{Z} \sim SN_n(\boldsymbol{\lambda})$, $\mathbf{A}$ uma matriz simétrica de dimensão $n \times n$ e $\mathbf{a} \in \mathbb{R}^n$. Então, a fgm de $Q = (\mathbf{Z} - \mathbf{a})^T \mathbf{A} (\mathbf{Z} - \mathbf{a})$ é

$$\begin{aligned} M_Q(s) &= 2|\mathbf{I}_n - 2s\mathbf{A}|^{-1/2} \exp\{\mathbf{a}^T [s\mathbf{A} + 2s^2\mathbf{A}(\mathbf{I}_n - 2s\mathbf{A})^{-1}\mathbf{A}]\mathbf{a}\} \\ &\quad \times \Phi_1\left(-2\frac{s\boldsymbol{\lambda}^T(\mathbf{I}_n - 2s\mathbf{A})^{-1}\mathbf{A}\mathbf{a}}{(1 + \boldsymbol{\lambda}^T(\mathbf{I}_n - 2s\mathbf{A})^{-1}\boldsymbol{\lambda})^{1/2}}\right), \quad s \in \mathbb{R}, \end{aligned} \quad (2.18)$$

com $|\mathbf{A}| = \det(\mathbf{A})$.

Demonstração. Note primeiro que $e^{s(\mathbf{Z}-\mathbf{a})^T \mathbf{A} (\mathbf{Z}-\mathbf{a})} \phi_n(\mathbf{z}) = |\mathbf{I}_n - 2s\mathbf{A}|^{-1/2} \exp\{\mathbf{a}^T [s\mathbf{A} + 2s^2\mathbf{A}(\mathbf{I}_n - 2s\mathbf{A})^{-1}\mathbf{A}]\mathbf{a}\} \phi_n(\mathbf{z} - 2s(\mathbf{I}_n - 2s\mathbf{A})^{-1}\mathbf{A}\mathbf{a}, (\mathbf{I}_n - 2s\mathbf{A})^{-1})$. Agora, da definição de fgm, temos para $s \in \mathbb{R}$ que

$$\begin{aligned} E(e^{sQ}) &= 2|\mathbf{I}_n - 2s\mathbf{A}|^{-1/2} \exp\{\mathbf{a}^T [s\mathbf{A} + 2s^2\mathbf{A}(\mathbf{I}_n - 2s\mathbf{A})^{-1}\mathbf{A}]\mathbf{a}\} \\ &\quad \times \int_{\mathbf{u} \in \mathbb{R}^n} \phi_n(\mathbf{u} | -2s(\mathbf{I}_n - 2s\mathbf{A})^{-1}\mathbf{A}\mathbf{a}, (\mathbf{I}_n - 2s\mathbf{A})^{-1}) \Phi_1(\boldsymbol{\lambda}^T \mathbf{u}) d\mathbf{u} \\ &= 2|\mathbf{I}_n - 2s\mathbf{A}|^{-1/2} \exp\{\mathbf{a}^T [s\mathbf{A} + 2s^2\mathbf{A}(\mathbf{I}_n - 2s\mathbf{A})^{-1}\mathbf{A}]\mathbf{a}\} E[\Phi_1(\boldsymbol{\lambda}^T \mathbf{Z})], \end{aligned}$$

onde $\mathbf{Z} \sim N_n(-2s(\mathbf{I}_n - 2s\mathbf{A})^{-1}\mathbf{A}\mathbf{a}, (\mathbf{I}_n - 2s\mathbf{A})^{-1})$. A prova segue do Lema 2.1 e algumas manipulações algébricas. $\square$

Duas importantes conseqüências da proposição acima, considerando $\mathbf{a} = \mathbf{0}$ e $\mathbf{A} = \mathbf{I}_n$, são dados no seguinte corolário.

Corolário 2.4. Seja $\mathbf{Z} \sim SN_n(\boldsymbol{\lambda})$. Então, sob as condições da Proposição 2.10,

a) $Q = \mathbf{Z}^T \mathbf{A} \mathbf{Z}$ tem fgm dada por

$$M_Q(s) = |\mathbf{I}_n - 2s\mathbf{A}|^{-1/2}, \quad \mathbf{I}_n - 2s\mathbf{A} > 0, \quad s \in \mathbb{R}. \quad (2.19)$$

b) $Q = \mathbf{Z}^T \mathbf{Z}$ tem distribuição qui-quadrado com n graus de liberdade, isto é,

$$Q = \mathbf{Z}^T \mathbf{Z} \sim \chi_n^2. \quad (2.20)$$

Onde a notação $\mathbf{A} > 0$ significa que $\mathbf{A}$ é definida positiva. Note que a distribuição das formas quadráticas centrais de um vetor aleatório normal-assimétrico padronizado não depende do parâmetro de assimetria $\boldsymbol{\lambda}$. Assim, propriedades de Q podem ser obtidas usando resultados conhecidos da distribuição normal multivariada; veja por exemplo,

Johnson & Kotz (1972). De $M_Q(s)$ em (2.19) e da relação em (A-21), encontramos que o r -ésimo cumulante de $\mathbf{Z}^T \mathbf{A} \mathbf{Z}$ é

$$K_r = 2^{r-1}(r-1)!\text{tr}(\mathbf{A})^r, \quad r = 1, 2, \dots \quad (2.21)$$

O fato apresentado em (2.20) é um caso especial de um resultado mais geral apresentado a seguir. A prova é decorrente da Proposição 2.9.

Corolário 2.5. *Seja $\mathbf{Y} \sim SN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda)$ e $\mathbf{C}$ uma matriz $n \times k$ ($k \leq n$) de posto completo. Então,*

$$Q = (\mathbf{C}^T \mathbf{Y} - \mathbf{C}^T \boldsymbol{\mu})^T (\mathbf{C}^T \boldsymbol{\Sigma} \mathbf{C})^{-1} (\mathbf{C}^T \mathbf{Y} - \mathbf{C}^T \boldsymbol{\mu}) \sim \chi_k^2. \quad (2.22)$$

Note neste último corolário que se $k = 1$, $\mathbf{C} = \mathbf{e}_j$, com $\mathbf{e}_j$ um vetor cujo j -ésimo elemento é igual a 1 e os outros iguais a 0 (base canônica de $\mathbb{R}^n$). Então,

$$(y_j - \mu_j)^2 / \sigma_j^2 \sim \chi_1^2, \quad (2.23)$$

onde σ_j^2 é o j -ésimo elemento da diagonal principal da matriz de dispersão $\boldsymbol{\Sigma}$ e μ_j o j -ésimo elemento do parâmetro de locação $\boldsymbol{\mu}$, $j = 1, \dots, n$.

Proposição 2.11. *Sejam $\mathbf{Y} \sim SN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda)$ e $\mathbf{A}$ uma matriz simétrica de ordem n . Então a fgm de $Q = \mathbf{Y}^T \mathbf{A} \mathbf{Y}$ é dada por*

$$M_Q(s) = 2|\mathbf{I}_n - 2s\mathbf{A}\boldsymbol{\Sigma}|^{-1/2} \exp\{\boldsymbol{\mu}^T [s\mathbf{A} + 2s^2\mathbf{A}(\boldsymbol{\Sigma}^{-1} - 2s\mathbf{A})^{-1}\mathbf{A}]\boldsymbol{\mu}\} \times \\ \Phi_1 \left(-2 \frac{s\boldsymbol{\lambda}^T (\mathbf{I}_n - 2s\boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2})^{-1}\boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\mu}}{(1 + \boldsymbol{\lambda}^T (\mathbf{I}_n - 2s\boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2})^{-1}\boldsymbol{\lambda})^{1/2}} \right), s \in \mathbb{R}. \quad (2.24)$$

Demonstração. Note que $Q = \mathbf{Y}^T \mathbf{A} \mathbf{Y} = (\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{Z})^T \mathbf{A} (\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{Z}) = (\mathbf{Z} - \bar{\boldsymbol{\mu}})^T \bar{\mathbf{A}} (\mathbf{Z} - \bar{\boldsymbol{\mu}})$, onde $\bar{\mathbf{A}} = \boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}$, $\bar{\boldsymbol{\mu}} = -\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}$ e $\mathbf{Z} \sim SN_n(\boldsymbol{\lambda})$. A prova segue da Proposição 2.10. $\square$

Os seguintes resultados estão relacionados com a distribuição de matrizes aleatórias normal-assimétrica.

Proposição 2.12. *Seja $\mathbf{Y} \sim SN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda)$. Então,*

$$\mathbf{Q} = (\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})^T \sim W_n(\boldsymbol{\Sigma}, 1), \quad (2.25)$$

isto é, $\mathbf{Q}$ tem distribuição Wishart com parâmetro de escala $\boldsymbol{\Sigma}$ e 1 grau de liberdade.

Demonstração. Note que se $\mathbf{Z} \sim SN_n(\boldsymbol{\lambda})$, então da definição da fgm para um matriz de variáveis aleatórias, temos

$$M_{\mathbf{Z}\mathbf{Z}^T}(\mathbf{S}) = E(e^{\text{tr}(\mathbf{S}\mathbf{Z}\mathbf{Z}^T)}) = E(e^{\text{tr}(\mathbf{Z}^T\mathbf{S}\mathbf{Z})}) = E(e^{\mathbf{Z}^T\mathbf{S}\mathbf{Z}}), \mathbf{S} \in \mathbb{R}^{n \times n},$$

de um modo similar à prova da Proposição 2.10, com $\mathbf{a} = \mathbf{0}$, segue que $\mathbf{Z}\mathbf{Z}^T$ tem fgm dada por

$$M_{\mathbf{Z}\mathbf{Z}^T}(\mathbf{S}) = |\mathbf{I}_n - 2\mathbf{S}|^{-1/2}.$$

A fgm acima não depende de $\boldsymbol{\lambda}$ e caracteriza a fdp de uma distribuição Wishart padronizada, mais especificamente $\mathbf{Z}\mathbf{Z}^T \sim W(\mathbf{I}_n, 1)$. Logo, como $\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{Z}$, uma pequena álgebra usando propriedades conhecidas da distribuição Wishart (veja, por exemplo, Mardia et al., 1979) completa a prova. □

O seguinte corolário é decorrente da Proposição 2.12 e Teorema 3.4.3 em Mardia et al. (1979).

Corolário 2.6. *Sejam $\mathbf{X}_1, \dots, \mathbf{X}_p$ vetores aleatórios independentes tais que $\mathbf{X}_i \sim SN_n(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}, \boldsymbol{\alpha}_i)$ e se $\mathbf{X}, \mathbf{U}$ são matrizes cujas linhas são $\mathbf{X}_1^T, \dots, \mathbf{X}_p^T$ e $\boldsymbol{\mu}_1^T, \dots, \boldsymbol{\mu}_p^T$, respectivamente. Então,*

$$(\mathbf{X} - \mathbf{U})^T(\mathbf{X} - \mathbf{U}) \sim W_n(\boldsymbol{\Sigma}, p).$$

Proposição 2.13. *Seja $\mathbf{Y} \sim SN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ independente de $\mathbf{X}_1, \dots, \mathbf{X}_p$, com $\mathbf{X}_i, \mathbf{X}$ e $\mathbf{U}$ como no Corolário 2.6. Se $[(\mathbf{X} - \mathbf{U})^T(\mathbf{X} - \mathbf{U})]^{-1}$ existe quase certamente, então,*

$$T^2 = p(\mathbf{Y} - \boldsymbol{\mu})^T[(\mathbf{X} - \mathbf{U})^T(\mathbf{X} - \mathbf{U})]^{-1}(\mathbf{Y} - \boldsymbol{\mu}) \sim T(n, p),$$

isto é, T^2 tem distribuição de Hotelling com parâmetros n e p .

Demonstração. Note pelo Corolário 2.6 que a distribuição de $[(\mathbf{X} - \mathbf{U})^T(\mathbf{X} - \mathbf{U})]$ não depende de $\boldsymbol{\lambda}, \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_p$. Similarmente, pela Proposição 2.10 a distribuição condicional

$$(\mathbf{Y} - \boldsymbol{\mu})^T[(\mathbf{X} - \mathbf{U})^T(\mathbf{X} - \mathbf{U})]^{-1}(\mathbf{Y} - \boldsymbol{\mu}) \mid ((\mathbf{X} - \mathbf{U})^T(\mathbf{X} - \mathbf{U})),$$

não depende de $\boldsymbol{\lambda}, \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_p$. Logo, sem perda de generalidade assumimos que $\boldsymbol{\lambda}, \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_p$ são iguais a zero. O resultado segue usando propriedades conhecidas da distribuição de Hotelling. □

Na seção seguinte apresentamos resultados relacionados aos momentos do vetor aleatório normal-assimétrico e suas formas quadráticas.

2.3.2 Momentos

Proposição 2.14. *Seja $\mathbf{Z} \sim SN_n(\boldsymbol{\lambda})$. Então,*

$$a) \mu_1^o = E[\mathbf{Z}] = \sqrt{\frac{2}{\pi}} \boldsymbol{\delta},$$

$$b) \mu_2^o = E[\mathbf{Z}\mathbf{Z}^T] = \mathbf{I}_n, \quad Var(\mathbf{Z}) = \mathbf{I}_n - \frac{2}{\pi} \boldsymbol{\delta}\boldsymbol{\delta}^T,$$

$$c) \mu_3^o = E[vec(\mathbf{Z}\mathbf{Z}^T)\mathbf{Z}^T] = \sqrt{\frac{2}{\pi}} [\boldsymbol{\delta} \otimes \mathbf{I}_n + vec(\mathbf{I}_n)\boldsymbol{\delta}^T + \mathbf{I}_n \otimes \boldsymbol{\delta} - \boldsymbol{\delta} \otimes (\boldsymbol{\delta}\boldsymbol{\delta}^T)],$$

$$d) \mu_4^o = E(vec(\mathbf{Z}\mathbf{Z}^T) \otimes vec(\mathbf{Z}\mathbf{Z}^T)^T) = \mathbf{I}_{n^2} + \mathbf{K}_{nn} + vec(\mathbf{I}_n)vec^T(\mathbf{I}_n),$$

onde $\otimes$, $\mathbf{K}_{nn}$ e vec são o produto de Kronecker, a matriz de permutação (dimensão $n^2 \times n^2$) e operador vetorial, respectivamente (veja Magnus & Neudecker, 1979 os detalhes desta notação) e $\boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{(1+\boldsymbol{\lambda}^T\boldsymbol{\lambda})^{1/2}}$.

Demonstração. De (2.15), calculamos as derivadas parciais até ordem 3 que estão relacionadas com os 3 primeiros momentos do vetor aleatório $\mathbf{Z} \sim SN_n(\boldsymbol{\lambda})$. Após uma longa mas não complicada álgebra, segue que (veja expressões em (A-2), (A-8) e (A-9))

$$\frac{\partial M(\mathbf{s})}{\partial \mathbf{s}} = 2 \exp\left\{\frac{1}{2}\mathbf{s}^T\mathbf{s}\right\} [\mathbf{s}\Phi_1(\boldsymbol{\delta}^T\mathbf{s}) + \boldsymbol{\delta}\phi_1(\boldsymbol{\delta}^T\mathbf{s})], \quad (2.26)$$

$$\begin{aligned} \frac{\partial^2 M(\mathbf{s})}{\partial \mathbf{s}\partial \mathbf{s}^T} &= 2 \exp\left\{\frac{1}{2}\mathbf{s}^T\mathbf{s}\right\} \{ \Phi_1(\boldsymbol{\delta}^T\mathbf{s})(\mathbf{I}_n + \mathbf{s} \otimes \mathbf{s}^T) \} + \phi_1(\boldsymbol{\delta}^T\mathbf{s}) [\mathbf{s} \otimes \boldsymbol{\delta}^T + \boldsymbol{\delta} \otimes \mathbf{s}^T \\ &\quad - \boldsymbol{\delta} \otimes \boldsymbol{\delta}^T(\boldsymbol{\delta}^T\mathbf{s})], \end{aligned} \quad (2.27)$$

$$\begin{aligned} \frac{\partial M^3(\mathbf{s})}{\partial \mathbf{s}\partial \mathbf{s}^T\partial \mathbf{s}} &= 2 \exp\left\{\frac{1}{2}\mathbf{s}^T\mathbf{s}\right\} \Phi(\boldsymbol{\delta}^T\mathbf{s}) \{ \mathbf{s} \otimes \mathbf{I}_n + vec(\mathbf{I}_n)\mathbf{s}^T + (\mathbf{I}_n \otimes \mathbf{s})[\mathbf{I}_n + \mathbf{s}\mathbf{s}^T] \} \\ &\quad + 2 \exp\left\{\frac{1}{2}\mathbf{s}^T\mathbf{s}\right\} \phi(\boldsymbol{\delta}^T\mathbf{s}) \{ \boldsymbol{\delta} \otimes \mathbf{I}_n + vec(\mathbf{I}_n)\boldsymbol{\delta}^T + (\mathbf{I}_n \otimes \mathbf{s})[\boldsymbol{\delta} \otimes \mathbf{s}^T + \mathbf{s} \otimes \boldsymbol{\delta}^T \\ &\quad - (\boldsymbol{\delta} \otimes \boldsymbol{\delta}^T)(\boldsymbol{\delta}^T\mathbf{s})] + (\mathbf{I}_n \otimes \boldsymbol{\delta})[\mathbf{I}_n + \mathbf{s} \otimes \mathbf{s}^T - (\boldsymbol{\delta} \otimes \mathbf{s}^T)(\boldsymbol{\delta}^T\mathbf{s}) - \boldsymbol{\delta} \otimes \boldsymbol{\delta}^T \\ &\quad - (\mathbf{s} \otimes \boldsymbol{\delta}^T)(\boldsymbol{\delta}^T\mathbf{s}) + (\boldsymbol{\delta} \otimes \boldsymbol{\delta}^T)(\boldsymbol{\delta}^T\mathbf{s})^2] \}. \end{aligned} \quad (2.28)$$

Fazendo $\mathbf{s} = \mathbf{0}$ em (2.26)-(2.28), os resultados de (a), (b) e (c) seguem. Para a obtenção de (d) veja, por exemplo, Magnus & Neudecker (1979, p. 388), lembrando que os momentos de ordem par coincidem com os respectivos momentos da distribuição normal padrão. $\square$

Proposição 2.15. *Seja $\mathbf{Y} \sim SN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. Então,*

$$a) \mu_1 = E(\mathbf{Y}) = \mu + \sqrt{\frac{2}{\pi}} \Sigma^{1/2} \delta,$$

$$b) \mu_2 = E(\mathbf{Y}\mathbf{Y}^T) = \Sigma + \mu\mu^T + \sqrt{\frac{2}{\pi}}(\mu\delta^T \Sigma^{1/2} + \Sigma^{1/2} \delta\mu^T),$$

$$c) \Omega = Var(\mathbf{Y}) = \Sigma - \frac{2}{\pi} \Sigma^{1/2} \delta\delta^T \Sigma^{1/2},$$

d)

$$\begin{aligned} \mu_3 &= E(vec(\mathbf{Y}\mathbf{Y}^T)\mathbf{Y}^T) = \Sigma \otimes \mu + \mu \otimes \Sigma + vec(\Sigma) \otimes \mu^T + \mu \otimes \mu^T \otimes \mu \\ &\quad + \sqrt{\frac{2}{\pi}}[(\Sigma^{1/2} \delta) \otimes \Sigma + vec(\Sigma) \delta^T \Sigma^{1/2} + (\mathbf{I}_n \otimes (\Sigma^{1/2} \delta)) \Sigma \\ &\quad - (\Sigma^{1/2} \delta) \otimes (\delta^T \Sigma^{1/2}) \otimes (\Sigma^{1/2} \delta) + (\Sigma^{1/2} \delta) \otimes \mu^T \otimes \mu + \mu \otimes (\delta^T \Sigma^{1/2}) \otimes \mu \\ &\quad + \mu \otimes \mu^T \otimes (\Sigma^{1/2} \delta)]. \end{aligned}$$

Demonstração. A prova segue da Proposição 2.14, fazendo $\mathbf{Y} = \mu + \Sigma^{1/2} \mathbf{Z}$. $\square$

A proposição seguinte apresenta os dois primeiros momentos das formas quadráticas de um vetor aleatório normal-assimétrico padronizado.

Proposição 2.16. *Sejam $\mathbf{Z} \sim SN_n(\lambda)$ e $\mathbf{A}, \mathbf{B}$ duas matrizes simétricas de dimensão $n \times n$. Então*

$$(a) E[\mathbf{Z}^T \mathbf{A} \mathbf{Z}] = tr(\mathbf{A}),$$

$$(b) E[(\mathbf{Z}^T \mathbf{A} \mathbf{Z})^3] = 8tr(\mathbf{A}^3),$$

$$(c) Var[\mathbf{Z}^T \mathbf{A} \mathbf{Z}] = 2tr(\mathbf{A}^2),$$

$$(d) Cov(\mathbf{Z}^T \mathbf{A} \mathbf{Z}, \mathbf{Z}^T \mathbf{B} \mathbf{Z}) = 2tr(\mathbf{A} \mathbf{B}),$$

$$(e) Cov((\mathbf{Z}^T \mathbf{A} \mathbf{Z})^2, \mathbf{Z}^T \mathbf{B} \mathbf{Z}) = 4tr(\mathbf{A})tr(\mathbf{A} \mathbf{B}) + 8tr(\mathbf{A}^2 \mathbf{B}),$$

$$(f) Cov((\mathbf{Z}^T \mathbf{A} \mathbf{Z})^3, \mathbf{Z}^T \mathbf{B} \mathbf{Z}) = 6[tr(\mathbf{A})]^2 tr(\mathbf{A} \mathbf{B}) + 12tr(\mathbf{A}^2)tr(\mathbf{A} \mathbf{B}) + 24tr(\mathbf{A})tr(\mathbf{A}^2 \mathbf{B}) + 48tr(\mathbf{A}^3 \mathbf{B}),$$

onde $tr(\mathbf{A})$ denota o traço de $\mathbf{A}$.

Demonstração. A prova segue do Corolário 2.4, pois uma vez que a fgm das formas quadráticas não depende do parâmetro de assimetria, então podemos considerá-lo, sem perda de generalidade, como sendo igual a zero e usar os resultados referentes à distribuição normal encontrados, por exemplo, em Ghazal (1996) $\square$

O corolário seguinte apresenta os dois primeiros momentos das formas quadráticas de um vetor aleatório normal-assimétrico qualquer. A prova segue da Proposição 2.16 e do fato de que $\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{Z}$, quando $\mathbf{Y} \sim SN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$.

Corolário 2.7. *Sejam $\mathbf{Y} \sim SN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ e $\mathbf{A}, \mathbf{B}$ duas matrizes simétricas de dimensão $n \times n$. Então*

$$\begin{aligned}
(a) \quad E(\mathbf{Y}^T \mathbf{A} \mathbf{Y}) &= \text{tr}(\mathbf{A} \boldsymbol{\Sigma}) + \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} + 2\sqrt{\frac{2}{\pi}} \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta}, \\
(b) \quad \text{Var}[\mathbf{Y}^T \mathbf{A} \mathbf{Y}] &= 2\text{tr}((\mathbf{A} \boldsymbol{\Sigma})^2) + 4\boldsymbol{\mu}^T (\mathbf{A} \boldsymbol{\Sigma} \mathbf{A}) (\boldsymbol{\mu} + 2\sqrt{\frac{2}{\pi}} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta}) - 2\sqrt{\frac{2}{\pi}} [2(\boldsymbol{\delta}^T \boldsymbol{\Sigma}^{1/2} \mathbf{A} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta}) \\
&\quad \times (\boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta}) + 2\sqrt{\frac{2}{\pi}} (\boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta})^2], \\
(c) \quad \text{Cov}(\mathbf{Y}^T \mathbf{A} \mathbf{Y}, \mathbf{Y}^T \mathbf{B} \mathbf{Y}) &= 2\text{tr}(\mathbf{A} \boldsymbol{\Sigma} \mathbf{B} \boldsymbol{\Sigma}) + 2\boldsymbol{\mu}^T (\mathbf{A} \boldsymbol{\Sigma} \mathbf{B} + \mathbf{B} \boldsymbol{\Sigma} \mathbf{A}) (\boldsymbol{\mu} + 2\sqrt{\frac{2}{\pi}} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta}) \\
&\quad - 2\sqrt{\frac{2}{\pi}} [(\boldsymbol{\delta}^T \boldsymbol{\Sigma}^{1/2} \mathbf{A} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta}) (\boldsymbol{\mu}^T \mathbf{B} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta}) + (\boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta}) (\boldsymbol{\delta}^T \boldsymbol{\Sigma}^{1/2} \mathbf{B} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta}) \\
&\quad + 2\sqrt{\frac{2}{\pi}} (\boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta}) (\boldsymbol{\mu}^T \mathbf{B} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta})].
\end{aligned}$$

Note que quando $\boldsymbol{\delta} = \mathbf{0}$ (ou equivalentemente $\boldsymbol{\lambda} = \mathbf{0}$), os resultados do corolário acima se reduz aos obtidos para o caso normal (veja Muirhead, 1982, pag. 47).

2.3.3 Distribuição Marginal e Independência

Denotemos por $(\mathbf{Z}_i, i = 1, 2)$ e por $(\boldsymbol{\delta}_i, i = 1, 2)$ a partição do vetor aleatório normal-assimétrico e a partição induzida no parâmetro $\boldsymbol{\delta} = \boldsymbol{\lambda}/(1 + \boldsymbol{\lambda}^T \boldsymbol{\lambda})^{1/2}$, respectivamente, onde $\mathbf{Z}_1 \in \mathbb{R}^{n_1}$ e $\mathbf{Z}_2 \in \mathbb{R}^{n_2}$, com $\sum_{i=1}^2 n_i = n$. Considerando esta notação, no próximo resultado mostramos que a distribuição normal-assimétrica é fechada sob marginalização.

Proposição 2.17. *Assuma que $\mathbf{Z} = (\mathbf{Z}_i, i = 1, 2) \sim SN_{n_i}(\boldsymbol{\lambda})$. Então $\mathbf{Z}_i \sim SN_{n_i}(\boldsymbol{\lambda}_i^*)$, onde $\boldsymbol{\lambda}_i^* = \frac{\boldsymbol{\delta}_i}{1 - \boldsymbol{\delta}_i^T \boldsymbol{\delta}_i}$, $i = 1, 2$, isto é,*

$$f_{\mathbf{Z}_i}(\mathbf{z}_i) = 2\phi_{n_i}(\mathbf{z}_i)\Phi_1(\boldsymbol{\lambda}_i^{*T} \mathbf{z}_i), \quad \mathbf{z}_i \in \mathbb{R}^{n_i}, \quad i = 1, 2.$$

Demonstração. A prova é direta de (2.17) tomando $\mathbf{u} = \mathbf{0}$ e $\boldsymbol{\Delta} = [\mathbf{I}_{n_1}, \mathbf{0}]$ para $i = 1$ e $\boldsymbol{\Delta} = [\mathbf{0}, \mathbf{I}_{n_2}]$ para $i = 2$. □

Como notado por Arellano-Valle & Genton (2003), não é possível obter independência entre os elementos de um vetor aleatório normal-assimétrico. No entanto, podemos examinar a independência entre blocos de uma transformação linear da forma $\mathbf{C}^T \mathbf{Y}$, com $\mathbf{C}$ como na Proposição 2.9. Para estabelecer independência entre blocos de $\mathbf{C}^T \mathbf{Y}$, um papel importante desempenha a componente $\Phi_1(\cdot)$ da expressão (2.6). Visto que $\Phi_1(u+v)$ não pode ser fatorado como o produto de $\Phi_1(u)\Phi_1(v)$, para que exista independência mútua, no máximo um dos blocos de $\mathbf{C}^T \mathbf{Y}$ deve ser uma normal-assimétrica própria, enquanto os outros deveriam ter parâmetro de assimetria nulo (variáveis normais). Isto é resumido na seguinte proposição.

Proposição 2.18. *Seja $\mathbf{Y} \sim SN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, $\mathbf{C}$ ($n \times k$) uma matriz de posto completo k ($k \leq n$) e h um inteiro positivo. Adicionalmente, suponha que $\boldsymbol{\lambda}$ e $\mathbf{C}$ são particionadas em $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^T, \dots, \boldsymbol{\lambda}_h^T)^T$ e $\mathbf{C} = (\mathbf{C}_1, \dots, \mathbf{C}_h)$, onde $\mathbf{C}_i$ de dimensão $n \times r_i$, $i = 1, \dots, h$, com $\sum_{i=1}^h r_i = k$. Se simultaneamente*

$$(i) \quad \mathbf{C}_i^T \boldsymbol{\Sigma} \mathbf{C}_j = \mathbf{0}, \quad i \neq j,$$

$$(ii) \quad \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{1/2} \mathbf{C}_i \neq \mathbf{0}, \quad \text{para no máximo um } i,$$

então, $\mathbf{C}_i^T \mathbf{Y}$ é independente de $\mathbf{C}_j^T \mathbf{Y}$, $i \neq j$, $i, j = 1, \dots, h$.

Demonstração. De (2.16) temos para $\mathbf{s} = (\mathbf{s}_1^T, \dots, \mathbf{s}_h^T)^T \in \mathbb{R}^k$ que

$$\begin{aligned} M_{\mathbf{C}^T \mathbf{Y}}(\mathbf{s}) &= 2e^{\mathbf{s}^T \mathbf{C}^T \boldsymbol{\mu} + \frac{1}{2} \mathbf{s}^T \mathbf{C}^T \boldsymbol{\Sigma} \mathbf{C} \mathbf{s}} \Phi_1(\boldsymbol{\delta}^T \boldsymbol{\Sigma}^{1/2} \mathbf{C} \mathbf{s}) \\ &= 2 \exp \left(\sum_{i=1}^h \mathbf{s}_i^T \mathbf{C}_i^T \boldsymbol{\mu} + \frac{1}{2} \left[\sum_{i=1}^h \mathbf{s}_i^T \mathbf{C}_i^T \boldsymbol{\Sigma} \mathbf{C}_i \mathbf{s}_i + 2 \underbrace{\sum_{i=1}^h \sum_{j=1}^h \mathbf{s}_i^T \mathbf{C}_i^T \boldsymbol{\Sigma} \mathbf{C}_j \mathbf{s}_j}_{i > j} \right] \right) \\ &\quad \times \Phi_1 \left(\sum_{i=1}^h \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{1/2} \mathbf{C}_i \mathbf{s}_i \right). \end{aligned}$$

Por (i) temos $M_{\mathbf{C}^T \mathbf{Y}}(\mathbf{s}) = 2e^{\sum_{i=1}^h \mathbf{s}_i^T \mathbf{C}_i^T \boldsymbol{\mu} + \frac{1}{2} \sum_{i=1}^h \mathbf{s}_i^T \mathbf{C}_i^T \boldsymbol{\Sigma} \mathbf{C}_i \mathbf{s}_i} \Phi_1 \left(\sum_{i=1}^h \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{1/2} \mathbf{C}_i \mathbf{s}_i \right)$.

Note que se $\boldsymbol{\delta}^T \boldsymbol{\Sigma}^{1/2} \mathbf{C}_i = \mathbf{0}$, para todo i , $i = 1, \dots, h$, então

$$M_{\mathbf{C}^T \mathbf{Y}}(\mathbf{s}) = 2e^{\sum_{i=1}^h \mathbf{s}_i^T \mathbf{C}_i^T \boldsymbol{\mu} + \frac{1}{2} \sum_{i=1}^h \mathbf{s}_i^T \mathbf{C}_i^T \boldsymbol{\Sigma} \mathbf{C}_i \mathbf{s}_i} = \prod_{i=1}^h M_{\mathbf{C}_i^T \mathbf{Y}_i}(\mathbf{s}_i),$$

implicando que todas os blocos da combinação linear são normalmente distribuídos. Um

resultado similar é obtido quando $\delta^T \Sigma^{1/2} \mathbf{C}_k \neq 0$ para um, e somente um k , $k = 1, \dots, h$. $\square$

A seguir estudamos condições para determinar quando uma função linear de vetores aleatórios normal-assimétrico é independente de uma forma quadrática de vetores aleatórios normal-assimétrico. Também fornecemos condições sob as quais duas formas quadráticas são independentes.

Proposição 2.19. *Sejam $\mathbf{Y} \sim SN_n(\mathbf{0}, \Sigma, \lambda)$, $\mathbf{h} \in \mathbb{R}^n$ e $\mathbf{A} (n \times n)$ uma matriz simétrica. A forma linear $\mathbf{h}^T \mathbf{Y}$ é independente da forma quadrática $\mathbf{Y}^T \mathbf{A} \mathbf{Y}$ se, e somente se, $\mathbf{A} \Sigma \mathbf{h} = \mathbf{0}$ e $\mathbf{A} \Sigma^{1/2} \lambda = \mathbf{0}$.*

Demonstração. Note que o Corolário 2.3 e da Proposição 2.11, temos $\mu = \mathbf{0}$, que

$$M_{\mathbf{h}^T \mathbf{Y}}(s_1) = 2e^{\frac{1}{2} s_1^2 \mathbf{h}^T \Sigma \mathbf{h}} \Phi_1(\delta^T \Sigma^{1/2} \mathbf{h} s_1) \text{ e } M_{\mathbf{Y}^T \mathbf{A} \mathbf{Y}}(s_2) = |\mathbf{I}_n - 2s_2 \mathbf{A} \Sigma|, s_1, s_2 \in \mathbb{R}.$$

Considere $\mathbf{s} = (s_1, s_2)^T \in \mathbb{R}^2$, $\bar{\mathbf{h}} = s_1 \Sigma^{1/2} \mathbf{h}$, $\bar{\mathbf{A}} = s_2 \Sigma^{1/2} \mathbf{A} \Sigma^{1/2}$ e $\mathbf{Z} = \Sigma^{-1/2} \mathbf{Y}$. Claramente $\mathbf{Z} \sim SN_n(\lambda)$. A fgm conjunta de $\mathbf{h}^T \mathbf{Y}$ e $\mathbf{Y}^T \mathbf{A} \mathbf{Y}$ é

$$\begin{aligned} M_{\mathbf{h}^T \mathbf{Y}, \mathbf{Y}^T \mathbf{A} \mathbf{Y}}(\mathbf{s}) &= E[e^{s_1 \mathbf{h}^T \mathbf{Y} + s_2 \mathbf{Y}^T \mathbf{A} \mathbf{Y}}] = e^{-\frac{1}{4} \bar{\mathbf{h}}^T \bar{\mathbf{A}}^{-1} \bar{\mathbf{h}}} E[e^{(\mathbf{Z} + \frac{1}{2} \bar{\mathbf{A}}^{-1} \bar{\mathbf{h}})^T \bar{\mathbf{A}} (\mathbf{Z} + \frac{1}{2} \bar{\mathbf{A}}^{-1} \bar{\mathbf{h}})}] \\ &= e^{-\frac{1}{4} \bar{\mathbf{h}}^T \bar{\mathbf{A}}^{-1} \bar{\mathbf{h}}} M_{(\mathbf{Z} + \frac{1}{2} \bar{\mathbf{A}}^{-1} \bar{\mathbf{h}})^T \bar{\mathbf{A}} (\mathbf{Z} + \frac{1}{2} \bar{\mathbf{A}}^{-1} \bar{\mathbf{h}})}(1). \end{aligned}$$

Usando o resultado da Proposição 2.10, com $\mathbf{a} = -\frac{1}{2} \bar{\mathbf{A}}^{-1} \bar{\mathbf{h}}$ e $\mathbf{A} = \bar{\mathbf{A}}$, e após algumas manipulações algébricas, segue que

$$\begin{aligned} M_{\mathbf{h}^T \mathbf{Y}, \mathbf{Y}^T \mathbf{A} \mathbf{Y}}(\mathbf{s}) &= 2|\mathbf{I}_n - 2s_2 \mathbf{A} \Sigma|^{-1/2} \exp\left\{\frac{s_1^2}{2} \mathbf{h}^T (\Sigma^{-1} - 2s_2 \mathbf{A})^{-1} \mathbf{h}\right\} \\ &\quad \times \Phi_1\left(s_1 \frac{\lambda^T \Sigma^{-1/2} (\Sigma^{-1} - 2s_2 \mathbf{A})^{-1} \mathbf{h}}{(1 + \lambda^T \Sigma^{-1/2} (\Sigma^{-1} - 2s_2 \mathbf{A})^{-1} \Sigma^{-1/2} \lambda)^{1/2}}\right). \end{aligned} \quad (2.29)$$

Agora, note que

$$(\Sigma^{-1} - 2s_2 \mathbf{A})^{-1} = \Sigma \sum_{j=0}^{\infty} (2s_2)^j (\mathbf{A} \Sigma)^j, \quad (2.30)$$

Para $\mathbf{I}_n - 2s_2 \mathbf{A} \Sigma > 0$ (veja Horn & Johnson, 1996, pag. 301). Finalmente, de (2.29) e (2.30) o resultado segue. $\square$

Proposição 2.20. *Sejam $\mathbf{Y} \sim SN_n(\mathbf{0}, \Sigma, \lambda)$ e $\mathbf{A} (n \times n)$, $\mathbf{B} (n \times n)$ matrizes simétricas. As formas quadráticas $\mathbf{Y}^T \mathbf{A} \mathbf{Y}$ e $\mathbf{Y}^T \mathbf{B} \mathbf{Y}$ são independentes se, e somente se, $\mathbf{A} \Sigma \mathbf{B} = \mathbf{0}$.*

Demonstração. Novamente considere $\mathbf{s} = (s_1, s_2)^T \in \mathbb{R}^2$, $\bar{\mathbf{A}} = \Sigma^{1/2} \mathbf{A} \Sigma^{1/2}$, $\bar{\mathbf{B}} = \Sigma^{1/2} \mathbf{B} \Sigma^{1/2}$ e $\mathbf{Z} = \Sigma^{-1/2} \mathbf{Y}$. A fgm conjunta de $\mathbf{Y}^T \mathbf{A} \mathbf{Y}$ e $\mathbf{Y}^T \mathbf{B} \mathbf{Y}$ é

$$M_{\mathbf{Y}^T \mathbf{A} \mathbf{Y}, \mathbf{Y}^T \mathbf{B} \mathbf{Y}}(\mathbf{s}) = E[e^{s_1 \mathbf{Z}^T \bar{\mathbf{A}} \mathbf{Z} + s_2 \mathbf{Z}^T \bar{\mathbf{B}} \mathbf{Z}}] = E[e^{\mathbf{Z}^T (s_1 \bar{\mathbf{A}} + s_2 \bar{\mathbf{B}}) \mathbf{Z}}] = M_{\mathbf{Z}^T (s_1 \bar{\mathbf{A}} + s_2 \bar{\mathbf{B}}) \mathbf{Z}}(1).$$

Do resultado (a) no Corolário 2.4 segue que

$$M_{\mathbf{Y}^T \mathbf{A} \mathbf{Y}, \mathbf{Y}^T \mathbf{B} \mathbf{Y}}(\mathbf{s}) = |\mathbf{I}_n - 2(s_1 \bar{\mathbf{A}} + s_2 \bar{\mathbf{B}})|^{-1/2}, \quad s_1, s_2 \in \mathbb{R}.$$

Como $M_{\mathbf{Y}^T \mathbf{A} \mathbf{Y}}(s_1) = |\mathbf{I}_n - 2s_1 \bar{\mathbf{A}}|^{-1/2}$ e $M_{\mathbf{Y}^T \mathbf{B} \mathbf{Y}}(s_2) = |\mathbf{I}_n - 2s_2 \bar{\mathbf{B}}|^{-1/2}$, para haver independência devemos ter que, $\bar{\mathbf{A}} \bar{\mathbf{B}} = \mathbf{0}$, implicando que $\mathbf{A} \Sigma \mathbf{B} = \mathbf{0}$. A prova da condição necessária segue o mesmo argumento. $\square$

Corolário 2.8. *Sob as condições da Proposição 2.20 as formas quadráticas $\mathbf{Y}^T \mathbf{A}_i \mathbf{Y}$ $i = 1, \dots, m$, são mutuamente independentes se, e somente se, $\mathbf{A}_k \Sigma \mathbf{A}_s = \mathbf{0}$, $k \neq s$.*

2.3.4 Distribuições Condicionais

Usando a notação da seção anterior mostraremos no resultado seguinte que a distribuição normal-assimétrico não é fechada sob condicionamento. Contudo, suas distribuições condicionais têm a forma de uma família mais geral definida em Arellano-Valle et al. (2002).

Proposição 2.21. *Seja $\mathbf{Z} = (\mathbf{Z}_i, i = 1, 2) \sim SN_n(\boldsymbol{\lambda})$. Então a densidade condicional de $\mathbf{Z}_1$ dado $\mathbf{Z}_2 = \mathbf{z}_2$ é dada por*

$$f_{\mathbf{Z}_1 | \mathbf{Z}_2 = \mathbf{z}_2}(\mathbf{z}_1) = \frac{\phi_{n_1}(\mathbf{z}_1) \Phi_1(\delta_1^T \mathbf{z}_1 | -\delta_2^T \mathbf{z}_2, 1 - \delta^T \delta)}{\Phi_1(\delta_2^T \mathbf{z}_2 | 0, 1 - \delta_2^T \delta_2)}. \quad (2.31)$$

Demonstração. Considerando que $\Phi_1(\boldsymbol{\lambda}^T \mathbf{Z}) = \Phi_1(\delta^T \mathbf{Z} | 0, 1 - \delta^T \delta) = \Phi_1(\delta_1^T \mathbf{z}_1 + \delta_2^T \mathbf{z}_2 | 0, 1 - \delta^T \delta) = \Phi_1(\delta_1^T \mathbf{z}_1 | -\delta_2^T \mathbf{z}_2, 1 - \delta^T \delta)$ temos

$$\begin{aligned} f_{\mathbf{Z}_1 | \mathbf{Z}_2 = \mathbf{z}_2}(\mathbf{z}_1) &= \frac{f_{\mathbf{Z}}(\mathbf{z}_1, \mathbf{z}_2)}{f_{\mathbf{Z}_2}(\mathbf{z}_2)} = \frac{2\phi_{n_1}(\mathbf{z}_1)\phi_{n_2}(\mathbf{z}_2)\Phi_1(\boldsymbol{\lambda}^T \mathbf{Z})}{2\phi_{n_2}(\mathbf{z}_2)\Phi_1(\boldsymbol{\lambda}_2^* \mathbf{z}_2)} \\ &= \frac{\phi_{n_1}(\mathbf{z}_1)\Phi_1(\delta_1^T \mathbf{z}_1 | -\delta_2^T \mathbf{z}_2, 1 - \delta^T \delta)}{\Phi_1(\boldsymbol{\lambda}_2^* \mathbf{z}_2)}, \end{aligned}$$

onde $\boldsymbol{\lambda}_2^* = \delta_2 / (1 - \delta_2^T \delta_2)^{1/2}$. $\square$

Os lemas que apresentamos a seguir são úteis para encontrar os momentos condicionais de $\mathbf{Z}_1$ dado $\mathbf{Z}_2 = \mathbf{z}_2$. Note que se $\mathbf{X}$ ($m \times 1$) é um vetor aleatório com densidade $f(\mathbf{x})$, então

$$f_*(\mathbf{x}|\mathbf{x}_o) = \frac{f(\mathbf{x})}{P(\mathbf{X} \leq \mathbf{x}_o)} \mathbb{I}_{\{\mathbf{x} \leq \mathbf{x}_o\}}, \quad (2.32)$$

é a densidade condicional de $\mathbf{X}$ dado $\{\mathbf{X} \leq \mathbf{x}_o\}$, para um vetor fixo $\mathbf{x}_o \in \mathbb{R}^m$, onde $\mathbb{I}_{\{\mathbf{A}\}}$ é o indicador de $\mathbf{A}$.

Lema 2.3. *Seja $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Então, para um vetor fixo $\mathbf{a}$ de dimensão k , uma matriz fixa $\mathbf{B}$ de dimensão $k \times n$, e uma função real integrável g , temos*

$$E(g(\mathbf{Y})\Phi_k(\mathbf{a} + \mathbf{B}\mathbf{Y}|\boldsymbol{\eta}, \boldsymbol{\Omega})) = \Phi_k(\mathbf{a} - \boldsymbol{\eta} | -\mathbf{B}\boldsymbol{\mu}, \boldsymbol{\Omega} + \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T) E[h(\mathbf{X})|\mathbf{X} \leq \mathbf{a} - \boldsymbol{\eta}], \quad (2.33)$$

onde $h(\mathbf{X}) = E(g(\mathbf{Y})|\mathbf{X} = \mathbf{x})$, com $\mathbf{X} \sim N_k(-\mathbf{B}\boldsymbol{\mu}, \boldsymbol{\Omega} + \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T)$ e $\mathbf{Y}|\mathbf{X} = \mathbf{x} \sim N_n(\boldsymbol{\mu} - \boldsymbol{\Sigma}\mathbf{B}^T(\boldsymbol{\Omega} + \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T)^{-1}(\mathbf{x} + \mathbf{B}\boldsymbol{\mu}), (\boldsymbol{\Sigma}^{-1} + \mathbf{B}^T\boldsymbol{\Omega}^{-1}\mathbf{B})^{-1})$.

Demonstração. Sendo g uma função real integrável, temos por definição de esperança que $E(g(\mathbf{Y})\Phi_k(\mathbf{a} + \mathbf{B}\mathbf{Y}|\boldsymbol{\eta}, \boldsymbol{\Omega}))$, é igual a

$$\begin{aligned} & \int_{\mathbb{R}^n} g(\mathbf{y})\Phi_k(\mathbf{a} + \mathbf{B}\mathbf{y}|\boldsymbol{\eta}, \boldsymbol{\Omega})\phi_n(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})d\mathbf{y} \\ &= \int_{\mathbb{R}^n} g(\mathbf{y})\Phi_k(\mathbf{a} - \boldsymbol{\eta} | -\mathbf{B}\mathbf{y}, \boldsymbol{\Omega})\phi_n(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})d\mathbf{y} \\ &= \int_{\mathbb{R}^n} \left[g(\mathbf{y})\phi_n(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \int_{\mathbf{x} \leq \mathbf{a} - \boldsymbol{\eta}} \phi_k(\mathbf{x} | -\mathbf{B}\mathbf{y}, \boldsymbol{\Omega})d\mathbf{x} \right] d\mathbf{y} \\ &= \int_{\mathbb{R}^n} \int_{\mathbf{x} \leq \mathbf{a} - \boldsymbol{\eta}} g(\mathbf{y})\phi_n(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\phi_k(\mathbf{x} | -\mathbf{B}\mathbf{y}, \boldsymbol{\Omega})d\mathbf{x}d\mathbf{y} \\ &= \int_{\mathbf{x} \leq \mathbf{a} - \boldsymbol{\eta}} \int_{\mathbb{R}^n} g(\mathbf{y})\phi_n(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\phi_k(\mathbf{x} | -\mathbf{B}\mathbf{y}, \boldsymbol{\Omega})d\mathbf{y}d\mathbf{x} \\ &= \int_{\mathbf{x} \leq \mathbf{a} - \boldsymbol{\eta}} \int_{\mathbb{R}^n} g(\mathbf{y})f_{\mathbf{Y}}(\mathbf{y})f_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x})d\mathbf{y}d\mathbf{x}, \end{aligned}$$

onde $f_{\mathbf{Y}}(\mathbf{y}) = \phi_n(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e $f_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x}) = \phi_k(\mathbf{x} | -\mathbf{B}\mathbf{y}, \boldsymbol{\Omega})$, i.e., $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e $[\mathbf{X}|\mathbf{Y} = \mathbf{y}] \sim N_k(-\mathbf{B}\mathbf{y}, \boldsymbol{\Omega})$. Usando que $f_{\mathbf{Y}}(\mathbf{y})f_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x})f_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y})$, onde $\mathbf{X} \sim N_k(-\mathbf{B}\boldsymbol{\mu}, \boldsymbol{\Omega} + \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T)$ e $P(\mathbf{X} \leq \mathbf{a} - \boldsymbol{\eta}) = \Phi_k(\mathbf{a} - \boldsymbol{\eta} | -\mathbf{B}\boldsymbol{\mu}, \boldsymbol{\Omega} + \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T)$, e que $[\mathbf{Y}|\mathbf{X} = \mathbf{x}] \sim N_n(\boldsymbol{\mu} - \boldsymbol{\Sigma}\mathbf{B}^T(\boldsymbol{\Omega} + \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T)^{-1}(\mathbf{x} + \mathbf{B}\boldsymbol{\mu}), (\boldsymbol{\Sigma}^{-1} + \mathbf{B}^T\boldsymbol{\Omega}^{-1}\mathbf{B})^{-1})$, temos que a expressão acima é

equivalente a

$$\begin{aligned} & \Phi_k(\mathbf{a} - \boldsymbol{\eta} | -\mathbf{B}\boldsymbol{\mu}, \boldsymbol{\Omega} + \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T) \int_{\mathbb{R}^k} \frac{f_{\mathbf{X}}(\mathbf{x})}{P(\mathbf{X} \leq \mathbf{a} - \boldsymbol{\eta})} \mathbb{I}_{\{\mathbf{x} \leq \mathbf{a} - \boldsymbol{\eta}\}} \int_{\mathbb{R}^n} g(\mathbf{y}) f_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y}) d\mathbf{y} d\mathbf{x}, \\ &= \Phi_k(\mathbf{a} - \boldsymbol{\eta} | -\mathbf{B}\boldsymbol{\mu}, \boldsymbol{\Omega} + \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T) \int_{\mathbb{R}^k} f_*(\mathbf{x}|\mathbf{a} - \boldsymbol{\eta}) E(g(\mathbf{Y})|\mathbf{X} = \mathbf{x}) d\mathbf{x}, \end{aligned}$$

onde, como definido em (2.32), $f_*(\mathbf{x}|\mathbf{a} - \boldsymbol{\eta})$ é a densidade condicional de $\mathbf{X}$ dado $\{\mathbf{X} \leq \mathbf{a} - \boldsymbol{\eta}\}$, concluindo a prova. $\square$

Lema 2.4. *Seja $X \sim N(\eta, \tau^2)$. Então, para $a \in \mathbb{R}$ constante, temos*

$$\begin{aligned} E(X|X \leq a) &= \eta - \frac{\phi_1(\frac{a-\eta}{\tau})}{\Phi_1(\frac{a-\eta}{\tau})}, \\ \text{Var}(X|X \leq a) &= \left[1 - \frac{\phi_1(\frac{a-\eta}{\tau})}{\Phi_1(\frac{a-\eta}{\tau})} \left(\frac{a-\eta}{\tau} + \frac{\phi_1(\frac{a-\eta}{\tau})}{\Phi_1(\frac{a-\eta}{\tau})} \right) \right] \tau^2. \end{aligned}$$

Demonstração. Veja Johnson et al. (1994), Seção 10.1. $\square$

Proposição 2.22. *Seja $\mathbf{Z} \sim (\mathbf{Z}_i, i = 1, 2) \sim SN_n(\boldsymbol{\lambda})$. Então,*

$$E[\mathbf{Z}_1|\mathbf{Z}_2 = \mathbf{z}_2] = -\boldsymbol{\delta}_1(1 - \boldsymbol{\delta}_2^T \boldsymbol{\delta}_2)^{-1} \frac{\phi_1(\boldsymbol{\lambda}_2^{*T} \mathbf{z}_2)}{\Phi_1(\boldsymbol{\lambda}_2^{*T} \mathbf{z}_2)}, \quad (2.34)$$

e

$$\text{Var}[\mathbf{Z}_1|\mathbf{Z}_2 = \mathbf{z}_2] = (\mathbf{I}_{n_1} + \boldsymbol{\delta}_1(1 - \boldsymbol{\delta}_2^T \boldsymbol{\delta}_2)^{-1} \boldsymbol{\delta}_1^T)^{-1} + (1 - \boldsymbol{\delta}_2^T \boldsymbol{\delta}_2)^{-2} \boldsymbol{\delta}_1 \boldsymbol{\delta}_1^T \text{Var}(S|S \leq \boldsymbol{\delta}_2^T \mathbf{z}_2), \quad (2.35)$$

onde

$$\text{Var}(S|S \leq \boldsymbol{\delta}_2^T \mathbf{z}_2) = \left[1 - \frac{\phi_1(\boldsymbol{\lambda}_2^{*T} \mathbf{z}_2)}{\Phi_1(\boldsymbol{\lambda}_2^{*T} \mathbf{z}_2)} \left(\boldsymbol{\lambda}_2^{*T} \mathbf{z}_2 + \frac{\phi_1(\boldsymbol{\lambda}_2^{*T} \mathbf{z}_2)}{\Phi_1(\boldsymbol{\lambda}_2^{*T} \mathbf{z}_2)} \right) \right] (1 - \boldsymbol{\delta}_2^T \boldsymbol{\delta}_2),$$

com $\boldsymbol{\lambda}_2^* = \boldsymbol{\delta}_2 / (1 - \boldsymbol{\delta}_2^T \boldsymbol{\delta}_2)^{1/2}$.

Demonstração. A prova segue usando os lemas 2.3 e 2.4, com $g(\mathbf{y}) = \mathbf{a}^T \mathbf{y}$ e $g(\mathbf{y}) = (\mathbf{a}^T \mathbf{y})^2$, para $\mathbf{a} \in \mathbb{R}^n$. $\square$

Com o intuito de estudar a distribuição da soma e a distribuição conjunta de vetores aleatórios normal-assimétrico, na seção seguinte estudamos uma nova classe de distribuições normal-assimétrica multivariada proposta por Arellano-Valle & Genton (2003).

2.3.5 Distribuição Normal-Assimétrica Fundamental

Seja $\mathbf{Z}^*$ um vetor aleatório $n \times 1$ e seja $\mathbf{D}$ uma matriz de correlações $n \times m$ (que é simétrica quando $m = n$) tal que $\mathbf{I}_m - \mathbf{D}^T \mathbf{D}$ e $\mathbf{I}_n - \mathbf{D} \mathbf{D}^T$ são matrizes definidas positivas, i.e., $\|\mathbf{D}\mathbf{a}\| \leq 1$ e $\|\mathbf{D}^T \mathbf{b}\| \leq 1$ para pares de vetores unitários $\mathbf{a} \in \mathbb{R}^m$ e $\mathbf{b} \in \mathbb{R}^n$, onde $\|\cdot\|$ denota a norma de um vetor.

Definição 2.2. Um vetor aleatório n -dimensional $\mathbf{Z}^*$ segue uma distribuição normal-assimétrica fundamental (SNF) com matriz de assimetria $\mathbf{D} \in \mathbb{R}^{n \times m}$ se sua fdp é dada por

$$f_{\mathbf{Z}^*}(\mathbf{z}) = 2^m \phi_n(\mathbf{z}) \Phi_m(\mathbf{D}^T \mathbf{z} | \mathbf{0}, \mathbf{I}_m - \mathbf{D}^T \mathbf{D}), \quad \mathbf{z} \in \mathbb{R}^n, \quad (2.36)$$

onde $\mathbf{D}$ é tal que $\|\mathbf{D}\mathbf{a}\| < 1$, para todo vetor unitário $\mathbf{a} \in \mathbb{R}^m$. Usamos a notação $\mathbf{Z}^* \sim SNF_{n,m}(\mathbf{D})$ e $\mathbf{Z}^* \sim SNF_n(\mathbf{D})$ quando $m = n$.

Observação 2.5. Se a matriz $\mathbf{D} = \boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{(1 + \boldsymbol{\lambda}^T \boldsymbol{\lambda})^{1/2}} \in \mathbb{R}^n$, então (2.36) se reduz a (2.7), isto é, $\mathbf{Z}^* \sim SN_n(\boldsymbol{\lambda})$.

A seguir apresentamos algumas propriedades desta distribuição. Detalhes e provas dos resultados podem ser encontrados em Arellano-Valle & Genton (2003).

Proposição 2.23. (Representação estocástica) Seja $\mathbf{Z}^* \sim SNF_{n,m}(\mathbf{D})$. Então

$$\mathbf{Z}^* \stackrel{d}{=} \mathbf{D}|\mathbf{X}| + (\mathbf{I}_n - \mathbf{D} \mathbf{D}^T)^{1/2} \mathbf{X}_1, \quad (2.37)$$

onde $|\mathbf{X}| = (|X_1|, \dots, |X_m|)^T \sim HN_m(\mathbf{0}, \mathbf{I}_m)$ e $\mathbf{X}_1 \sim N_n(\mathbf{0}, \mathbf{I}_n)$, $\mathbf{X}$ e $\mathbf{X}_1$ independentes.

Da proposição acima, como $E[|\mathbf{X}|] = \sqrt{\frac{2}{\pi}} \mathbf{1}_m$ e $Var(|\mathbf{X}|) = (1 - 2/\pi) \mathbf{I}_m$, segue que

$$E(\mathbf{Z}^*) = \sqrt{\frac{2}{\pi}} \mathbf{D} \mathbf{1}_m \quad \text{e} \quad Var(\mathbf{Z}^*) = \mathbf{I}_n - \frac{2}{\pi} \mathbf{D} \mathbf{D}^T. \quad (2.38)$$

Proposição 2.24. (fgm) Se $\mathbf{Z}^* \sim SNF_{n,m}(\mathbf{D})$, então sua fgm é dada por

$$M_{\mathbf{Z}^*}(\mathbf{s}) = 2^m e^{\frac{1}{2} \mathbf{s}^T \mathbf{s}} \Phi_m(\mathbf{D}^T \mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^n. \quad (2.39)$$

De maneira similar ao caso da distribuição normal-assimétrica multivariada, podemos estender a densidade em (2.36) para o modelo de locação-escala correspondente, denotada

por $SNF_{n,m}(\mu, \Omega, D)$, considerando $Y^* = \mu + \Omega^{1/2}Z^*$, onde $\mu \in \mathbb{R}^n$ e Ω uma matriz definida positiva de dimensão $n \times n$. A densidade e a fgm correspondente, são dadas na seguinte proposição.

Proposição 2.25. (Extensão para locação-escala) *Se $Y^* \sim SNF_{n,m}(\mu, \Omega, D)$. Então*

$$f_{Y^*}(y) = 2^m \phi_n(y|\mu, \Omega) \Phi_m(D^T \Omega^{-1/2}(y - \mu)|0, I_m - D^T D), \quad y \in \mathbb{R}^n, \quad (2.40)$$

e

$$M_{Y^*}(s) = 2^m e^{s^T \mu + \frac{1}{2} s^T \Omega s} \Phi_m(D^T \Omega^{1/2} s), \quad s \in \mathbb{R}^n. \quad (2.41)$$

De (2.38) segue que

$$E(Y^*) = \mu + \sqrt{\frac{2}{\pi}} \Omega^{1/2} D \mathbf{1}_m \quad e \quad Var(Y^*) = \Omega - \frac{2}{\pi} \Omega^{1/2} D D^T \Omega^{1/2} \quad (2.42)$$

No contexto de distribuição SNF podemos agora apresentar alguns resultados relacionados com a soma e a distribuição conjunta de vetores aleatórios independentes normal-assimétrico. Estes resultados são dados a seguir:

Proposição 2.26. *Sejam $Y_1, \dots, Y_p$ vetores aleatórios independentes com $Y_i \sim SN_n(\mu_i, \Sigma_i, \lambda_i)$, $i = 1, \dots, p$. Então,*

$$\frac{1}{p} \sum_{i=1}^p Y_i \sim SNF_{n,p}(\mu, \Sigma, D), \quad (2.43)$$

onde $\mu = \frac{1}{n} \sum_{i=1}^p \mu_i$, $\Sigma = \frac{1}{p^2} \sum_{i=1}^p \Sigma_i$ e $D = (\sum_{i=1}^p \Sigma_i)^{-1/2} (\Sigma_1^{1/2} \delta_1, \dots, \Sigma_p^{1/2} \delta_p)$, com $\delta_i = \lambda_i / (1 + \lambda_i^T \lambda_i)^{1/2}$, $i = 1, \dots, p$.

Demonstração. De (2.16) temos para $s \in \mathbb{R}^n$ que

$$\begin{aligned} M_{\frac{1}{p} \sum_{i=1}^p Y_i}(s) &= \prod_{i=1}^p M_{Y_i}\left(\frac{1}{p}s\right) = 2^p e^{s^T \left(\frac{1}{p} \sum_{i=1}^p \mu_i + \frac{1}{2p^2} s^T \left(\sum_{i=1}^p \Sigma_i\right) s\right)} \prod_{i=1}^p \Phi_1\left(\frac{1}{p} \delta_i^T \Sigma_i^{1/2} s\right) \\ &= 2^p e^{s^T \left(\frac{1}{p} \sum_{i=1}^p \mu_i + \frac{1}{2p^2} s^T \left(\sum_{i=1}^p \Sigma_i\right) s\right)} \Phi_p\left(\left(\Sigma_1^{1/2} \delta_1, \dots, \Sigma_p^{1/2} \delta_p\right)^T \left(\sum_{i=1}^p \Sigma_i\right)^{-1/2} \left(\frac{\sum_{i=1}^p \Sigma_i}{p^2}\right)^{1/2} s\right). \end{aligned}$$

Logo, a prova é verificada usando a Proposição 2.25. $\square$

Para o caso i.i.d. temos o seguinte corolário para a soma de vetores aleatórios normal-assimétrico.

Corolário 2.9. Se $\mathbf{Y}_1, \dots, \mathbf{Y}_p$ são vetores aleatórios independentes e identicamente distribuídos com $\mathbf{Y}_i \sim SN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, $i = 1, \dots, p$. Então

$$\frac{1}{p} \sum_{i=1}^p \mathbf{Y}_i \sim SNF_{n,p}(\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{p}, \mathbf{D}), \quad (2.44)$$

onde $\mathbf{D} = \frac{1}{\sqrt{p}} \mathbf{1}_p^T \otimes \boldsymbol{\delta}$, com $\boldsymbol{\delta} = \boldsymbol{\lambda}/(1 + \boldsymbol{\lambda}^T \boldsymbol{\lambda})^{1/2}$ e $\mathbf{1}_p$ é o vetor p -dimensional com elementos iguais a 1.

A seguir, usaremos a notação $\oplus$ para indicar o operador soma direta, que dá como resultado uma matriz bloco diagonal, ou seja, $\oplus_{i=1}^p \mathbf{A}_i = \text{diag}(\mathbf{A}_1, \dots, \mathbf{A}_p)$. É fácil ver que $\oplus_{i=1}^p \mathbf{A} = \mathbf{I}_m \otimes \mathbf{A}$, onde $\mathbf{I}_m$ é a matriz identidade de dimensão $m \times m$.

Proposição 2.27. Se $\mathbf{Y}_1, \dots, \mathbf{Y}_p$ são vetores aleatórios independentes com $\mathbf{Y}_i \sim SN_{n_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \boldsymbol{\lambda}_i)$, $i = 1, \dots, p$. Então a distribuição conjunta de $\mathbf{Y}_1, \dots, \mathbf{Y}_p$ é

$$\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_p^T)^T \sim SNF_{r,p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}), \quad (2.45)$$

onde $r = \sum_{i=1}^p n_i$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^T, \dots, \boldsymbol{\mu}_p^T)^T$, $\boldsymbol{\Sigma} = \oplus_{i=1}^p \boldsymbol{\Sigma}_i$, $\mathbf{D} = \oplus_{i=1}^p \boldsymbol{\delta}_i$, com $\boldsymbol{\delta}_i = \boldsymbol{\lambda}_i/(1 + \boldsymbol{\lambda}_i^T \boldsymbol{\lambda}_i)^{1/2}$, $i = 1, \dots, p$.

Demonstração. Para $\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_p^T)^T$, a densidade de $\mathbf{y}$ é dada por

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= \prod_{i=1}^p 2\phi_{n_i}(\mathbf{y}_i | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \Phi_1(\boldsymbol{\lambda}_i^T \boldsymbol{\Sigma}_i^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}_i)) \\ &= 2^p \phi_r(\mathbf{y} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \Phi_p([\oplus_{i=1}^p \boldsymbol{\lambda}_i^T][\oplus_{i=1}^p \boldsymbol{\Sigma}_i^{-1/2}](\mathbf{y} - \boldsymbol{\mu})). \end{aligned}$$

Note que $\oplus_{i=1}^p \boldsymbol{\lambda}_i^T = (\oplus_{i=1}^p \frac{1}{(1 - \boldsymbol{\delta}_i^T \boldsymbol{\delta}_i)^{1/2}})(\oplus_{i=1}^p \boldsymbol{\delta}_i^T)$ e a matriz diagonal $\oplus_{i=1}^p \frac{1}{(1 - \boldsymbol{\delta}_i^T \boldsymbol{\delta}_i)^{1/2}} = \mathbf{I}_p - \oplus_{i=1}^p \boldsymbol{\delta}_i^T \oplus_{i=1}^p \boldsymbol{\delta}_i$. Logo,

$$f_{\mathbf{Y}}(\mathbf{y}) = 2^p \phi_r(\mathbf{y} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \Phi_p(\oplus_{i=1}^p \boldsymbol{\delta}_i^T \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}) | \mathbf{0}, \mathbf{I}_p - \oplus_{i=1}^p \boldsymbol{\delta}_i^T \oplus_{i=1}^p \boldsymbol{\delta}_i),$$

concluindo desta forma a prova. □

Uma vez que $\oplus_{i=1}^p \mathbf{M} = \mathbf{I}_p \otimes \mathbf{M}$, no caso i.i.d. podemos simplificar a distribuição conjunta apresentada em (2.45), como segue.

Corolário 2.10. Sejam $\mathbf{Y}_1, \dots, \mathbf{Y}_p$ vetores aleatórios independentes e identicamente distribuídos com $\mathbf{Y}_i \sim SN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, $i = 1, \dots, p$. Então, a distribuição conjunta de

$\mathbf{Y}_1, \dots, \mathbf{Y}_p$ é

$$\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_p^T)^T \sim SNF_{r,p}(\boldsymbol{\mu}_f, \boldsymbol{\Sigma}_f, \mathbf{D}_f), \quad (2.46)$$

onde $r = np$, $\boldsymbol{\mu}_f = \mathbf{1}_p \otimes \boldsymbol{\mu}$, $\boldsymbol{\Sigma}_f = \mathbf{I}_p \otimes \boldsymbol{\Sigma}$, $\mathbf{D}_f = \mathbf{I}_p \otimes \boldsymbol{\delta}$, com $\boldsymbol{\delta} = \boldsymbol{\lambda}/(1 + \boldsymbol{\lambda}^T \boldsymbol{\lambda})^{1/2}$

Um outro caso especial de (2.40), que pode ser considerado como uma extensão da definição encontrada em Sahu et al. (2003), ocorre quando $\boldsymbol{\Omega} = \boldsymbol{\Sigma} + \boldsymbol{\Delta}\boldsymbol{\Delta}^T$ e $\mathbf{D} = (\boldsymbol{\Sigma} + \boldsymbol{\Delta}\boldsymbol{\Delta}^T)^{-1/2}\boldsymbol{\Delta}$, onde $\boldsymbol{\Sigma}$ ($n \times n$) é uma matriz simétrica definida positiva e $\boldsymbol{\Delta} \in \mathbb{R}^{n \times k}$ é tal que $\|\boldsymbol{\Delta}\mathbf{a}\| \leq 1$, para um vetor unitário $\mathbf{a} \in \mathbb{R}^k$. Assim, de (2.40) após uma simples álgebra, obtemos a seguinte densidade:

$$f_{\mathbf{Y}}(\mathbf{y}) = 2^k \phi_n(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{\Delta}\boldsymbol{\Delta}^T) \Phi_k(\boldsymbol{\Delta}^T(\boldsymbol{\Sigma} + \boldsymbol{\Delta}\boldsymbol{\Delta}^T)^{-1}(\mathbf{y} - \boldsymbol{\mu})|\mathbf{0}, \mathbf{I}_k - \boldsymbol{\Delta}^T(\boldsymbol{\Sigma} + \boldsymbol{\Delta}\boldsymbol{\Delta}^T)^{-1}\boldsymbol{\Delta}). \quad (2.47)$$

Para formalizar a densidade acima, da qual um caso especial será usada no Capítulo 4, apresentamos a seguinte definição.

Definição 2.3. Um vetor aleatório n -dimensional $\mathbf{Y}$ segue uma distribuição normal-assimétrica (SN^*) com vetor de locação $\boldsymbol{\mu} \in \mathbb{R}^n$, matriz de dispersão $\boldsymbol{\Sigma}$ (uma matriz $n \times n$ definida positiva) e matriz de assimetria $\boldsymbol{\Delta} \in \mathbb{R}^{n \times k}$, se sua fdp é dada por

$$\begin{aligned} f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}) &= 2^k \phi_n(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{\Delta}\boldsymbol{\Delta}^T) \\ &\times \Phi_k(\boldsymbol{\Delta}^T(\boldsymbol{\Sigma} + \boldsymbol{\Delta}\boldsymbol{\Delta}^T)^{-1}(\mathbf{y} - \boldsymbol{\mu})|\mathbf{0}, (\mathbf{I}_k + \boldsymbol{\Delta}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta})^{-1}). \end{aligned} \quad (2.48)$$

Usamos a notação $\mathbf{Y} \sim SN_{n,k}^*(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$ e $\mathbf{Y} \sim SN_n^*(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$ quando $n = k$.

Observação 2.6. Note que para $\boldsymbol{\Delta} = \mathbf{0}$, (2.48) se reduz à usual distribuição $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Mais ainda, como $(\mathbf{I}_k + \boldsymbol{\Delta}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta})^{-1} = \mathbf{I}_k - \boldsymbol{\Delta}^T(\boldsymbol{\Sigma} + \boldsymbol{\Delta}\boldsymbol{\Delta}^T)^{-1}\boldsymbol{\Delta}$ temos, para $k = n$, $\boldsymbol{\Delta} = \text{diag}(\delta_1, \dots, \delta_n)$ e $\boldsymbol{\Sigma}$ diagonal, que (2.48) se reduz ao modelo normal-assimétrico considerado em Sahu et al. (2003).

2.4 Aplicação a Modelos de Regressão Multivariados

Suponha que temos m indivíduos sendo cada um observado em n ocasiões (dados balanceados). A resposta observada para o j -ésimo indivíduo será formada pelo vetor ($n \times 1$) $\mathbf{Y}_j = (Y_{j1}, \dots, Y_{jn})^T$. Além disso, assumiremos o modelo linear

$$\mathbf{Y}_j = \mathbf{X}_j\boldsymbol{\beta} + \boldsymbol{\epsilon}_j, \quad j = 1, \dots, m, \quad (2.49)$$

em que $\mathbf{X}_j$ é a matriz de planejamento conhecida de dimensão $(n \times p)$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ e os $\boldsymbol{\epsilon}_j$ são vetores n -dimensionais de erros independentes. Cada $\boldsymbol{\epsilon}_j$ supostamente segue uma distribuição $SN_n(\mathbf{0}, \boldsymbol{\Psi}, \boldsymbol{\lambda})$, cuja função densidade é da forma

$$f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_j) = 2\phi_n(\boldsymbol{\epsilon}_j|\mathbf{0}, \boldsymbol{\Psi})\Phi_1(\boldsymbol{\lambda}^T \boldsymbol{\Psi}^{-1/2} \boldsymbol{\epsilon}_j), \quad j = 1, \dots, m. \quad (2.50)$$

Podemos parametrizar $\boldsymbol{\Psi}$ assumindo que $\boldsymbol{\Psi} = \sigma^2 \mathbf{R}$, em que $\sigma^2 > 0$ é um parâmetro de escala e $\mathbf{R} = \mathbf{R}(\boldsymbol{\rho})$ é uma matriz de correlação parametrizada em um conjunto de parâmetros $\boldsymbol{\rho} = (\rho_1, \dots, \rho_q)^T$, que não depende de $\boldsymbol{\beta}$. O modelo definido em (2.49)-(2.50) será chamado de *modelo linear normal-assimétrico multivariado* e têm a característica que a média dos erros aleatórios é diferente de zero. Algumas alternativas para contornar este problema serão dadas no capítulo 3. O interesse é fazer inferências sobre o vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2, \boldsymbol{\rho}^T, \boldsymbol{\lambda}^T)^T$.

O logaritmo da função verossimilhança para o modelo (2.49)-(2.50), denotado por $\ell(\boldsymbol{\theta})$, pode ser escrito como

$$\begin{aligned} \ell(\boldsymbol{\theta}) = & -\frac{m}{2} \log \sigma^2 - \frac{m}{2} \log |\mathbf{R}| - \frac{1}{2\sigma^2} \sum_{j=1}^m \{(\mathbf{Y}_j - \mathbf{X}_j \boldsymbol{\beta})^T \mathbf{R}^{-1} (\mathbf{Y}_j - \mathbf{X}_j \boldsymbol{\beta})\} \\ & + \sum_{j=1}^m \Phi_1\left(\frac{1}{\sigma} \boldsymbol{\lambda}^T \mathbf{R}^{-1/2} (\mathbf{Y}_j - \mathbf{X}_j \boldsymbol{\beta})\right). \end{aligned} \quad (2.51)$$

Note que não existem soluções explícitas para o problema de maximização da função de verossimilhança acima. Esta pode ser maximizada numericamente usando programas como, por exemplo, Matlab, R, Ox e S-plus. Estes programas contém rotinas prontas para tratar problemas de maximização (ou minimização) de qualquer função. No entanto, a menos que valores iniciais adequados sejam dados, não há garantia de se obter um máximo global, sendo necessários a implementação de métodos mais confiáveis como, por exemplo, o algoritmo EM. A seguir, apresentamos os passos necessários para a implementação do algoritmo EM para encontrar os EMV dos parâmetros do modelo definido em (2.49)-(2.50). Antes relatamos alguns resultados adicionais, necessários para esta implementação.

Proposição 2.28. *Suponha que $\mathbf{Y}|T = t \sim N_n(\boldsymbol{\mu} + t\mathbf{d}, \boldsymbol{\Psi})$ e $T \sim HN_1(0, 1)$. Seja $\boldsymbol{\Sigma} = \boldsymbol{\Psi} + \mathbf{d}\mathbf{d}^T$. Então, a distribuição conjunta de $(\mathbf{Y}^T, T)^T$ pode ser escrita como*

$$f_{\mathbf{Y}, T}(\mathbf{y}, t) = 2\phi_n(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\phi_1(t|\eta, \tau^2)\mathbb{I}\{t > 0\}, \quad (2.52)$$

onde

$$\eta = \frac{\mathbf{d}^T \Psi^{-1}(\mathbf{y} - \boldsymbol{\mu})}{1 + \mathbf{d}^T \Psi^{-1} \mathbf{d}} \text{ e } \tau^2 = \frac{1}{1 + \mathbf{d}^T \Psi^{-1} \mathbf{d}}. \quad (2.53)$$

Demonstração. De fato, a densidade conjunta de $\mathbf{Y}$ e T é

$$f_{\mathbf{Y},T}(\mathbf{y}, t) = 2\phi_n(\mathbf{y}|\boldsymbol{\mu} + \mathbf{d}t, \Psi)\phi_1(t)\mathbb{I}\{t > 0\}.$$

Do Lema 2.2 segue que

$$\phi_n(\mathbf{y}|\boldsymbol{\mu} + \mathbf{d}t, \Psi)\phi_1(t) = \phi_n(\mathbf{y}|\boldsymbol{\mu}, \Sigma)\phi_1(t|\eta, \tau^2),$$

concluindo a prova. $\square$

Note-se que a distribuição marginal de $\mathbf{Y}$ segue de (2.52) após integrar sobre t e é dada por

$$f_{\mathbf{Y}}(\mathbf{y}) = 2\phi_n(\mathbf{y}|\boldsymbol{\mu}, \Sigma)\Phi_1\left(\frac{\eta}{\tau}\right). \quad (2.54)$$

O próximo resultado está relacionado com as propriedades da distribuição normal truncada.

Lema 2.5. *Seja $X \sim N(\eta, \tau^2)$. Então, para a constante real , segue que*

$$E(X|X > a) = \eta + \frac{\phi_1(\frac{a-\eta}{\tau})}{1 - \Phi_1(\frac{a-\eta}{\tau})}\tau,$$

$$E(X^2|X > a) = \eta^2 + \tau^2 + \frac{\phi_1(\frac{a-\eta}{\tau})}{1 - \Phi_1(\frac{a-\eta}{\tau})}(\eta + a)\tau.$$

Demonstração. Veja Johnson et al. (1994), Seção 10.1. $\square$

Proposição 2.29. *Sob as condições da Proposição 2.28,*

$$E(T^k|\mathbf{y}) = E(X^k|X > 0),$$

onde $X \sim N_1(\eta, \tau^2)$, com η e τ^2 dados em (2.53). Particularmente,

$$E(T|\mathbf{y}) = \eta + \frac{\phi_1(\frac{\eta}{\tau})}{\Phi_1(\frac{\eta}{\tau})}\tau, \quad (2.55)$$

e

$$E(T^2|\mathbf{y}) = \eta^2 + \tau^2 + \frac{\phi_1(\frac{\eta}{\tau})}{\Phi_1(\frac{\eta}{\tau})} \tau \eta. \quad (2.56)$$

Demonstração. Note que podemos escrever

$$E(T^k|\mathbf{y}) = \int_0^\infty t^k f(t|\mathbf{y}) dt = \frac{1}{f_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\lambda})} \int_0^\infty t^k f_{\mathbf{Y},T}(\mathbf{y}, t|\boldsymbol{\theta}, \boldsymbol{\lambda}) dt.$$

De (2.52) e (2.54), segue que

$$E(T^k|\mathbf{y}) = \frac{1}{\Phi_1(\frac{\eta}{\tau})} \int_0^\infty t^k \phi_1(t|\eta, \tau^2) dt = E(X^k|X > 0),$$

em que $X \sim N_1(\eta, \tau^2)$ e $\Phi_1(\frac{\eta}{\tau}) = P(X > 0)$. Assim, (2.55) e (2.56) seguem do Lema 2.5 com $a=0$, concluindo a prova. $\square$

Para implementar o algoritmo EM, e assim encontrar as estimativas de máxima verossimilhança do vetor de parâmetros $\boldsymbol{\theta}$, da Proposição 2.4 temos que

$$\boldsymbol{\epsilon}_j \stackrel{d}{=} \sigma \mathbf{R}^{1/2} \boldsymbol{\delta} |X_{oj}| + \sigma \mathbf{R}^{1/2} (\mathbf{I}_n - \boldsymbol{\delta} \boldsymbol{\delta}^T)^{1/2} \mathbf{X}_{1j}, \quad (2.57)$$

onde $X_{oj} \stackrel{\text{iid}}{\sim} N(0, 1)$, $\mathbf{X}_{1j} \stackrel{\text{iid}}{\sim} N_n(0, \mathbf{I}_n)$, com X_{oj} e $\mathbf{X}_{1j}$ independentes, $j = 1, \dots, m$, e $\boldsymbol{\delta} = \boldsymbol{\lambda}/(1 + \boldsymbol{\lambda}^T \boldsymbol{\lambda})^{1/2}$. Logo, substituindo (2.57) em (2.49) segue que

$$\mathbf{Y}_j = \mathbf{X}_j \boldsymbol{\beta} + \bar{\boldsymbol{\delta}}_x T_j + \mathbf{r}_j, \quad (2.58)$$

onde

$$\bar{\boldsymbol{\delta}}_x = \sigma \mathbf{R}^{1/2} \boldsymbol{\delta}, \quad T_j = |X_{oj}| \quad \text{e} \quad \mathbf{r}_j = \sigma \mathbf{R}^{1/2} (\mathbf{I}_n - \boldsymbol{\delta} \boldsymbol{\delta}^T)^{1/2} \mathbf{X}_{1j}, \quad j = 1, \dots, m.$$

Note que $E[\mathbf{r}_j] = \mathbf{0}$, e como $E[T_j] = \sqrt{2/\pi}$, o termo $\sqrt{2/\pi} \bar{\boldsymbol{\delta}}_x$ pode ser usado para corrigir o intercepto do modelo, de modo que os parâmetros do modelo tenham a mesma interpretação que no modelo linear usual, onde o erro aleatório tem média zero.

Logo, (2.58) implica que o modelo definido por (2.49)-(2.50) pode ser escrito como

$$\mathbf{Y}_j | T_j = t_j \stackrel{\text{ind.}}{\sim} N_n(\mathbf{X}_j \boldsymbol{\beta} + \bar{\boldsymbol{\delta}}_x t_j, \sigma^2 \mathbf{R} - \bar{\boldsymbol{\delta}}_x \bar{\boldsymbol{\delta}}_x^T) \quad \text{e} \quad T_j \sim HN(0, 1), \quad j = 1, \dots, m. \quad (2.59)$$

Sejam $\mathbf{y}^T = (\mathbf{y}_1^T, \dots, \mathbf{y}_m^T)^T$ e $\mathbf{t} = (t_1, \dots, t_m)^T$. Como uma consequência direta da

Proposição 2.28 e de algumas manipulações algébricas temos o seguinte resultado.

Proposição 2.30. *Sob (2.59), segue que a log-verossimilhança completa associada a $(\mathbf{y}, \mathbf{t})$ no modelo linear normal-assimétrico (2.49)-(2.50) pode ser escrita como*

$$\ell_c(\boldsymbol{\theta}) \propto -\frac{m}{2} \log(\sigma^2 \tau^2 |\mathbf{R}|) - \frac{1}{2\sigma^2} \sum_{j=1}^m (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})^T \mathbf{R}^{-1} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}) - \frac{1}{2\tau^2} \sum_{j=1}^m (t_j - \eta_j)^2, \quad (2.60)$$

onde, por (2.53),

$$\eta_j = \frac{\boldsymbol{\delta}^T (\mathbf{I}_n - \boldsymbol{\delta} \boldsymbol{\delta}^T)^{-1} \mathbf{R}^{-1/2} (\mathbf{Y}_j - \mathbf{X}_j \boldsymbol{\beta})}{\sigma (1 + \boldsymbol{\delta}^T (\mathbf{I}_n - \boldsymbol{\delta} \boldsymbol{\delta}^T)^{-1} \boldsymbol{\delta})} \text{ e } \tau^2 = \frac{1}{1 + \boldsymbol{\delta}^T (\mathbf{I}_n - \boldsymbol{\delta} \boldsymbol{\delta}^T)^{-1} \boldsymbol{\delta}}, \quad j = 1, \dots, m. \quad (2.61)$$

Fazendo $\hat{t}_j = E(T_j | \hat{\boldsymbol{\theta}}, \mathbf{Y}_j = \mathbf{y}_j)$ e $\hat{t}_j^2 = E(T_j^2 | \hat{\boldsymbol{\theta}}, \mathbf{Y}_j = \mathbf{y}_j)$, da Proposição 2.29 temos que

$$\hat{t}_j = \hat{\eta}_j + \frac{\phi_1(\frac{\eta_j}{\tau_j})}{\Phi_1(\frac{\eta_j}{\tau_j})} \hat{\tau}_j, \quad (2.62)$$

$$\hat{t}_j^2 = \hat{\eta}_j^2 + \hat{\tau}_j^2 + \frac{\phi_1(\frac{\eta_j}{\tau_j})}{\Phi_1(\frac{\eta_j}{\tau_j})} \hat{\tau}_j \hat{\eta}_j, \quad (2.63)$$

onde η_j e τ^2 são como em (2.61).

Portanto, se $\boldsymbol{\theta}^{(s)}$ maximiza (2.51) na s -ésima iteração, então na iteração $s + 1$ o algoritmo EM procede como segue:

Etapa - E: Calcule $\hat{t}_j(\boldsymbol{\theta}^{(s)})$ e $\hat{t}_j^2(\boldsymbol{\theta}^{(s)})$, $j = 1, \dots, m$, a partir de (2.62) e (2.63), respectivamente, com $\boldsymbol{\theta} = \boldsymbol{\theta}^{(s)}$.

Etapa - M: Calcule $(\boldsymbol{\theta} = \boldsymbol{\theta}^{(s+1)})$ a partir de

$$\begin{aligned} \boldsymbol{\beta}^{(s+1)} &= \left[\sum_{j=1}^m \mathbf{X}_j^T (\mathbf{R}^{(s)})^{-1/2} (\mathbf{I}_n + \tau^2 f(\boldsymbol{\delta}^{(s)})) (\mathbf{R}^{(s)})^{-1/2} \mathbf{X}_j \right]^{-1} \\ &\times \sum_{j=1}^m [\mathbf{X}_j^T (\mathbf{R}^{(s)})^{-1/2} (\mathbf{I}_n + \tau^2 f(\boldsymbol{\delta}^{(s)})) (\mathbf{R}^{(s)})^{-1/2} \mathbf{y}_j - \sigma^{(s)} \hat{t}_j^{(s)} \mathbf{X}_j (\mathbf{R}^{(s)})^{-1/2} (\mathbf{I}_n - \boldsymbol{\delta} \boldsymbol{\delta}^T) \boldsymbol{\delta}], \end{aligned} \quad (2.64)$$

$$\text{e } \boldsymbol{\nu}^{(s+1)} = \operatorname{argmax}_{\boldsymbol{\nu}} [\ell_c(\boldsymbol{\beta}^{(s+1)}, \boldsymbol{\nu}^{(s)})], \text{ com } \boldsymbol{\nu} = (\sigma^2, \boldsymbol{\rho}^T, \boldsymbol{\delta}^T)^T, \quad (2.65)$$

onde $f(\delta) = (\mathbf{I}_n - \delta\delta^T)^{-1}\delta\delta^T(\mathbf{I}_n - \delta\delta^T)^{-1}$.

Note que para $\delta = \mathbf{0}$, (2.64) se reduz à estimação de mínimos quadrados ponderados de β no modelo linear geral. Este processo iterativo requer valores iniciais $\hat{\beta}^{(0)}$, $\hat{\sigma}^{2(o)}$, $\hat{\rho}^{(o)}$ e $\hat{\lambda}^{(o)}$, que podem ser as estimativas sob o modelo linear normal, com $\lambda = \mathbf{0}$. Para resolver (2.65) podemos usar algum algoritmo de maximização, tal como quase-Newton ou escore de Fisher. Estes algoritmos, com algumas modificações, encontram-se implementados em sistemas como Ox, R e S-plus. No Ox encontramos, por exemplo, a rotina MaxBFGS que usa uma versão modificada do algoritmo quase-Newton.

Para escolher o modelo que melhor explique a variabilidade presente nos dados, propomos avaliar a série de ajustes inspecionando critérios de informação, tais como, o critério de informação de Akaike (AIC, $-\ell(\hat{\theta})/N + P/N$), critério de informação bayesiano (BIC, $-\ell(\hat{\theta})/N + 0,5 \log(N)P/N$) e critério de informação de Hannan-Quinn (HQ, $-\ell(\hat{\theta})/N + \log(\log(N))P/N$), onde P é o número de parâmetros livres no modelo, $N = n \times m$ e $\ell(\theta)$ como em (2.51).

2.4.1 Estudo de Simulação

Nesta seção realizamos um estudo de simulação para investigar o desempenho do algoritmo e método proposto. O objetivo é basicamente avaliar o efeito da correlação na estimativa dos parâmetros, especificamente nos parâmetros de assimetria. Os dados foram gerados de acordo com o seguinte modelo linear normal-assimétrico multivariado:

$$y_{ij} = \beta_1 + \beta_2 x_{ij} + \epsilon_{ij},$$

onde $i = 1, \dots, 3$, isto é, $n = 3$, $\epsilon_j = (\epsilon_{1j}, \epsilon_{2j}, \epsilon_{3j})^T \sim SN_3(\mathbf{0}, \sigma^2 \mathbf{R}, \lambda \mathbf{1}_3)$ e valores fixados dos parâmetros $\beta_1 = 1$, $\beta_2 = 1$, $\sigma^2 = 1$. Os valores das covariáveis x_{ij} , $j = 1, \dots, m$, $i = 1, \dots, 3$, foram gerados de uma distribuição uniforme no intervalo $[0, 1]$. Os tamanhos amostrais usados foram $m = 30, 50$ e 100 , que são comuns em amostras de estudos longitudinais. Os erros aleatórios normal-assimétrico foram gerados usando a representação estocástica (2.12) e investigamos os resultados para diferentes valores de ρ e λ , assim como

estruturas uniforme e AR(1) para a matriz de correlações $\mathbf{R}$, isto é,

$$\mathbf{R} = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix} \quad \text{ou} \quad \mathbf{R} = \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix},$$

respectivamente. Para cada combinação de parâmetros, 1000 amostras independentes foram geradas. As simulações foram realizadas usando o programa Matlab, sendo a maximização de (2.65) efetuada através da rotina *fmincon*.

As Tabelas 2.1 - 2.3 resumem os resultados das simulações Monte Carlo em que as amostras com $\hat{\lambda} > 40$ foram eliminadas. Podemos notar nessas tabelas que quando o tamanho da amostra é pequeno ($m = 30$), existe um número considerável de amostras em que o parâmetro de assimetria estimado é maior que 40, especialmente quando geramos amostras com $\lambda \geq 2$. Notamos também neste caso que a estimação de σ^2 é ligeiramente viciada à medida que λ cresce. No caso em que $m = 100$, a estimação de σ^2 é altamente viciada quando a correlação é alta ($\rho = 0.9$) e independente do valor de λ . Em todos os casos os parâmetros β_1 e β_2 são bem estimados, qualquer que seja o tamanho da amostra. Um comportamento similar para os parâmetros β_1 e β_2 é detectado quando calculamos as médias Monte Carlo sem eliminar as amostras em que $\hat{\lambda} > 40$. Uma observação importante que pode-se resgatar deste estudo é que em geral e independente do tamanho de amostra, o aumento do parâmetro de assimetria não afeta a estimação de ρ , mas o contrario não é valido, a estimação de λ é severamente afetada quando a correlação é alta ($\rho = 0.9$).

Tabela 2.1: Médias das estimativas dos parâmetros em 1000 amostras simuladas para diferentes valores de λ e ρ , com $m = 30$. $\mathbf{R}$ é a matriz de correlações e N.C. é a porcentagem de amostras onde o parâmetro de assimetria estimado é menor que 40.

λ	ρ	$\mathbf{R}$ com estrutura uniforme						$\mathbf{R}$ com estrutura AR(1)					
		β_1	β_2	σ^2	ρ	λ	N.C.	β_1	β_2	σ^2	ρ	λ	N.C.
0.5	0	1.0061	0.9977	1.0244	0.0076	0.8436	88.4000	0.9947	1.0067	0.9995	0.0024	0.7217	92.4000
	0.3	1.0077	0.9983	1.0296	0.2939	0.7883	95.0000	0.9492	1.0510	1.0286	0.2912	0.8698	95.4000
	0.6	1.0033	1.0038	1.0428	0.5775	0.7534	95.2000	0.9920	1.0061	1.0710	0.5982	0.8448	95.4000
	0.9	0.9734	1.0077	1.4075	0.9062	1.1053	94.2000	0.9963	0.9717	1.5169	0.9126	1.3619	94.6000
1	0	1.0294	0.9857	0.9714	-0.0428	1.3146	85.4000	1.0306	0.9760	0.9717	-0.0226	1.3212	87.8000
	0.3	1.0187	1.0017	0.9423	0.2226	1.2265	90.6000	1.0340	0.9710	0.9757	0.2731	1.3473	89.4000
	0.6	1.0163	1.0051	0.9314	0.5238	1.3002	92.0000	1.0730	0.9423	0.9597	0.5519	1.2785	90.2000
	0.9	1.0169	0.9917	1.1626	0.8818	1.5814	86.2000	0.9657	1.0390	1.2158	0.8894	1.5916	84.0000
1.5	0	0.9969	1.0163	0.9516	-0.0377	1.8958	78.4000	1.0357	0.9711	0.9607	-0.0417	1.8252	75.0000
	0.3	1.0153	1.0023	0.9337	0.2229	1.7973	81.2000	0.9828	1.0305	0.9407	0.2381	1.8134	83.0000
	0.6	1.0270	0.9987	0.8950	0.5132	1.7333	83.6000	1.0410	0.9712	0.9326	0.5480	1.8421	82.0000
	0.9	0.9843	1.0235	1.0744	0.8822	2.2670	77.8000	1.0250	0.9871	1.0712	0.8844	2.3108	76.0000
2	0	1.0299	0.9861	0.9464	-0.0699	2.1457	72.6000	1.0279	0.9797	0.9562	-0.0287	2.5251	70.4000
	0.3	1.0153	1.0004	0.9500	0.2327	2.3050	73.0000	0.9794	1.0330	0.9375	0.2365	2.2396	68.6000
	0.6	1.0135	1.0045	0.9125	0.5211	2.3595	74.4000	0.9910	1.0240	0.9240	0.5470	2.3725	76.0000
	0.9	1.0008	1.0158	0.9859	0.8752	2.6039	74.6000	1.0126	1.0007	1.0063	0.8826	2.5697	69.8000
2.5	0	1.0016	1.0129	0.9247	-0.0624	2.5525	59.6000	1.0227	0.9857	0.9623	-0.0435	2.8711	59.2000
	0.3	1.0510	0.9647	0.9104	0.2178	2.6077	62.6000	1.0588	0.9520	0.9357	0.2506	2.6054	62.4000
	0.6	1.0243	0.9951	0.8790	0.5110	2.7035	73.8000	0.9738	1.0415	0.8992	0.5287	2.6820	71.8000
	0.9	1.0236	0.9899	0.9703	0.8779	3.1870	67.2000	0.9951	1.0134	1.0136	0.8877	3.1009	61.2000
3	0	1.0225	0.9902	0.9420	-0.0529	3.0143	53.8000	1.0406	0.9676	0.9284	-0.0401	3.1442	49.6000
	0.3	1.0153	1.0028	0.9031	0.2124	3.0914	59.0000	1.0004	1.0118	0.9255	0.2465	3.0446	53.8000
	0.6	1.0537	0.9677	0.8748	0.5160	3.0301	67.2000	1.0272	0.9880	0.8941	0.5335	2.9509	60.0000
	0.9	1.0131	1.0034	0.9385	0.8758	3.3791	57.0000	1.0090	1.0028	0.9796	0.8834	3.4560	53.0000

Tabela 2.2: Médias das estimativas dos parâmetros em 1000 amostras simuladas para diferentes valores de λ e ρ , com $m = 50$. $\mathbf{R}$ é a matriz se correlações e N.C. é a porcentagem de amostras onde o parâmetro de assimetria estimado é menor que 40

λ	ρ	$\mathbf{R}$ com estrutura uniforme						$\mathbf{R}$ com estrutura AR(1)					
		β_1	β_2	σ^2	ρ	λ	N.C.	β_1	β_2	σ^2	ρ	λ	N.C.
0.5	0	1.0504	0.9561	1.0070	0.0026	0.7585	97.6000	1.0167	0.9838	1.0070	0.0085	0.6360	98.2000
	0.3	1.0228	0.9791	1.0436	0.3124	0.7941	98.4000	1.0103	0.9897	1.0419	0.3121	0.7057	99.2000
	0.6	1.0010	1.0019	1.0508	0.5938	0.7048	99.4000	0.9919	1.0054	1.0531	0.6029	0.7577	99.2000
	0.9	0.9993	0.9924	1.2841	0.9026	0.8863	99.4000	1.0082	0.9719	1.3185	0.9074	0.9940	98.4000
1	0	1.0124	0.9942	0.9958	-0.0044	1.3999	95.8000	0.9859	1.0166	0.9982	-0.0050	1.2751	95.2000
	0.3	0.9587	1.0555	0.9724	0.2517	1.2411	95.8000	1.0058	0.9999	0.9851	0.2755	1.3900	94.4000
	0.6	1.0426	0.9748	0.9485	0.5451	1.1481	98.0000	1.0288	0.9810	0.9901	0.5721	1.2585	97.2000
	0.9	1.0077	0.9978	1.1233	0.8911	1.4288	97.4000	1.0120	0.9916	1.1276	0.8915	1.5348	96.2000
1.5	0	0.9902	1.0160	0.9724	-0.0223	2.0035	85.4000	1.0062	0.9962	0.9849	-0.0109	1.9940	88.2000
	0.3	1.0349	0.9740	0.9600	0.2567	1.9251	91.4000	1.0159	0.9886	0.9891	0.2683	1.9867	93.2000
	0.6	1.0163	0.9936	0.9630	0.5614	1.8896	94.2000	1.0027	1.0045	0.9604	0.5676	1.9631	94.4000
	0.9	0.9851	1.0241	1.0263	0.8861	2.1287	92.2000	1.0016	1.0014	1.0748	0.8902	2.1688	92.0000
2	0	1.0308	0.9762	0.9724	-0.0362	2.6152	81.4000	0.9612	1.0424	0.9898	-0.0223	2.4675	80.8000
	0.3	0.9807	1.0258	0.9774	0.2641	2.6564	87.0000	0.9930	1.0099	0.9778	0.2775	2.7567	86.8000
	0.6	1.0137	0.9935	0.9647	0.5672	2.4911	89.6000	1.0019	1.0034	0.9785	0.5773	2.6738	89.8000
	0.9	1.0025	1.0043	1.0255	0.8888	2.7778	89.8000	1.0111	0.9946	1.0381	0.8920	2.9094	86.4000
2.5	0	0.9801	1.0273	0.9631	-0.0270	3.1870	73.8000	1.0128	0.9912	0.9734	-0.0169	3.2135	75.0000
	0.3	1.0128	0.9917	0.9676	0.2660	3.1312	80.4000	1.0137	0.9887	0.9774	0.2786	3.1820	78.0000
	0.6	1.0313	0.9778	0.9462	0.5603	3.1311	85.6000	0.9941	1.0126	0.9538	0.5633	3.1277	85.0000
	0.9	0.9941	1.0095	1.0621	0.8934	3.6427	80.0000	0.9976	1.0081	1.0238	0.8921	3.3574	82.0000
3	0	0.9863	1.0175	0.9612	-0.0201	3.5950	68.4000	1.0009	1.0031	0.9737	-0.0270	3.6046	72.6000
	0.3	0.9937	1.0120	0.9681	0.2631	3.7089	75.6000	1.0189	0.9864	0.9713	0.2709	3.6492	74.8000
	0.6	1.0039	1.0039	0.9444	0.5651	3.5081	80.2000	1.0146	0.9906	0.9622	0.5728	3.7596	78.8000
	0.9	0.9940	1.0114	1.0053	0.8913	3.9302	75.8000	1.0172	0.9862	1.0012	0.8929	4.1630	76.0000

Tabela 2.3: Médias das estimativas dos parâmetros em 1000 amostras simuladas para diferentes valores de λ e ρ , com $m = 100$. $\mathbf{R}$ é a matriz se correlações e N.C. é a porcentagem de amostras onde o parâmetro de assimetria estimado é menor que 40

λ	ρ	R com estrutura uniforme						R com estrutura AR(1)					
		β_1	β_2	σ^2	ρ	λ	N.C.	β_1	β_2	σ^2	ρ	λ	N.C.
0.5	0	1.0165	0.9906	1.0079	0.0058	0.5551	100.0000	1.0377	0.9655	1.0090	-0.0007	0.5497	100.0000
	0.3	1.0193	0.9890	1.0066	0.2902	0.5314	100.0000	1.0196	0.9830	1.0034	0.3032	0.5529	100.0000
	0.6	0.9669	1.0406	1.0237	0.5893	0.5567	100.0000	1.0119	0.9893	1.0353	0.6021	0.5868	100.0000
	0.9	1.0015	0.9924	1.1881	0.9021	0.7172	99.4000	0.9999	0.9913	1.2222	0.9029	0.7280	100.0000
	1	1.0172	0.9890	0.9867	-0.0109	1.1029	99.6000	1.0040	0.9981	0.9921	-0.0166	1.0978	99.4000
1.5	0	0.9968	1.0120	0.9900	0.2732	1.1071	99.8000	1.0149	0.9884	0.9905	0.2885	1.1052	99.6000
	0.3	1.0116	0.9988	0.9768	0.5741	1.0545	100.0000	1.0099	0.9972	0.9833	0.5821	1.0958	100.0000
	0.6	1.0049	0.9910	1.2428	0.8999	1.5122	99.0000	0.9924	1.0085	1.1582	0.8955	1.4006	99.0000
	0.9	1.0022	0.9994	0.9951	-0.0074	1.7966	97.6000	1.0269	0.9731	1.0033	-0.0006	1.8759	99.2000
	1	1.0017	1.0018	0.9887	0.2817	1.7410	99.8000	1.0161	0.9852	0.9905	0.2955	1.8040	99.4000
2	0	1.0050	0.9994	0.9810	0.5761	1.6773	99.4000	1.0063	0.9968	0.9820	0.5851	1.7047	100.0000
	0.3	1.0037	0.9906	1.2579	0.9030	2.3045	97.6000	0.9823	1.0126	1.2661	0.9030	2.3608	96.2000
	0.6	1.0064	0.9943	0.9901	-0.0061	2.5787	97.4000	1.0085	0.9916	0.9951	-0.0030	2.4017	97.0000
	0.9	1.0284	0.9730	0.9869	0.2882	2.4360	98.6000	1.0201	0.9804	1.0014	0.2982	2.4643	98.8000
	1	1.0117	0.9909	0.9899	0.5865	2.4253	97.6000	1.0266	0.9759	0.9906	0.5887	2.3463	98.6000
2.5	0	0.9886	1.0052	1.2874	0.9055	3.1908	94.0000	1.0002	0.9938	1.2842	0.9075	3.0794	92.6000
	0.3	0.9688	1.0330	0.9876	-0.0112	2.9956	94.4000	0.9963	1.0039	1.0051	-0.0079	3.3106	95.0000
	0.6	0.9838	1.0174	0.9869	0.2869	3.0792	95.8000	1.0189	0.9820	0.9901	0.2930	3.1633	96.0000
	0.9	0.9943	1.0073	0.9946	0.5905	3.1277	98.0000	0.9970	1.0038	0.9910	0.5909	3.1188	98.4000
	1	0.9944	1.0015	1.2627	0.9046	3.9829	88.8000	0.9973	0.9978	1.2855	0.9077	3.8651	88.0000
3	0	1.0094	0.9917	0.9924	-0.0101	3.6516	91.4000	0.9945	1.0063	0.9903	-0.0036	3.8568	89.4000
	0.3	1.0376	0.9632	0.9893	0.2916	3.7928	93.0000	0.9949	1.0063	0.9837	0.2885	3.7304	92.8000
	0.6	1.0163	0.9853	0.9942	0.5886	3.8033	93.6000	0.9930	1.0089	0.9789	0.5908	3.6196	95.4000
	0.9	0.9954	1.0008	1.2621	0.9057	4.7777	85.2000	0.9964	0.9990	1.2314	0.9051	4.6297	86.2000

2.4.2 Aplicação

Como um exemplo consideremos os seguintes dados tomados de Barnett (1969), relativos a um estudo médico para avaliar a capacidade vital pulmonar usando dois instrumentos. Os instrumentos foram comparados num grupo comum de $m = 72$ pacientes, quando operados por um operador especializado e por um não especializado. As quatro combinações instrumento-operador serão referidas como $V1$, $V2$, $V3$ e $V4$, isto é, $n = 4$. Estes dados também foram examinados em um outro contexto por Bolfarine & Galea (1995). A Figura 2.3 mostra os histogramas das 4 combinações dos instrumentos em estudo, indicando claramente que os dados são assimétricos e que um modelo normal-assimétrico poderia ser adequado para modelar os dados. Assim, consideramos o seguinte modelo linear normal-assimétrico

$$y_j = \mu + \epsilon_j, \quad \epsilon_j \sim SN_4(\mathbf{0}, \sigma^2 \mathbf{R}, \boldsymbol{\lambda}), \quad j = 1, \dots, 72,$$

em que $m = 72$, $n = 4$ e para propósitos ilustrativos suponhamos que a matriz de correlações $\mathbf{R}$ tem estrutura uniforme. Para fins de comparação, o modelo também foi ajustado supondo que $\epsilon_j \sim N_n(\mathbf{0}, \sigma^2 \mathbf{R})$, $j = 1, \dots, m$. Os resultados dos ajustes sob as duas suposições podem ser vistos na Tabela 2.5. Os critérios de informação AIC, BIC e HQ, mostrados na parte inferior desta tabela, permitem afirmar que o modelo normal-assimétrico explica melhor a variabilidade presente nos dados.

Uma conclusão similar é obtida considerando o teste paramétrico para normalidade, que é dado pela estatística da razão de verossimilhanças com a hipótese nula $H_0 : \boldsymbol{\lambda} = \mathbf{0}$, isto é, pela seguinte estatística de teste

$$2\{\ell(\hat{\boldsymbol{\mu}}, \hat{\sigma}^2, \hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\lambda}}) - \ell(\tilde{\boldsymbol{\mu}}, \tilde{\sigma}^2, \tilde{\boldsymbol{\rho}}, \mathbf{0})\},$$

onde $\tilde{\boldsymbol{\theta}}$ indica o EMV de $\boldsymbol{\theta}$ sob a suposição de normalidade. O valor observado desta estatística de teste é 18,8377 que associado com o valor do quantil da distribuição $\chi_4^2(0.95) = 9.48$, nos leva a rejeitar a hipótese de que os dados seguem uma distribuição normal.

Na Figura 2.4 apresentamos os gráficos Q-Q dos resíduos r_j , como em (2.58), sob os dois modelos ajustados. Estes gráficos também parecem indicar que o modelo normal-assimétrico é mais apropriado, com presença de alguns outliers.

Tabela 2.4: Dados de Barnett. Estatísticas descritivas das medições dos instrumentos

	V1	V2	V3	V4
Mínimo	9.200	9.600	6.400	6.000
1º quantil	18.100	16.600	15.225	13.750
Média	22.461	21.757	21.486	21.022
Mediana	21.250	19.850	19.400	19.450
3º quantil	27.500	28.500	27.175	26.825
Máximo	40.400	41.800	41.500	39.600
Desvio padrão	7.698	7.922	8.939	8.554
Assimetria	0.503	0.669	0.5221	0.488
Curtose	-0.302	-0.280	-0.443	-0.692

Figura 2.3: Dados de Barnett. Histogramas das medições dos 4 instrumentos

Tabela 2.5: Resultado dos ajustes para os dados de Barnett. D.P. são estimativas dos desvios padrão dos estimadores correspondentes. P.C. são as estimativas corrigidas como explicado na Seção 2.4.

Parâmetros	Modelo normal-assimétrico		Modelo normal		P.C.
	Estimativas	D.P.	Estimativas	D.P.	
μ_1	14.5017	0.6375	22.4614	0.9598	22.7019
μ_2	13.4853	0.5567	21.7573	0.6797	22.0071
μ_3	13.9203	0.8740	21.4864	0.9589	21.7151
μ_4	13.1353	0.6679	21.0225	0.9133	21.2608
σ^2	130.8461	1.2003	67.8150	1.0018	-
ρ	0.9700	0.0037	0.9433	0.0071	-
λ_1	1.7397	0.5826	-	-	-
λ_2	2.4812	0.6635	-	-	-
λ_3	0.8048	0.4885	-	-	-
λ_4	1.5675	0.6370	-	-	-
-log-verossimilhança	1070.20		1085.00		
AIC	3.75		3.79		
BIC	3.81		3.83		
HQ	3.77		3.81		

Figura 2.4: Gráficos Q-Q dos resíduos para os dados de Barnett. (a) Modelo normal-assimétrico. (b) Modelo normal.

Capítulo 3

Modelos Lineares Mistos Normal-Assimétrico: Enfoque Clássico

3.1 Introdução

Como discutido no Capítulo 1, a normalidade dos erros e efeitos aleatórios é uma suposição rotineira em modelos lineares mistos (MLM's), mas esta pode não ser realista, obscurecendo características importantes da variação entre e intra-unidades amostrais. Embora inferência nos efeitos fixos possa ser robusta à não normalidade dos efeitos aleatórios (Butler & Louis, 1992; Verbeke & Lesaffre, 1997), é natural estar interessado sobre a eficiência e validade de inferências nos efeitos individuais. Mais ainda, estimação dos efeitos aleatórios sob suposições menos restritivas pode fornecer considerável discernimento. Por exemplo, assimetria estimada pode sugerir a exclusão de covariáveis importantes e revelar características críticas da heterogeneidade inerente entre sujeitos. Assim, propostas recentes estão sendo feitas baseadas em substituir a suposição de normalidade por suposições fracas de que os efeitos aleatórios têm uma densidade suave ("smooth") que pode ser assimétrica. Do ponto de vista prático, isto envolve representar a densidade por uma forma flexível e estimar a densidade juntamente com outros parâmetros adicionais do modelo. Por exemplo, Zhang & Davidian (2001) usam a representação semi-não-paramétrica (SNP) da densidade proposta por Gallant & Nychka (1987) para caracterizar a densidade dos efeitos aleatórios. Uma característica interessante desta representação é que a verossimilhança do modelo pode ser expressa em forma fechada. Alternativamente, Verbeke &

Lessaffre (1996) adotam uma representação via mistura de normais e tratam inferências usando o algoritmo EM (veja também Verbeke & Molenberghs, 2000, cap. 12). Magder & Zeger (1996) usam uma forma de máxima verossimilhança não-paramétrica baseada em densidades normais e usam técnicas um tanto *ad hoc* para assinalar ajustes. Tao et al. (1999) estimam a densidade dos efeitos aleatórios via algoritmos recursivos preditivos.

Neste capítulo relaxamos as suposições típicas do MLM, considerando que tanto os erros como os efeitos aleatórios seguem uma distribuição da família normal-assimétrica, que inclui a distribuição normal como caso especial e fornece flexibilidade para captar uma ampla variedade de comportamentos não normais. A densidade marginal das quantidades observadas (verossimilhança observada) é obtida integrando sobre os efeitos aleatórios, embora as estimativas de máxima verossimilhança possam ser obtidas diretamente maximizando a função log-verossimilhança, as representações estocásticas propostas no capítulo anterior permitem a implementação do algoritmo EM, o qual parece ser mais efetivo em termos de valores iniciais e, mais ainda, não envolve expressões complexas como o caso da função de verossimilhança observada.

O plano deste capítulo é como segue. Na Seção 3.2 apresentamos a versão univariada do estudo, considerando que tanto os erros como os efeitos aleatórios seguem uma distribuição normal-assimétrica de Azzalini (1985). Chamamos este modelo de modelo com interceptos aleatórios normal-assimétrica (MIASN). Na Seção 3.3 generalizamos os resultados obtidos para o caso multivariado e fazemos um estudo detalhado dos procedimentos de estimação nesta classe de modelos. Na Seção 3.4 apresentamos um estudo de simulação onde mostramos que existe perda de eficiência na estimação de certos parâmetros do modelo quando assimetria está presente, relativo a considerar normalidade (simetria) na distribuição dos dados. Na última seção apresentamos uma aplicação dos métodos propostos a um conjunto de dados de colesterol, previamente analisada por Zhang & Davidian (2001) usando técnicas semi-paramétricas.

3.2 O Modelo com Interceptos Aleatórios Normal-Assimétrico

Para fixar idéias, nesta seção centramos nossa atenção em um modelo linear misto relativamente simples, chamado de modelo de interceptos aleatórios e definido por

$$Y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + b_j + \epsilon_{ij}, \quad (3.1)$$

onde Y_{ij} é a resposta para a observação i no sujeito j , $i = 1, \dots, n_j$, $j = 1, \dots, m$, $\mathbf{x}_{ij}$ é um vetor de covariáveis conhecidas de dimensão $(p \times 1)$, $\boldsymbol{\beta}$ é o vetor desconhecido dos coeficientes da regressão e b_j é o efeito aleatório, supostamente independente do erro aleatório ϵ_{ij} . Assumiremos também que

$$b_j \sim SN_1(0, \sigma_b^2, \lambda_b) \text{ e } \epsilon_{ij} \sim SN_1(0, \sigma_e^2, \lambda_e), \quad (3.2)$$

onde SN_1 denota a distribuição normal-assimétrica univariada definida em (2.5).

Agrupando os Y_{ij} no vetor $\mathbf{Y}_j$ de dimensão $(n_j \times 1)$, os ϵ_{ij} no vetor $\boldsymbol{\epsilon}_j$ de dimensão $(n_j \times 1)$ e empilhando as linhas $\mathbf{x}_{ij}^T$ na matriz $\mathbf{X}_j$ de dimensão $(n_j \times p)$, temos o seguinte modelo

$$\mathbf{Y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{1}_{n_j} b_j + \boldsymbol{\epsilon}_j, \quad j = 1, \dots, m, \quad (3.3)$$

onde, de (3.2) e do Corolário 2.10, temos que

$$b_j \sim SN_1(0, \sigma_b^2, \lambda_b) \text{ e } \boldsymbol{\epsilon}_j \sim SNF_{n_j}(\mathbf{0}, \sigma_e^2 \mathbf{I}_{n_j}, \delta_e \mathbf{I}_{n_j}), \quad (3.4)$$

com $\mathbf{1}_{n_j}$ o vetor de uns de dimensão $(n_j \times 1)$, $\mathbf{I}_{n_j}$ a matriz identidade de dimensão $(n_j \times n_j)$ e $\delta_e = \lambda_e / (1 + \lambda_e^2)^{1/2}$. Como $[\mathbf{Y}_j | b_j] \stackrel{d}{=} (\mathbf{X}_j \boldsymbol{\beta} + \mathbf{1}_{n_j} b_j) + \boldsymbol{\epsilon}_j$, da Proposição 2.25, segue que o modelo definido em (3.3)-(3.4) pode ser expresso como

$$\mathbf{Y}_j | b_j \sim SNF_{n_j}(\mathbf{X}_j \boldsymbol{\beta} + \mathbf{1}_{n_j} b_j, \sigma_e^2 \mathbf{I}_{n_j}, \delta_e \mathbf{I}_{n_j}), \quad (3.5)$$

$$b_j \sim SN_1(0, \sigma_b^2, \lambda_b), \quad j = 1, \dots, m. \quad (3.6)$$

O principal interesse é fazer inferências sobre o vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma_e^2, \sigma_b^2)$ e $\boldsymbol{\lambda} = (\lambda_e, \lambda_b)^T$. Note que a densidade condicional do vetor aleatório $[\mathbf{Y}_j | b_j]$, após algumas

manipulações algébricas, pode alternativamente ser escrita como

$$f(\mathbf{y}_j|b_j, \boldsymbol{\beta}, \sigma_e^2, \lambda_e) = 2^{n_j} \phi_{n_j}(\mathbf{y}_j|\mathbf{X}_j\boldsymbol{\beta} + \mathbf{1}_{n_j}b_j, \sigma_e^2\mathbf{I}_{n_j})\Phi_{n_j}\left(\frac{\lambda_e}{\sigma_e}(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta} - \mathbf{1}_{n_j}b_j)\right), \quad (3.7)$$

para $j = 1, \dots, m$.

Como notado no Capítulo 1, procedimentos de inferência clássica neste tipo de modelo são baseados na verossimilhança do modelo marginal, obtida integrando a densidade conjunta sobre os efeitos aleatórios. Para simplificar a notação e obter a densidade marginal de $\mathbf{Y}_j$, requerida para derivar a função log-verossimilhança marginal, omitimos o índice j , correspondente ao j -ésimo grupo (ou indivíduo). De (3.5)-(3.6) e da densidade normal-assimétrica univariada em (2.5) (ou da Definição 2.1, com $n = 1$), segue que a densidade marginal de $\mathbf{Y}$ é obtida calculando-se a seguinte integral

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\lambda}) &= \int_{\mathbb{R}} f(\mathbf{y}|b, \boldsymbol{\beta}, \sigma_e^2, \lambda_e) f(b|\sigma_b^2, \lambda_b) db \\ &= \int_{\mathbb{R}} 2^{n+1} \phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta} + \mathbf{1}_n b, \sigma_e^2\mathbf{I}_n) \Phi_n\left(\frac{\lambda_e}{\sigma_e}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{1}_n b)\right) \\ &\quad \times \phi_1(b|0, \sigma_b^2) \Phi_1\left(\frac{\lambda_b}{\sigma_b}b\right) db. \end{aligned} \quad (3.8)$$

Provaremos a seguir que a solução desta integral é fechada, e para tal, fazemos uso da seguinte proposição.

Proposição 3.1. *Sob a notação considerada em (3.8), segue que*

$$\phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta} + \mathbf{1}_n b, \sigma_e^2\mathbf{I}_n) \phi_1(b|0, \sigma_b^2) = \phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}) \phi_1(b|\mu_1, \Lambda) \quad (3.9)$$

e

$$\Phi_n\left(\frac{\lambda_e}{\sigma_e}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{1}_n b)\right) \Phi_1\left(\frac{\lambda_b}{\sigma_b}b\right) = \Phi_{n+1}(-\Gamma b| -\boldsymbol{\mu}_2, \mathbf{I}_{n+1}), \quad (3.10)$$

onde

$$\mu_1 = \frac{1}{\sigma_e^2} \Lambda \mathbf{1}_n^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad \boldsymbol{\Sigma} = \sigma_e^2 \mathbf{I}_n + \sigma_b^2 \mathbf{1}_n \mathbf{1}_n^T, \quad \Lambda = \frac{\sigma_e^2 \sigma_b^2}{\sigma_e^2 + n \sigma_b^2}, \quad (3.11)$$

$$\boldsymbol{\mu}_2 = \begin{pmatrix} \frac{\lambda_e}{\sigma_e}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ 0 \end{pmatrix} \quad e \quad \boldsymbol{\Gamma} = \begin{pmatrix} \frac{\lambda_e}{\sigma_e} \mathbf{1}_n \\ \frac{\lambda_b}{\sigma_b} \end{pmatrix}. \quad (3.12)$$

Demonstração. O resultado em (3.9) segue facilmente do Lema 2.2. O resultado em (3.10) é provado notando-se que se o vetor aleatório $\mathbf{U} \sim N_n(0, \mathbf{I}_n)$ é independente da variável aleatória $V \sim N_1(0, 1)$, então (3.10) pode ser escrito como

$$\begin{aligned} P(\mathbf{U} \leq \frac{\lambda_e}{\sigma_e}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{1}_n b) | V \leq \frac{\lambda_b}{\sigma_b} b) &= P(\mathbf{U} + \frac{\lambda_e}{\sigma_e} \mathbf{1}_n b \leq \frac{\lambda_e}{\sigma_e}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) | V - \frac{\lambda_b}{\sigma_b} b \leq 0) \\ &= P(\mathbf{T} \leq \boldsymbol{\mu}_2), \end{aligned}$$

onde $\mathbf{T} = \mathbf{W} + \Gamma b$, com $\mathbf{W} = (\mathbf{U}^T, V)^T \sim N_{n+1}(\mathbf{0}, \mathbf{I}_{n+1})$, $\boldsymbol{\mu}_2$ e Γ como definidos em (3.12). Assim, levando em conta que $\mathbf{T} \sim N_{n+1}(\Gamma b, \mathbf{I}_{n+1})$, temos que

$$P(\mathbf{T} \leq \boldsymbol{\mu}_2) = \Phi_{n+1}(\boldsymbol{\mu}_2 | \Gamma b, \mathbf{I}_{n+1}) = \Phi_{n+1}(-\Gamma b | -\boldsymbol{\mu}_2, \mathbf{I}_{n+1}),$$

concluindo a prova. $\square$

A solução da integral (3.8) e conseqüentemente a densidade marginal de $\mathbf{Y}_j$, $j = 1, \dots, m$ é dada no seguinte teorema.

Teorema 3.1. *Seja $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{1}_n b + \boldsymbol{\epsilon}$, onde $b \sim SN_1(0, \sigma_b^2, \lambda_b)$ e os elementos do vetor $\boldsymbol{\epsilon}$ são i.i.d. $SN_1(0, \sigma_e^2, \lambda_e)$. Então, a densidade marginal de $\mathbf{Y}$ é dada por*

$$f_{\mathbf{Y}}(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\lambda}) = 2^{n+1} \phi_n(\mathbf{y} | \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}) \Phi_{n+1}(\boldsymbol{\mu}_2 - \Gamma \boldsymbol{\mu}_1 | \mathbf{0}, \mathbf{I}_{n+1} + \Lambda \Gamma \Gamma^T), \quad (3.13)$$

onde $\boldsymbol{\mu}_1$, $\boldsymbol{\Sigma}$, Λ , $\boldsymbol{\mu}_2$ e Γ são dados em (3.11) e (3.12).

Demonstração. De (3.8), (3.9) e (3.10) segue que

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\lambda}) &= \int_{\mathbb{R}} 2^{n+1} \phi_n(\mathbf{y} | \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}) \phi(b | \mu_1, \Lambda) \Phi_{n+1}(-\Gamma b | -\boldsymbol{\mu}_2, \mathbf{I}_{n+1}) db \\ &= 2^{n+1} \phi_n(\mathbf{y} | \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}) E[\Phi_{n+1}(-\Gamma N | -\boldsymbol{\mu}_2, \mathbf{I}_{n+1})], \end{aligned}$$

onde $N \sim N_1(\mu_1, \Lambda)$. A prova segue do Lema 2.1. $\square$

Note que a densidade em (3.13) pertence à classe de distribuições normal-assimétrica fundamental, introduzida por Arellano-Valle & Genton (2003) e definida em (2.36). A função verossimilhança para $\boldsymbol{\theta}$ e $\boldsymbol{\lambda}$ dada a amostra observada $\mathbf{y}_1, \dots, \mathbf{y}_m$ pode ser escrita como

$$L(\boldsymbol{\theta}, \boldsymbol{\lambda} | \mathbf{y}_1, \dots, \mathbf{y}_m) = \prod_{j=1}^m f_{\mathbf{Y}}(\mathbf{y}_j | \boldsymbol{\theta}, \boldsymbol{\lambda}).$$

onde $f_Y(\mathbf{y}_j|\boldsymbol{\theta}, \boldsymbol{\lambda})$ é a densidade marginal que segue de (3.13) incorporando o índice j . Assim, denotando a função log-verossimilhança por $\ell(\boldsymbol{\theta}, \boldsymbol{\lambda})$, esta pode ser escrita como

$$\begin{aligned} \ell(\boldsymbol{\theta}, \boldsymbol{\lambda}) \propto & -\frac{1}{2} \sum_{j=1}^m \log |\boldsymbol{\Sigma}_j| - \frac{1}{2} \sum_{j=1}^m \{(\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})^T \boldsymbol{\Sigma}_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})\} \\ & + \sum_{j=1}^m \log \Phi_{n_j+1}(\boldsymbol{\mu}_{2j} - \boldsymbol{\Gamma}_j \boldsymbol{\mu}_{1j} | \mathbf{0}, \mathbf{I}_{n_j+1} + \boldsymbol{\Lambda}_j \boldsymbol{\Gamma}_j \boldsymbol{\Gamma}_j^T), \end{aligned} \quad (3.14)$$

onde

$$\boldsymbol{\mu}_{1j} = \frac{1}{\sigma_e^2} \boldsymbol{\Lambda}_j \mathbf{1}_{n_j}^T (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}), \quad \boldsymbol{\Sigma}_j = \sigma_e^2 \mathbf{I}_{n_j} + \sigma_b^2 \mathbf{1}_{n_j} \mathbf{1}_{n_j}^T, \quad \boldsymbol{\Lambda}_j = \frac{\sigma_e^2 \sigma_b^2}{\sigma_e^2 + n_j \sigma_b^2}, \quad (3.15)$$

$$\boldsymbol{\mu}_{2j} = \begin{pmatrix} \frac{\lambda_e}{\sigma_e} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}) \\ 0 \end{pmatrix} \quad \text{e} \quad \boldsymbol{\Gamma} = \begin{pmatrix} \frac{\lambda_e}{\sigma_e} \mathbf{1}_{n_j} \\ -\frac{\lambda_b}{\sigma_b} \end{pmatrix}. \quad (3.16)$$

Para reduzir a dimensão da função de assimetria, $\Phi_{n_j+1}(\cdot)$, em (3.13), no seguinte corolário consideramos o caso em que $\lambda_e = 0$ (ou seja, consideramos que os erros têm distribuição normal). Outros comentários relativos a procedimentos inferenciais serão abordados na Seção 3.3.

Corolário 3.1. *Sob as condições do Teorema 3.1, se $\lambda_e = 0$, então*

$$f_Y(\mathbf{y}|\boldsymbol{\theta}, \lambda_b) = 2\phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})\Phi_1(\bar{\boldsymbol{\lambda}}_b^T \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})), \quad (3.17)$$

isto é,

$$\mathbf{Y} \sim SN_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}, \bar{\boldsymbol{\delta}}_b), \quad \text{com} \quad \bar{\boldsymbol{\lambda}}_b = \frac{\lambda_b \sigma_b \boldsymbol{\Sigma}^{-1/2} \mathbf{1}_n}{(1 + \frac{\lambda_b^2}{\sigma_b^2} \boldsymbol{\Lambda})^{1/2}}.$$

Demonstração. Note que para $\lambda_e = 0$,

$$\boldsymbol{\mu}_2 - \boldsymbol{\Gamma} \boldsymbol{\mu}_1 = \begin{pmatrix} 0 \\ \frac{\lambda_b \boldsymbol{\Lambda}}{\sigma_b \sigma_e^2} \mathbf{1}_n^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \end{pmatrix}, \quad \mathbf{I}_{n+1} + \boldsymbol{\Lambda} \boldsymbol{\Gamma} \boldsymbol{\Gamma}^T = \begin{pmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & 1 + \frac{\lambda_b^2}{\sigma_b^2} \boldsymbol{\Lambda} \end{pmatrix}.$$

Assim, a parte assimétrica de (3.13) pode ser escrita como

$$\begin{aligned} \Phi_{n+1}(\boldsymbol{\mu}_2 - \boldsymbol{\Gamma} \boldsymbol{\mu}_1 | \mathbf{0}, \mathbf{I}_{n+1} + \boldsymbol{\Lambda} \boldsymbol{\Gamma} \boldsymbol{\Gamma}^T) &= \Phi_{n+1}((\mathbf{I}_{n+1} + \boldsymbol{\Lambda} \boldsymbol{\Gamma} \boldsymbol{\Gamma}^T)^{-1/2} (\boldsymbol{\mu}_2 - \boldsymbol{\Gamma} \boldsymbol{\mu}_1)) \\ &= \frac{1}{2^n} \Phi_1 \left((1 + \frac{\lambda_b^2}{\sigma_b^2} \boldsymbol{\Lambda})^{-1/2} \frac{\lambda_b \boldsymbol{\Lambda}}{\sigma_b \sigma_e^2} \mathbf{1}_n^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right). \end{aligned}$$

e a prova segue notando-se que $\Lambda \mathbf{1}_n^T = \sigma_b^2 \sigma_e^2 \mathbf{1}_n^T \Sigma^{-1}$. $\square$

De um modo similar a (3.14), segue que para $\lambda_e = 0$ e dada uma amostra observada, a função log-verossimilhança marginal é dada por

$$\begin{aligned} \ell(\boldsymbol{\theta}, \boldsymbol{\lambda}) \propto & -\frac{1}{2} \sum_{j=1}^m \log |\Sigma_j| - \frac{1}{2} \sum_{j=1}^m \{(\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})^T \Sigma_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})\} \\ & + \sum_{j=1}^m \log \Phi_1(\bar{\boldsymbol{\lambda}}_{b_j}^T \Sigma_j^{-1/2} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})), \end{aligned} \quad (3.18)$$

onde $\bar{\boldsymbol{\lambda}}_{b_j} = \frac{\lambda_b \sigma_b \Sigma_j^{-1/2} \mathbf{1}_{n_j}}{(1 + \frac{\lambda_b^2}{\sigma_b^2} \Lambda_j)^{1/2}}$, com Λ_j e Σ_j como em (3.15).

Note que a função log-verossimilhança (3.18) pode ser maximizada diretamente via algoritmos de maximização numérica. A matriz de covariância assintótica dos estimadores de máxima verossimilhança pode ser estimada usando a matriz hessiana que pode ser calculada também numericamente usando programas computacionais tais como Ox, R, S-plus e Matlab. No entanto, é possível que a maximização direta possua alguns problemas computacionais, pois $\ell(\boldsymbol{\theta}, \boldsymbol{\lambda})$, envolve o termo $\log(\Phi_1(x))$, que pode não existir, por exemplo, para valores de x menores que -37 . Além disso, este procedimento parece não ser robusto em relação aos valores iniciais, isto é, a menos que valores iniciais adequados sejam dados, a maximização direta pode tipicamente não convergir. Na seção seguinte apresentamos o algoritmo EM para calcular os EMV da densidade obtida no Corolário 3.1. Como é próprio do algoritmo EM, as estimativas são mais confiáveis, mas eventualmente pode requerer maior tempo computacional para alcançar a convergência.

3.2.1 Algoritmo EM para encontrar os EMV no MIASN

Para implementar o algoritmo EM, consideramos o modelo definido em (3.3)-(3.4) com $\lambda_e = 0$, isto é,

$$\mathbf{Y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{1}_{n_j} b_j + \boldsymbol{\epsilon}_j, \quad (3.19)$$

onde

$$b_j \sim SN_1(0, \sigma_b^2, \lambda_b) \text{ e } \boldsymbol{\epsilon}_j \sim N_{n_j}(\mathbf{0}, \sigma_e^2 \mathbf{I}_{n_j}), \quad j = 1, \dots, m. \quad (3.20)$$

Logo, (3.20) conjuntamente com a Proposição 2.4 implica que

$$b_j = \sigma_b \delta_b |X_{oj}| + \sigma_b (1 - \delta_b^2)^{1/2} X_{1j}, \quad j = 1, \dots, m, \quad (3.21)$$

em que $X_{ij} \stackrel{\text{iid}}{\sim} N(0, 1)$, para $i = 0, 1, j = 1, \dots, m$, e $\delta_b = \lambda_b / (1 + \lambda_b^2)^{1/2}$. Mais ainda, a independência entre b_j e ϵ_j implica que $\mathbf{V}_j = (X_{oj}, X_{1j})^T$ e ϵ_j são independentes. Logo, substituindo (3.21) em (3.19) temos que

$$\mathbf{Y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{1}_{n_j} \bar{\delta}_b T_j + \mathbf{r}_j, \quad (3.22)$$

onde

$$\bar{\delta}_b = \sigma_b \delta_b, \quad T_j = |X_{oj}| \quad \text{e} \quad \mathbf{r}_j = \epsilon_j + \mathbf{1}_{n_j} \sigma_b (1 - \delta_b^2)^{1/2} X_{1j},$$

são tais que

$$\mathbf{r}_j \stackrel{\text{ind.}}{\sim} N_{n_j}(0, \sigma_e^2 \mathbf{I}_{n_j} + \mathbf{1}_{n_j} (\sigma_b^2 - \bar{\delta}_b^2) \mathbf{1}_{n_j}^T) \quad \text{e} \quad T_j \stackrel{\text{iid}}{\sim} HN(0, 1), \quad (3.23)$$

ambos independentes, $j = 1, \dots, m$. Assim, (3.22) e (3.23) implicam que o modelo definido em (3.19)-(3.20) pode ser escrito como

$$\mathbf{Y}_j | T_j = t_j \stackrel{\text{ind.}}{\sim} N_{n_j}(\mathbf{X}_j \boldsymbol{\beta} + \mathbf{d}_j t_j, \boldsymbol{\Psi}_j) \quad \text{e} \quad T_j \stackrel{\text{iid}}{\sim} HN(0, 1), \quad j = 1, \dots, m, \quad (3.24)$$

onde $\mathbf{d}_j = \bar{\delta}_b \mathbf{1}_{n_j}$, $\boldsymbol{\Psi}_j = \boldsymbol{\Sigma}_j - \mathbf{d}_j \mathbf{d}_j^T$ e $\boldsymbol{\Sigma}_j$ como definido em (3.15).

Seja $\mathbf{y}^T = (\mathbf{y}_1^T, \dots, \mathbf{y}_m^T)^T$ e $\mathbf{t} = (t_1, \dots, t_m)^T$. Como uma consequência direta da Proposição 2.28 e algumas manipulações algébricas, temos o seguinte resultado.

Proposição 3.2. *Sob (3.24) segue que a log-verossimilhança completa associada a $(\mathbf{y}, \mathbf{t})$, $j = 1, \dots, m$, pode ser escrita como*

$$\ell_c(\boldsymbol{\theta}, \lambda_b) \propto -\frac{1}{2} \sum_{j=1}^m \log |\boldsymbol{\Psi}_j| - \frac{1}{2} \sum_{j=1}^m (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})^T \boldsymbol{\Sigma}_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}) - \frac{1}{2} \sum_{j=1}^m \frac{(t_j - \eta_j)^2}{\tau_j^2}, \quad (3.25)$$

onde, por (2.53),

$$\eta_j = \frac{\mathbf{d}_j \boldsymbol{\Psi}_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})}{1 + \mathbf{d}_j^T \boldsymbol{\Psi}_j^{-1} \mathbf{d}_j} \quad \text{e} \quad \tau_j^2 = \frac{1}{1 + \mathbf{d}_j^T \boldsymbol{\Psi}_j^{-1} \mathbf{d}_j}, \quad (3.26)$$

com $\mathbf{d}_j = \sigma_b \delta_b \mathbf{1}_{n_j}$ e $\boldsymbol{\Psi}_j = \boldsymbol{\Sigma}_j - \mathbf{d}_j \mathbf{d}_j^T = \sigma_e^2 \mathbf{I}_{n_j} + \sigma_b^2 (1 - \delta_b^2) \mathbf{1}_{n_j} \mathbf{1}_{n_j}^T$.

Fazendo $\boldsymbol{\theta}_c = (\boldsymbol{\theta}^T, \lambda_b)^T$, $\hat{t}_j = E(T_j | \hat{\boldsymbol{\theta}}_c, \mathbf{Y}_j = \mathbf{y}_j)$ e $\hat{t}_j^2 = E(T_j^2 | \hat{\boldsymbol{\theta}}_c, \mathbf{Y}_j = \mathbf{y}_j)$, da

Proposição 2.29 temos que

$$\hat{t}_j = \hat{\eta}_j + \frac{\phi_1(\frac{\eta_j}{\tau_j})}{\Phi_1(\frac{\eta_j}{\tau_j})} \hat{\tau}_j, \quad (3.27)$$

$$\hat{t}_j^2 = \hat{\eta}_j^2 + \hat{\tau}_j^2 + \frac{\phi_1(\frac{\eta_j}{\tau_j})}{\Phi_1(\frac{\eta_j}{\tau_j})} \hat{\tau}_j \hat{\eta}_j, \quad (3.28)$$

onde η_j e τ_j^2 são como em (3.26), $j=1, \dots, m$.

Portanto, se $\theta_c^{(s)} = (\theta^T, \lambda_b)^T$ maximiza (3.25) na s -ésima iteração, então na iteração $s+1$ o algoritmo EM procede como segue:

Etapa E: Calcule $\hat{t}_j(\theta_c^{(s)})$ e $\hat{t}_j^2(\theta_c^{(s)})$, $j = 1, \dots, m$, a partir de (3.27) e (3.28), respectivamente, com $\theta_c = \theta_c^{(s)}$.

Etapa M: Calcule $(\theta_c = \theta_c^{(s+1)})$ a partir de

$$\begin{aligned} \beta^{(s+1)} &= \left[\sum_{j=1}^m \mathbf{X}_j^T (\Sigma_j^{-1} + \tau_j^2 \Psi_j^{-1} \mathbf{d}_j \mathbf{d}_j^T \Psi_j^{-1}) \mathbf{X}_j \right]^{-1} \\ &\times \sum_{j=1}^m [\mathbf{X}_j^T (\Sigma_j^{-1} + \tau_j^2 \Psi_j^{-1} \mathbf{d}_j \mathbf{d}_j^T \Psi_j^{-1}) \mathbf{y}_j - \hat{t}_j \mathbf{X}_j^T \Psi_j^{-1} \mathbf{d}_j], \end{aligned} \quad (3.29)$$

e

$$\nu^{(s+1)} = \operatorname{argmax}_{\nu} [\ell_c(\beta^{(s+1)}, \nu^{(s)})], \text{ com } \nu = (\sigma_e^2, \sigma_b^2, \lambda_b)^T. \quad (3.30)$$

Detalhes sobre a convergência e valores iniciais para a implementação do algoritmo serão discutidos na seção seguinte.

3.3 O Modelo Linear Misto Normal-Assimétrico

A fim de permitir propriedades de simetria-assimetria para caracterizar dados de medidas repetidas, dados agrupados e dados longitudinais, nesta seção estendemos o MLM definido em (1.1)-(1.2), considerando o seguinte modelo linear misto normal-assimétrico (MLMSN):

$$\mathbf{Y}_j = \mathbf{X}_j \beta + \mathbf{Z}_j \mathbf{b}_j + \epsilon_j, \quad (3.31)$$

com

$$\mathbf{b}_j \stackrel{\text{iid}}{\sim} SN_q(\mathbf{0}, \mathbf{D}, \boldsymbol{\lambda}_b) \text{ e } \boldsymbol{\epsilon}_j \stackrel{\text{iid}}{\sim} SN_{n_j}(\mathbf{0}, \boldsymbol{\psi}_j, \boldsymbol{\lambda}_{ej}), \quad (3.32)$$

onde $\mathbf{b}_j$ é independente de $\boldsymbol{\epsilon}_j$, $j = 1, \dots, m$. Usando a Proposição 2.5, segue que o modelo definido acima pode ser representado hierarquicamente como

$$\mathbf{Y}_j | \mathbf{b}_j \stackrel{\text{ind.}}{\sim} SN_{n_j}(\mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \mathbf{b}_j, \boldsymbol{\psi}_j, \boldsymbol{\lambda}_{ej}), \quad (3.33)$$

$$\mathbf{b}_j \stackrel{\text{iid}}{\sim} SN_q(\mathbf{0}, \mathbf{D}, \boldsymbol{\lambda}_b), \quad j = 1, \dots, m, \quad (3.34)$$

Note de (3.31)-(3.32) e da Proposição 2.15 que

$$E[\mathbf{Y}_j] = \mathbf{X}_j \boldsymbol{\beta} + E[\mathbf{Z}_j \mathbf{b}_j + \boldsymbol{\epsilon}_j] = \mathbf{X}_j \boldsymbol{\beta} + \sqrt{\frac{2}{\pi}} (\mathbf{Z}_j \mathbf{D}^{1/2} \boldsymbol{\delta}_b + \boldsymbol{\psi}_j^{1/2} \boldsymbol{\delta}_{ej}),$$

onde $\boldsymbol{\delta}_b = \boldsymbol{\lambda}_b(1 + \boldsymbol{\lambda}_b^T \boldsymbol{\lambda}_b)^{-1/2}$ e $\boldsymbol{\delta}_{ej} = \boldsymbol{\lambda}_{ej}(1 + \boldsymbol{\lambda}_{ej}^T \boldsymbol{\lambda}_{ej})^{-1/2}$, $j = 1, \dots, m$. O segundo termo da direita na equação acima pode ser considerado para obter uma interpretação correta dos parâmetros do modelo (regressão clássica). Contudo, na prática podemos em geral resgatar a interpretação usual corrigindo o intercepto do modelo ajustado, como será visto na Seção 3.3.1. O principal interesse neste caso é fazer inferências sobre o vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T)^T$ e $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_b^T, \boldsymbol{\lambda}_{e1}^T, \dots, \boldsymbol{\lambda}_{em}^T)^T$ (veja Capítulo 1, para detalhes desta notação). Como discutido em Verbeke & Molenberghs (2000), a menos que os dados sejam analisados numa estrutura bayesiana, inferências neste tipo de modelo são baseadas na distribuição marginal da resposta $\mathbf{Y}_j$. A densidade marginal de $\mathbf{Y}_j$ é obtida no seguinte teorema, cuja prova encontra-se no Apêndice B.

Teorema 3.2. *Seja $\mathbf{Y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \mathbf{b}_j + \boldsymbol{\epsilon}_j$, onde $\mathbf{b}_j \stackrel{\text{iid}}{\sim} SN_q(\mathbf{0}, \mathbf{D}, \boldsymbol{\lambda}_b)$ e $\boldsymbol{\epsilon}_j \stackrel{\text{iid}}{\sim} SN_{n_j}(\mathbf{0}, \boldsymbol{\psi}_j, \boldsymbol{\lambda}_{ej})$ são independentes. Então, a distribuição marginal de $\mathbf{Y}_j$ é dada por*

$$f_{\mathbf{Y}_j}(\mathbf{y}_j | \boldsymbol{\theta}, \boldsymbol{\lambda}) = 2^2 \phi_{n_j}(\mathbf{y}_j | \mathbf{X}_j \boldsymbol{\beta}, \boldsymbol{\Sigma}_j) \Phi_2 \left((\boldsymbol{\mu}_{2j} - \boldsymbol{\Gamma}_j \boldsymbol{\mu}_{1j})(\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}) | \mathbf{0}, \mathbf{I}_2 + \boldsymbol{\Gamma}_j \boldsymbol{\Lambda}_j \boldsymbol{\Gamma}_j^T \right), \quad (3.35)$$

onde

$$\boldsymbol{\mu}_{1j} = \boldsymbol{\Lambda}_j \mathbf{Z}_j^T \boldsymbol{\psi}_j^{-1}, \quad \boldsymbol{\Sigma}_j = \boldsymbol{\psi}_j + \mathbf{Z}_j \mathbf{D} \mathbf{Z}_j^T, \quad \boldsymbol{\Lambda}_j = (\mathbf{D}^{-1} + \mathbf{Z}_j^T \boldsymbol{\psi}_j^{-1} \mathbf{Z}_j)^{-1}, \quad (3.36)$$

$$\boldsymbol{\mu}_{2j} = \begin{pmatrix} \boldsymbol{\lambda}_{ej}^T \boldsymbol{\psi}_j^{-1/2} \\ 0 \end{pmatrix} \quad \text{e} \quad \boldsymbol{\Gamma}_j = \begin{pmatrix} \boldsymbol{\lambda}_{ej}^T \boldsymbol{\psi}_j^{-1/2} \mathbf{Z}_j \\ -\boldsymbol{\lambda}_b^T \mathbf{D}^{-1/2} \end{pmatrix}, \quad j = 1, \dots, m. \quad (3.37)$$

Note que a densidade (3.35) não está na classe definida aqui, pois o fator de assimetria é de dimensão 2. Contudo, está na classe de distribuições normal-assimétrica fundamental (Arellano-Valle & Genton, 2003), definida no Capítulo 2 (veja também a relação (2.9) na Observação 2.2). O resultado apresentado no Teorema 3.2 é importante porque evita usar técnicas numéricas mais complexas tal como integração Monte Carlo para tratar inferências neste tipo de modelo, pois permite obter uma forma fechada para a densidade marginal de $\mathbf{Y}_j$, $j = 1, \dots, m$, facilitando a implementação direta de inferências usando técnicas padrões de otimização numérica. Assim, denotando a função log-verossimilhança por $\ell(\boldsymbol{\theta}, \boldsymbol{\lambda})$, esta pode ser escrita como

$$\begin{aligned} \ell(\boldsymbol{\theta}, \boldsymbol{\lambda}) \propto & -\frac{1}{2} \sum_{j=1}^m \log |\boldsymbol{\Sigma}_j| - \frac{1}{2} \sum_{j=1}^m \{(\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})^T \boldsymbol{\Sigma}_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})\} \\ & + \sum_{j=1}^m \log \Phi_2 \left((\boldsymbol{\mu}_{2j} - \boldsymbol{\Gamma}_j \boldsymbol{\mu}_{1j})(\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}) | \mathbf{0}, \mathbf{I}_2 + \boldsymbol{\Gamma}_j \boldsymbol{\Lambda}_j \boldsymbol{\Gamma}_j^T \right), \end{aligned} \quad (3.38)$$

onde $\boldsymbol{\mu}_{1j}$, $\boldsymbol{\mu}_{2j}$, $\boldsymbol{\Sigma}_j$, $\boldsymbol{\Gamma}_j$ e $\boldsymbol{\Lambda}_j$ são definidos em (3.36) e (3.37), $j=1, \dots, m$.

Chamamos atenção para o fato de que nenhuma solução explícita pode ser encontrada para o problema de maximização da função log-verossimilhança acima, contudo esta deve ser maximizada numericamente, usando técnicas implementadas em diversos programas computacionais. Alguns casos especiais são de interesse prático. Por exemplo, as situações em que $\boldsymbol{\lambda}_{e1} = \dots = \boldsymbol{\lambda}_{em} = \mathbf{0}$ ou $\boldsymbol{\lambda}_b = \mathbf{0}$. Estas situações serão tratadas a seguir. As provas são dadas no Apêndice B

Corolário 3.2. *Sob as condições do Teorema 3.2, segue que*

(i) *Se $\boldsymbol{\lambda}_{e_j} = \mathbf{0}$, $j = 1, \dots, m$, então*

$$f_{\mathbf{Y}_j}(\mathbf{y}_j | \boldsymbol{\theta}, \boldsymbol{\lambda}_b) = 2\phi_{n_j}(\mathbf{y}_j | \mathbf{X}_j \boldsymbol{\beta}, \boldsymbol{\Sigma}_j) \Phi_1 \left(\bar{\boldsymbol{\lambda}}_{b_j}^T \boldsymbol{\Sigma}_j^{-1/2} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}) \right), \quad (3.39)$$

ou seja,

$$\mathbf{Y}_j \sim SN_{n_j}(\mathbf{X}_j \boldsymbol{\beta}, \boldsymbol{\Sigma}_j, \bar{\boldsymbol{\lambda}}_{b_j}), \quad \text{com} \quad \bar{\boldsymbol{\lambda}}_{b_j} = \frac{\boldsymbol{\Sigma}_j^{-1/2} \mathbf{Z}_j \mathbf{D}^{1/2} \boldsymbol{\lambda}_b}{\sqrt{1 + \boldsymbol{\lambda}_b^T \mathbf{D}^{-1/2} \boldsymbol{\Lambda}_j \mathbf{D}^{-1/2} \boldsymbol{\lambda}_b}}.$$

(ii) Se $\lambda_b = \mathbf{0}$, então

$$f_{\mathbf{Y}_j}(\mathbf{y}_j|\theta, \lambda_{e_j}) = 2\phi_{n_j}(\mathbf{y}_j|\mathbf{X}_j\beta, \Sigma_j)\Phi_1\left(\bar{\lambda}_{e_j}^T \Sigma_j^{-1/2}(\mathbf{y}_j - \mathbf{X}_j\beta)\right), \quad (3.40)$$

ou seja,

$$\mathbf{Y}_j \sim SN_{n_j}(\mathbf{X}_j\beta, \Sigma_j, \bar{\lambda}_{e_j}), \quad \text{com} \quad \bar{\lambda}_{e_j} = \frac{\Sigma_j^{-1/2} \psi_j^{1/2} \lambda_{e_j}}{\sqrt{1 + \lambda_{e_j}^T \psi_j^{-1/2} \mathbf{Z}_j \Lambda_j \mathbf{Z}_j^T \psi_j^{-1/2} \lambda_{e_j}}}.$$

A seguir, apresentamos algoritmos tipo EM para cálculo de EMV dos parâmetros dos modelos do Corolário 3.2.

3.3.1 Algoritmo EM quando $\lambda_{e1} = \dots = \lambda_{em} = \mathbf{0}$

Sob esta suposição, temos o seguinte MLMSN:

$$\mathbf{Y}_j = \mathbf{X}_j\beta + \mathbf{Z}_j\mathbf{b}_j + \epsilon_j, \quad (3.41)$$

com

$$\epsilon_j \stackrel{\text{ind.}}{\sim} N_{n_j}(\mathbf{0}, \psi_j), \quad \mathbf{b}_j \stackrel{\text{ind.}}{\sim} SN_q(\mathbf{0}, \mathbf{D}, \lambda_b) \quad \text{e} \quad (3.42)$$

ϵ_j independente de $\mathbf{b}_j$, $j = 1, \dots, m$. É claro que (3.42) conjuntamente com a Proposição 2.4 implica que

$$\mathbf{b}_j \stackrel{d}{=} \mathbf{D}^{1/2} \delta_b |X_{0j}| + \mathbf{D}^{1/2} (\mathbf{I}_q - \delta_b \delta_b^T)^{1/2} \mathbf{X}_{1j}, \quad j = 1, \dots, m, \quad (3.43)$$

onde $X_{0j} \stackrel{\text{iid}}{\sim} N_1(0, 1)$, $\mathbf{X}_{1j} \stackrel{\text{iid}}{\sim} N_q(\mathbf{0}, \mathbf{I}_q)$, com X_{0j} e $\mathbf{X}_{1j}$ independentes $j = 1, \dots, m$, e $\delta_b = \frac{\lambda_b}{\sqrt{1 + \lambda_b^T \lambda_b}}$. Além disso, a independência entre $\mathbf{b}_j$ e ϵ_j implica que $\mathbf{V}_j = (X_{0j}, \mathbf{X}_{1j}^T)^T$ e ϵ_j , são independentes, $j = 1, \dots, m$. Assim, substituindo (3.43) em (3.41), temos que

$$\mathbf{Y}_j = \mathbf{X}_j\beta + \mathbf{Z}_j \bar{\delta}_b t_j + \mathbf{r}_j, \quad (3.44)$$

onde

$$\bar{\delta}_b = \mathbf{D}^{1/2} \delta_b, \quad t_j = |X_{0j}| \quad \text{e} \quad \mathbf{r}_j = \epsilon_j + \mathbf{Z}_j \mathbf{D}^{1/2} (\mathbf{I}_q - \delta_b \delta_b^T)^{1/2} \mathbf{X}_{1j},$$

da qual temos que

$$\mathbf{r}_j \stackrel{\text{ind.}}{\sim} N_{n_j}(\mathbf{0}, \boldsymbol{\psi}_j + \mathbf{Z}_j(\mathbf{D} - \bar{\boldsymbol{\delta}}_b \bar{\boldsymbol{\delta}}_b^T) \mathbf{Z}_j^T), \quad t_j \stackrel{\text{iid}}{\sim} HN(0, 1), \quad (3.45)$$

ambas independentes, $j = 1, \dots, m$. Note que $\mathbf{r}_j$ tem média zero, de modo que pode ser usado, por exemplo, para fazer análises de resíduos e assim, avaliar a adequabilidade do modelo. Além disso, o segundo termo no lado direito da equação (3.44) tem média $\sqrt{\frac{2}{\pi}} \mathbf{Z}_j \bar{\boldsymbol{\delta}}_b$, que pode ser usado eventualmente para corrigir o intercepto do modelo de modo que os efeitos fixos tenham a mesma interpretação que no usual MLM (média populacional).

Portanto, (3.44) e (3.45) implicam que o modelo definido por (3.41)-(3.42) pode ser escrito como

$$\mathbf{Y}_j | t_j \stackrel{\text{ind.}}{\sim} N_{n_j}(\boldsymbol{\mu}_j + \mathbf{d}_j t_j, \boldsymbol{\Psi}_j) \quad \text{e} \quad t_j \stackrel{\text{iid}}{\sim} HN_1(0, 1), \quad j = 1, \dots, m, \quad (3.46)$$

onde

$$\boldsymbol{\mu}_j = \mathbf{X}_j \boldsymbol{\beta}, \quad \mathbf{d}_j = \mathbf{Z}_j \bar{\boldsymbol{\delta}}_b, \quad \boldsymbol{\Psi}_j = \boldsymbol{\Sigma}_j - \mathbf{d}_j \mathbf{d}_j^T \quad \text{e} \quad \boldsymbol{\Sigma}_j = \boldsymbol{\psi}_j + \mathbf{Z}_j \mathbf{D} \mathbf{Z}_j^T. \quad (3.47)$$

Note que em (3.47), $\boldsymbol{\mu}_j$ e $\boldsymbol{\Sigma}_j$ são a média marginal e a matriz de variâncias-covariâncias, respectivamente, sob o usual MLM.

Sejam $\mathbf{y}^T = (\mathbf{y}_1^T, \dots, \mathbf{y}_m^T)^T$ e $\mathbf{t} = (t_1, \dots, t_m)^T$. Como uma consequência direta da Proposição 2.28 e usando uma álgebra simples, temos o seguinte resultado.

Proposição 3.3. *Sob (3.46) segue que a função log-verossimilhança completa associada com $(\mathbf{y}, \mathbf{t})$ no MLMSN (3.41)-(3.42) pode ser escrita como*

$$\ell_c(\boldsymbol{\theta}, \boldsymbol{\lambda}_b) \propto -\frac{1}{2} \sum_{j=1}^m \log |\boldsymbol{\Psi}_j| - \frac{1}{2} \sum_{j=1}^m (\mathbf{y}_j - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1} (\mathbf{y}_j - \boldsymbol{\mu}_j) - \frac{1}{2} \sum_{j=1}^m \frac{(t_j - \eta_j)^2}{\tau_j^2}, \quad (3.48)$$

onde, por (2.53),

$$\eta_j = \frac{\mathbf{d}_j^T \boldsymbol{\Psi}_j^{-1} (\mathbf{y}_j - \boldsymbol{\mu}_j)}{1 + \mathbf{d}_j^T \boldsymbol{\Psi}_j^{-1} \mathbf{d}_j} \quad \text{e} \quad \tau_j^2 = \frac{1}{1 + \mathbf{d}_j^T \boldsymbol{\Psi}_j^{-1} \mathbf{d}_j}, \quad (3.49)$$

com $\boldsymbol{\mu}_j$, $\mathbf{d}_j$, $\boldsymbol{\Sigma}_j$ e $\boldsymbol{\Psi}_j$ definidos em (3.47).

Fazendo $\boldsymbol{\theta}_c = (\boldsymbol{\theta}^T, \boldsymbol{\lambda}_b^T)^T$, $\hat{t}_j = E(T_j | \hat{\boldsymbol{\theta}}_c, \mathbf{Y}_j = \mathbf{y}_j)$ e $\hat{t}_j^2 = E(T_j^2 | \hat{\boldsymbol{\theta}}_c, \mathbf{Y}_j = \mathbf{y}_j)$, obtemos

da Proposição 2.29 que

$$\widehat{t}_j = \widehat{\eta}_j + \frac{\phi_1(\frac{\eta_j}{\tau_j})}{\Phi_1(\frac{\eta_j}{\tau_j})} \widehat{\tau}_j, \quad (3.50)$$

$$\widehat{t}_j^2 = \widehat{\eta}_j^2 + \widehat{\tau}_j^2 + \frac{\phi_1(\frac{\eta_j}{\tau_j})}{\Phi_1(\frac{\eta_j}{\tau_j})} \widehat{\tau}_j \widehat{\eta}_j, \quad (3.51)$$

onde η_j e τ_j^2 são como em (3.49). Assim, temos o seguinte algoritmo EM:

passo-E: Dado $\boldsymbol{\theta}_c = \widehat{\boldsymbol{\theta}}_c$, Calcular $\widehat{t}_j$ e $\widehat{t}_j^2$ for $j = 1, \dots, m$, usando (3.50) e (3.51), respectivamente.

passo-M: Atualizar $\widehat{\boldsymbol{\theta}}_c$ maximizando $E[\ell_c(\boldsymbol{\theta}_c) | \mathbf{y}, \widehat{\boldsymbol{\theta}}_c]$ (substituindo previamente $\widehat{t}_j$ e $\widehat{t}_j^2$) sobre $\boldsymbol{\theta}_c$, que leva a

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= \left[\sum_{j=1}^m \mathbf{X}_j^T (\widehat{\boldsymbol{\Sigma}}_j^{-1} + \widehat{\tau}_j^2 \widehat{\boldsymbol{\Psi}}_j^{-1} \widehat{\mathbf{d}}_j \widehat{\mathbf{d}}_j^T \widehat{\boldsymbol{\Psi}}_j^{-1}) \mathbf{X}_j \right]^{-1} \\ &\quad \times \sum_{j=1}^m [\mathbf{X}_j^T (\widehat{\boldsymbol{\Sigma}}_j^{-1} + \widehat{\tau}_j^2 \widehat{\boldsymbol{\Psi}}_j^{-1} \widehat{\mathbf{d}}_j \widehat{\mathbf{d}}_j^T \widehat{\boldsymbol{\Psi}}_j^{-1}) \mathbf{y}_j - \widehat{t}_j \mathbf{X}_j^T \widehat{\boldsymbol{\Psi}}_j^{-1} \widehat{\mathbf{d}}_j], \end{aligned} \quad (3.52)$$

e

$$\widehat{\boldsymbol{\nu}} = \operatorname{argmax}_{\boldsymbol{\nu}} [\ell_c(\widehat{\boldsymbol{\beta}}, \boldsymbol{\nu})], \text{ com } \boldsymbol{\nu} = (\boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T, \boldsymbol{\lambda}_b^T)^T, \quad (3.53)$$

onde $\ell_c(\widehat{\boldsymbol{\beta}}, \boldsymbol{\nu})$ é (3.48) avaliado nos valores atualizados $\widehat{\boldsymbol{\beta}}$, $t_j = \widehat{t}_j$ e $t_j^2 = \widehat{t}_j^2$, $j = 1, \dots, m$.

3.3.2 Algoritmo EM quando $\boldsymbol{\lambda}_b = \mathbf{0}$

Similarmente, considerando o caso em que $\boldsymbol{\lambda}_b = \mathbf{0}$, isto é, o MLM em (3.41) com as suposições

$$\boldsymbol{\epsilon}_j \stackrel{\text{ind.}}{\sim} SN_{n_j}(\mathbf{0}, \boldsymbol{\psi}_j, \boldsymbol{\lambda}_{ej}) \quad \text{e} \quad \mathbf{b}_j \stackrel{\text{iid}}{\sim} N_q(\mathbf{0}, \mathbf{D}), \quad j = 1, \dots, m, \quad (3.54)$$

todos independentes, podemos escrever

$$\mathbf{Y}_j = \mathbf{X}_j \boldsymbol{\beta} + \boldsymbol{\psi}_j^{1/2} \boldsymbol{\delta}_{ej} t_j + \mathbf{r}_j,$$

onde

$$t_j = |X_{1j}|, \quad \delta_{ej} = \frac{\lambda_{ej}}{(1 + \lambda_{ej}\lambda_{ej}^T)^{1/2}} \text{ e } \mathbf{r}_j = \mathbf{Z}_j\mathbf{b}_j + \psi_j^{1/2}(\mathbf{I}_{n_j} - \delta_{ej}\delta_{ej}^T)^{1/2}\mathbf{X}_{0j},$$

isto é,

$$\mathbf{Y}_j|t_j \stackrel{\text{ind.}}{\sim} N_{n_j}(\boldsymbol{\mu}_j + \mathbf{d}_j t_j, \boldsymbol{\Psi}_j) \text{ e } t_j \sim HN_1(0, 1), \quad (3.55)$$

onde

$$\boldsymbol{\mu}_j = \mathbf{X}_j\boldsymbol{\beta}, \quad \mathbf{d}_j = \psi_j^{1/2}\boldsymbol{\delta}_{ej}, \quad \boldsymbol{\Psi}_j = \boldsymbol{\Sigma}_j - \mathbf{d}_j\mathbf{d}_j^T \text{ e } \boldsymbol{\Sigma}_j = \psi_j + \mathbf{Z}_j\mathbf{D}\mathbf{Z}_j^T \quad j = 1, \dots, m. \quad (3.56)$$

Como uma consequência dos resultados acima, segue da Proposição 2.28 que

Proposição 3.4. *Sob (3.55), segue que a log-verossimilhança completa associada com $(\mathbf{y}, \mathbf{t})$ no MLMSN (3.41) e (3.54) pode ser escrita como*

$$\ell_c(\boldsymbol{\theta}, \boldsymbol{\lambda}_e) \propto -\frac{1}{2} \sum_{j=1}^m \log |\boldsymbol{\Psi}_j| - \frac{1}{2} \sum_{j=1}^m (\mathbf{y}_j - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1} (\mathbf{y}_j - \boldsymbol{\mu}_j) - \frac{1}{2} \sum_{j=1}^m \frac{(t_j - \eta_j)^2}{\tau_j^2}, \quad (3.57)$$

onde

$$\eta_j = \frac{\mathbf{d}_j^T \boldsymbol{\Psi}_j^{-1} (\mathbf{y}_j - \boldsymbol{\mu}_j)}{1 + \mathbf{d}_j^T \boldsymbol{\Psi}_j^{-1} \mathbf{d}_j} \text{ e } \tau_j^2 = \frac{1}{1 + \mathbf{d}_j^T \boldsymbol{\Psi}_j^{-1} \mathbf{d}_j},$$

com $\boldsymbol{\mu}_j$, $\mathbf{d}_j$, $\boldsymbol{\Sigma}_j$ e $\boldsymbol{\Psi}_j$ definidos em (3.56).

O algoritmo EM neste caso procede como em (3.50)-(3.53), com η_j e τ_j^2 como em (3.58). Note que em ambos os casos o **passo M** requer maximização numérica, que pode ser facilmente implementada usando programas computacionais como Ox, R e Matlab com as rotinas *maxBFGS*, *optim* e *fmincon*, respectivamente. Os valores iniciais podem ser escolhidos como sendo as correspondentes estimativas sob normalidade, em que o parâmetro de assimetria é igual a $\mathbf{0}$, e como recomendado na literatura, é prudente executar o algoritmo EM várias vezes com diferentes valores iniciais.

Seguindo a idéia de outros autores (exemplo, Zhang & Davidian, 2001), aqui também propomos avaliar a série de ajustes inspecionando critérios de informação, tais como, AIC $(-\ell(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}})/N + P/N)$, BIC $(-\ell(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}})/N + 0.5 \log(N)P/N)$ e HQ $(-\ell(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}})/N + \log(\log(N))P/N)$, onde P é o número de parâmetros livres no modelo e $N = \sum_{j=1}^m n_j$, como será mostrado rapidamente na seção seguinte.

Tabela 3.1: Resultados de Monte Carlo baseados em 100 conjuntos de dados simulados com distribuição $Gama(4, 1)$ para os efeitos aleatórios. Média MC e D.P. MC são a média e desvio padrão das estimativas, Média E.P. é a média dos erros padrão estimados baseados na matriz hessiana. IC é o intervalo confiança empírico das estimativas com probabilidade de 95%. Os verdadeiros valores dos parâmetros estão entre parênteses.

Parâmetros	Média MC	D.P. MC	Média E.P.	IC
(a) Cenário normal-assimétrico				
α	1.5447	0.3349	-	-
$\beta_1(2)$	2.0008	0.0106	0.0108	0.95
$\beta_2(1)$	0.9731	0.2081	0.2083	0.95
$\sigma_e^2(0.25)$	0.2490	0.0135	0.0119	0.93
$E[\alpha + b](4)$	4.0083	0.1636	-	-
$Var[\alpha + b](4)$	3.9853	0.5399	-	-
(b) Cenário normal				
α	4.0213	0.1828	-	-
$\beta_1(2)$	2.0008	0.0106	0.0116	0.98
$\beta_2(1)$	0.9336	0.2638	0.3035	0.93
$\sigma_e^2(0.25)$	0.2490	0.0135	0.0125	0.93
$E[\alpha + b](4)$	4.0213	0.1828	-	-
$Var[\alpha + b](4)$	3.9409	0.5839	-	-

3.4 Estudo de Simulação

Para analisar a performance do modelo e métodos propostos, realizamos dois estudos de simulação. Tomamos o MLM como sendo igual a

$$Y_{ij} = \alpha + t_{ij}\beta_1 + w_j\beta_2 + b_j + e_{ij}, \quad (3.58)$$

onde para $i = 1, \dots, 5$, $t_{ij} = i - 3$, $\beta_1 = 2$, $\beta_2 = 1$, e $e_{ij} \sim N(0; 0, 5^2)$. Em cada simulação, 100 conjuntos de dados foram simulados de (3.58) segundo especificações adicionais descritas abaixo. Como uma vantagem da representação normal-assimétrica é a sua flexibilidade para acomodar assimetria, em uma primeira simulação geramos $\alpha + b_j$ de uma distribuição $Gama(4, 1)$, com fdp igual a $(1/6)x^3 \exp(-x)$, produzindo alta assimetria como sugerido pela Figura 3.1. Sob estas especificações, $E[\alpha + b_j] = 4$ e $Var[\alpha + b_j] = 4$. Note que t_{ij} representa uma covariável com valores mudando dentro de indivíduos (ou nas unidades de cada grupo) e fixa para todos os indivíduos (grupos), enquanto w_j pode ser visto como um indicador de tratamento. Tomamos $m = 200$ com $w_j = 1$ se $j \leq 100$ e $w_j = 0$ se $j > 100$. Para cada um dos 100 conjuntos de dados gerados, (3.58) foi ajus-

tado duas vezes sob as suposições estudadas previamente, isto é, com a densidade de b_j representada pela distribuição normal-assimétrica e também pela distribuição normal.

Tabela 3.2: Resultados de Monte Carlo baseados em 100 conjuntos de dados com distribuição $normal(0, 4)$ para os efeitos aleatórios . As entradas são como na Tabela 3.1.

Parâmetros	Média MC	D.P. MC	Média E.P.	IC
(a) Cenário normal-assimétrico				
α (4)	3.6766	1.1681	1.3058	0.60
β_1 (2)	2.0009	0.0103	0.0108	0.96
β_2 (1)	1.0161	0.3137	0.2953	0.92
σ_e^2 (0.25)	0.2482	0.0143	0.0121	0.89
$E[\alpha + b]$ (4)	3.9799	0.1916	-	-
$Var[\alpha + b]$ (4)	3.9368	0.4306	-	-
(b) Cenário normal				
α (4)	3.9780	0.1923	0.2056	0.96
β_1 (2)	2.0009	0.0103	0.0111	0.96
β_2 (1)	1.0207	0.3167	0.2932	0.95
σ_e^2 (0.25)	0.2482	0.0143	0.0124	0.90
$E[\alpha + b]$ (4)	3.9780	0.1923	-	-
$Var[\alpha + b]$ (4)	3.9347	0.4312	-	-

Para os resultados do algoritmo EM, nenhum dos critérios (AIC, BIC e HQ), selecionou a especificação normal como melhor modelo, mostrando a capacidade destes métodos de seleção para detectar afastamentos da normalidade e sugerindo forte evidência de assimetria. A Tabela 3.1 mostra resultados numéricos para as estimativas onde normalidade e distribuição normal-assimétrica são consideradas. Na maior parte, os parâmetros estimados são não viciados. Na maioria dos casos a média dos erros padrão estimados (baseados na matrix hessiana da log-verossimilhança) estão próximos dos desvios padrão estimados. Como encontrado por outros autores (por Exemplo, Tao et al., 1999; Zhang & Davidian, 2001), eficiência na estimação de β_2 (parâmetro associado com w_j) é degradada quando normalidade é assumida relativa a permitir uma representação mais flexível via distribuição normal-assimétrica. Como um dos principais interesses de tal análise pode ser a avaliação do efeito de tratamento, este resultado sugere que adotar normalidade nas suposições do modelo pode levar a inferências ineficientes nos parâmetros fixos que são de interesse fundamental. Note que alguma perda de eficiência é também detectada associada a estimação da variância inter-indivíduos $Var[b_j]$. Assim, embora

estimação não viciada seja possível sob normalidade, esta falha por não levar em conta as verdadeiras características dos efeitos aleatórios, acarretando inferências menos precisas sobre os parâmetros de interesse.

A vantagem de estimar a densidade dos efeitos aleatórios também pode ser apreciada na Figura 3.1. A Figura 3.1(a) apresenta as médias de Monte Carlo das densidades estimadas nos 100 conjuntos de dados, comparando à verdadeira densidade (gama), o ajuste normal e o ajuste normal-assimétrico. Esta figura mostra que a flexibilidade adicional proporcionada pela distribuição normal-assimétrica é suficiente para captar as verdadeiras características dos efeitos aleatórios. Esta observação é adicionalmente suportada pela Figura 3.1(b), na qual mostramos as 100 densidades estimadas dos ajustes sob distribuição normal-assimétrica. O segundo estudo de simulação foi idêntico ao primeiro, exceto que

Figura 3.1: Resultados de simulação baseados em 100 conjuntos de dados simulados. (a) Verdadeira densidade (linha cheia) e média de Monte Carlo das 100 densidade estimadas usando normalidade (linha tracejada) e usando normal-assimétrica (linha pontilhada) (b) Densidades estimadas para os ajustes sob distribuição normal-assimétrica.

a distribuição de $\alpha + b_j$ foi $N(4, 4)$, isto é, $\alpha = 4$ e $b_j \sim N(0, 4)$. Aqui então, não existe necessidade de maior flexibilidade na distribuição dos efeitos aleatórios, e esperamos que os métodos propostos identifiquem isso. Os critérios AIC, BIC e HQ selecionam corretamente o modelo normal 86%, 98% e 93% das vezes, respectivamente. Os resultados da simulação Monte Carlo são apresentados na Tabela 3.2, a qual mostra que não existe perda de eficiência quando a distribuição normal-assimétrica é assumida. Resultados similares podem ser encontrados em Hu, Tsiatis & Davidian (1998) e Zhang & Davidian (2001).

3.5 Aplicação

Nesta seção ilustramos a utilidade dos métodos propostos aplicando-os a um conjunto de dados longitudinais coletado como parte do conhecido estudo de colesterol em Framingham (*Framingham cholesterol study*). O conjunto de dados inclui níveis de colesterol em diferentes instantes de tempo, idade no início do estudo (*baseline*) e sexo para $m = 200$ indivíduos selecionados aleatoriamente e reportado em Zhang & Davidian (2001). Aqui adotamos o mesmo modelo linear usado por estes autores, dado por

$$y_{ij} = \beta_o + \beta_1 \text{sexo}_j + \beta_2 \text{idade}_j + \beta_3 t_{ij} + b_{oj} + b_{1j} t_{ij} + \epsilon_{ij}, \quad (3.59)$$

onde y_{ij} é o nível de colesterol (mg/dL) dividido por 100 no i -ésimo tempo para o sujeito j e t_{ij} é $(\text{tempo} - 5)/10$, com tempo medido em anos; idade_j é a idade no início do estudo (em anos); sexo_j é o indicador do sexo (0 = *feminino*, 1 = *masculino*). Assim, $\mathbf{x}_{ij} = (1, \text{idade}_j, \text{sexo}_j, t_{ij})^T$, $\mathbf{b}_j = (b_{oj}, b_{1j})^T$ e $\mathbf{Z}_j = (1, t_{ij})^T$. A Figura 3.2(a) apresenta o histograma dos níveis de colesterol, que claramente indica presença de assimetria e parece adequado considerar um modelo normal-assimétrico para analisar o conjunto de dados. Zhang & Davidian (2001) mediante um estudo informal, similar ao apresentado no capítulo seguinte, analisam estes dados e mostram que o comportamento assimétrico é parcialmente explicado pelas covariáveis apresentadas e que os efeitos aleatórios podem não ser normalmente distribuídos. Baseados nestas informações, três modelos estatísticos, diferindo na distribuição dos erros e efeitos aleatórios foram analisados. Estes modelos são:

Modelo 1: Um modelo com distribuição normal para os erros e distribuição normal-assimétrica multivariada para os efeitos aleatórios $\mathbf{b}_j$, com $\boldsymbol{\lambda}_b = (\lambda_{b1}, \lambda_{b2})^T$;

Modelo 2: Um modelo com distribuição normal-assimétrica multivariada para os erros $\boldsymbol{\epsilon}_j = (\epsilon_{1j}, \dots, \epsilon_{n_{jj}})^T$, com o mesmo parâmetro de assimetria (λ_e) entre grupos e distribuição normal multivariada para os efeitos aleatórios $\mathbf{b}_j$; e

Modelo 3: Um modelo puramente gaussiano.

Em todos os casos consideramos $\boldsymbol{\psi}_j = \sigma_e^2 \mathbf{I}_{n_j}$, $j = 1, \dots, 200$ (modelo de independência condicional). A Tabela 3.3 apresenta os resultados obtidos usando o algoritmo EM para os três modelos descritos acima. E.P. são os erros padrão assintóticos obtidos numericamente

Figura 3.2: (a) Histograma dos níveis de colesterol (mg/dL) para os 200 sujeitos do estudo em Framingham. (b) Gráfico de contornos da densidade estimada de b_j . (c) e (d) Densidades marginais estimadas para os componentes de b_j (linha cheia) e a densidade normal estimada (linha pontilhada).

via matriz hessiana. Quando consideramos o **Modelo 2** (erros com distribuição normal-assimétrica), a assimetria não é detectada e as estimativas dos parâmetros são próximas das obtidas sob a suposição de normalidade (**Modelo 3**). Os critérios AIC, BIC e HQ indicam que o **Modelo 1** apresenta o melhor ajuste, verificando a suposição de que os dados têm um comportamento assimétrico. A estimativas dos coeficientes β_1, β_2 e β_3 também diferem, refletindo o comportamento visto no estudo de simulação. As Figuras 3.2(b)-(d) mostram as densidades estimadas dos efeitos aleatórios, que são obviamente assimétricas, isto é mais evidente no gráfico de contornos mostrado na Figura 3.2(b). As figuras 3.2(c) e (d) apresentam as densidades marginais estimada de b_{0j} e b_{1j} , respectivamente. Note que a distribuição dos b_{1j} (inclinações) é aparentemente normal, enquanto que a forma da densidade dos interceptos (b_{0j}), claramente indica presença de assimetria, como Zhang & Davidian (2001) concluem usando uma representação semi-não-paramétrica (SNP).

Tabela 3.3: Resultados dos ajustes dos modelos 1, 2 e 3 para os dados de colesterol. d_{11} , d_{12} e d_{22} são os elementos distintos da matriz de variâncias-covariâncias $\mathbf{D}^{1/2}$

Parâmetros	Modelo 1		Modelo 2		Modelo 3	
	Estimativas	E.P	Estimativas	E.P.	Estimativas	E.P.
β_0	1.3755	0.1418	1.5955	0.1568	1.5968	0.1543
β_1	-0.0591	0.0534	-0.0630	0.0554	-0.0630	0.0568
β_2	0.0136	0.0034	0.0184	0.0035	0.0184	0.0037
β_3	0.2281	0.0511	0.2817	0.0240	0.2817	0.0242
σ_e^2	0.0434	0.0025	0.0434	0.0024	0.0434	0.0024
d_{11}	0.5600	0.0418	0.3716	0.0199	0.3715	0.0201
d_{12}	0.0700	0.0317	0.0563	0.0173	0.0563	0.0179
d_{22}	0.1924	0.0311	0.1868	0.0308	0.1868	0.0329
λ_{b1}	2.9947	0.7789	-	-	-	-
λ_{b2}	0.0000	0.4814	-	-	-	-
λ_e	-	-	0.0140	0.9890	-	-
- log-verossimilhança	-666.1717		-659.7560		-659.7560	
AIC	-0.6285		-0.6233		-0.6243	
BIC	-0.6057		-0.6020		-0.6053	
HQ	-0.6195		-0.6152		-0.6171	

Capítulo 4

Modelos Lineares Mistos Normal-Assimétrico: Enfoque Bayesiano

4.1 Introdução

Avanços em computação bayesiana e técnicas Monte Carlo Via Cadeia de Markov (MCMC) estão estendendo o alcance de modelos estatísticos que podem ser usados para ajustar dados de interesse prático. A principal motivação do presente capítulo vem desta observação. A idéia é propor uma versão de distribuição normal-assimétrica multivariada como alternativa à usada no Capítulo 3, que seja conveniente para implementação bayesiana em modelos lineares mistos normal-assimétrico (MLMSN), e conseqüentemente em outros tipos de modelos estatísticos.

Modelos lineares mistos são amplamente usados para análise de dados onde as respostas estão agrupadas, de modo que não é razoável assumir independência entre observações de um mesmo grupo. Por exemplo, considere medidas repetidas tomadas para cada sujeito em um estudo longitudinal, ou observações tomadas em membros de uma mesma família em um estudo genético. Na maioria das aplicações biológicas, físicas e econômicas se assume que a distribuição dos efeitos e erros aleatórios é gaussiana, mas esta suposição pode ser demasiadamente restritiva para fornecer uma representação adequada da estrutura que está presente nos dados, fazendo-a vulnerável à presença de assimetria. Alguns trabalhos usando distribuições simétricas e distribuições com caudas pesadas, tal como a

distribuição t de Student, foram sugeridos para estimação robusta em MLM (Pinheiro, et al., 2001). Liu (1996) discute a classe de distribuições robustas conhecidas como independentes/normais (Lange & Sinsheimer, 1993) e usa esta em modelos de regressão multivariada dentro de uma estrutura bayesiana. Rosa et al. (2003) abordam uma estrutura bayesiana para tratar inferências em um MLM com erros independentes/normais, esta classe inclui as versões univariada e multivariada da distribuições t de Student, slash e normal contaminada. Azzalini & Capitanio (1999), Azzalini & Kotz (2002), Capitanio et al. (2002), Sahu & Dey (2001), Sahu et al. (2003), entre outros, estudaram modelos de regressão com distribuições assimétricas, mas não existe literatura no contexto de modelos lineares com efeitos aleatórios normais assimétricos. Aqui, propomos uma extensão da versão típica do MLM assumindo que tanto os erros quanto os efeitos aleatórios seguem uma distribuição normal-assimétrica e usamos o esquema MCMC para tratar inferências em uma perspectiva bayesiana. A densidade normal-assimétrica usada nesta seção estende a densidade normal-assimétrica introduzida por Sahu et al. (2003).

O plano deste capítulo é como segue. Na Seção 4.2 apresentamos a versão univariada do estudo, considerando que tanto os erros como os efeitos aleatórios seguem uma distribuição normal-assimétrica de Azzalini (1985) vista em (2.5), os resultados e métodos propostos são ilustrados aplicando eles ao conjunto de dados de Barnett (veja, Capítulo 2). Na Seção 4.3 introduzimos uma nova classe de distribuição normal-assimétrica que é conveniente para análise bayesiana em MLMSN. Propriedades como momentos e a representação estocástica desta distribuição são discutidos. A densidade marginal do vetor de dados observáveis $\mathbf{Y}$ é também obtida em forma fechada, tal que inferência clássica neste tipo de modelos pode ser tratada usando técnicas de otimização usuais. Logo, fornecemos as distribuições condicionais completas que são necessárias para implementar o amostrador de Gibbs, fazendo da inferência bayesiana uma alternativa viável para tais modelos. Na Seção 4.4 apresentamos um estudo de simulação na qual fornecemos uma completa estratégia inferencial para análises de possíveis dados assimétricos, mostrando, por exemplo, que os critérios de informação usuais como AIC, BIC e HQ podem ser usados para detectar afastamentos da normalidade e escolher o modelo que melhor represente os dados sob investigação. Finalmente, uma análise bayesiana de um conjunto de dados longitudinais balanceados é apresentada na Seção 4.5.

4.2 O Modelo com Interceptos Aleatórios Normal-Assimétrico Bayesiano

Consideremos novamente o modelo definido em (3.1)-(3.2), isto é, o modelo linear

$$Y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + b_j + \epsilon_{ij}, \quad (4.1)$$

com

$$b_j \sim SN_1(0, \sigma_b^2, \lambda_b) \text{ e } \epsilon_{ij} \sim SN_1(0, \sigma_e^2, \lambda_e), \quad i = 1, \dots, n_j, \quad j = 1, \dots, m, \quad (4.2)$$

chamaremos a este modelo de MIASN por brevidade. Como discutido no Capítulo 3, este modelo pode ser escrito hierarquicamente como

$$\mathbf{Y}_j | \boldsymbol{\beta}, b_j, \sigma_e^2, \lambda_e \stackrel{\text{ind.}}{\sim} SNF_{n_j}(\mathbf{X}_j \boldsymbol{\beta} + \mathbf{1}_{n_j} b_j, \sigma_e^2 \mathbf{I}_{n_j}, \delta_e \mathbf{I}_{n_j}), \quad (4.3)$$

$$b_j | \sigma_b^2, \lambda_b \stackrel{\text{iid}}{\sim} SN_1(0, \sigma_b^2, \lambda_b), \quad j = 1, \dots, m, \quad (4.4)$$

onde $\delta_e = \lambda_e / (1 + \lambda_e^2)^{1/2}$, SNF denota a distribuição normal-assimétrica fundamental definida na Seção 2.3.5 e o interesse é fazer inferências sobre o vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma_e^2, \sigma_b^2)$ e $\boldsymbol{\lambda} = (\lambda_e, \lambda_b)^T$.

De (3.7), a densidade condicional (nos efeitos aleatórios $\mathbf{b} = (b_1, \dots, b_m)^T$) das quantidades observadas $\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_m^T)^T$ é

$$\begin{aligned} f(\mathbf{y} | \mathbf{b}, \boldsymbol{\beta}, \sigma_e^2, \lambda_e) &= \prod_{j=1}^m 2^{n_j} \phi_{n_j}(\mathbf{y}_j | \mathbf{X}_j \boldsymbol{\beta} + \mathbf{1}_{n_j} b_j, \sigma_e^2 \mathbf{I}_{n_j}) \Phi_{n_j} \left(\frac{\lambda_e}{\sigma_e} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta} - \mathbf{1}_{n_j} b_j) \right) \\ &= 2^N \phi_N(\mathbf{y} | \mathbf{X} \boldsymbol{\beta} + \mathbf{J} \mathbf{b}, \sigma_e^2 \mathbf{I}_N) \Phi_N \left(\frac{\lambda_e}{\sigma_e} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta} - \mathbf{J} \mathbf{b}) \right), \end{aligned} \quad (4.5)$$

onde $\mathbf{X} = (\mathbf{X}_1^T, \dots, \mathbf{X}_m^T)^T$, $\mathbf{J}$ é uma matriz bloco diagonal com elementos $\mathbf{1}_{n_1}, \dots, \mathbf{1}_{n_m}$ e $N = \sum_{j=1}^m n_j$. Como visto no Capítulo 3, procedimentos de inferência clássica são difíceis de tratar neste tipo de modelos quando tanto os erros quanto os efeitos aleatórios têm distribuição normal-assimétrica, pois a função de verossimilhança marginal depende da função $\Phi_{n_j+1}(\cdot)$ que envolve uma integral de dimensão $n_j + 1$. Uma alternativa é usar procedimentos bayesianos de estimação, discutidos a seguir.

4.2.1 Modelagem Bayesiana para o MIASN

Uma parte fundamental da análise bayesiana é especificar distribuições a priori para todos os parâmetros desconhecidos do modelo. A fim de garantir distribuições a posteriori próprias, adotamos distribuições a priori próprias para todas as quantidades desconhecidas do modelo. Assim, consideramos uma distribuição normal multivariada para o vetor de parâmetros β , com densidade

$$\beta \sim N_p(\beta_o, S_\beta). \quad (4.6)$$

Para o parâmetro de escala, σ_e^2 , consideramos a distribuição gamma inversa, $GI(\frac{\tau_e}{2}, \frac{T_e}{2})$, com densidade

$$\pi(\sigma_e^2 | \tau_e, T_e) = \frac{\left(\frac{T_e}{2}\right)^{\frac{\tau_e}{2}}}{\Gamma(\frac{\tau_e}{2})} \left(\frac{1}{\sigma_e^2}\right)^{\frac{\tau_e}{2}+1} \exp\left\{-\frac{T_e}{2\sigma_e^2}\right\}. \quad (4.7)$$

Para o parâmetro de escala do efeito aleatório consideramos também a distribuição $GI(\frac{\tau_b}{2}, \frac{T_b}{2})$, com densidade

$$\pi(\sigma_b^2 | \tau_b, T_b) = \frac{\left(\frac{T_b}{2}\right)^{\frac{\tau_b}{2}}}{\Gamma(\frac{\tau_b}{2})} \left(\frac{1}{\sigma_b^2}\right)^{\frac{\tau_b}{2}+1} \exp\left\{-\frac{T_b}{2\sigma_b^2}\right\}. \quad (4.8)$$

Para os parâmetros de assimetria λ_e e λ_b duas famílias de distribuições a priori interessantes podem ser assumidas, uma relacionada com a reparametrização usada na representação estocástica de uma variável aleatória normal-assimétrica, que depende de $-1 < \delta = \lambda/(1 + \lambda^2)^{1/2} < 1$ e a outra geralmente usada na literatura (Sahu et al., 2003). Assim, podemos considerar as seguintes famílias de distribuições a priori:

$$\delta_e = \lambda_e/(1 + \lambda_e^2)^{1/2} \sim U(0, 1) \text{ e } \delta_b = \lambda_b/(1 + \lambda_b^2)^{1/2} \sim U(0, 1), \quad (4.9)$$

onde $U(a, b)$ indica a distribuição uniforme no intervalo $[a, b]$ e

$$\lambda_e \sim N_1(\mu_e, \gamma_e^2) \mathbb{I}_{\{\lambda_e > 0\}} \text{ e } \lambda_b \sim N_1(\mu_b, \gamma_b^2) \mathbb{I}_{\{\lambda_b > 0\}}, \quad (4.10)$$

onde $\mathbb{I}_{\{A\}}$ é o indicador de A e todos os hiperparâmetros supõe-se conhecidos. O truncamento das densidades considera que a distribuição dos dados é assimétrica à direita (positiva). Uma representação similar pode ser adotada quando a assimetria dos dados é à esquerda (negativa).

Considerando (4.5) e (4.6)-(4.10), temos que a distribuição a posteriori conjunta de

todas as quantidades envolvidas é dada por

$$\begin{aligned}
\pi(\boldsymbol{\beta}, \mathbf{b}, \lambda_e, \lambda_b, \sigma_e^2, \sigma_b^2 \mid \mathbf{y}) &\propto \phi_n(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta} + \mathbf{J}\mathbf{b}, \sigma_e^2 \mathbf{I}_n) \Phi_n\left(\frac{\lambda_e}{\sigma_e}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{J}\mathbf{b})\right) \\
&\times \phi_m(\mathbf{b} \mid \mathbf{0}, \sigma_b^2 \mathbf{I}_m) \Phi_m\left(\frac{\lambda_b}{\sigma_b} \mathbf{b}\right) \\
&\times \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \mathbf{S}_\beta^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right\} \\
&\times (\sigma_e^2)^{-(\frac{T_e}{2}+1)} \exp\left\{-\frac{T_e}{2\sigma_e^2}\right\} (\sigma_b^2)^{-(\frac{T_b}{2}+1)} \exp\left\{-\frac{T_b}{2\sigma_b^2}\right\} \\
&\times \pi(\lambda_e, \lambda_b),
\end{aligned} \tag{4.11}$$

em que

$$\pi(\lambda_e, \lambda_b) \propto \exp\left\{-\frac{1}{2\gamma_e^2}(\lambda_e - \mu_e)^2\right\} \exp\left\{-\frac{1}{2\gamma_b^2}(\lambda_b - \mu_b)^2\right\} \mathbb{I}_{\{\lambda_e > 0\}} \mathbb{I}_{\{\lambda_b > 0\}},$$

ou

$$\pi(\lambda_e, \lambda_b) \propto \mathbb{I}_{\{0 < \delta_e < 1\}} \mathbb{I}_{\{0 < \delta_b < 1\}},$$

dependendo do tipo de distribuição a priori a ser considerada e supondo independência entre λ_e e λ_b . Note em (4.11) que é complicado obter analiticamente a distribuição a posteriori marginal das quantidades envolvidas de interesse. Assim, para conduzir uma análise bayesiana precisamos implementar o esquema MCMC. A seguir apresentamos os passos necessários à implementação do amostrador de Gibbs, que é um caso especial do MCMC e necessita da especificação das distribuições a posteriori condicionais para cada parâmetro. Como veremos, para implementar este procedimento precisamos também recorrer ao algoritmo de Metropolis-Hasting (Chib & Greenberg, 1995).

4.2.2 O Esquema MCMC

A fim de especificar o modelo (4.3)-(4.4) em uma estrutura conveniente para implementar o procedimento MCMC, usamos as representações estocásticas dadas em (2.4) e (2.37), tal que estas distribuições assimétricas podem ser representadas hierarquicamente como segue:

$$\begin{aligned}
\mathbf{Y}_j \mid b_j, \boldsymbol{\beta}, \sigma_e^2, \lambda_e, \mathbf{w}_{ej} &\stackrel{\text{ind.}}{\sim} N_{n_j}(\mathbf{X}_j \boldsymbol{\beta} + \mathbf{1}_{n_j} b_j + \sigma_e \delta_e \mathbf{w}_{ej}, \sigma_e^2 (1 - \delta_e^2) \mathbf{I}_{n_j}), \\
\mathbf{W}_{ej} &\stackrel{\text{ind.}}{\sim} N_{n_j}(0, \mathbf{I}_{n_j}) \mathbb{I}_{\{\mathbf{w}_{ej} > 0\}} \text{ e}
\end{aligned}$$

$$\begin{aligned}
b_j | \sigma_b^2, \lambda_b, w_{bj} &\stackrel{\text{ind.}}{\sim} N_1(\sigma_b \delta_b w_{bj}, \sigma_b^2(1 - \delta_b^2)), \\
W_{bj} &\stackrel{\text{ind.}}{\sim} HN_1(0, 1), \quad j = 1, \dots, m.
\end{aligned}$$

A especificação do modelo completo, dependendo das distribuições a priori dos parâmetros de assimetria considerados, é então dada como segue:

Considerando distribuições a priori uniformes para os parâmetros de assimetria

$$\begin{aligned}
\mathbf{Y}_j | b_j, \beta, \sigma_e^2, \lambda_e, \mathbf{w}_{ej} &\sim N_{n_j}(\mathbf{X}_j \beta + \mathbf{1}_j b_j + \sigma_e \delta_e \mathbf{w}_{ej}, \sigma_e^2(1 - \delta_e^2) \mathbf{I}_{n_j}), \\
\mathbf{W}_{ej} &\sim N_{n_j}(0, \mathbf{I}) \mathbb{I}_{\{\mathbf{w}_{ej} > 0\}}, \\
b_j | \sigma_e, \lambda_b, w_{bj} &\sim N_1(\sigma_b \delta_b w_{bj}, \sigma_b^2(1 - \delta_b^2)), \\
W_{bj} &\sim HN_1(0, 1), \\
\beta &\sim N_p(\beta_0, \mathbf{S}_\beta), \\
\sigma_e^2 &\sim IG\left(\frac{\tau_e}{2}, \frac{T_e}{2}\right), \\
\sigma_b^2 &\sim IG\left(\frac{\tau_b}{2}, \frac{T_b}{2}\right), \\
\delta_e = \lambda_e / (1 + \lambda_e^2)^{1/2} &\sim U(0, 1), \\
\delta_b = \lambda_b / (1 + \lambda_b^2)^{1/2} &\sim U(0, 1).
\end{aligned} \tag{4.12}$$

Considerando distribuições a priori normais para os parâmetros de assimetria

$$\begin{aligned}
\mathbf{Y}_j | b_j, \beta, \sigma_e^2, \lambda_e, \mathbf{w}_{ej} &\sim N_{n_j}(\mathbf{X}_j \beta + \mathbf{1}_j b_j + \sigma_e \delta_e \mathbf{w}_{ej}, \sigma_e^2(1 - \delta_e^2) \mathbf{I}_{n_j}), \\
\mathbf{W}_{ej} &\sim N_{n_j}(0, \mathbf{I}) \mathbb{I}_{\{\mathbf{w}_{ej} > 0\}}, \\
b_j | \sigma_e, \lambda_b, w_{bj} &\sim N_1(\sigma_b \delta_b w_{bj}, \sigma_b^2(1 - \delta_b^2)), \\
W_{bj} &\sim HN_1(0, 1), \\
\beta &\sim N_p(\beta_0, \mathbf{S}_\beta), \\
\sigma_e^2 &\sim IG\left(\frac{\tau_e}{2}, \frac{T_e}{2}\right), \\
\sigma_b^2 &\sim IG\left(\frac{\tau_b}{2}, \frac{T_b}{2}\right), \\
\lambda_e &\sim N_1(\mu_e, \gamma_e^2) \mathbb{I}_{\{\lambda_e > 0\}}, \\
\lambda_b &\sim N_1(\mu_b, \gamma_b^2) \mathbb{I}_{\{\lambda_b > 0\}}.
\end{aligned} \tag{4.13}$$

As representações dadas em (4.12) e (4.13) são importantes porque permitem traduzir facilmente para o código BUGS (usado pelo software WinBUGS). As distribuições condicionais completas, necessárias para implementar o amostrador de Gibbs em ambos os casos, são obtidas de forma direta e são dadas nas seguintes proposições.

Proposição 4.1. *Sob o modelo completo em (4.12), as distribuições condicionais completas de β , b_j , $\mathbf{W}_{ej}$, W_{bj} , σ_e^2 , σ_b^2 , δ_e e δ_b , $j = 1, \dots, m$, são dadas por*

$$\beta | \dots \sim N_p(\mathbf{A}_\beta^{-1} \mathbf{a}_\beta, \mathbf{A}_\beta^{-1}),$$

$$\text{onde } \mathbf{A}_\beta = \mathbf{S}_\beta^{-1} + \sum_{j=1}^m \frac{\mathbf{X}_j^T \mathbf{X}_j}{\sigma_e^2(1-\delta_e^2)} \text{ e } \mathbf{a}_\beta = \mathbf{S}_\beta^{-1} \beta_0 + \frac{1}{\sigma_e^2(1-\delta_e^2)} \sum_{j=1}^m \mathbf{X}_j^T (\mathbf{y}_j - \mathbf{1}_j b_j - \sigma_e \delta_e \mathbf{w}_{ej});$$

$$b_j | \dots \sim N_1(A_{bj}^{-1} a_j, A_{bj}^{-1}),$$

$$\text{onde } A_{bj} = \frac{n_j}{\sigma_e^2(1-\delta_e^2)} + \frac{1}{\sigma_b^2(1-\delta_b^2)} \text{ e } a_j = \frac{\mathbf{1}_j^T (\mathbf{y}_j - \mathbf{X}_j \beta - \sigma_e \delta_e \mathbf{w}_{ej})}{\sigma_e^2(1-\delta_e^2)} + \frac{\delta_b}{\sigma_b(1-\delta_b^2)} w_{bj}, j = 1, \dots, m;$$

$$\mathbf{W}_{ej} | \dots \sim N_{n_j}(\mathbf{A}_{w_{ej}}^{-1} \mathbf{a}_{w_{ej}}, \mathbf{A}_{w_{ej}}^{-1}) \mathbb{I}_{\{\mathbf{w}_{ej} > 0\}},$$

$$W_{bj} | \dots \sim N_1(A_{wbj}^{-1} a_{wbj}, A_{wbj}^{-1}) \mathbb{I}_{\{w_{bj} > 0\}},$$

$$\text{onde } \mathbf{A}_{w_{ej}} = \left(\frac{\delta_e^2}{1-\delta_e^2} + 1\right) \mathbf{I}_{n_j}, \quad \mathbf{a}_{w_{ej}} = \frac{\delta_e}{\sigma_e(1-\delta_e^2)} (\mathbf{y}_j - \mathbf{X}_j \beta - \mathbf{1}_j b_j), \quad A_{wbj} = \left(1 + \frac{\delta_b^2}{1-\delta_b^2}\right) \text{ e } a_{wbj} = \frac{\delta_b}{\sigma_b(1-\delta_b^2)} b_j, \quad j = 1, \dots, m;$$

$$\sigma_e^2 | \dots \propto \frac{1}{(\sigma_e^2)^{\frac{N+\tau_e}{2}+1}} \exp\left\{-\frac{1}{2\sigma_e^2} [T_e + \frac{1}{1-\delta_e^2} \sum_{j=1}^m [\mu_{j\sigma e}^T \mu_{j\sigma e} - 2\sigma_e \delta_e \mu_{j\sigma e}^T \mathbf{w}_{ej}]]\right\}$$

$$\text{onde } N = \sum_{j=1}^m n_j \text{ e } \mu_{j\sigma e} = \mathbf{y}_j - \mathbf{X}_j \beta - \mathbf{1}_{n_j} b_j;$$

$$\sigma_b^2 | \dots \propto \frac{1}{(\sigma_b^2)^{\frac{m+\tau_b}{2}+1}} \exp\left\{-\frac{1}{2\sigma_b^2} [T_b + \frac{1}{1-\delta_b^2} \sum_{j=1}^m (b_j^2 - 2\sigma_b \delta_b w_{bj} b_j)]\right\}$$

$$\delta_e | \dots \propto \frac{1}{(1-\delta_e^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma_e^2(1-\delta_e^2)} \sum_{j=1}^m (\mu_{j\delta e} - \sigma_e \delta_e \mathbf{w}_{ej})^T (\mu_{j\delta e} - \sigma_e \delta_e \mathbf{w}_{ej})\right\} \mathbb{I}_{\{0 < \delta_e < 1\}}$$

$$\text{onde } \mu_{j\delta e} = \mathbf{y}_j - \mathbf{X}_j \beta - \mathbf{1}_{n_j} b_j;$$

$$\delta_b | \dots \propto \frac{1}{(1-\delta_b^2)^{m/2}} \exp\left\{-\frac{1}{2\sigma_b^2(1-\delta_b^2)} \sum_{j=1}^m (b_j - \sigma_b \delta_b w_{bj})^2\right\} \mathbb{I}_{\{0 < \delta_b < 1\}}$$

A notação " $\beta|...$ " é usada para denotar a distribuição condicional de β dados todos os outros parâmetros e de modo similar para as outras distribuições condicionais.

Proposição 4.2. *Sob o modelo completo em (4.13), as distribuições condicionais completas para β , b_j , $\mathbf{W}_{ej}$, W_{bj} , σ_e^2 , σ_b^2 , λ_e e λ_b , $j = 1, \dots, m$, são dadas por*

$$\beta|... \sim N_p(\mathbf{A}_\beta^{-1}\mathbf{a}_\beta, \mathbf{A}_\beta^{-1}),$$

$$\text{onde } \mathbf{A}_\beta = \mathbf{S}_\beta^{-1} + \sum_{j=1}^m \frac{\mathbf{X}_j^T \mathbf{X}_j}{\sigma_e^2(1-\delta_e^2)} \quad e \quad \mathbf{a}_\beta = \mathbf{S}_\beta^{-1}\beta_0 + \frac{1}{\sigma_e^2(1-\delta_e^2)} \sum_{j=1}^m \mathbf{X}_j^T (\mathbf{y}_j - \mathbf{1}_j b_j - \sigma_e \delta_e \mathbf{w}_{ej});$$

$$b_j|... \sim N_1(A_{bj}^{-1}a_j, A_{bj}^{-1}),$$

$$\text{onde } A_{bj} = \frac{n_j}{\sigma_e^2(1-\delta_e^2)} + \frac{1}{\sigma_b^2(1-\delta_b^2)} \quad e \quad a_j = \frac{\mathbf{1}_j^T (\mathbf{y}_j - \mathbf{X}_j \beta - \sigma_e \delta_e \mathbf{w}_{ej})}{\sigma_e^2(1-\delta_e^2)} + \frac{\delta_b}{\sigma_b(1-\delta_b^2)} w_{bj}, \quad j = 1, \dots, m;$$

$$\mathbf{W}_{ej}|... \sim N_{n_j}(\mathbf{A}_{w_{ej}}^{-1} \mathbf{a}_{w_{ej}}, \mathbf{A}_{w_{ej}}^{-1}) \mathbb{I}_{\{\mathbf{w}_{ej} > 0\}},$$

$$W_{bj}|... \sim N_1(A_{wbj}^{-1} a_{wbj}, A_{wbj}^{-1}) \mathbb{I}_{\{w_{bj} > 0\}},$$

$$\text{onde } \mathbf{A}_{w_{ej}} = (\frac{\delta_e^2}{1-\delta_e^2} + 1)\mathbf{I}_{n_j}, \quad \mathbf{a}_{w_{ej}} = \frac{\delta_e}{\sigma_e(1-\delta_e^2)} (\mathbf{y}_j - \mathbf{X}_j \beta - \mathbf{1}_j b_j), \quad A_{wbj} = (1 + \frac{\delta_b^2}{1-\delta_b^2}) \quad e \\ a_{wbj} = \frac{\delta_b}{\sigma_b(1-\delta_b^2)} b_j, \quad j = 1, \dots, m;$$

$$\sigma_e^2|... \propto \frac{1}{(\sigma_e^2)^{\frac{N+\tau_e}{2}+1}} \exp\{-\frac{1}{2\sigma_e^2}[T_e + \frac{1}{1-\delta_e^2} \sum_{j=1}^m [\mu_{j\sigma e}^T \mu_{j\sigma e} - 2\sigma_e \delta_e \mu_{j\sigma e}^T \mathbf{w}_{ej}]]\}$$

$$\text{onde } N = \sum_{j=1}^m n_j \quad e \quad \mu_{j\sigma e} = \mathbf{y}_j - \mathbf{X}_j \beta - \mathbf{1}_{n_j} b_j;$$

$$\sigma_b^2|... \propto \frac{1}{(\sigma_b^2)^{\frac{m+\tau_b}{2}+1}} \exp\{-\frac{1}{2\sigma_b^2}[T_b + \frac{1}{1-\delta_b^2} \sum_{j=1}^m (b_j^2 - 2\sigma_b \delta_b w_{bj} b_j)]\}$$

$$\lambda_e|... \propto (1 + \lambda_e^2)^{n/2} \exp\{-\frac{(1 + \lambda_e^2)}{2\sigma_e^2} \sum_{j=1}^m (\mu_{j\lambda e} - \frac{\sigma_e \lambda_e}{(1 + \lambda_e^2)^{1/2}} \mathbf{w}_{ej})^T (\mu_{j\lambda e} - \frac{\sigma_e \lambda_e}{(1 + \lambda_e^2)^{1/2}} \mathbf{w}_{ej}) \\ - \frac{(\lambda_e^2 - \mu_e)^2}{2\gamma_e^2}\} \mathbb{I}_{\{\lambda_e > 0\}}$$

onde $\mu_{j\lambda_e} = y_j - \mathbf{X}_j\beta - \mathbf{1}_{n_j}b_j$;

$$\lambda_b | \dots \propto (1 + \lambda_b^2)^{m/2} \exp\left\{-\frac{(1 + \lambda_b^2)}{2\sigma_b^2} \sum_{j=1}^m \left(b_j - \frac{\sigma_b \lambda_b}{(1 + \lambda_b^2)^{1/2}} w_{bj}\right)^2 - \frac{(\lambda_b - \mu_b)^2}{2\gamma_b^2}\right\} \mathbb{I}_{\{\lambda_b > 0\}}$$

As provas das proposições acima são diretamente derivadas trabalhando sobre a densidade a posteriori conjunta em (4.11) e usando repetidamente o Lema 2.2. Note que algumas das condicionais completas não podem ser amostradas diretamente, por exemplo, a distribuição a posteriori condicional completa para o parâmetro σ_e^2 requer algoritmos, tal como o Metropolis-Hasting (veja, Chib & Greenberg, 1995).

4.2.3 Aplicação

Consideremos novamente o conjunto de dados de Barnett (Barnett, 1969) usado como ilustração no Capítulo 2. Desta vez propomos usar o seguinte modelo de interceptos aleatórios:

$$Y_{ij} = \beta_i + b_j + \epsilon_{ij}, \quad (4.14)$$

para $i = 1, \dots, 4$ e $j = 1, \dots, 72$. Cinco modelos estatísticos diferindo nas distribuições a priori adotadas, na distribuição dos erros aleatórios e dos efeitos aleatórios foram testados. Estes modelos foram:

Modelo 1: Um modelo com distribuição normal-assimétrica para os erros e normal-assimétrica para os efeitos aleatórios com distribuições a priori normais truncadas para os parâmetros de assimetria.

Modelo 2: Um modelo com distribuição normal para os erros e normal-assimétrica para os efeitos aleatórios com distribuição a priori normal truncada para o parâmetro de assimetria.

Modelo 3: Um modelo com distribuição normal-assimétrica para os erros e normal-assimétrica para os efeitos aleatórios com distribuições a priori uniformes para os parâmetros de assimetria.

Modelo 4: Um modelo com distribuição normal para os erros e normal-assimétrica para os efeitos aleatórios com distribuição a priori uniforme para o parâmetro de assimetria.

Modelo 5: Um modelo puramente gaussiano.

O conjunto de dados foi ajustado de acordo com as estratégias apresentada nas proposições 4.1 e 4.2, e usando o software WinBUGS. As seguintes especificações relativas às distribuições a priori foram adotadas:

$$\beta_i \sim N_1(0, 10^2), i = 1, \dots, 4, \sigma_e^2 \sim GI(0.01, 0.01), \sigma_b^2 \sim GI(0.01, 0.01),$$

$$\lambda_e \sim N_1(0, 10^2)\mathbb{I}_{\lambda_e > 0}, \lambda_b \sim N_1(0, 10^2)\mathbb{I}_{\lambda_b > 0},$$

e adicionalmente as respectivas distribuições a priori uniformes para λ_e e λ_b discutidas na seção anterior. Usamos 50000 iterações descartando as primeiras 20000 iterações para fazer inferências. Para evitar problemas de correlação nas cadeias geradas, o valor do *lag* considerado foi igual a 6. Assim, o tamanho efetivo das cadeias analisadas foi de 5000. O código BUGS para o ajuste do **Modelo 1** encontra-se no Apêndice C. O código para os outros modelos é direto bastando modificar algumas linhas do código fornecido.

Os resultados dos ajustes são dados nas Tabelas 4.1 e 4.2. Para comparar os cinco modelos informalmente calculamos o número efetivo de parâmetros ρ_D 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 software WinBUGS pode ser usado para comparar modelos complexos, embora esteja sendo duramente criticado e não seja tão eficiente para detectar afastamentos da normalidade de acordo com as simulações feitas na Seção 4.4. Usando os valores do DIC mostrados na parte inferior das tabelas, vemos que os modelos assimétricos melhoram o correspondente modelo normal. Notamos que os modelos assimétricos 1 e 3 são os que melhor se ajustam ao comportamento dos dados. Os resultados usando distribuições a priori uniformes e normais são bem similares especialmente nos β_i , $i = 1, \dots, 4$, que são de interesse principal. Esta observação, deve-se a que estamos usando distribuições a priori com grande variância e esperasse que os resultados sejam robustos respeito da a priori considerada. Os históricos das cadeias de Markov geradas para todos os parâmetros do **Modelo 1** são apresentados na Figura 4.1. Note-se que a convergência é alcançada rapidamente.

Figura 4.1: Histórico das cadeias de Markov geradas sob o Modelo 1 para todos os parâmetros

Tabela 4.1: Médias a posteriori, desvios padrão (D.P.) e intervalos de credibilidade de 95% para os parâmetros dos modelos 1 e 2 ajustados aos dados de Barnett. EMV são as estimativas de máxima verossimilhança usando o procedimento da Seção 3.2.1.

Parâmetro	Distribuições a priori normais para os parâmetros de assimetria									
	Modelo 1					Modelo 2				
	Média	D.P.	2.5%	97.5%	Média	D.P.	2.5%	97.5%	Estimativa	D.P.
β_1	10.260	1.452	7.652	13.380	11.210	1.117	9.270	13.640	12.331	0.869
β_2	9.627	1.407	7.138	12.670	10.510	1.120	8.572	12.930	11.627	0.945
β_3	9.312	1.441	6.702	12.430	10.240	1.118	8.287	12.600	11.356	0.870
β_4	8.852	1.431	6.270	9.780	11.960	1.118	7.839	12.190	10.892	0.814
σ_c^2	5.597	1.679	3.575	9.894	3.943	0.380	3.257	4.737	3.845	0.999
σ_b^2	192.60	43.56	120.60	290.90	198.70	44.80	126.00	302.90	166.59	49.930
λ_b	8.806	5.324	2.469	22.510	8.927	5.0560	2.815	21.470	4.607	4.005
λ_c	0.893	0.717	0.047	2.561	-	-	-	-	-	-
DIC	1730.940					1786.260				
ρ_D	100.250					74.844				

Tabela 4.2: Médias a posteriori, desvios padrão (D.P.) e intervalos de credibilidade de 95% para os parâmetros dos modelos 3, 4 e 5 ajustados aos dados de Barnett.

Distribuições a priori uniformes para os parâmetros de assimetria												
Parâmetro	Modelo 3				Modelo 4				Modelo 5			
	Média	D.P.	2.5%	97.5%	Média	D.P.	2.5%	97.5%	Média	D.P.	2.5%	97.5%
β_1	11.880	1.512	8.825	15.310	12.240	1.329	9.631	14.820	21.680	1.004	19.700	23.620
β_2	11.210	1.503	8.231	14.640	11.540	1.330	8.932	14.130	20.980	1.007	18.980	22.930
β_3	10.910	1.504	7.893	14.360	11.270	1.333	8.692	13.890	20.720	1.007	18.710	22.660
β_4	10.440	1.513	7.423	13.910	10.810	1.330	8.194	13.440	20.250	1.004	18.260	22.230
σ_e^2	4.836	1.097	3.461	7.692	3.932	0.387	3.233	4.758	3.934	0.390	3.222	4.757
σ_b^2	162.50	39.05	97.87	250.00	170.80	40.04	106.40	260.90	67.35	11.67	48.18	93.50
λ_b	4.202	2.751	1.348	11.110	0.959	0.037	0.856	0.998	-	-	-	-
λ_e	0.573	0.424	0.025	1.567	-	-	-	-	-	-	-	-
DIC	1737.680				1784.700				1784.850			
ρ_D	70.732				75.285				75.196			

4.3 O Modelo Linear Misto Normal-Assimétrico Bayesiano

Antes de definir o modelo linear misto normal-assimétrico (MLMSN) que será usado nesta parte do trabalho, apresentamos uma breve discussão de uma nova distribuição normal-assimétrica multivariada que basicamente é um caso particular da densidade apresentada em (2.48) e que em certo modo generaliza a distribuição normal-assimétrica introduzida por Sahu et al. (2003). Esta nova distribuição normal assimétrica é conveniente para análise de qualquer modelo de uma perspectiva bayesiana.

Consideremos a densidade em (2.48). A seguir consideramos um caso especial desta densidade, com $k = n$ e $\Delta = \text{diag}(\delta)$, $\delta = (\delta_1, \dots, \delta_n)^T \in \mathbb{R}^n$ (mas Σ não é diagonal como é implicitamente assumido por Sahu et al., 2003) e em tal caso escrevemos $\Delta\Delta^T = \Delta^2$. Assim, para $\delta = 0$ recuperamos a distribuição normal original; para valores positivos de δ obtemos uma distribuição assimétrica positiva (à direita) e para valores negativos obtemos uma distribuição assimétrica negativa (à esquerda). Aqui, denotaremos esta distribuição por SN^* .

Muitas propriedades probabilísticas desta distribuição podem ser derivadas dos resultados apresentados em Arellano-Valle e Genton (2003) (veja Capítulo 2). Daqui segue, por exemplo, a representação estocástica, dada a seguir, de um vetor aleatório normal-assimétrico (veja também Sahu et al., 2003).

Proposição 4.3. *Seja $Y \sim SN_n^*(\mu, \Sigma, \Delta)$. Então*

$$Y \stackrel{d}{=} \Delta|X_0| + X_1, \quad (4.15)$$

onde $X_0 \sim N_n(0, I_n)$, $X_1 \sim N_n(\mu, \Sigma)$ e " $\stackrel{d}{=}$ " significa "distribuído como", com X_0 e X_1 independentes.

Demonstração. Seja $U = \Delta|X_0| + X_1$. Como $U | T = t \sim N_n(\Delta t + \mu, \Sigma)$, onde $T = |X_0| \sim HN_n(0, I_n)$, então pelo Lema 2.2 e o fato de que $(I_n + \Delta\Sigma^{-1}\Delta)^{-1}\Delta\Sigma^{-1} =$

$\Delta(\Sigma + \Delta^2)^{-1}$, segue que

$$\begin{aligned}
f_U(\mathbf{w}) &= \int_{\mathbb{R}_+^n} \phi_n(\mathbf{w}|\Delta \mathbf{t} + \boldsymbol{\mu}, \Sigma) 2^n \phi_n(\mathbf{t}) d\mathbf{t} \\
&= 2^n \int_{\mathbb{R}_+^n} \phi_n(\mathbf{w}|\boldsymbol{\mu}, \Sigma + \Delta^2) \\
&\quad \times \phi_n(\mathbf{t} | (\mathbf{I}_n + \Delta \Sigma^{-1} \Delta)^{-1} \Delta \Sigma^{-1} (\mathbf{w} - \boldsymbol{\mu}), (\mathbf{I}_n + \Delta \Sigma^{-1} \Delta)^{-1}) d\mathbf{t} \\
&= 2^n \phi_n(\mathbf{w}|\boldsymbol{\mu}, \Sigma + \Delta^2) \Phi_n(\Delta(\Sigma + \Delta^2)^{-1}(\mathbf{w} - \boldsymbol{\mu})|\mathbf{0}, (\mathbf{I}_n + \Delta \Sigma^{-1} \Delta)^{-1}),
\end{aligned}$$

Ou seja, $\mathbf{U} \stackrel{d}{=} \mathbf{Y} \sim SN_n^*(\boldsymbol{\mu}, \Sigma, \Delta)$, que conclui a prova. $\square$

Uma vez que assumimos que Δ é diagonal, notamos de (4.15) que a introdução de assimetria não afeta a estrutura de correlação do modelo normal original (veja também Sahu et al., 2003). Os valores das correlações são afetados, mas a estrutura permanece a mesma. Contudo, isto não é verdade para as distribuições usadas no Capítulo 3 e também para a classe de distribuições normal-assimétrica multivariada de Azzalini & Dalla-Valle (1996), pois nestes casos os parâmetros de assimetria afetam a estrutura de correlação. Na Figura 4.2 apresentamos alguns contornos da densidade associada à distribuição normal-assimétrica bivariada $SN_2^*(\mathbf{0}, \Sigma, \text{diag}(\boldsymbol{\delta}))$ para diferentes valores de Σ e $\boldsymbol{\delta}$. Note que estes contornos não são elípticos e de fato são diferentes dos encontrados em Sahu et al. (2003).

Uma consequência direta da Proposição 4.3, relacionada com os momentos do vetor aleatório normal-assimétrico (SN^*), é dada a seguir.

Corolário 4.1. *Seja $\mathbf{Y} \sim SN_n^*(\boldsymbol{\mu}, \Sigma, \Delta)$. Então,*

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}} \boldsymbol{\delta} \text{ and } V[\mathbf{Y}] = \Sigma + \left(1 - \frac{2}{\pi}\right) \Delta^2.$$

O MLMSN que abordamos nesta parte do trabalho estende o MLM tradicional definido em (1.1)-(1.2), considerando a seguinte relação linear

$$\mathbf{Y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \mathbf{b}_j + \boldsymbol{\epsilon}_j,$$

com as suposições de que

$$\mathbf{b}_j \stackrel{\text{iid}}{\sim} SN_q^*(\mathbf{0}, \mathbf{D}, \Delta_b) \text{ e } \boldsymbol{\epsilon}_j \stackrel{\text{ind.}}{\sim} SN_{n_j}^*(\mathbf{0}, \boldsymbol{\psi}_j, \Delta_{\epsilon_j}), \quad j = 1, \dots, m.$$

Figura 4.2: Contornos da distribuição normal-assimétrica bivariada em (2.48), com $\boldsymbol{\mu} = (0, 0)^T$,

$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ e $\boldsymbol{\delta} = (\delta_1, \delta_2)^T$ para diferentes valores de δ_1, δ_2 e ρ . ($\boldsymbol{\Delta}$ é diagonal)

(a) Para $\boldsymbol{\delta} = (0, 3)^T$ e $\rho = 0, 0,5$ e $0,9$, respectivamente.

(b) Para $\boldsymbol{\delta} = (2, 3)^T$ e $\rho = 0, 0,5$ e $0,9$, respectivamente.

(c) Para $\boldsymbol{\delta} = (-2, 2)^T$ e $\rho = 0, 0,5$ e $0,9$, respectivamente.

ambos independentes, as quais levam à seguinte representação hierárquica do modelo

$$\mathbf{Y}_j | \mathbf{b}_j, \boldsymbol{\beta}, \boldsymbol{\psi}_j, \boldsymbol{\Delta}_{ej} \stackrel{\text{ind.}}{\sim} SN_{n_j}^*(\mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \mathbf{b}_j, \boldsymbol{\psi}_j, \boldsymbol{\Delta}_{ej}), \quad (4.16)$$

$$\mathbf{b}_j | \mathbf{D}, \boldsymbol{\Delta}_b \stackrel{\text{ind.}}{\sim} SN_q^*(\mathbf{0}, \mathbf{D}, \boldsymbol{\Delta}_b), \quad j = 1, \dots, m, \quad (4.17)$$

onde $\boldsymbol{\Delta}_{ej}$ e $\boldsymbol{\Delta}_b$ são matrizes diagonais com elementos $\delta_{e_{j1}}, \dots, \delta_{e_{jn_j}}$ e $\delta_1, \dots, \delta_q$, respectivamente, correspondentes aos parâmetros de assimetria. Dependendo do contexto várias suposições podem ser feitas sobre o parâmetro $\boldsymbol{\psi}_j$. A seguir, assumiremos que $\boldsymbol{\psi}_j = \sigma_\epsilon^2 \mathbf{I}_{n_j}$

e $\Delta_{ej} = \delta_e \mathbf{I}_{n_j}$, $j = 1, \dots, m$. Em tal caso, temos do Corolário 4.1 que

$$E[\mathbf{Y}_j | \mathbf{b}_j, \boldsymbol{\beta}, \sigma_e^2, \delta_e] = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \mathbf{b}_j + \delta_e \left(\frac{2}{\pi} \right)^{1/2} \mathbf{1}_{n_j},$$

e

$$\text{Var}[\mathbf{Y}_j | \mathbf{b}_j, \boldsymbol{\beta}, \sigma_e^2, \delta_e] = \left(\sigma_e^2 + \delta_e^2 \left(1 - \frac{2}{\pi} \right) \right) \mathbf{I}_{n_j}, \quad j = 1, \dots, m,$$

onde $\mathbf{1}_k$ denota o vetor k -dimensional de uns. Além disso, a densidade condicional (nos efeitos aleatórios $\mathbf{b} = (\mathbf{b}_1^T, \dots, \mathbf{b}_m^T)^T$) das variáveis observáveis $\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_m^T)^T$ é dada por

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y} | \mathbf{b}, \boldsymbol{\beta}, \sigma_e^2, \delta_e) &= \prod_{j=1}^m 2^{n_j} \phi_{n_j}(\mathbf{y}_j | \mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \mathbf{b}_j, (\delta_e^2 + \sigma_e^2) \mathbf{I}_{n_j}) \\ &\quad \times \Phi_{n_j} \left(\delta_e (\delta_e^2 + \sigma_e^2)^{-1} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta} - \mathbf{Z}_j \mathbf{b}_j) | \mathbf{0}, \frac{\sigma_e^2}{\delta_e^2 + \sigma_e^2} \mathbf{I}_{n_j} \right) \\ &= 2^N \phi_N(\mathbf{y} | \mathbf{X} \boldsymbol{\beta} + \mathbf{Z} \mathbf{b}, (\delta_e^2 + \sigma_e^2) \mathbf{I}_N) \\ &\quad \times \Phi_N \left(\frac{\delta_e}{\sigma_e} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}) | \mathbf{0}, (\sigma_e^2 + \delta_e^2) \mathbf{I}_N \right), \end{aligned} \quad (4.18)$$

onde $\mathbf{X} = (\mathbf{X}_1^T, \dots, \mathbf{X}_m^T)^T$, $\mathbf{Z}$ é uma matriz bloco diagonal com elementos $\mathbf{Z}_1, \dots, \mathbf{Z}_m$ e $N = \sum_{j=1}^m n_j$. Neste caso, o principal interesse é fazer inferência sobre o vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma_e^2, \boldsymbol{\alpha}^T)^T$ e $\boldsymbol{\lambda} = (\delta_e, \boldsymbol{\delta}_b^T)^T$, com $\boldsymbol{\delta}_b = (\delta_1, \dots, \delta_q)^T$ (veja notação no Capítulo 1).

A densidade marginal de $\mathbf{Y}_j$ é obtida do seguinte teorema. A prova é dada no Apêndice C.

Teorema 4.1. *Seja $\mathbf{Y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \mathbf{b}_j + \boldsymbol{\epsilon}_j$, onde $\mathbf{b}_j \sim SN_q^*(\mathbf{0}, \mathbf{D}, \Delta_b)$ e $\boldsymbol{\epsilon}_j \sim SN_{n_j}^*(\mathbf{0}, \sigma_e^2 \mathbf{I}_{n_j}, \delta_e \mathbf{I}_{n_j})$ são independentes. Então, a densidade marginal de $\mathbf{Y}_j$ é dada por*

$$f_{\mathbf{Y}_j}(\mathbf{y}_j | \boldsymbol{\theta}, \boldsymbol{\lambda}) = 2^{n_j+q} \phi_{n_j}(\mathbf{y}_j | \mathbf{X}_j \boldsymbol{\beta}, \Sigma_j) \Phi_{n_j+q}(\boldsymbol{\mu}_{2j} - \Gamma_j \boldsymbol{\mu}_{1j} | \mathbf{0}, \mathbf{R}_j + \Gamma_j \Lambda_j \Gamma_j^T), \quad (4.19)$$

onde

$$\Sigma_j = (\delta_e^2 + \sigma_e^2) \mathbf{I}_{n_j} + \mathbf{Z}_j (\Delta_b^2 + \mathbf{D}) \mathbf{Z}_j^T, \quad \boldsymbol{\mu}_{1j} = \frac{\Lambda_j \mathbf{Z}_j^T (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})}{\delta_e^2 + \sigma_e^2},$$

$$\begin{aligned}\boldsymbol{\mu}_{2j} &= \begin{pmatrix} \frac{\delta_e}{\sqrt{\sigma_e^2(\delta_e^2 + \sigma_e^2)}}(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta}) \\ \mathbf{0}_q \end{pmatrix}, \quad \boldsymbol{\Gamma}_j = \begin{pmatrix} \frac{\delta_e}{\sqrt{\sigma_e^2(\delta_e^2 + \sigma_e^2)}}\mathbf{Z}_j \\ -\boldsymbol{\Delta}_b(\boldsymbol{\Delta}_b^2 + \mathbf{D})^{-1} \end{pmatrix}, \\ \mathbf{R}_j &= \begin{pmatrix} \mathbf{I}_{n_j} & \mathbf{0} \\ \mathbf{0} & (\mathbf{I}_q + \boldsymbol{\Delta}_b\mathbf{D}^{-1}\boldsymbol{\Delta}_b)^{-1} \end{pmatrix}, \quad \boldsymbol{\Lambda}_j = \left((\boldsymbol{\Delta}_b^2 + \mathbf{D})^{-1} + \frac{\mathbf{Z}_j^T\mathbf{Z}_j}{\delta_e^2 + \sigma_e^2} \right)^{-1}, \quad j = 1, \dots, m.\end{aligned}\tag{4.20}$$

Assim, denotando a função log-verossimilhança para $\boldsymbol{\theta}$ e $\boldsymbol{\lambda}$ dada a amostra observada $\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_m^T)^T$ por $\ell(\boldsymbol{\theta}, \boldsymbol{\lambda}|\mathbf{y})$, esta pode ser escrita como

$$\begin{aligned}\ell(\boldsymbol{\theta}, \boldsymbol{\lambda}|\mathbf{y}) &\propto -\frac{1}{2} \sum_{j=1}^m \log |\boldsymbol{\Sigma}_j| - \frac{1}{2} \sum_{j=1}^m \{(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta})^T \boldsymbol{\Sigma}_j^{-1} (\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta})\} \\ &\quad + \sum_{j=1}^m \log \Phi_{n_j+q}(\boldsymbol{\mu}_{2j} - \boldsymbol{\Gamma}_j\boldsymbol{\mu}_{1j}|\mathbf{0}, \mathbf{R}_j + \boldsymbol{\Gamma}_j\boldsymbol{\Lambda}_j\boldsymbol{\Gamma}_j^T),\end{aligned}\tag{4.21}$$

onde $\boldsymbol{\mu}_{1j}$, $\boldsymbol{\mu}_{2j}$, $\boldsymbol{\Sigma}_j$, $\boldsymbol{\Gamma}_j$, $\boldsymbol{\Lambda}_j$ e $\mathbf{R}_j$ estão definidos em (4.20).

O resultado apresentado no Teorema 4.1 é importante porque permite encontrar os valores dos critérios de informação usuais (AIC, BIC e HQ) que podem ser usados para detectar afastamentos da normalidade, como será visto no estudo de simulação apresentado na Seção 4.4. Além disso, de modo similar aos resultados do Capítulo 3, a forma fechada desta densidade permite tratar métodos clássicos usando técnicas padrão de otimização, por exemplo, no software R, podemos usar a livreria *mvtnorm*, que trata a distribuição normal multivariada, conjuntamente com a rotina *optim*.

4.3.1 Modelagem Bayesiana para o MLMSN

Considere a notação da Seção 4.2.1. A fim de garantir distribuições a posteriori próprias, consideramos que $\boldsymbol{\beta} \sim N_p(\boldsymbol{\beta}_0, \mathbf{S}_\beta)$, $\sigma_e^2 \sim GI(\frac{\tau_e}{2}, \frac{T_e}{2})$ (gama inversa). A família de distribuições a priori para a matriz de escala dos efeitos aleatórios é escolhida principalmente por facilidade computacional. Assim, usamos a distribuição Wishart inversa, $IW_{\tau_b}(\mathbf{T}_b)$, (Odell & Fieveson, 1966), com densidade

$$\begin{aligned}\pi(\mathbf{D}|\tau_b, \mathbf{T}_b) &= \left(2^{\frac{\tau_b q}{2}} \pi^{\frac{q(q-1)}{4}} \prod_{i=1}^q \Gamma\left(\frac{\tau_b + 1 - i}{2}\right) \right)^{-1} |\mathbf{T}_b|^{-\frac{\tau_b}{2}} |\mathbf{D}|^{-\frac{1}{2}(\tau_b + q + 1)} \\ &\quad \times \exp \left\{ -\frac{1}{2} \text{tr}(\mathbf{T}_b^{-1} \mathbf{D}^{-1}) \right\},\end{aligned}\tag{4.22}$$

onde τ_b é um inteiro e é referido ser o parâmetro de forma, correspondente aos graus de liberdade da correspondente variável aleatória Wishart e representa o grau de incerteza da distribuição a priori. A matriz $\mathbf{T}_b$ descreve a estrutura de variâncias-covariâncias das variáveis, e q é o número de variáveis aleatórias correlacionadas. A média de $\mathbf{D}$ é $\mathbf{T}_b^{-1}/(\tau_b - q - 1)$ (Tanner, 1996).

Consideramos que os parâmetros de assimetria δ_e e $\delta_b = (\delta_1, \dots, \delta_q)^T$ têm uma distribuição normal truncada, com densidades

$$\pi(\delta_e | \mu_e, \gamma_e^2) \propto \exp \left\{ -\frac{1}{2\gamma_e^2} (\delta_e - \mu_e)^2 \right\} \mathbb{I}_{\{\delta_e > 0\}}, \quad (4.23)$$

e

$$\pi(\delta_b | \mu_b, \gamma_b) \propto \exp \left\{ -\frac{1}{2} (\delta_b - \mu_b)^T \gamma_b^{-1} (\delta_b - \mu_b) \right\} \mathbb{I}_{\{\delta_b > 0\}}, \quad (4.24)$$

onde γ_b é uma matriz definida positiva e $\mathbb{I}_A$ é o indicador de A . Distribuições a priori similares para δ_e e δ_b podem ser adotadas no caso em que temos assimetria negativa.

De (4.18) e das distribuições a priori definidas acima, temos que a densidade a posteriori conjunta de todas as quantidades envolvidas é dada por

$$\begin{aligned} \pi(\beta, \mathbf{b}, \delta_e, \delta_b, \sigma_e^2, \mathbf{D} \mid \mathbf{y}) &\propto \phi_n(\mathbf{y} | \mathbf{X}\beta + \mathbf{Z}\mathbf{b}, (\delta_e^2 + \sigma_e^2)\mathbf{I}_n) \\ &\times \Phi_n \left(\frac{\delta_e}{\sigma_e} (\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{b}) | \mathbf{0}, (\sigma_e^2 + \delta_e^2)\mathbf{I}_n \right) \\ &\times \prod_{j=1}^m [\phi_q(\mathbf{b}_j | \mathbf{0}, \mathbf{D} + \Delta_b^2) \\ &\times \Phi_q(\Delta_b(\Delta_b^2 + \mathbf{D})^{-1}\mathbf{b}_j | \mathbf{0}, (\mathbf{I}_q + \Delta_b\mathbf{D}^{-1}\Delta_b)^{-1})] \\ &\times \exp \left\{ -\frac{1}{2} (\beta - \beta_0) \mathbf{S}_\beta^{-1} (\beta - \beta_0) \right\} \\ &\times |\mathbf{D}|^{-\frac{1}{2}(\tau_b + q + 1)} \exp \left\{ -\frac{1}{2} \text{tr}(\mathbf{T}_b^{-1}\mathbf{D}^{-1}) \right\} (\sigma_e^2)^{-(\frac{\tau_e}{2} + 1)} \\ &\times \exp \left\{ -\frac{T_e}{2\sigma_e^2} \right\} \exp \left\{ -\frac{1}{2\gamma_e^2} (\delta_e - \mu_e)^2 \right\} \mathbb{I}_{\{\delta_e > 0\}} \\ &\times \exp \left\{ -\frac{1}{2} (\delta_b - \mu_b)^T \gamma_b^{-1} (\delta_b - \mu_b) \right\} \mathbb{I}_{\{\delta_b > 0\}}, \end{aligned} \quad (4.25)$$

onde todos os hiperparâmetros são especificados (conhecidos). Assim, para implementar o

amostrador de Gibbs necessitamos obter as distribuições a posteriori condicionais. Estas serão tratadas na seção seguinte.

4.3.2 O Esquema MCMC

A fim de especificar o modelo (4.16)-(4.17) numa estrutura conveniente para implementar o MCMC, usamos a representação estocástica em (4.15), de modo que as distribuições normal-assimétrica definidas aqui podem ser representadas hierarquicamente como:

$$\begin{aligned} \mathbf{Y}_j | \mathbf{b}_j, \boldsymbol{\beta}, \sigma_e^2, \delta_e, \mathbf{w}_{ej} &\stackrel{\text{ind.}}{\sim} N_{n_j}(\mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \mathbf{b}_j + \delta_e \mathbf{w}_{ej}, \sigma_e^2 \mathbf{I}_{n_j}), \\ \mathbf{W}_{ej} &\stackrel{\text{ind.}}{\sim} N_{n_j}(0, \mathbf{I}_{n_j}) \mathbb{I}_{\{\mathbf{w}_{ej} > 0\}} \quad \text{e} \end{aligned}$$

$$\begin{aligned} \mathbf{b}_j | \mathbf{D}, \boldsymbol{\Delta}_b, \mathbf{w}_{bj} &\stackrel{\text{ind.}}{\sim} N_q(\boldsymbol{\Delta}_b \mathbf{w}_{bj}, \mathbf{D}), \\ \mathbf{W}_{bj} &\stackrel{\text{ind.}}{\sim} N_q(0, \mathbf{I}) \mathbb{I}_{\{\mathbf{w}_{bj} > 0\}}, \end{aligned}$$

$j = 1, \dots, m$.

O esquema MCMC, tal como o amostrador de Gibbs, é facilitado usando a distribuição a posteriori aumentada, resultante da mistura de modelos. A especificação do modelo completo é dada como segue:

$$\begin{aligned} \mathbf{Y}_j | \mathbf{b}_j, \boldsymbol{\beta}, \sigma_e^2, \delta_e, \mathbf{w}_{ej} &\sim N_{n_j}(\mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \mathbf{b}_j + \delta_e \mathbf{w}_{ej}, \sigma_e^2 \mathbf{I}_{n_j}), \\ \mathbf{W}_{ej} &\sim N_{n_j}(0, \mathbf{I}) \mathbb{I}_{\{\mathbf{w}_{ej} > 0\}}, \\ \mathbf{b}_j | \mathbf{D}, \boldsymbol{\Delta}_b, \mathbf{w}_{bj} &\sim N_q(\boldsymbol{\Delta}_b \mathbf{w}_{bj}, \mathbf{D}), \\ \mathbf{W}_{bj} &\sim N_q(0, \mathbf{I}) \mathbb{I}_{\{\mathbf{w}_{bj} > 0\}}, \\ \boldsymbol{\beta} &\sim N_p(\boldsymbol{\beta}_0, \mathbf{S}_\beta), \\ \sigma_e^2 &\sim IG\left(\frac{\tau_e}{2}, \frac{T_e}{2}\right), \\ \mathbf{D} &\sim IW_{\tau_b}(\mathbf{T}_b), \\ \delta_e &\sim N(\mu_e, \gamma_e) \mathbb{I}_{\{\delta_e > 0\}}, \\ \boldsymbol{\delta}_b &\sim N_q(\boldsymbol{\mu}_b, \boldsymbol{\gamma}_b) \mathbb{I}_{\{\boldsymbol{\delta}_b > 0\}}. \end{aligned} \tag{4.26}$$

Esta representação é importante porque permite escrever facilmente o código BUGS. As distribuições condicionais completas necessárias para implementar o amostrador de Gibbs

são derivadas diretamente da especificação acima. Estas são dadas na seguinte Proposição.

Proposição 4.4. *Sob o modelo completo em (4.26), as distribuições a posteriori completas de β , $\mathbf{b}_j$, $\mathbf{W}_{ej}$, $\mathbf{W}_{bj}$, σ_e^2 , $\mathbf{D}$, δ_e e δ_b , $j = 1, \dots, m$, são dadas por*

$$\beta | \dots \sim N_p(\mathbf{A}_\beta^{-1} \mathbf{a}_\beta, \mathbf{A}_\beta^{-1}),$$

$$\text{onde } \mathbf{A}_\beta = \mathbf{S}_\beta^{-1} + \sum_{j=1}^m \frac{\mathbf{x}_j^T \mathbf{x}_j}{\sigma_e^2} \text{ e } \mathbf{a}_\beta = \mathbf{S}_\beta^{-1} \beta_0 + \sum_{j=1}^m \frac{\mathbf{x}_j^T (\mathbf{y}_j - \mathbf{z}_j \mathbf{b}_j - \delta_e \mathbf{w}_{ej})}{\sigma_e^2};$$

$$\mathbf{b}_j | \dots \sim N_q(\mathbf{A}_{bj}^{-1} \mathbf{a}_j, \mathbf{A}_{bj}^{-1}),$$

$$\text{onde } \mathbf{A}_{bj} = \frac{\mathbf{z}_j^T \mathbf{z}_j}{\sigma_e^2} + \mathbf{D}^{-1} \text{ e } \mathbf{a}_j = \frac{\mathbf{z}_j (\mathbf{y}_j - \mathbf{x}_j \beta - \delta_e \mathbf{w}_{ej})}{\sigma_e^2} + \mathbf{D}^{-1} \Delta_b \mathbf{w}_{bj}, \quad j = 1, \dots, m;$$

$$\mathbf{W}_{ej} | \dots \sim N_{n_j}(\mathbf{A}_{w_{ej}}^{-1} \mathbf{a}_{w_{ej}}, \mathbf{A}_{w_{ej}}^{-1}) \mathbb{I}_{\{\mathbf{w}_{ej} > 0\}},$$

$$\mathbf{W}_{bj} | \dots \sim N_q(\mathbf{A}_{w_{bj}}^{-1} \mathbf{a}_{w_{bj}}, \mathbf{A}_{w_{bj}}^{-1}) \mathbb{I}_{\{\mathbf{w}_{bj} > 0\}},$$

$$\text{onde } \mathbf{A}_{w_{ej}} = (\frac{\delta_e^2}{\sigma_e^2} + 1) \mathbf{I}_{n_j}, \quad \mathbf{a}_{w_{ej}} = \frac{\delta_e}{\sigma_e^2} (\mathbf{y}_j - \mathbf{x}_j \beta - \mathbf{z}_j \mathbf{b}_j), \quad \mathbf{A}_{w_{bj}} = (\Delta_b \mathbf{D}^{-1} \Delta_b + \mathbf{I}_q) \\ \text{e } \mathbf{a}_{w_{bj}} = \Delta_b \mathbf{D}^{-1} \mathbf{b}_j, \quad j = 1, \dots, m;$$

$$\sigma_e^2 | \dots \sim IG(\frac{N + \tau_e}{2}, \frac{T_e + \sum_{j=1}^m \mu_{j\sigma}^T \mu_{j\sigma}}{2}),$$

$$\text{onde } N = \sum_{j=1}^m n_j \text{ e } \mu_{j\sigma} = \mathbf{y}_j - \mathbf{x}_j \beta - \mathbf{z}_j \mathbf{b}_j - \delta_e \mathbf{w}_{ej};$$

$$\mathbf{D} | \dots \sim IW_q(\tau_b + m, T_b^{-1} + \sum_{j=1}^m \mu_{jD}^T \mu_{jD}),$$

$$\text{onde } \mu_{jD} = \mathbf{b}_j - \Delta_b \mathbf{w}_{jb};$$

$$\delta_e | \dots \sim N(\mu_{\delta_e}, \sigma_{\delta_e}^2) \mathbb{I}_{\{\delta_e > 0\}},$$

$$\text{onde } \mu_{\delta_e} = \sigma_{\delta_e}^2 [\frac{\mu_e}{\gamma_e^2} + \sum_{j=1}^m \frac{\mathbf{w}_{ej}^T (\mathbf{y}_j - \mathbf{x}_j \beta - \mathbf{z}_j \mathbf{b}_j)}{\sigma_e^2}] \text{ e } \sigma_{\delta_e}^2 = (\frac{1}{\gamma_e^2} + \sum_{j=1}^m \frac{\mathbf{w}_{ej}^T \mathbf{w}_{ej}}{\sigma_e^2})^{-1};$$

$$\delta_b | \dots \sim N_q(\mathbf{A}_{\delta_b}^{-1} \mathbf{a}_{\delta_b}, \mathbf{A}_{\delta_b}^{-1}) \mathbb{I}_{\{\delta_b > 0\}},$$

com $\mathbf{A}_{\delta_b} = \gamma_b^{-1} + \sum_{j=1}^m [\text{diag}(\mathbf{w}_{bj}) \mathbf{D}^{-1} \text{diag}(\mathbf{w}_{bj})]$, $\mathbf{a}_{\delta_b} = \gamma_b^{-1} \mu_b + \sum_{j=1}^m \text{diag}(\mathbf{w}_{bj}) \times \mathbf{D}^{-1} \mathbf{b}_j$ e $\text{diag}(\mathbf{a})$ é uma matriz diagonal com elementos diagonais sendo as componentes de $\mathbf{a}$.

A notação " $\beta|...$ " é usada para denotar a distribuição condicional de β dados todos os outros parâmetros e de modo similar para as outras distribuições condicionais. O algoritmo requer valores iniciais para todas as variáveis do modelo e as iterações geram amostras das distribuições condicionais acima até alcançar convergência, que pode ser verificada, por exemplo, usando as metodologias desenvolvidas por Gelman & Rubin (1992). Para dados desbalanceados este método pode ser facilmente implementado usando linguagens de programação, tais como R, Ox, Matlab. Para dados balanceados o amostrador de Gibbs pode ser implementado usando o programa WinBUGS.

similar aos procedimentos do Capítulo 3, propomos avaliar a série de ajustes inspecionando os critérios de informação tais como AIC ($-\ell(\hat{\theta}, \hat{\lambda}|\mathbf{y})/N + P/N$), BIC ($-\ell(\hat{\theta}, \hat{\lambda}|\mathbf{y})/N + 0.5 \log(N)P/N$) e HQ ($-\ell(\hat{\theta}, \hat{\lambda}|\mathbf{y})/N + \log(\log(N))P/N$), onde P é o número de parâmetros livres no modelo $N = \sum_{j=1}^m n_j$ e $(\hat{\theta}, \hat{\lambda})$ é o valor de (θ, λ) na convergência, que pode ser a média ou mediana da cadeia gerada. Aqui, $\ell(\hat{\theta}, \hat{\lambda}|\mathbf{y})$ pode ser avaliada diretamente usando (4.21). Chamamos atenção para o fato de que de acordo com simulações conduzidas, o critério DIC (Spiegelhalter et al., 2002) falha em detectar afastamentos da normalidade e não será considerado aqui. Portanto, ao menos neste caso, o DIC não é recomendado para selecionar MLM com comportamento assimétricos.

4.4 Estudo de Simulação

Para avaliar a performance do modelo e dos métodos propostos, conduzimos dois estudos de simulação. O modelo linear misto usado é similar ao apresentado na simulação do capítulo anterior e é dado por

$$Y_{ij} = \alpha + t_{ij}\beta_1 + w_j\beta_2 + b_j + e_{ij}, \quad (4.27)$$

onde para $i = 1, \dots, 5$, $t_{ij} = i - 3$, $\beta_1 = 2$, $\beta_2 = 1$, e $e_{ij} \sim N(0, 0.5^2)$, isto é, $\delta_e = 0$. Em cada simulação, 100 conjuntos de dados foram simulados de (4.27) segundo especificações descritas abaixo e (4.27) foi ajustado segundo a estratégia proposta na Proposição 4.4 usando o programa R juntamente com o pacote R2WinBUGS (Gelman, 2004) e as seguintes especificações relativas às distribuições a priori:

$$\begin{aligned} \alpha &\sim N_1(0, 10^2), \quad \beta_1 \sim N_1(0, 10^2), \quad \beta_2 \sim N_1(0, 10^2), \quad \sigma_e^2 \sim IG(0.01, 0.01), \\ \sigma_b^2 &\sim IG(0.01, 0.01), \quad \delta_b \sim N_1(0, 10^2) \mathbb{I}_{\{\delta_b > 0\}}. \end{aligned}$$

Tabela 4.3: Resultados de Monte Carlo baseados em 100 conjuntos de dados com a distribuição $Gama(4, 1)$ para os efeitos aleatórios. Média MC, D.P. MC e Mediana MC são as médias Monte Carlo das médias, desvios padrão e medianas, respectivamente, das cadeias geradas para estimar os parâmetros; 5% q. e 95% q. são o quantil da ordem 5 e 95 para a mediana a posteriori das cadeias geradas. (*) indica que a densidade da variável resposta é a SN^* definida em (2.48) com $k = n = 1$.

Parâmetros	Média MC	D.P. MC	Mediana MC	5% q.	95% q.
(a) Cenário normal-assimétrico (*)					
α	1.5345	0.2124	1.5328	1.1386	1.8876
β_1 (2)	2.0002	0.0116	2.0001	1.9820	2.0160
β_2 (1)	0.9985	0.2195	0.9986	0.6013	1.3284
σ_e^2 (0.25)	0.2502	0.0126	0.2498	0.2289	0.2686
σ_b^2	0.4553	0.2329	0.4158	0.1049	0.7586
δ_b	3.1347	0.2474	3.1312	2.7497	3.5187
$E[\alpha + b]$ (4)	4.0356	-	4.0311	3.7704	4.3252
$Var[\alpha + b]$ (4)	4.0501	-	4.0023	3.3442	4.7976
(b) Cénario normal					
α	3.9430	0.2040	3.9431	3.6237	4.2224
β_1 (2)	2.0002	0.0112	2.0002	1.9820	2.0160
β_2 (1)	1.0152	0.2812	1.0159	0.5120	1.5092
σ_e^2 (0.25)	0.2501	0.0124	0.2496	0.2287	0.2693
σ_b^2	4.0229	0.4124	3.9965	3.2560	4.9749
$E[\alpha + b]$ (4)	3.9430	-	3.9431	3.6237	4.2224
$Var[\alpha + b]$ (4)	4.0229	-	3.9965	3.2560	4.9749

Usamos 14000 iterações depois de descartar as 6000 primeiras para fazer inferências. Para evitar problemas de correlação na cadeia gerada o valor do *lag* foi igual a 7, isto é, o tamanho efetivo da cadeia analisada foi igual a 2000.

Uma vantagem da representação normal-assimétrica é sua flexibilidade para modelar assimetria. Em uma primeira simulação geramos $\alpha + b_j$ de uma distribuição $Gama(4, 1)$ com fdp $(1/6)x^3 \exp(-x)$, produzindo alta assimetria como sugerido pela Figura 4.3. Sob estas especificações, $E[\alpha + b_j] = 4$ e $Var[\alpha + b_j] = 4$. Note que t_{ij} representa uma co-variável com valores mudando nas unidades de cada grupo e fixa para todos os grupos, enquanto w_j pode ser visto como um indicador de tratamento. Tomamos $m = 200$ com $w_j = 1$ se $j \leq 100$ e $w_j = 0$ se $j > 100$. Para cada um dos 100 conjuntos de dados gerados, (4.27) foi ajustado duas vezes, uma com densidade para b_j representada pela distribuição

Tabela 4.4: Resultados de Monte Carlo baseados em 100 conjuntos de dados simulados com distribuição $N_1(0, 4)$ para os efeitos aleatórios. As entradas são como na Tabela 4.3.

Parâmetros	Média MC	D.P. MC	Mediana MC	5% q.	95% q.
(a) Cenário normal-assimétrico (*)					
α (4)	3.1370	0.4294	3.1445	2.2542	3.8490
β_1 (2)	2.0011	0.0112	2.0011	1.9820	2.0172
β_2 (1)	1.0442	0.2844	1.0441	0.6398	1.5169
σ_e^2 (0.25)	0.2512	0.0126	0.2508	0.2341	0.2748
σ_b^2	3.4223	0.5613	3.4256	2.3365	4.3111
δ_b	1.0890	0.5037	1.0679	0.4446	2.2057
$E[\alpha + b]$ (4)	4.0060	-	3.9965	3.6694	4.2978
$Var[\alpha + b]$ (4)	3.9998	-	3.9545	3.385	4.5175
(b) Cénario normal					
α	4.0131	0.1797	4.013	3.6885	4.3269
β_1 (2)	2.0008	0.0111	2.001	1.9819	2.0172
β_2 (1)	1.0315	0.2574	1.018	0.6266	1.4711
σ_e^2 (0.25)	0.2502	0.0122	0.250	0.2335	0.2736
σ_b^2	3.9931	0.4058	3.967	3.3796	4.5502
$E[\alpha + b]$ (4)	4.0131	-	4.013	3.6885	4.3269
$Var[\alpha + b]$ (4)	3.9931	-	3.967	3.3796	4.5502

normal-assimétrica (SN^*) e a outra representada pela distribuição normal. Para avaliar o objetivo uso dos critérios de informação, os modelos preferidos pelo AIC, BIC e HQ foi registrado.

Para os resultados do amostrador de Gibbs, nenhum dos critérios (AIC, BIC, HQ) selecionou a especificação normal em nenhum dos 100 conjuntos de dados, demonstrando a habilidade destes critérios de seleção em detectar afastamentos da normalidade e sugerindo forte evidência de assimetria. A Tabela 4.3 apresenta os resultados numéricos para as estimativas onde normalidade e assimetria foram consideradas (SN^*). A maior parte da estimativas dos parâmetros é consistente. Note que há perda de eficiência na estimação de β_2 quando normalidade é assumida (alto D.P. MC) comparando-se com a distribuição normal-assimétrica. Uma vez que um dos principais interesses da análise pode ser a avaliação do efeito do tratamento, este estudo sugere que adotar a distribuição normal para a variável resposta pode levar a inferências ineficientes sobre os efeitos fixos. Note também que há perda de eficiência na estimação de $E[\alpha + b_j]$. Assim, embora estimação consistente seja possível sob normalidade, esta falha por tomar em conta as ver-

dadeiras características dos efeitos aleatórios levando a inferências menos precisas sobre os parâmetros ou quantidades de interesse.

A vantagem em estimar a densidade dos efeitos aleatórios pode ser apreciada na Figura 4.3. A Figura 4.3(b) apresenta as médias de Monte Carlo das densidades a posteriori estimadas a partir dos 100 conjuntos de dados, comparando-as com a verdadeira densidade (gama), o ajuste normal e o ajuste sob a distribuição normal-assimétrica. Esta figura mostra que a flexibilidade adicional propiciada pela distribuição normal-assimétrica é suficiente para captar as verdadeiras características dos efeitos aleatórios. Esta observação é enriquecida pela Figura 4.3(a), onde mostramos as 100 densidades a posteriori estimadas sob distribuição normal-assimétrica.

Figura 4.3: Resultados de simulação baseados em 100 conjuntos de dados gerados com distribuição gama para os efeitos aleatórios e medianas a posteriori estimadas. (a) Densidades a posteriori estimadas para os 100 ajustes normal-assimétrica (SN^*). (b) Verdadeira densidade (linha cheia) e média de Monte Carlo das densidades a posteriori estimadas para os 100 conjuntos de dados usando distribuição normal (linha pontilhada) e usando distribuição normal-assimétrica (linha tracejada)

O segundo estudo de simulação foi idêntico ao primeiro, exceto que a distribuição de $\alpha + b_j$ foi $N(4, 4)$, isto é, $\alpha = 4$ e $b_j \sim N(0, 4)$. Aqui não existe necessidade de maior

Figura 4.4: Resultados de simulação baseados em 100 conjuntos de dados gerados com distribuição normal para os efeitos aleatórios e medianas a posteriori estimadas. (a) Densidades a posteriori estimadas para os 100 ajustes normal-assimétrica. (b) Verdadeira densidade (linha cheia) e média de Monte Carlo das densidades a posteriori estimadas para os 100 conjuntos de dados: usando distribuição normal (linha pontilhada) e usando distribuição normal-assimétrica (linha tracejada)

flexibilidade na distribuição dos efeitos aleatórios, e esperamos que os métodos propostos identifiquem isso. Os critérios AIC, BIC e HQ selecionam corretamente o modelo normal em 97%, 100% e 96% das vezes, respectivamente. Os resultados da simulação Monte Carlo são mostrados na Tabela 4.4, que indica não haver perda de eficiência na estimação quando usamos a distribuição normal-assimétrica (SN^*). Todas as densidades a posteriori estimadas são essencialmente normais como mostrado na Figura 4.4. Resultados similares podem ser encontrados no Capítulo 3, em Hu, Tsiatis, & Davidian (1998) e em Zhang & Davidian (2001), estes usando máxima verossimilhança.

4.5 Aplicação

Nesta seção ilustramos os métodos propostos baseados na distribuição normal-assimétrica SN^* , aplicando-os novamente ao conjunto de dados de colesterol (*Framingham cholesterol data*). O estudo em Framingham é talvez um dos mais utilizados para identificar a relação entre vários fatores de risco e ocorrência de doenças, e também para caracterizar a história

natural do processo de doenças circulatórias crônicas. Os dados em vários aspectos foram e continuam sendo coletados a cada dois anos num grupo de indivíduos. Iniciou-se em 1948 em Framingham, localizada a 21 milhas de Boston (EUA) com metas limitadas a fatores de risco em doenças coronárias como estudos dos níveis de colesterol, hábitos de fumo e pressão sanguínea como . Nestes últimos anos as metas estão sendo estendidas para entender numerosos fatores etiológicos de várias doenças.

A fim de implementar o amostrador de Gibbs usando o programa WinBUGS, o conjunto de "dados balanceados" usado nesta seção foi obtida como uma parte do conjunto de dados usados por Zhang e Davidian (2001) (veja também aplicação no Capítulo 3). O conjunto de dados inclui níveis de colesterol sobre o tempo para 133 indivíduos, idade ao início do estudo e sexo. A Figura 4.5 mostra os níveis de colesterol ao longo do tempo e o histograma dos níveis de colesterol para os indivíduos selecionados. A Figura 4.5(a) sugere

Figura 4.5: Dados de colesterol. (a) Níveis de colesterol para 133 indivíduos do estudo em Framingham com trajetórias de 5 indivíduos selecionados aleatoriamente. (b) Histograma dos níveis de colesterol para os 133 indivíduos.

que o nível de colesterol cresce com o tempo para a maioria de sujeitos e com marcante variação entre sujeitos. O histograma da Figura 4.5(b) mostra uma ligeira assimetria na distribuição dos dados. O índice de assimetria para os níveis de colesterol é 0.66. Com o intuito de investigar empiricamente as causas da assimetria, ajustamos os perfis individuais usando uma regressão simples sobre o tempo para cada sujeito. A Figura 4.6 mostra o histograma dos interceptos e inclinações individuais estimados por mínimos quadrados. A Figura 4.6 sugere que, embora as estimativas das inclinações aparentem ser normalmente

distribuídas, a variação nos interceptos pode não ser normalmente distribuída. Baseados

Figura 4.6: Ajuste dos perfis individuais usando uma regressão linear simples sobre o tempo. (a) Histograma dos interceptos individuais estimados (b) Histograma das inclinações individuais estimadas.

nestas observações, consideramos o seguinte modelo linear misto:

$$y_{ij} = \beta_o + \beta_1 \text{sexo}_j + \beta_2 \text{idade}_j + b_{oj} + b_{1j} t_{ij} + \epsilon_{ij}, \quad (4.28)$$

$j = 1, \dots, 133$ e $i = 1, \dots, 6$, onde y_{ij} , t_{ij} , idade_j e sexo_j como na Seção 4.5. Três modelos estatísticos, diferindo na distribuição dos erros e efeitos aleatórios, foram analisados. Estes modelos foram:

Modelo 1: Um modelo com distribuição normal-assimétrica (SN^*) para os erros $\epsilon_j = (\epsilon_{1j}, \dots, \epsilon_{6j})^T$ e distribuição normal-assimétrica multivariada (SN^*) para os efeitos aleatórios;

Modelo 2: Um modelo com distribuição normal multivariada para os erros $\epsilon_j = (\epsilon_{1j}, \dots, \epsilon_{6j})^T$ e distribuição normal-assimétrica multivariada (SN^*) para os efeitos aleatórios;

Modelo 3: Um modelo puramente gaussiano.

Distribuições a priori para os parâmetros envolvidos foram as mesmas para os três modelos para fins de comparação. Os efeitos fixos $(\beta_o, \beta_1, \beta_2)$ foram especificados com distribuições independentes $N(0, 10^2)$. Para os parâmetros de escala da distribuição dos

erros tomamos a distribuição $\sigma_e^2 \sim IG(0.01, 0.01)$. Igualmente, os hiperparâmetros para a distribuição a priori da matriz de escala dos efeitos aleatórios (D_{11} , D_{12} , D_{22}) foram tomados como sendo $\tau_b = 2$ e $\mathbf{T}_b = \mathbf{I}_2$. Finalmente, para os parâmetros de assimetria (δ_e , δ_{b1} , δ_{b2}), atribuímos distribuições normais truncadas com média 0 e variância 10^2 . Assim, esperamos que os resultados obtidos sejam um tanto robustos, pois as distribuições a priori selecionadas têm variâncias grandes.

O amostrador de Gibbs foi implementado usando o programa WinBUGS. O código BUGS usado para ajustar o **Modelo 1** é fornecido no Apêndice C. Usamos 25000 iterações, descartando as primeiras 10000 iterações para fazer inferências a posteriori. Para evitar problemas de correlação nas cadeias geradas, o valor do *lag* usado foi igual a 5. Os resultados para os parâmetros estimados (média a posteriori) estão na Tabela 4.5. Note que os valores dos critérios de informação AIC, BIC e HQ mostrados na parte inferior da Tabela 4.5 favorecem os modelos normal-assimétrico, justificando a suposição de que os dados têm um comportamento assimétrico. Analisando esta tabela, os critérios AIC e HQ indicam que o **Modelo 1** apresenta o melhor ajuste para os dados. Na Figura 4.7 apresentamos gráficos Q-Q dos resíduos sob o **Modelo 1** (melhor ajuste) e **Modelo 3** (normalidade). Os dois gráficos são semelhantes e parecem indicar presença de *outliers*. O histórico das cadeias de Markov para os parâmetros sob o **Modelo 1** é apresentado na Figura 4.8. Note que visualmente não existe problemas de convergência nas cadeias geradas.

Figura 4.7: Gráficos Q-Q dos resíduos dos ajustes sob os (a) **modelo 1** e (b) **modelo 3** com envelope gerado.

Tabela 4.5: Médias a posteriori, desvios padrão (D.P.) e intervalos de credibilidade de 95% para os parâmetros sob os três modelos ajustados para o conjunto de dados de colesterol.

Parâmetros	Modelo 1					Modelo 2					Modelo 3					
	Média	D.P.	2.5%	97.5%	Média	D.P.	2.5%	97.5%	Média	D.P.	2.5%	97.5%	Média	D.P.	2.5%	97.5%
β_o	1.287	0.199	0.911	1.682	1.396	0.187	1.016	1.757	1.532	0.189	1.153	1.906				
β_1	-0.006	0.064	-0.131	0.120	-0.009	0.066	-0.137	0.116	-0.031	0.066	-0.160	0.105				
β_2	0.016	0.004	0.008	0.025	0.017	0.004	0.009	0.025	0.018	0.004	0.010	0.027				
σ_e^2	0.027	0.006	0.016	0.041	0.042	0.002	0.037	0.047	0.043	0.003	0.038	0.049				
D_{11}	0.099	0.028	0.051	0.155	0.100	0.029	0.050	0.157	0.138	0.019	0.106	0.179				
D_{12}	0.013	0.011	-0.008	0.038	0.014	0.012	-0.007	0.039	0.034	0.019	-0.002	0.073				
D_{22}	0.060	0.014	0.038	0.092	0.060	0.014	0.037	0.091	0.132	0.023	0.092	0.184				
δ_{b1}	0.275	0.143	0.013	0.493	0.291	0.135	0.025	0.502	-	-	-	-				
δ_{b2}	0.305	0.039	0.231	0.384	0.300	0.038	0.225	0.378	-	-	-	-				
δ_e	0.198	0.052	0.068	0.275	-	-	-	-	-	-	-	-				
-log-verossimilhança	104.4511					106.5964					139.8230					
AIC	0.1434					0.1448					0.1839					
BIC	0.1727					0.1712					0.2045					
HQ	0.1427					0.1442					0.1835					

Figura 4.8: Histórico das cadeias de Markov para os parâmetros geradas sob o Modelo 1 .

Capítulo 5

Modelo Linear Misto com Erros de Medida

Existe uma vasta literatura em modelos de regressão de observações independentes com covariáveis medidas com erro, seja no caso linear (Fuller, 1987 e Cheng & Van Ness, 1999) e no caso não-linear (Carroll, et al., 1995). Enquanto vários enfoques estão sendo desenvolvidos para tratar dados correlacionados que seguem a estrutura de um MLM, poucos trabalhos estão contribuindo na análise de dados correlacionados quando covariáveis medidas com erro estão presentes. Como discutido por Davidian & Giltinan (1995) em um de seus problemas abertos (Capítulo 12, p. 328.), "é freqüente em casos práticos que os valores das covariáveis em indivíduos sejam medidas com erros e a inferência destes modelos ainda não está desenvolvida adequadamente". Erros de medida são problemas freqüentes em diversas áreas de pesquisa (social, médica, industrial) e existe um consenso sobre a importância de incorporá-los nos modelos estatísticos.

Em um contexto geral existem diversos estudos que mostram que o principal efeito de erro de medida na estimação é o viés nos coeficientes estimados e, conseqüentemente, as propostas existentes buscam corrigir este efeito. Em modelos clássicos de regressão linear simples o viés causado pelo erro de medida é sempre em forma de atenuação, ou seja, o coeficiente estimado preserva o sinal do parâmetro (assintoticamente), mas é viesado na direção de zero. Contudo, de acordo com Carroll, et al. (1995), em modelos mais gerais este tipo de atenuação pode não valer. Um outro efeito importante mas geralmente ignorado, é o aumento na variabilidade dos estimadores dos parâmetros modelo.

Na literatura clássica de modelos com erros de medida existem duas abordagens usadas frequentemente para fazer inferências, determinadas pelas propriedades assumidas para as verdadeiras covariáveis não observadas $\mathbf{X}$. Nos *Modelos funcionais* as verdadeiras covariáveis $\mathbf{X}$ são tratadas como uma sequência de constantes fixas desconhecidas, chamadas de parâmetros incidentais. Nos *modelos estruturais* as covariáveis $\mathbf{X}$ são consideradas aleatórias.

Entre as técnicas existentes para tratar tanto modelos funcionais como estruturais, existe o método de simulação-extrapolação (SIMEX), proposto por Cook & Stefansky (1994). Este método é fácil de implementar e produz estimadores aproximadamente consistentes (Carroll et al., 1995). A maior crítica ao método é seu elevado custo computacional. O método do escore corrigido (EC, Nakamura, 1990) representa uma técnica aplicável tanto na modelagem estrutural como funcional e produz estimadores consistentes e assintoticamente normais (Gimenez & Bolfarine, 1997). Contudo, em algumas situações, a obtenção do escore corrigido pode ser difícil, ou pode não ser possível. A combinação de efeitos aleatórios e erros de medida para modelos lineares é motivo de um número considerável de trabalhos recentes. Trabalhos relacionados com métodos para dados longitudinais com erros de medida podem ser encontrados em Palta e Lin (1996). Wang et al. (1998) e Lin & Carroll (1999) investigam o vício dos parâmetros estimados e testam os componentes de variância em modelos lineares mistos generalizados com erros de medida usando o método SIMEX. Recentemente, Xu - Ping et al. (2002) usam a função escore corrigido para estudar modelos lineares mistos com erros de medida, supondo que os parâmetros da matriz de variâncias-covariâncias são conhecidos ou substituídos por estimadores consistentes em aplicações.

Neste capítulo desejamos comparar, através de simulação do modelo linear com efeito aleatório e erros de medida (MLMeM), dois métodos de estimação: escore corrigido (EC) e simulação-extrapolação (SIMEX). Para cada método investigamos também o efeito do aumento da variância do erro de medida nas estimativas dos coeficientes da regressão e dos parâmetros de escala da distribuição do efeito aleatório. Adicionalmente, no caso de um MLMeM estrutural com uma covariável medida com erro, fornecemos os passos necessários para a implementação do algoritmo EM. Para complementar encontramos a densidade marginal deste modelo sob a suposição de distribuição normal-assimétrica em todas as variáveis consideradas no modelo.

5.1 Método de Simulação-Extrapolação

Consideremos o caso em que $\mathbf{X} = (X_1, \dots, X_m)^T$ é sujeita a erro e $\mathbf{W} = (W_1, \dots, W_m)^T$ é o valor observado de $\mathbf{X}$. Isto é, num modelo com erro aditivo temos a seguinte especificação:

$$W_j = X_j + \delta_j, \quad j = 1, \dots, m, \quad (5.1)$$

em que $\delta_j \sim N(0, \Delta)$ com Δ conhecida ou estimável. A idéia central do método SIMEX é a possibilidade de determinar experimentalmente, via simulação, o efeito do erro de medida sobre os estimadores. Assim, o SIMEX pode ser definido como um método de estimação e redução do viés devido ao erro de medida, baseado em simulação, onde nenhuma suposição é feita a respeito da distribuição da covariável não observada X_j , $j = 1, \dots, m$. Esta técnica proposta por Cook & Stefanski (1994) é aplicável a métodos mais gerais de estimação. Carroll, et al. (1995, Cap. 4) apresentam uma descrição completa deste método e fornecem diversas referências. Estes autores argumentam que o SIMEX representa um método que reproduz réplicas do erro de medida (erro) da mesma forma que o bootstrap reproduz réplicas da amostra.

5.1.1 Procedimento de Estimação

Consideremos a especificação (5.1) para descrever este procedimento. O algoritmo é composto de dois passos: simulação e extrapolção. No passo simulação, para algum $t_s \geq 0$ definimos

$$W_{b,j}(t_s) = W_j + t_s^{1/2} J_{b,j}, \quad j = 1, \dots, m, \quad b = 1, \dots, B, \quad (5.2)$$

sendo $J_{b,1}, \dots, J_{b,m}$ pseudo-erros de medida, gerados independentemente conforme uma distribuição normal com média zero e variância Δ , independente de todos os dados observados. Obtemos então estimativas ingênuas (naive) dos parâmetros usando, $W_{b,1}, \dots, W_{b,m}$ no lugar da covariável não observada. Seja $\hat{\theta}_b(t_s)$ o estimador obtido para o vetor de parâmetros θ nesta situação e seja $\hat{\theta}(t_s)$ a média das estimativas obtidas com respeito a B repetições de um experimento com uma mesma quantidade de erro de medida, ou seja,

$$\hat{\theta}(t_s) = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b(t_s).$$

Realizamos este procedimento para uma sequência crescente $t_1, \dots, t_S$. Deste modo são criados conjuntos de dados com variâncias do erro de medida sucessivamente maiores.

A etapa de simulação corresponde então à obtenção de seqüência de estimativas

$$\hat{\theta}(t_1), \dots, \hat{\theta}(t_S).$$

Com base em evidências obtidas por simulação, Carroll, et al. (1995) recomendam a utilização de t_s no intervalo $[0, t_S]$, com $1 \leq t_S \leq 2$.

Para cada componente do vetor de parâmetros estimados, $\hat{\theta}(t_s)$, os valores resultantes das médias são modelados como função de t_s , para $s = 1, \dots, S$. Estimada esta função, extrapolamos o valor estimado para $t = -1$, que corresponde à etapa de extrapolação.

O estimador SIMEX, denotado por $\hat{\theta}_{simex}$, corresponde ao vetor resultante de valores extrapolados. O argumento para justificar esta escolha, é explicado pelo fato de que, em um modelo com erro de medida aditivo, a variância do erro em $W_{b,j}(t)$ é $(1+t)\Delta$, que se torna zero quando $t = -1$. Em muitos problemas de interesse, a forma da verdadeira função de extrapolação pode ser adequadamente modelada através de um modelo de regressão.

5.1.2 Propriedades Assintóticas do Estimador

De acordo com Cook & Stefanski (1994), o método SIMEX produz, em problemas de regressão, estimadores aproximadamente consistentes. Carroll et al. (1996) apresentam resultados assintóticos para o estimador SIMEX usando a abordagem de equações de estimação. Assumindo amostras independentes e identicamente distribuídas, e considerando que a verdadeira função de extrapolação é usada, estes autores mostram a normalidade assintótica para estimadores SIMEX obtidos como soluções de equações de estimação não viciadas. Também propõem um estimador para a matriz de covariâncias assintótica, sendo os resultados válidos para Δ conhecido ou estimado a partir de replicações. A teoria assintótica é tipicamente considerada com respeito ao tamanho de amostra, supondo B fixo. Stefanski & Cook (1995) descrevem como a matriz de covariâncias do estimador SIMEX pode ser obtida. Seja $\mathbf{G}_1(t_s) = \frac{1}{B} \sum_{b=1}^B \mathbf{G}_1(\mathbf{W}_b(t_s))$, com $\mathbf{G}_1(\mathbf{W}_b(t_s))$ representando a matriz de covariâncias de $\theta_b(t_s)$ estimada, com base no ajuste do modelo ingênuo (sem considerar erros de medida), que usa $\mathbf{W}_b(t_s)$ no lugar de $\mathbf{X}$. Seja $\mathbf{G}_2(t_s)$ a matriz de covariâncias amostral de $\hat{\theta}_b(t_s)$, ou seja,

$$\mathbf{G}_2(t_s) = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b(t_s) - \hat{\theta}(t_s))(\hat{\theta}_b(t_s) - \hat{\theta}(t_s))^T. \quad (5.3)$$

Definimos então $G(t_s) = G_1(t_s) - G_2(t_s)$. Usando o mesmo procedimento dado acima, ajustamos $G(t_s)$ como função de t_s , através de um modelo de regressão linear, quadrático ou não linear, com base em uma sequência para t_s , $s = 1, \dots, S$, e extrapolamos esta função de volta para $t = -1$. A matriz resultante da extrapolação corresponde à estimativa da matriz de covariâncias de $\hat{\theta}_{simex}$. Deve-se enfatizar que este procedimento é aproximado, tendo em vista que é válido em grandes amostras. As justificativas teóricas para este procedimento de estimação são dadas em Carroll et al. (1995).

5.2 Método do Escore Corrigido

Baseando-se no fato de que um estimador consistente de um parâmetro pode ser obtido como solução de uma equação de estimação não viciada, Nakamura (1990) propõe o método do escore corrigido para estimação e testes de hipóteses em modelos com erros nas variáveis, que pode ser aplicado a modelos funcionais e estruturais. A partir de uma correção da log-verossimilhança, Nakamura (1990) define uma função escore dependendo das covariáveis observadas, cuja esperança em relação à distribuição dos erros de medida, coincide com a função escore usual baseada nas verdadeiras covariáveis desconhecidas. A seguir descrevemos o método que visa obter estimadores consistentes e assintoticamente normais.

5.2.1 Procedimento de Estimação

Seja $\theta = (\theta_1, \dots, \theta_p)$ o vetor $p \times 1$ de parâmetros desconhecidos que pertence ao espaço paramétrico Θ . Denotemos por $\ell(\theta|X, Y)$, $S(\theta|X, Y)$ e $I(\theta|X, Y)$ a log-verossimilhança, a função escore e matriz de informação observada $(-\partial S(\theta|X, Y)/\partial \theta)$, respectivamente, de θ dados X e Y , onde $\theta \in F$, com F um subconjunto convexo aberto de Θ . Sejam θ_o o verdadeiro valor do parâmetro ($\theta_o \in F$) e E^+ a esperança com relação ao vetor aleatório Y . Então,

$$E^+\{S(\theta_o|X, Y)\} = 0. \quad (5.4)$$

Quando X é sujeita a erro e W é o valor observado de X , chamamos a $S(\theta|W, Y)$ a função escore ingênua, onde $\theta \in F$. Então, em geral $E^+\{S(\theta_o|W, Y)\} \neq 0$, implicando em $\tilde{\theta}_w$ solução de $S(\tilde{\theta}_w|W, Y) = 0$ não é necessariamente um estimador consistente de θ . Para corrigir este problema, Nakamura (1990) propõe a log-verossimilhança corrigida $\ell^*(\theta|W, Y)$, que satisfaz

$$E^*(\ell^*(\theta|W, Y)) = \ell(\theta|X, Y), \quad (5.5)$$

onde E^* denota a média condicional com relação a $\mathbf{W}$ dados $\mathbf{X}$ e $\mathbf{Y}$ e $\boldsymbol{\theta} \in F$. Se $\mathbf{S}^*(\boldsymbol{\theta}|\mathbf{W}, \mathbf{Y}) = \partial \ell^*(\boldsymbol{\theta}|\mathbf{W}, \mathbf{Y})/\partial \boldsymbol{\theta}$ e $\mathbf{I}^*(\boldsymbol{\theta}|\mathbf{W}, \mathbf{Y}) = \partial \mathbf{S}^*(\boldsymbol{\theta}|\mathbf{W}, \mathbf{Y})/\partial \boldsymbol{\theta}$ são a função escore corrigido e a matriz de informação corrigida, respectivamente, e se E^* e $\partial \boldsymbol{\theta}$ são permutáveis, então

$$E^*\{\mathbf{S}^*(\boldsymbol{\theta}|\mathbf{W}, \mathbf{Y})\} = \mathbf{S}(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Y}) \text{ e } E^*\{\mathbf{I}^*(\boldsymbol{\theta}|\mathbf{W}, \mathbf{Y})\} = \mathbf{I}(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Y}), \quad (5.6)$$

finalmente chamamos de estimador corrigido de $\boldsymbol{\theta}$ a $\boldsymbol{\theta}^*$ tal que $\mathbf{S}^*(\boldsymbol{\theta}^*|\mathbf{W}, \mathbf{Y}) = 0$, com $\mathbf{I}^*(\boldsymbol{\theta}^*|\mathbf{W}, \mathbf{Y})$ sendo definida positiva.

Como é notado por Nakamura (1990), a verossimilhança corrigida ℓ^* e o estimador corrigido têm boas propriedades. Se $E = E^+ E^*$ denota a esperança global (incondicional), então segue de (5.4) e (5.6) que

$$E\{\mathbf{S}^*(\boldsymbol{\theta}_o|\mathbf{W}, \mathbf{Y})\} = E^+ E^*\{\mathbf{S}^*(\boldsymbol{\theta}_o|\mathbf{W}, \mathbf{Y})\} = E^+\{\mathbf{S}(\boldsymbol{\theta}_o|\mathbf{X}, \mathbf{Y})\} = 0.$$

Isto indica que a função escore corrigido é uma função escore não viciada (Stefanski & Carroll, 1987; Nakamura, 1990). Em particular, como mostrado por Nakamura (1990), o estimador $\boldsymbol{\theta}^*$ tem propriedades de consistência e normalidade assintótica, sob certas condições de regularidade. A seguir, considerando que o vetor de parâmetros $\boldsymbol{\theta} \in F$, apresentamos alguns resultados assintóticos relacionados com este método.

5.2.2 Resultados Assintóticos

Para $\mathbf{W} = (W_1, \dots, W_m)^T$, $\mathbf{Y} = (Y_1, \dots, Y_m)^T$ e $\mathbf{X} = (X_1, \dots, X_m)^T$ matrizes ou vetores de variáveis independentes, seja $\mathbf{S}^*(\boldsymbol{\theta}|\mathbf{W}, \mathbf{Y}) = \sum_{j=1}^m \mathbf{S}_j^*(\boldsymbol{\theta}|W_j, Y_j)$ satisfazendo as condições acima. Definimos $\mathbf{I}_m^*$ a matriz de informação observada corrigida, $\boldsymbol{\Gamma}_m^*$ a matriz de covariâncias do escore corrigido e $\boldsymbol{\Lambda}_m^*$ a matriz de informação esperada corrigida, conforme segue

$$\mathbf{I}_m^*(\boldsymbol{\theta}|\mathbf{W}, \mathbf{Y}) = -\frac{\partial \mathbf{S}^*(\boldsymbol{\theta}|\mathbf{W}, \mathbf{Y})}{\partial \boldsymbol{\theta}} = \sum_{j=1}^m \mathbf{I}_j^*(\boldsymbol{\theta}|W_j, Y_j),$$

$$\boldsymbol{\Gamma}_m^*(\boldsymbol{\theta}, \mathbf{X}) = \sum_{j=1}^m E\{\mathbf{S}_j^*(\boldsymbol{\theta}|W_j, Y_j) \mathbf{S}_j^{*T}(\boldsymbol{\theta}|W_j, Y_j)\}, \quad (5.7)$$

$$\boldsymbol{\Lambda}_m^*(\boldsymbol{\theta}, \mathbf{X}) = E(\mathbf{I}_m^*(\boldsymbol{\theta}|\mathbf{W}, \mathbf{Y})). \quad (5.8)$$

Sejam $\bar{\mathbf{I}}_m^* = \mathbf{I}_m^*/m$, $\bar{\mathbf{\Lambda}}_m^* = \mathbf{\Lambda}_m^*/m$ e $\bar{\mathbf{\Gamma}}_m^* = \mathbf{\Gamma}_m^*/m$. A seguinte proposição resume os principais resultados assintóticos sobre o EC e sobre o estimador dos parâmetros obtido como raiz da função escore corrigido.

Proposição 5.1. *Seja $\mathbf{S}^*(\boldsymbol{\theta}|\mathbf{W}, \mathbf{Y}) = \sum_{j=1}^m \mathbf{S}_j^*(\boldsymbol{\theta}|W_j, Y_j)$ uma função escore corrigido. Sob condições gerais de regularidade temos:*

- Com probabilidade tendendo um quando $n \rightarrow \infty$, $\mathbf{S}^* = 0$ tem uma raiz consistente, $\boldsymbol{\theta}^*$,
- $\boldsymbol{\theta}^*$ é assintoticamente normal com média $\boldsymbol{\theta}$ e matriz de covariâncias dada por $m^{-1}\bar{\mathbf{\Sigma}}_m(\boldsymbol{\theta}, X)$, sendo

$$\bar{\mathbf{\Sigma}}_m(\boldsymbol{\theta}, X) = [\bar{\mathbf{\Lambda}}_m^*(\boldsymbol{\theta}, X)]^{-1} \bar{\mathbf{\Gamma}}_m^*(\boldsymbol{\theta}, X) [\bar{\mathbf{\Lambda}}_m^{*T}(\boldsymbol{\theta}, X)]^{-1}. \quad (5.9)$$

As condições de regularidade e a prova da proposição acima são dadas em Gimenez & Bolfarine (1997). Um estimador consistente para $\bar{\mathbf{\Sigma}}_m^*$ é dado pelo estimador sanduíche definido como

$$\mathbf{G}_m^*(\boldsymbol{\theta}|\mathbf{W}, \mathbf{Y}) = [\bar{\mathbf{I}}_m^*(\boldsymbol{\theta}|\mathbf{W}, \mathbf{Y})]^{-1} \frac{\hat{\mathbf{\Gamma}}_m^*(\boldsymbol{\theta}|\mathbf{W}, \mathbf{Y})}{m} [\bar{\mathbf{I}}_m^{*T}(\boldsymbol{\theta}|\mathbf{W}, \mathbf{Y})]^{-1}, \quad (5.10)$$

onde $\hat{\mathbf{\Gamma}}_m^*(\boldsymbol{\theta}|\mathbf{W}, \mathbf{Y}) = \sum_{j=1}^m \mathbf{S}_j^*(\boldsymbol{\theta}|W_j, Y_j) \mathbf{S}_j^{*T}(\boldsymbol{\theta}|W_j, Y_j)$, com todas as matrizes avaliadas em $\boldsymbol{\theta} = \boldsymbol{\theta}^*$. Chamamos a atenção para o fato de que Gimenez & Bolfarine (1997) apresentam este resultado assintótico em um âmbito mais geral, no entanto o resultado segue sendo válido para as condições apresentadas em Nakamura (1990).

5.3 Aplicação do EC e SIMEX no MLMeM

Nosso interesse agora é a inferência no modelo linear definido em (1.1)-(1.2), quando $\mathbf{X}_j$ não pode ser medido exatamente, mas em seu lugar observamos $\mathbf{W}_j$, $j = 1, \dots, m$. Isto é, num modelo com erro de medida aditivo temos o seguinte MLMeM:

$$\mathbf{Y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \mathbf{b}_j + \boldsymbol{\epsilon}_j, \quad (5.11)$$

$$\mathbf{W}_j = \mathbf{X}_j + \boldsymbol{\delta}_j, \quad j = 1, \dots, m, \quad (5.12)$$

onde $\mathbf{b}_j \stackrel{\text{iid}}{\sim} N_q(\mathbf{0}, \mathbf{D})$, $\boldsymbol{\epsilon}_j \stackrel{\text{ind.}}{\sim} N_{n_j}(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_{n_j})$, $\mathbf{X}_j$ e $\boldsymbol{\delta}_j$ são todos independentes. A matriz de erros de medida $\boldsymbol{\delta}_j = (\delta_{j1}, \dots, \delta_{jn_j})^T$ tem dimensão $n_j \times p$, com $\delta_{ji} \stackrel{\text{iid}}{\sim} N_p(\mathbf{0}, \boldsymbol{\Delta})$.

para $i = 1, \dots, n_j$ e $j = 1, \dots, m$. Consideramos que Δ é conhecida ou pode ser estimada através de repetições do experimento. Para este estudo não fazemos suposições com relação à verdadeira covariável $\mathbf{X}_j$, considerando que o modelo pode ser estrutural ou funcional. Note que no caso em que somente um subconjunto de covariáveis são medidas com erro, podemos considerar os respectivos elementos de Δ como sendo zero em aplicações. O interesse neste caso é fazer inferência sobre o vetor de parâmetros $\theta = (\beta^T, \sigma^T)^T$ com σ definido em (1.13).

Para uma amostra observada, seja $\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_m^T)^T$, $\mathbf{b} = (\mathbf{b}_1^T, \dots, \mathbf{b}_m^T)^T$, $\mathbf{W} = (\mathbf{W}_1^T, \dots, \mathbf{W}_m^T)^T$, $\mathbf{X} = (\mathbf{X}_1^T, \dots, \mathbf{X}_m^T)^T$ e $\ell(\theta|\mathbf{X}, \mathbf{y})$ a função log-verossimilhança marginal do MLM em (5.11). Temos que

$$\ell(\theta|\mathbf{X}, \mathbf{y}) = \sum_{j=1}^m \ell_j(\theta|\mathbf{X}_j, \mathbf{y}_j), \quad (5.13)$$

onde

$$\ell_j(\theta|\mathbf{X}_j, \mathbf{y}_j) = -\frac{1}{2} \log |\Sigma_j| - \frac{1}{2} (\mathbf{y}_j - \mathbf{X}_j \beta)^T \Sigma_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \beta),$$

com $\Sigma_j = \sigma_e^2 \mathbf{I}_{n_j} + \mathbf{Z}_j \mathbf{D} \mathbf{Z}_j^T = \sigma_e^2 \mathbf{I}_{n_j} + \sum_{k=1}^q \sum_{s=1}^q [\mathbf{D}]_{ks} \mathbf{Z}_j^k (\mathbf{Z}_j^s)^T$ (veja notação na Seção 1.2.2). O resultado seguinte será de grande utilidade para encontrar a log-verossimilhança marginal corrigida $\ell^*(\theta|\mathbf{W}, \mathbf{y})$.

Proposição 5.2. *Sejam $\mathbf{X}_j \stackrel{\text{ind.}}{\sim} N_p(\mu_j, \mathbf{V})$, $j = 1, \dots, n$, $\mathbf{A}$ uma matriz simétrica de dimensão $n \times n$, $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)^T$ e $\mu = (\mu_1, \dots, \mu_n)^T$ matrizes de dimensão $n \times p$. Então,*

$$E[\mathbf{X}^T \mathbf{A} \mathbf{X}] = \text{tr}(\mathbf{A}) \mathbf{V} + \mu^T \mathbf{A} \mu.$$

Demonstração. Considerando a notação da Seção 1.2.2, escrevemos

$$\begin{aligned} E[\mathbf{X}^T \mathbf{A} \mathbf{X}] &= E\left(\sum_{k=1}^n \sum_{s=1}^n [A]_{ks} \mathbf{X}^k \mathbf{X}^{sT}\right) = E\left(\sum_{k=1}^n [A]_{kk} \mathbf{X}^k \mathbf{X}^{sT}\right) + E\left(\sum_{k=1}^n \sum_{s \neq k}^n [A]_{ks} \mathbf{X}^k \mathbf{X}^{sT}\right) \\ &= \sum_{k=1}^n [A]_{kk} (\mathbf{V} + \mu_k \mu_k^T) + E\left(\sum_{k=1}^n \sum_{s \neq k}^n [A]_{ks} \mathbf{X}^k \mathbf{X}^{sT}\right) \\ &= \mathbf{V} \sum_{k=1}^n [A]_{kk} + \sum_{k=1}^n \sum_{s=1}^n [A]_{ks} \mu_k \mu_s^T \\ &= \text{tr}(\mathbf{A}) \mathbf{V} + \mu^T \mathbf{A} \mu. \end{aligned}$$

□

Sob as condições do MLMeM definido acima, denotando por E^* a esperança condicional de $\mathbf{W}$ dados $\mathbf{X}$ e $\mathbf{Y}$ e usando o resultado da Proposição 5.2, temos o seguinte resultado:

$$E^*(\mathbf{W}_j^T \Sigma_j^{-1} \mathbf{W}_j) = \mathbf{X}_j^T \Sigma_j^{-1} \mathbf{X}_j + \text{tr}(\Sigma_j^{-1}) \Delta, \quad j = 1, \dots, m. \quad (5.14)$$

Para Δ dada, de (5.13) e usando (5.14), temos que a única log-verossimilhança marginal corrigida para o modelo definido em (5.11)-(5.12), é dada por

$$\ell^*(\boldsymbol{\theta} | \mathbf{W}, \mathbf{Y}) = \sum_{j=1}^m \ell_j^*(\boldsymbol{\theta} | \mathbf{W}_j, \mathbf{Y}_j), \quad (5.15)$$

em que

$$\ell_j^*(\boldsymbol{\theta} | \mathbf{W}_j, \mathbf{y}_j) = -\frac{1}{2} \log |\Sigma_j| - \frac{1}{2} (\mathbf{y}_j - \mathbf{W}_j \boldsymbol{\beta})^T \Sigma_j^{-1} (\mathbf{y}_j - \mathbf{W}_j \boldsymbol{\beta}) + \frac{1}{2} \text{tr}(\Sigma_j^{-1}) \boldsymbol{\beta}^T \Delta \boldsymbol{\beta},$$

depende unicamente dos parâmetros e das variáveis observáveis.

A função escore corrigido para $\boldsymbol{\theta}$ é então obtida como

$$\mathbf{S}^*(\boldsymbol{\theta} | \mathbf{W}, \mathbf{y}) = \frac{\partial \ell^*(\boldsymbol{\theta} | \mathbf{W}, \mathbf{y})}{\partial \boldsymbol{\theta}} = \sum_{j=1}^m \frac{\partial \ell_j^*(\boldsymbol{\theta} | \mathbf{W}_j, \mathbf{y}_j)}{\partial \boldsymbol{\theta}},$$

onde, considerando a notação da Seção 1.2.2 e as fórmulas de derivadas matriciais dadas no Apêndice A, temos

$$\begin{aligned} \frac{\partial \ell_j^*(\boldsymbol{\theta} | \mathbf{W}_j, \mathbf{y}_j)}{\partial \boldsymbol{\beta}} &= \mathbf{W}_j^T \Sigma_j^{-1} (\mathbf{y}_j - \mathbf{W}_j \boldsymbol{\beta}) + \text{tr}(\Sigma_j^{-1}) \Delta \boldsymbol{\beta}, \\ \frac{\partial \ell_j^*(\boldsymbol{\theta} | \mathbf{W}_j, \mathbf{Y}_j)}{\partial \sigma_k^2} &= -\frac{1}{2} \text{tr}(\Sigma_j^{-1} \mathbb{G}_j^k) + \frac{1}{2} (\mathbf{y}_j - \mathbf{W}_j \boldsymbol{\beta})^T \Sigma_j^{-1} \mathbb{G}_j^k \Sigma_j^{-1} (\mathbf{y}_j - \mathbf{W}_j \boldsymbol{\beta}) \\ &\quad - \frac{1}{2} \text{tr}(\Sigma_j^{-1} \mathbb{G}_j^k \Sigma_j^{-1}) \boldsymbol{\beta}^T \Delta \boldsymbol{\beta}, \end{aligned} \quad (5.16)$$

para $k = 0, \dots, p_1$.

Usando a relação em (5.14) é fácil verificar que $E^*(\mathbf{S}^*(\boldsymbol{\theta} | \mathbf{W}, \mathbf{y})) = \mathbf{S}(\boldsymbol{\theta} | \mathbf{X}, \mathbf{y})$, isto é, E^* e $\partial \boldsymbol{\theta}$ são operações permutáveis, indicando que as condições de Nakamura (1990) são satisfeitas para o modelo considerado.

A matriz de informação corrigida $\mathbf{I}^*(\boldsymbol{\theta}|\mathbf{W}, \mathbf{y}) = -\sum_{j=1}^m \frac{\partial^2 \ell_j^*(\boldsymbol{\theta}|\mathbf{W}_j, \mathbf{y}_j)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}$ tem elementos (veja o Apêndice A)

$$\begin{aligned}\frac{\partial^2 \ell_j^*(\boldsymbol{\theta}|\mathbf{W}_j, \mathbf{y}_j)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} &= -\mathbf{W}_j^T \boldsymbol{\Sigma}_j^{-1} \mathbf{W}_j + \text{tr}(\boldsymbol{\Sigma}_j^{-1}) \boldsymbol{\Delta}, \\ \frac{\partial^2 \ell_j^*(\boldsymbol{\theta}|\mathbf{W}_j, \mathbf{y}_j)}{\partial \sigma_k^2 \partial \boldsymbol{\beta}} &= -\mathbf{W}_j^T \boldsymbol{\Sigma}_j^{-1} \mathbb{G}_j^k \boldsymbol{\Sigma}_j^{-1} (\mathbf{y}_j - \mathbf{W}_j \boldsymbol{\beta}) - \text{tr}(\boldsymbol{\Sigma}_j^{-1} \mathbb{G}_j^k \boldsymbol{\Sigma}_j^{-1}) \boldsymbol{\Delta} \boldsymbol{\beta}, \\ \frac{\partial^2 \ell_j^*(\boldsymbol{\theta}|\mathbf{W}_j, \mathbf{y}_j)}{\partial \sigma_k^2 \partial \sigma_s^2} &= \frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_j^{-1} \mathbb{G}_j^s \boldsymbol{\Sigma}_j^{-1} \mathbb{G}_j^k) - (\mathbf{y}_j - \mathbf{W}_j \boldsymbol{\beta})^T (\boldsymbol{\Sigma}_j^{-1} \mathbb{G}_j^k \boldsymbol{\Sigma}_j^{-1} \mathbb{G}_j^s \boldsymbol{\Sigma}_j^{-1}) (\mathbf{y}_j - \mathbf{W}_j \boldsymbol{\beta}) \\ &\quad + \text{tr}(\boldsymbol{\Sigma}_j^{-1} \mathbb{G}_j^k \boldsymbol{\Sigma}_j^{-1} \mathbb{G}_j^s \boldsymbol{\Sigma}_j^{-1}) \boldsymbol{\beta}^T \boldsymbol{\Delta} \boldsymbol{\beta},\end{aligned}\tag{5.17}$$

com $k, s = 0, \dots, p_1$.

A obtenção do estimador corrigido, $\boldsymbol{\theta}^*$, solução de $\mathbf{S}^* = \mathbf{0}$ não têm solução explícita, mas pode ser facilmente manipulado com programas computacionais como, por exemplo, S-plus, Matlab e Ox. A raiz apropriada considera que $\mathbf{I}^*(\boldsymbol{\theta}^*|\mathbf{W}, \mathbf{Y})$ é definida positiva. Como valores iniciais podem ser usados as respectivas estimativas ingênuas.

Como nosso interesse é comparar os distintos métodos de estimação nesta classe de modelos, apresentamos agora os passos necessários para implementar o algoritmo SIMEX no MLMem definido em (5.11)-(5.12).

Etapa de simulação

- 1.- fixar B suficientemente grande,
- 2.- fixar $\lambda_1 \in [0, 2]$,
- 3.- para cada $b = 1, \dots, B$:
 - 3.1.- gerar $\mathbf{J}_{b,j}$ com distribuição $N_p(\mathbf{0}, \boldsymbol{\Delta})$, $j = 1, \dots, m$, todos independentes e calcular

$$\mathbf{W}_{b,j} = \mathbf{W}_j + \lambda_1^{1/2} \mathbf{J}_{b,j},$$

- 3.2.- usando o algoritmo EM dado na Seção 1.2.1, obter as estimativas ingênuas $\hat{\boldsymbol{\theta}}_b(\lambda_1)$, substituindo $\mathbf{W}_{b,j}$ por $\mathbf{X}_j$, $j = 1, \dots, m$,
- 4.- calcular $\hat{\boldsymbol{\theta}}(\lambda_1) = \frac{1}{B} \sum_{j=1}^B \hat{\boldsymbol{\theta}}_b(\lambda_1)$ e

5.- repetir os passos 2-4 para $m = 1, \dots, M$, com $\lambda_m \in [0, 2]$, M inteiro arbitrário.

Etapa de extrapolação

6.- Relacionar $\hat{\theta}(\lambda)$ e λ através de um modelo de regressão (geralmente quadrático).

7.- Fazer $\lambda = -1$ nas regressões estimadas, obtendo deste modo o estimador $\hat{\theta}_{simex}$.

Muitas vezes também é razoável considerar que as verdadeiras covariáveis $\mathbf{X}_j, j = 1, \dots, m$, são variáveis aleatórias não correlacionadas com certas média e matriz de variâncias-covariâncias, sendo possível neste caso encontrar os EMV do vetor de parâmetros θ via algum algoritmo de maximização. Na seguinte seção apresentamos o algoritmo EM para encontrar os EMV em um MLMeM estrutural restrito ao caso em que uma covariável é medida com erro.

5.3.1 Comparando os Métodos EC e SIMEX através de Simulação

Realizamos agora um estudo de simulação para investigar o desempenho dos métodos EC e SIMEX descritos nas seções anteriores. Desejamos basicamente avaliar a mudança no comportamento das estimativas dos parâmetros com aumento da variância do erro de medida.

Os dados foram gerados de acordo com o modelo (5.11)-(5.12), com $p = 2$, $\mathbf{Z}_j = \mathbf{1}_{n_j}$ e selecionando valores fixos dos parâmetros $\beta_1 = 1$, $\beta_2 = 2$, $\sigma_b^2 = 0.5^2$ e $\sigma_e^2 = 0.6^2$. As covariáveis $x_{ij}, i = 1, \dots, n_j, j = 1, \dots, m$, foram geradas de uma distribuição normal padrão $N(0, 1)$. A matriz de covariâncias do erro de medida $\Delta = \text{diag}(\sigma_\delta^2, \sigma_\delta^2)$ foi obtida para diferentes valores de σ_δ^2 . Para obter os estimadores SIMEX usamos, na etapa de extrapolação, funções quadráticas para todos os parâmetros. As variâncias assintóticas dos estimadores SIMEX e EC foram calculadas baseadas em (5.3) e (5.9), respectivamente.

Investigamos os resultados para duas situações de tamanhos de amostra em que $N = \sum_{j=1}^m n_j = 200$. Um considerando $m = 25$ e $n_j = 8$ e outra considerando $m = 5$ e $n_j = 40$. As simulações foram realizadas usando o programa Matlab, sendo a solução de (5.16) obtida através da rotina *fmincon*.

Tabela 5.1: Resultados da simulação do modelo linear misto com erros nas variáveis baseada em 1000 repetições de amostras com $m = 5$ e $n_1 = \dots = n_5 = 40$. EC : estimador baseado na função escore corrigido. INGENUO : estimador baseado em W e VER X: estimador baseado no verdadeiro valor X .

		$\sigma_b^2 = 0.25$	$\sigma_e^2 = 0.36$	$\beta_1 = 1$	$\beta_2 = 2$
EC	Média	0.2463	0.3386	1.0021	2.0045
	Desvio padrão	0.1521	0.0769	0.0657	0.0660
	Erro padrão	0.1228	0.0666	0.0562	0.0573
INGENUO	Média	0.2543	0.7615	0.9169	1.8360
	Desvio padrão	0.1521	0.0754	0.0585	0.0581
VER X	Média	0.2542	0.3564	0.9994	2.0005
	Desvio padrão	0.1470	0,0342	0,0434	0,044
SIMEX (quadrático)	Média	0.2533	0.3729	0.9953	1.9947
	Desvio padrão	0.1716	0.0840	0.0650	0.0682
	Erro padrão	0.0439	0.0840	0.0650	0.0682

A Tabela 5.1 resume parte dos resultados das simulações, para um valor fixo da variância do erro de medida ($\sigma_\delta^2 = 0.3^2$), $m = 5$ e $n_1 = \dots = n_5 = 40$, que é comum em amostras de estudos longitudinais. Neste caso 1000 repetições foram realizadas. O estimador corrigido, o estimador usual ignorando erro de medida em W (ingênuo) e o estimador baseado na função escore do verdadeiro valor X são investigados. Note que o estimador obtido pelo método do escore corrigido (EC) é bem comportado, todos os valores médios de θ estão próximos dos verdadeiros valores. Em contraste, o estimador ingênuo (obtido sem considerar erro de medição) é viciado devido à inconsistência. Os estimadores baseados no verdadeiro X (EMV) são obviamente consistentes. O estimador SIMEX parece ser bem comportado, embora a variância assintótica de $\hat{\sigma}_b^2$ esteja afastada da variância baseada nas 1000 repetições.

Nas Figuras 5.1 e 5.2 apresentamos o comportamento dos estimadores EC e SIMEX, respectivamente, considerando $m = 25$, $n_1 = \dots, = n_{25} = 8$ e diferentes valores para σ_δ^2 . Na Figura 5.1 apresentamos o comportamento dos estimadores corrigidos (EC) para os elementos do parâmetro θ , assim como seus desvios padrão assintóticos estimados, as mesmas são comparadas com o estimador ingênuo. Note que o EC e SIMEX do parâmetro de dispersão σ_b^2 é confundido com o estimador ingênuo, indicando que num processo de estimação este pode ser usado para conjuntamente estimar os outros parâmetros do modelo. Nos outros parâmetros do modelo simulado o estimador ingênuo se torna muito viesado com o aumento de σ_δ^2 , enquanto que o EC é bem comportado, embora as variâncias

Figura 5.1: Comportamento do EC de $\theta = (\sigma_b^2, \sigma_e^2, \beta_1, \beta_2)'$ com o aumento da variância do de erro de medida, calculadas com base em médias de 1000 réplicas de amostras de tamanho $m = 25$ e $n_1 = \dots = n_{25} = 8$ ($N = 200$).

assintóticas sejam maiores que as do estimador ingênuo. Na Figura 5.2 apresentamos o comportamento dos estimadores SIMEX para o vetor de parâmetros θ , assim como as estimativas das variâncias assintóticas dos estimadores. Notamos novamente que as estimativas SIMEX de σ_b^2 são próximas às estimativas ingênuas. Um comentário importante que se resgata deste gráfico é que o estimador SIMEX é estável apenas para valores pequenos do erro de medida. Ou seja, sofre de atenuação a medida que o erro de medição aumenta.

Figura 5.2: Comportamento do estimador SIMEX de $\theta = (\sigma_b^2, \sigma_e^2, \beta_1, \beta_2)'$ com o aumento da variância do de erro de medida, calculadas com base em 1000 replicas de amostras de tamanho $m = 25$ e $n_1 = \dots = n_{25} = 8$ ($N = 200$).

5.4 Algoritmo EM no MLMeM Estrutural

Consideremos o modelo MLM com uma covariável medida com erro. Sem perda de generalidade escrevemos

$$\mathbf{Y}_j = \mathbf{X}_{A_j}\boldsymbol{\beta}_1 + \beta_p\mathbf{X}_{p_j} + \mathbf{Z}_j\mathbf{b}_j + \boldsymbol{\epsilon}_j, \quad (5.18)$$

$$\mathbf{W}_{p_j} = \mathbf{X}_{p_j} + \boldsymbol{\delta}_{p_j}, \quad j = 1, \dots, m, \quad (5.19)$$

onde $\mathbf{b}_j \stackrel{\text{iid}}{\sim} N_q(\mathbf{0}, \mathbf{D})$, $\boldsymbol{\epsilon}_j \stackrel{\text{ind.}}{\sim} N_{n_j}(\mathbf{0}, \sigma_e^2 \mathbf{I}_{n_j})$, $\mathbf{X}_{A_j}$ é a matriz de covariáveis medidas sem erro de dimensão $[n_j \times (p-1)]$, $\mathbf{X}_{p_j}$ é o vetor de covariáveis medidas com erro de dimensão $(n_j \times 1)$ com distribuição $N_{n_j}(\mu_x \mathbf{1}_{n_j}, \sigma_x^2 \mathbf{I}_{n_j})$, $\boldsymbol{\beta}_1 = (\beta_1, \dots, \beta_{p-1})^T$ é o vetor de efeitos fixos de dimensão $[(p-1) \times 1]$, β_p (também efeito fixo) é o parâmetro escalar correspondente à covariável medida com erro e $\boldsymbol{\delta}_{p_j} \sim N(0, \sigma_\delta^2 \mathbf{I}_{n_j})$, com σ_δ^2 conhecido ou estimável por meio de replicas. A fim de que o modelo seja identificável consideramos que $\boldsymbol{\epsilon}_j$, $\boldsymbol{\delta}_{p_j}$, $\mathbf{X}_{p_j}$ e $\mathbf{b}_j$ são mutuamente independentes, $j = 1, \dots, m$.

O modelo definido acima pode ser especificado como

$$\mathbf{Y}_j^* = \mathbf{X}_{o_j}\boldsymbol{\beta}_1 + \boldsymbol{\beta}_o \otimes \mathbf{X}_{p_j} + \mathbf{U}_{o_j}\mathbf{b}_j + \boldsymbol{\Delta}_j, \quad (5.20)$$

em que $\mathbf{Y}_j^* = (\mathbf{Y}_j^T, \mathbf{W}_{p_j}^T)^T$, $\boldsymbol{\beta}_o = (\beta_p, 1)^T$, $\boldsymbol{\Delta}_j = (\boldsymbol{\epsilon}_j^T, \boldsymbol{\delta}_{p_j}^T)^T$ e as matrizes $\mathbf{X}_{o_j} = \begin{pmatrix} \mathbf{X}_{A_j} \\ \mathbf{0} \end{pmatrix}$ têm dimensão $[2n_j \times (p-1)]$ e $\mathbf{U}_{o_j} = \begin{pmatrix} \mathbf{Z}_j \\ \mathbf{0} \end{pmatrix}$ é de dimensão $(2n_j \times q)$.

Note que

$$\mathbf{Y}_j^* | \mathbf{X}_{p_j}, \mathbf{b}_j \stackrel{\text{ind.}}{\sim} N_{2n_j}(\boldsymbol{\mu}_j, \mathbf{V}_j), \quad j = 1, \dots, m,$$

onde $\boldsymbol{\mu}_j = \mathbf{X}_{o_j}\boldsymbol{\beta}_1 + \boldsymbol{\beta}_o \otimes \mathbf{X}_{p_j} + \mathbf{U}_{o_j}\mathbf{b}_j$, $\mathbf{V}_j = \text{diag}(\sigma_e^2, \sigma_\delta^2) \otimes \mathbf{I}_{n_j}$, com $\mathbf{X}_{p_j}$ e $\mathbf{b}_j$ independentes, $j = 1, \dots, m$.

Sejam $\mathbf{y}^* = (\mathbf{y}_1^{*T}, \dots, \mathbf{y}_m^{*T})^T$, $\mathbf{b} = (\mathbf{b}_1^T, \dots, \mathbf{b}_m^T)^T$ e $\mathbf{X}_p = (\mathbf{X}_{p_1}, \dots, \mathbf{X}_{p_m})$. Sob as considerações acima, a função log-verossimilhança para os dados completos (ver Capítulo 1, Seção 1.2.1) no MLMeM estrutural representada por (5.20) é dada por

$$\begin{aligned}
\ell_c(\theta|\mathbf{y}^*, \mathbf{b}, \mathbf{X}_p) &= \sum_{j=1}^m \left[-\frac{1}{2} \log |\mathbf{V}_j| - \frac{1}{2} (\mathbf{y}_j^* - \boldsymbol{\mu}_j)^T \mathbf{V}_j^{-1} (\mathbf{y}_j^* - \boldsymbol{\mu}_j) \right. \\
&\quad \left. - \frac{n_j}{2} \log \sigma_x^2 - \frac{1}{2\sigma_x^2} (\mathbf{X}_{p_j} - \mu_x \mathbf{1}_{n_j})^T (\mathbf{X}_{p_j} - \mu_x \mathbf{1}_{n_j}) - \frac{1}{2} \log |\mathbf{D}| - \frac{1}{2} \mathbf{b}_j^T \mathbf{D}^{-1} \mathbf{b}_j \right] \\
&= \sum_{j=1}^m \left[-\frac{n_j}{2} \log \sigma_e^2 - \frac{1}{2\sigma_e^2} (\mathbf{y}_j - \mathbf{Z}_j \mathbf{b}_j)^T (\mathbf{y}_j - \mathbf{Z}_j \mathbf{b}_j) + \frac{1}{\sigma_e^2} \beta_p \mathbf{X}_{p_j}^T (\mathbf{y}_j - \mathbf{Z}_j \mathbf{b}_j) \right. \\
&\quad \left. - \frac{1}{2\sigma_e^2} \beta_p^2 \mathbf{X}_{p_j}^T \mathbf{X}_{p_j} + \frac{1}{\sigma_e^2} \beta_1^T \mathbf{X}_{A_j}^T (\mathbf{y}_j - \mathbf{X}_{p_j} \beta_p - \mathbf{Z}_j \mathbf{b}_j) - \frac{1}{2\sigma_e^2} \beta_1^T \mathbf{X}_{A_j}^T \mathbf{X}_{A_j} \beta_1 \right. \\
&\quad \left. - \frac{n_j}{2} \log \sigma_x^2 - \frac{1}{2\sigma_x^2} (\mathbf{X}_{p_j} - \mu_x \mathbf{1}_{n_j})^T (\mathbf{X}_{p_j} - \mu_x \mathbf{1}_{n_j}) - \frac{1}{2} \log |\mathbf{D}| - \frac{1}{2} \mathbf{b}_j^T \mathbf{D}^{-1} \mathbf{b}_j \right]. \quad (5.21)
\end{aligned}$$

Fazendo $\hat{\mathbf{b}}_j = E[\mathbf{b}_j|\theta = \hat{\theta}, \mathbf{y}^*]$, $\hat{\Omega}_j = Cov[\mathbf{b}_j|\theta = \hat{\theta}, \mathbf{y}^*]$, $\hat{\mathbf{X}}_{p_j} = E[\mathbf{X}_{p_j}|\theta = \hat{\theta}, \mathbf{y}^*]$ e $\hat{\Psi}_j = Cov[\mathbf{X}_{p_j}|\theta = \hat{\theta}, \mathbf{y}^*]$. Lembrando que $\mathbf{b}_j$ e $\mathbf{X}_{p_j}$ são independentes e das propriedades de distribuição normal multivariada (Muirhead, 1982) temos

$$\begin{aligned}
\hat{\Omega}_j &= \left(\hat{\mathbf{D}}^{-1} + \frac{1}{\hat{k}_o^2} \mathbf{Z}_j^T \mathbf{Z}_j \right)^{-1}, \\
\hat{\mathbf{b}}_j &= [\hat{\mathbf{D}} \mathbf{Z}_j^T \mathbf{0}_{q \times n_j}] \hat{\mathbf{D}}_{o_j}^{-1} (\mathbf{y}_j^* - \hat{\boldsymbol{\mu}}_j), \\
\hat{\mathbf{X}}_{p_j} &= \mu_x \mathbf{1}_{n_j} + \sigma_x^2 [\boldsymbol{\beta}_o \otimes \mathbf{I}_{n_j}] \hat{\mathbf{D}}_{o_j}^{-1} (\mathbf{y}_j^* - \hat{\boldsymbol{\mu}}_j), \\
\hat{\Psi}_j &= \sigma_x^2 \mathbf{I}_{n_j} + \sigma_x^4 [\boldsymbol{\beta}_o \otimes \mathbf{I}_{n_j}] \hat{\mathbf{D}}_{o_j}^{-1} [\boldsymbol{\beta}_o \otimes \mathbf{I}_{n_j}]^T, \quad (5.22)
\end{aligned}$$

$$\text{com } k_o = \frac{\sigma_e^2 \sigma_x^2 + (\beta_p \sigma_x^2 + \sigma_e^2) \sigma_\delta^2}{\sigma_x^2 + \sigma_\delta^2} \text{ e } \mathbf{D}_{o_j} = \begin{pmatrix} (\beta_p^2 \sigma_x^2 + \sigma_e^2) \mathbf{I}_{n_j} + \mathbf{Z}_j \mathbf{D} \mathbf{Z}_j^T & \beta_p \sigma_x^2 \mathbf{I}_{n_j} \\ \beta_p \sigma_x^2 \mathbf{I}_{n_j} & (\sigma_x^2 + \sigma_\delta^2) \mathbf{I}_{n_j} \end{pmatrix},$$

de dimensão $2n_j \times 2n_j$. Usando uma álgebra simples (veja o Apêndice A) temos que

$$\begin{aligned}
E(\ell_c(\theta|\mathbf{y}^*, \mathbf{b}, \mathbf{X}_p)|\mathbf{y}^*, \hat{\theta}) &= \\
&\sum_{j=1}^m \left[-\frac{n_j}{2} \log \sigma_e^2 - \frac{1}{2\sigma_e^2} \left((\mathbf{y}_j - \mathbf{Z}_j \hat{\mathbf{b}}_j)^T (\mathbf{y}_j - \mathbf{Z}_j \hat{\mathbf{b}}_j) + \text{tr}(\mathbf{Z}_j \hat{\Omega}_j \mathbf{Z}_j^T) \right) \right. \\
&\quad + \frac{1}{\sigma_e^2} \beta_p \hat{\mathbf{X}}_{p_j}^T (\mathbf{y}_j - \mathbf{Z}_j \hat{\mathbf{b}}_j) - \frac{1}{2\sigma_e^2} \beta_p^2 \left(\hat{\mathbf{X}}_{p_j}^T \hat{\mathbf{X}}_{p_j} + \text{tr}(\hat{\Psi}_j) \right) + \frac{1}{\sigma_e^2} \beta_1^T \mathbf{X}_{A_j}^T (\mathbf{y}_j - \hat{\mathbf{X}}_{p_j} \beta_p - \mathbf{Z}_j \hat{\mathbf{b}}_j) \\
&\quad - \frac{1}{2\sigma_e^2} \beta_1^T \mathbf{X}_{A_j}^T \mathbf{X}_{A_j} \beta_1 - \frac{n_j}{2} \log \sigma_x^2 - \frac{1}{2\sigma_x^2} \left((\hat{\mathbf{X}}_{p_j} - \mu_x \mathbf{1}_{n_j})^T (\hat{\mathbf{X}}_{p_j} - \mu_x \mathbf{1}_{n_j}) + \text{tr}(\hat{\Psi}_j) \right) \\
&\quad \left. - \frac{1}{2} \log |\mathbf{D}| - \frac{1}{2} \text{tr} \left(\mathbf{D}^{-1} (\hat{\mathbf{b}}_j^T \hat{\mathbf{b}}_j + \hat{\Omega}_j) \right) \right]. \quad (5.23)
\end{aligned}$$

Conforme comentado no Capítulo 1, quando o passo M no algoritmo EM é complexo de implementar é freqüente substituí-lo por uma seqüência de passos de maximização restrita, cada um dos quais maximiza $Q(\theta|\theta')$ sobre θ com alguma função de θ mantida fixa. Isto leva a uma simples extensão do algoritmo EM chamado de algoritmo ECM (Meng & Rubin, 1993). Assim, propomos o seguinte algoritmo ECM:

Etapa E : Dado $\theta = \hat{\theta}$, calcular $\hat{\mathbf{b}}_j$, $\hat{\Omega}_j$, $\hat{\mathbf{X}}_{p_j}$ e $\hat{\Psi}_j$ para $j = 1, \dots, m$, usando (5.22).

Etapa CM-1: Fixar $\beta_p = \hat{\beta}_p$ e atualizar β_1 maximizando (5.23) em relação a β_1 , da qual obtemos

$$\hat{\beta}_1 = \left(\sum_{j=1}^m \mathbf{X}_{A_j}^T \mathbf{X}_{A_j} \right)^{-1} \sum_{j=1}^m \mathbf{X}_{A_j}^T (\mathbf{Y}_j - \hat{\mathbf{X}}_{p_j} \hat{\beta}_p - \mathbf{Z}_j \hat{\mathbf{b}}_j),$$

Etapa CM-2 Fixar $\beta_1 = \hat{\beta}_1$ e atualizar β_p maximizando (5.23) sobre β_p , levando a

$$\hat{\beta}_p = \frac{\sum_{j=1}^m \hat{\mathbf{X}}_{p_j}^T (\mathbf{Y}_j - \mathbf{Z}_j \hat{\mathbf{b}}_j) - \hat{\beta}_1^T \mathbf{X}_{A_j}^T \hat{\mathbf{X}}_{p_j}}{\sum_{j=1}^m \text{tr}(\hat{\mathbf{X}}_{p_j} \hat{\mathbf{X}}_{p_j}^T + \hat{\Psi}_j)},$$

Etapa CM-3 Fixar $\beta_1 = \hat{\beta}_1$, $\beta_p = \hat{\beta}_p$ e atualizar σ_e^2 maximizando (5.23) sobre σ_e^2 , fornecendo

$$\begin{aligned} \hat{\sigma}_e^2 = & \frac{1}{N} \sum_{j=1}^m \left[(\mathbf{Y}_j - \mathbf{X}_{A_j} \hat{\beta}_1 - \mathbf{Z}_j \hat{\mathbf{b}}_j - \hat{\beta}_p \hat{\mathbf{X}}_{p_j})^T (\mathbf{Y}_j - \mathbf{X}_{A_j} \hat{\beta}_1 - \mathbf{Z}_j \hat{\mathbf{b}}_j - \hat{\beta}_p \hat{\mathbf{X}}_{p_j}) \right. \\ & \left. + \text{tr}(\mathbf{Z}_j \hat{\Omega}_j \mathbf{Z}_j^T + \hat{\beta}_p^2 \hat{\Psi}_j) \right], \end{aligned}$$

Etapa CM-4 Atualizar μ_x maximizando (5.23) sobre μ_x , fornecendo

$$\hat{\mu}_x = \frac{1}{N} \sum_{j=1}^m \mathbf{1}_{n_j}^T \hat{\mathbf{X}}_{p_j},$$

Etapa CM-5 Fixar $\mu_x = \hat{\mu}_x$ e atualizar σ_x^2 maximizando (5.23) em σ_x^2 , conduzindo a

$$\hat{\sigma}_x^2 = \frac{1}{N} \sum_{j=1}^m \text{tr} \left[(\hat{\mathbf{X}}_{p_j} - \hat{\mu}_x \mathbf{1}_{n_j})(\hat{\mathbf{X}}_{p_j} - \hat{\mu}_x \mathbf{1}_{n_j})^T + \hat{\Psi}_j \right]$$

Etapa CM-6 Atualizar $\mathbf{D}$ maximizando (5.23) em relação a $\mathbf{D}$, que fornece

$$\hat{\mathbf{D}} = \frac{1}{m} \sum_{j=1}^m (\hat{\mathbf{b}}_j \hat{\mathbf{b}}_j^T + \hat{\Omega}_j)$$

onde $N = \sum_{j=1}^m n_j$.

5.5 O MLMeM Estrutural Normal-Assimétrico

Consideremos novamente o MLMeM definido em (5.18)-(5.19), isto é, o modelo

$$\mathbf{Y}_j = \mathbf{X}_{A_j}\beta_1 + \beta_p\mathbf{X}_{p_j} + \mathbf{Z}_j\mathbf{b}_j + \epsilon_j, \quad (5.24)$$

$$\mathbf{W}_{p_j} = \mathbf{X}_{p_j} + \delta_{p_j}, \quad j = 1, \dots, m, \quad (5.25)$$

com as seguintes suposições sobre as distribuições das variáveis envolvidas:

$$\epsilon_j \sim SN_{n_j}(\mathbf{0}, \psi_j, \lambda_{e_j}), \quad (5.26)$$

$$\mathbf{b}_j \sim SN_q(\mathbf{0}, \mathbf{D}, \lambda_b), \quad (5.27)$$

$$\mathbf{X}_{p_j} \sim SN_{n_j}(\mu_{x_j}, \psi_{x_j}, \lambda_{x_j}), \quad (5.28)$$

$$\delta_{p_j} \sim SN_{n_j}(\mathbf{0}, \psi_{\delta_j}, \lambda_{\delta_j}), \quad j = 1, \dots, m \quad (5.29)$$

onde $\mathbf{D} = \mathbf{D}(\gamma)$, $\psi_j = \psi_j(\alpha)$, $\psi_{x_j} = \psi_j(\alpha_x)$ e $\psi_{\delta_j} = \psi_j(\alpha_\delta)$ são matrizes de dispersão dependentes de um número reduzido de parâmetros desconhecidos γ , α , α_x e α_δ , respectivamente. A fim de que o modelo seja identificável consideramos que os vetores aleatórios ϵ_j , $\mathbf{b}_j$, $\mathbf{X}_{p_j}$ e δ_{p_j} são mutuamente independentes, $j = 1, \dots, m$. Neste caso o interesse é fazer inferência sobre o vetor de parâmetros $\theta = (\gamma^T, \alpha^T, \alpha_x^T, \alpha_\delta^T)^T$ e $\lambda = (\lambda_b^T, \lambda_{e1}^T, \dots, \lambda_{em}^T, \lambda_{x1}^T, \dots, \lambda_{xm}^T, \lambda_{\delta1}^T, \dots, \lambda_{\delta m}^T)^T$. Inferências neste tipo de modelos são baseadas na densidade marginal das variáveis observáveis $\varpi_j = (\mathbf{Y}_j^T, \mathbf{W}_{p_j}^T)^T$, $j = 1, \dots, m$. A fim de determinar a densidade marginal de ϖ_j eliminamos o sub-índice j para facilitar a notação. Como $\mathbf{X}_p$ e $\mathbf{b}$ são vetores aleatórios independentes, a densidade marginal do vetor de variáveis observáveis ϖ é obtido calculando-se a seguinte integral:

$$\begin{aligned} f_{\varpi}(\varpi|\theta, \lambda) &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^q} f(\mathbf{y}, \mathbf{W}_p, \mathbf{X}_p, \mathbf{b}) d\mathbf{b} d\mathbf{X}_p \\ &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^q} f(\mathbf{y}|\mathbf{X}_p, \mathbf{b}, \mathbf{W}_p) f(\mathbf{X}_p, \mathbf{b}, \mathbf{W}_p) d\mathbf{b} d\mathbf{X}_p \\ &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^q} f(\mathbf{y}|\mathbf{X}_p, \mathbf{b}, \mathbf{W}_p) f(\mathbf{W}_p|\mathbf{X}_p) f(\mathbf{X}_p) f(\mathbf{b}) d\mathbf{b} d\mathbf{X}_p, \end{aligned}$$

visto que das condições de independência do modelo, dado $\mathbf{X}_p$, $\mathbf{W}_p$ é independente de $\mathbf{b}$. Note também que dados $\mathbf{X}_p$ e $\mathbf{b}$, $\mathbf{Y}$ é independente de $\mathbf{W}_p$ (não-diferencial, Arellano-Valle

& Bolfarine, 1998); logo,

$$\begin{aligned} f_{\varpi}(\varpi|\theta, \lambda) &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^q} f(y|\mathbf{X}_p, \mathbf{b}) f(\mathbf{W}_p|\mathbf{X}_p) f(\mathbf{X}_p) f(\mathbf{b}) d\mathbf{b} d\mathbf{X}_p \\ &= \int_{\mathbb{R}^n} \left[f(\mathbf{W}_p|\mathbf{X}_p) f(\mathbf{X}_p) \left(\int_{\mathbb{R}^q} f(y|\mathbf{X}_p, \mathbf{b}) f(\mathbf{b}) d\mathbf{b} \right) \right] d\mathbf{X}_p. \end{aligned} \quad (5.30)$$

Note primeiro que da Proposição 2.5 temos

$$\mathbf{Y}|\mathbf{X}_p, \mathbf{b} \sim SN_n(\mathbf{X}_A\beta_1 + \beta_p\mathbf{X}_p + \mathbf{Zb}, \psi, \lambda_e) \quad (5.31)$$

$$\mathbf{b} \sim SN_q(\mathbf{0}, \mathbf{D}, \lambda_b), \quad (5.32)$$

e como $\mathbf{X}_p$ e $\mathbf{b}$ são independentes, do Teorema 3.2 segue que a solução da integral

$$\int_{\mathbb{R}^q} f(y|\mathbf{X}_p, \mathbf{b}) f(\mathbf{b}) d\mathbf{b},$$

denotada por $f_{\mathbf{Y}|\mathbf{X}_p}(y|\theta, \lambda)$, é dada por

$$\begin{aligned} f_{\mathbf{Y}|\mathbf{X}_p}(y|\theta, \lambda) &= 2^2 \phi_n(y|\mathbf{X}_A\beta_1 + \beta_p\mathbf{X}_p, \Sigma) \\ &\quad \times \Phi_2((\mu_2 - \Gamma\mu_1)(y_j - \mathbf{X}_A\beta_1 - \beta_b\mathbf{X}_p)|\mathbf{0}, \mathbf{I}_2 + \Gamma\Lambda\Gamma^T), \end{aligned} \quad (5.33)$$

onde

$$\mu_1 = \Lambda\mathbf{Z}^T\psi^{-1}, \quad \Sigma = \psi + \mathbf{Z}\mathbf{D}\mathbf{Z}^T, \quad \Lambda = (\mathbf{D}^{-1} + \mathbf{Z}^T\psi^{-1}\mathbf{Z})^{-1}, \quad (5.34)$$

$$\mu_2 = \begin{pmatrix} \lambda_e^T \psi^{-1/2} \\ 0 \end{pmatrix} \quad \text{e} \quad \Gamma = \begin{pmatrix} \lambda_e^T \psi^{-1/2} \mathbf{Z} \\ -\lambda_b^T \mathbf{D}^{-1/2} \end{pmatrix}, \quad (5.35)$$

ou similarmente, após uma álgebra simples,

$$f_{\mathbf{Y}|\mathbf{X}_p}(y|\theta, \lambda) = 2^2 \phi_n(y|\mathbf{X}_A\beta_1 + \mathbf{X}_p\beta_p, \Sigma) \Phi_2(\mu_1^*(y - \mathbf{X}_A\beta_1) - \beta_p\mu_1^*\mathbf{X}_p|\mathbf{0}, \mathbf{I}_2 + \Gamma\Lambda\Gamma^T), \quad (5.36)$$

onde $\mu_1^* = \left[\begin{pmatrix} \lambda_e^T \psi^{-1/2} \\ 0_n^T \end{pmatrix} - \Gamma\Lambda\mathbf{Z}^T\psi^{-1} \right]$, com Σ , Λ e Γ como em (5.34) e (5.35), respectivamente.

Por outro lado, note também que

$$\begin{aligned}\mathbf{W}_p|\mathbf{X}_p &\sim SN_n(\mathbf{X}_p, \psi_\delta, \lambda_\delta), \\ \mathbf{X}_p &\sim SN_n(\mu_x, \psi_x, \lambda_x).\end{aligned}$$

Logo, substituindo as respectivas densidades na integral em (5.30) temos

$$\begin{aligned}f_Z(\mathbf{z}|\theta, \lambda) &= \int_{\mathbb{R}^n} 2^2 \phi_n(\mathbf{y}|\mathbf{X}_A\beta_1 + \mathbf{X}_p\beta_p, \Sigma) \Phi_2(\mu_1^*(\mathbf{y} - \mathbf{X}_A\beta_1) - \beta_p\mu_1^*\mathbf{X}_p|0, \mathbf{I}_2 + \Gamma\Lambda\Gamma^T) \\ &\times 2\phi_n(\mathbf{W}_p|\mathbf{X}_p, \psi_\delta) \Phi_1\left(\lambda_\delta^T\psi_\delta^{-1/2}(\mathbf{W}_p - \mathbf{X}_p)\right) \\ &\times 2\phi_n(\mathbf{X}_p|\mu_x, \psi_x) \Phi_1\left(\lambda_x^T\psi_x^{-1/2}(\mathbf{X}_p - \mu_x)\right) d\mathbf{X}_p.\end{aligned}\quad (5.37)$$

Para calcular a integral acima, fazemos uso das seguintes proposições:

Proposição 5.3. *Sob a notação em (5.37), segue que*

$$\begin{aligned}\phi_n(\mathbf{y}|\mathbf{X}_A\beta_1 + \mathbf{X}_p\beta_p, \Sigma)\phi_n(\mathbf{W}_p|\mathbf{X}_p, \psi_\delta)\phi_n(\mathbf{X}_p|\mu_x, \psi_x) &= \phi_n(\mathbf{W}_p|\mu_x, \psi_\delta + \psi_x) \\ &\times \phi_n(\mathbf{y}|\mathbf{X}_A\beta_1 + \beta_p\eta, \Sigma + \beta_p^2\Upsilon)\phi_n(\mathbf{X}_p|\eta + \beta_p\Lambda^*\Sigma^{-1}(\mathbf{y} - \mathbf{X}_A\beta_1 - \beta_p\eta, \Lambda^*)),\end{aligned}\quad (5.38)$$

onde $\eta = \mu_x + \Upsilon\psi_\delta^{-1}(\mathbf{W}_p - \mu_x)$, $\Lambda^* = (\Upsilon^{-1} + \beta_p^2\Sigma^{-1})^{-1}$, com $\Upsilon = (\psi_x^{-1} + \psi_\delta^{-1})^{-1}$.

Demonstração. A prova segue aplicando duas vezes o Lema 2.2. □

Proposição 5.4. *Sob a notação em (5.37), segue que*

$$\begin{aligned}\Phi_2(\mu_1^*(\mathbf{y} - \mathbf{X}_A\beta_1) - \beta_p\mu_1^*\mathbf{X}_p|0, \mathbf{I}_2 + \Gamma\Lambda\Gamma^T) \Phi_1\left(\lambda_\delta^T\psi_\delta^{-1/2}(\mathbf{W}_p - \mathbf{X}_p)\right) \\ \times \Phi_1\left(\lambda_x^T\psi_x^{-1/2}(\mathbf{X}_p - \mu_x)\right) = \Phi_4(-\Gamma^*\mathbf{X}_p|-\mu_3, \Omega),\end{aligned}\quad (5.39)$$

onde

$$\Gamma^* = \begin{pmatrix} \beta_p\mu_1^* \\ \lambda_\delta^T\psi_\delta^{-1/2} \\ -\lambda_x^T\psi_x^{-1/2} \end{pmatrix}, \quad \mu_3 = \begin{pmatrix} \mu_1^*(\mathbf{y} - \mathbf{X}_A\beta_1) \\ \lambda_\delta^T\psi_\delta^{-1/2}\mathbf{W}_p \\ -\lambda_x^T\psi_x^{-1/2}\mu_x \end{pmatrix} \quad e \quad \Omega = \text{diag}(\mathbf{I}_2 + \Gamma\Lambda\Gamma^T, \mathbf{I}_2), \quad (5.40)$$

com Γ , Λ , μ_1^* como em (5.34), (5.35), (5.36), respectivamente e $\text{diag}(\mathbf{a}, \mathbf{b})$ sendo uma matriz bloco diagonal cujos blocos são as matrizes $\mathbf{a}$ e $\mathbf{b}$.

Demonstração. Note que se $\mathbf{R} \sim N_2(0, \mathbf{I}_2 + \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^T)$, $U \sim N_1(0, 1)$ e $V \sim N_1(0, 1)$ são independentes, então (5.39) pode ser escrita como

$$\begin{aligned} P(\mathbf{R} \leq \boldsymbol{\mu}_1^*(\mathbf{y} - \mathbf{X}_A\boldsymbol{\beta}_1) - \beta_p\boldsymbol{\mu}_1^*\mathbf{X}_p)P(U \leq \boldsymbol{\lambda}_\delta^T\boldsymbol{\psi}_\delta^{-1/2}(\mathbf{W}_p - \mathbf{X}_p))P(V \leq \boldsymbol{\lambda}_x^T\boldsymbol{\psi}_x^{-1/2}(\mathbf{X}_p - \boldsymbol{\mu}_x)) \\ = P(\mathbf{T} \leq \boldsymbol{\mu}_3), \end{aligned}$$

onde $\mathbf{T} = \mathbf{S} + \mathbf{\Gamma}^*\mathbf{X}_p$, com $\mathbf{S} = (\mathbf{R}^T, U, V)^T$, $\boldsymbol{\mu}_3$ e $\mathbf{\Gamma}^*$ como definido em (5.40). Assim como $\mathbf{T} \sim N_4(\mathbf{\Gamma}^*\mathbf{X}_p, \boldsymbol{\Omega})$, temos que

$$P(\mathbf{T} \leq \boldsymbol{\mu}_3) = \Phi_4(\boldsymbol{\mu}_3 | \mathbf{\Gamma}^*\mathbf{X}_p, \boldsymbol{\Omega}),$$

concluindo a prova. $\square$

Teorema 5.1. *Seja $\mathbf{Y} = \mathbf{X}_A\boldsymbol{\beta}_1 + \mathbf{X}_p\boldsymbol{\beta}_p + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}$ e $\mathbf{W}_p = \mathbf{X}_p + \boldsymbol{\delta}_p$, onde $\boldsymbol{\epsilon} \sim SN_n(0, \boldsymbol{\psi}, \boldsymbol{\lambda}_\epsilon)$, $\mathbf{b} \sim SN_q(0, \mathbf{D}, \boldsymbol{\lambda}_b)$, $\mathbf{X}_p \sim SN_n(\boldsymbol{\mu}_x, \boldsymbol{\psi}_x, \boldsymbol{\lambda}_x)$ e $\boldsymbol{\delta}_p \sim SN_n(0, \boldsymbol{\psi}_\delta, \boldsymbol{\lambda}_\delta)$ são mutuamente independentes. Então, a densidade marginal das quantidades observáveis $\mathbf{Z} = (\mathbf{Y}^T, \mathbf{W}^T)^T$ é dada por*

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{z} | \boldsymbol{\theta}, \boldsymbol{\lambda}) &= 2^4 \phi_n(\mathbf{y} | \mathbf{X}_A\boldsymbol{\beta}_1 + \beta_p\boldsymbol{\eta}, \boldsymbol{\Sigma} + \beta_p^2\Upsilon) \phi_n(\mathbf{W}_p | \boldsymbol{\mu}_x, \boldsymbol{\psi}_x + \boldsymbol{\psi}_\delta) \\ &\times \Phi_4(\boldsymbol{\mu}_3 - \mathbf{\Gamma}^*[\boldsymbol{\eta} + \beta_p\boldsymbol{\Lambda}^*\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}_A\boldsymbol{\beta}_1 - \beta_p\boldsymbol{\eta})] | \mathbf{0}, \boldsymbol{\Omega} + \mathbf{\Gamma}^*\boldsymbol{\Lambda}^*\boldsymbol{\Gamma}^T), \end{aligned} \quad (5.41)$$

onde $\boldsymbol{\eta}$, $\boldsymbol{\Lambda}^*$, Υ , $\mathbf{\Gamma}^*$, $\boldsymbol{\mu}_3$ e $\boldsymbol{\Omega}$ como em (5.38) e (5.40).

Demonstração. De (5.37), (5.38) e (5.39) segue que

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{z} | \boldsymbol{\theta}, \boldsymbol{\lambda}) &= \int_{\mathbb{R}^n} 2^4 \phi_n(\mathbf{W}_p | \boldsymbol{\mu}_x, \boldsymbol{\psi}_\delta + \boldsymbol{\psi}_x) \phi_n(\mathbf{y} | \mathbf{X}_A\boldsymbol{\beta}_1 + \beta_p\boldsymbol{\eta}, \boldsymbol{\Sigma} + \beta_p^2\Upsilon) \\ &\times \phi_n(\mathbf{X}_p | \boldsymbol{\eta} + \beta_p\boldsymbol{\Lambda}^*\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}_A\boldsymbol{\beta}_1 - \beta_p\boldsymbol{\eta}), \boldsymbol{\Lambda}^*) \Phi_4(-\mathbf{\Gamma}^*\mathbf{X}_p | -\boldsymbol{\mu}_3, \boldsymbol{\Omega}) d\mathbf{X}_p \\ &= 2^4 \phi_n(\mathbf{W}_p | \boldsymbol{\mu}_x, \boldsymbol{\psi}_\delta + \boldsymbol{\psi}_x) \phi_n(\mathbf{y} | \mathbf{X}_A\boldsymbol{\beta}_1 + \beta_p\boldsymbol{\eta}, \boldsymbol{\Sigma} + \beta_p^2\Upsilon) E[\Phi_4(-\mathbf{\Gamma}^*\mathbf{N} | -\boldsymbol{\mu}_3, \boldsymbol{\Omega})], \end{aligned}$$

onde $\mathbf{N} \sim N_n(\boldsymbol{\eta} + \beta_p\boldsymbol{\Lambda}^*\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}_A\boldsymbol{\beta}_1 - \beta_p\boldsymbol{\eta}), \boldsymbol{\Lambda}^*)$. A prova conclui usando o Lema 2.1. $\square$

Procedimentos de inferência clássica podem ser tratados usando as técnicas discutidas nos capítulos prévios, desde que a verossimilhança marginal têm uma forma fechada, facilitando direta implementação.

Capítulo 6

Considerações Finais

6.1 Conclusões

Neste trabalho discutimos vários aspectos envolvendo modelos lineares mistos. Um dos principais aspectos abordados foi o desenvolvimento de técnicas de estimação em modelos lineares mistos normal-assimétrico, do ponto de vista clássico e bayesiano. Para estimação via máxima verossimilhança a densidade marginal das quantidades envolvidas foi obtida, de modo que os estimadores podem ser obtidos diretamente via alguma técnica de maximização usando programas computacionais como Matlab, R e Ox. Alternativamente, com o intuito de fazer a estimação mais confiável, o algoritmo EM foi proposto explorando propriedades probabilísticas da distribuição normal assimétrica introduzida. Recordamos ademais que procedimentos desta natureza não são encontrados na literatura. Verificamos também através de simulação que existe perda de eficiência na estimação de certos parâmetros fixos quando normalidade é assumida e assimetria está presente nos dados a serem analisados. Os resultados obtidos por esta técnica foram aplicados a um conjunto de dados reais onde estes coincidem com os relatados em Zhang & Davidian (2001), usando técnicas semi-paramétricas. Para desenvolver procedimentos bayesianos de estimação nos MLM's normal-assimétrico, definimos uma nova densidade que em certo sentido generaliza a distribuição normal-assimétrica introduzida por Sahu. et al. (2003) e que faz da inferência bayesiana uma alternativa viável para tratar tais modelos. Em ambos os casos (EMV e Bayesiano) mostramos que critérios de informação usuais, tais como, AIC, BIC e HQ, podem ser usados para detectar afastamentos da normalidade e apontar o modelo que melhor representa os dados. Em resumo, propusemos estratégias inferenciais úteis para abordar análises de possíveis dados com características assimétricas.

Outro aspecto abordado foi na área dos modelos lineares mistos com erros de medida, em que desenvolvemos procedimentos de estimação, tanto em modelos estruturais como funcionais, através dos métodos SIMEX e escore corrigido. Notamos que o método escore corrigido é bastante consistente e não é afetado pelo erro de medição, ao contrario ao método SIMEX em que o seu desempenho é severamente afetado quando as variâncias dos erros de medição são relativamente grandes. Uma observação importante é que os estimadores das matrizes de dispersão dos efeitos aleatórios independem dos erros de medição, sendo assim, é possível estimar estes parâmetros usando um modelo linear misto tradicional usando procedimentos incorporados em programas computacionais, tais como, SAS e S-plus; logo, basta abordar a estimação simplesmente considerando um modelo linear geral sem erro de medida. No contexto de modelos lineares com erros nas variáveis normal-assimétrico, derivamos a densidade marginal das quantidades envolvidas no caso em que uma covariável é medida com erro. Desta forma, inferência clássica pode ser implementada usando técnicas de maximização direta, evitando o uso de técnicas mais complexas como, por exemplo, integração Monte Carlo.

6.2 Perspectivas para Pesquisas Futuras

Antes de propor algumas perspectivas para trabalhos futuros gostaria apresentar algumas linhas de um comentário feito por A. Azzalini no artigo de Arnold and Beaver (2002).

...Another, very general, comment which I would like to make is not really related to the paper itself, but to the literature on "skewed models" as a whole. This has developed quite fast in the last five years or so, but the body of theoretical result which have been developed has not been widely applied to real problems. The practical usefulness of these stochastic models is the real benchmark to test whether these results are going to have actual relevance in statistical practice. This is a plea for more applied work to make use of available theoretical tools.

Parece claro, então, que existe pouca literatura de aplicações de distribuições assimétricas em casos práticos. Assim, seria possível formular inúmeras propostas, mas por brevidade gostaria de mencionar algumas que dariam início a um plano de pesquisa mais abrangente e lógico, num âmbito mais geral de modelos estatísticos. Muitas das idéias relatadas a seguir estão sendo já implementadas.

Motivados por este trabalho, várias linhas de pesquisa podem ser propostas para fazer estudos inferenciais em diversos modelos estatísticos existentes na literatura, considerando que determinadas variáveis aleatórias (ou todas) presentes no modelo, seguem uma distribuição simétrica-assimétrica ou simétrica-assimétrica flexível (Ma & Genton, 2004; Apêndice D). No caso de um modelo linear normal-assimétrico é possível derivar a verossimilhança marginal das quantidades observáveis de modo que inferências possam ser tratadas usando técnicas de otimização numérica e programas existentes (alternativamente, algoritmos de estimação como o EM podem ser implementados). Adicionalmente, no caso normal-assimétrico univariado (Azzalini, 1985) é possível derivar uma expressão algébrica da matriz de informação observada, de modo que permita, por exemplo, fazer análise de influência local no modelo linear considerado (Lachos et al., 2004). Já no caso multivariado esta técnica pode ser implementada usando a matriz hessiana, calculada numericamente em linguagens como R, Matlab e S-plus. No caso de um modelo linear normal-assimétrico flexível é possível implementar o algoritmo EM-Monte Carlo para o processo de estimação, a matriz de informação observada pode ser calculada usando o estimador Sanduíche via integração Monte Carlo (veja Genton, 2004). Em modelos com distribuições mais complexas (ex. t-assimétrica, t-assimétrica flexível) é necessário usar técnicas bayesianas como o esquema MCMC para a qual o sistema WinBUGS pode ser de grande ajuda. Como mostrado neste trabalho, critérios de informação como AIC, BIC e HQ podem ser usados para detectar afastamentos da simetria e selecionar o modelo que melhor caracteriza os dados sob investigação.

A seguir descrevemos alguns modelos interessantes onde as metodologias descritas acima (e neste trabalho) podem ser aplicadas. Apresentamos estes modelos sob as suposições típicas encontradas na literatura.

Modelo de Calibração Comparativa Estrutural

Suponha que temos a nossa disposição $p \geq 2$ utilizados instrumentos para medir uma característica de interesse x em um grupo de n unidades experimentais. Seja x_i o verdadeiro valor (desconhecido) na unidade i e y_{ij} o valor medido obtido com o instrumento j na unidade i , $i = 1, \dots, n$ e $j = 1, \dots, p$. O modelo de calibração comparativa estrutural é

definido pelo seguinte modelo linear

$$\mathbf{Y}_i = \mathbf{a} + \mathbf{b}x_i + \boldsymbol{\epsilon}_i, \quad (3.1)$$

onde $\mathbf{a} = (0, \boldsymbol{\alpha}^T)^T = (0, \alpha_2, \dots, \alpha_p)^T$ e $\mathbf{b} = (1, \boldsymbol{\beta}^T)^T = (1, \beta_2, \dots, \beta_p)^T$ são vetores $p \times 1$; $\mathbf{Y}_i = (y_{i1}, \dots, y_{ip})^T$ e $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{ip})^T$ são vetores aleatórios $p \times 1$. É comum assumir que

$$\boldsymbol{\epsilon}_i \stackrel{\text{iid}}{\sim} N_p(\mathbf{0}, D(\boldsymbol{\phi})) \text{ e } x_i \stackrel{\text{iid}}{\sim} N_1(\mu_x, \sigma_x^2), \quad (3.2)$$

$i = 1, \dots, n$, with $D(\boldsymbol{\phi}) = \text{diag}(\phi_1, \dots, \phi_p)$ e $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)$. Estudos inferenciais nesta classe de modelos podem ser encontrados em Galea (1995).

Modelo de Grubbs Estrutural

O modelo de Grubbs é um caso particular do modelo de calibração comparativa, no qual supõe-se que $\beta_2 = \dots = \beta_p = 1$, isto é, temos o seguinte modelo linear:

$$\mathbf{Y}_i = \mathbf{a} + \mathbf{1}_p x_i + \boldsymbol{\epsilon}_i, \quad (3.3)$$

onde $\mathbf{1}_p$ ($p \times 1$) é um vetor com todos os elementos igual a 1, com

$$\boldsymbol{\epsilon}_i \stackrel{\text{iid}}{\sim} N_p(\mathbf{0}, D(\boldsymbol{\phi})) \text{ e } x_i \stackrel{\text{iid}}{\sim} N_1(\mu_x, \sigma_x^2), \quad i = 1, \dots, n. \quad (3.4)$$

Estudos inferenciais neste tipo de modelo são tratados com detalhe em Lachos (2002).

Modelo de Regressão com Erros nas Variáveis com Intercepto Nulo

Este modelo linear é definido por

$$\begin{aligned} \mathbf{x}_i &= \boldsymbol{\xi}_i + \boldsymbol{\delta}_i, \\ \mathbf{y}_i &= \boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i, \\ \boldsymbol{\eta}_i &= \mathbf{X}_i \boldsymbol{\beta}_i, \quad i = 1, \dots, p, \end{aligned}$$

onde $\mathbf{x}_i = (x_{i1}, \dots, x_{in_i})^T$, $\mathbf{y}_i = (\mathbf{y}_{i1}^T, \mathbf{y}_{i2}^T)^T = (y_{i11}, \dots, y_{i1n_i}, y_{i21}, \dots, y_{i2n_i})$, $\boldsymbol{\beta} = (\beta_{1i}, \beta_{2i})^T$, $\boldsymbol{\xi}_i = (\xi_{i1}, \dots, \xi_{in_i})^T$, $\boldsymbol{\delta} = (\delta_{i1}, \dots, \delta_{in_i})^T$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_{1i}^T, \boldsymbol{\epsilon}_{2i}^T)^T = (\epsilon_{1i1}, \dots, \epsilon_{1in_i}, \epsilon_{2i1}, \dots, \epsilon_{2in_i})^T$

$$\mathbf{X}_i = \begin{bmatrix} \xi_i & 0 \\ 0 & \xi_i \end{bmatrix}. \text{ Tipicamente supõe-se que}$$

$$\delta_{ij} \stackrel{\text{iid}}{\sim} N_1(0, \sigma^2), \quad \epsilon_{kij} \stackrel{\text{ind.}}{\sim} N_1(0, \sigma_{\epsilon_i}^2), \quad \delta_{ij} \text{ e } \epsilon_{kij} \text{ são não correlacionados}$$

e independentes de

$$\xi_{ij} \stackrel{\text{iid}}{\sim} N_1(\mu_x, \sigma_x^2), \quad j = 1, \dots, n_i, \quad i = 1, \dots, p, \quad k = 1, 2.$$

Para mais detalhes veja Aoki (2001).

Modelo Probito com Erros nas Variáveis

Consideremos respostas binárias onde o i -ésimo indivíduo na amostra com fator de risco x_i têm probabilidade de sucesso, dada por

$$P[Y_i = 1|x_i] = \Phi(\alpha + \beta x_i), \quad (3.5)$$

onde x_i não é observado diretamente, mas medido com erro, dado pela equação

$$w_i = x_i + u_i, \quad i = 1, \dots, n, \quad (3.6)$$

onde $u_i \stackrel{\text{iid}}{\sim} N_1(0, \sigma_u^2)$ é independente de $x_i \stackrel{\text{iid}}{\sim} N_1(\mu_x, \sigma_x^2)$, $i=1, \dots, n$.

Detalhes podem ser vistos em Carroll et al. (1984).

Apêndice A

Resultados de Álgebra Matricial

As provas dos seguintes resultados podem ser encontradas em Graham (1981, Cap. 6) e Searle et al. (1992).

- Sejam $\mathbf{A}$ uma matriz simétrica constante de dimensão $n \times n$ e $\mathbf{x}$, $\mathbf{t}$, $\mathbf{a}$ vetores de dimensão $n \times 1$ ($\mathbf{x}$ aleatório). Então,

$$E[\mathbf{x}^T \mathbf{A} \mathbf{x}] = \text{tr}(\mathbf{A}(E[\mathbf{x}]E[\mathbf{x}^T] + \text{Var}(\mathbf{x}))) = E[\mathbf{x}^T] \mathbf{A} E[\mathbf{x}] + \text{tr}[\mathbf{A} \text{Cov}(\mathbf{x})] \quad (\text{A-1})$$

$$\frac{\partial \mathbf{t}^T \mathbf{A} \mathbf{t}}{\partial \mathbf{t}} = 2\mathbf{A} \mathbf{t}, \quad \frac{\partial \mathbf{a}^T \mathbf{t}}{\partial \mathbf{t}} = \mathbf{a}, \quad \frac{\partial \mathbf{A} \mathbf{t}}{\partial \mathbf{t}^T} = \mathbf{A} \text{ e } \frac{\partial \mathbf{A} \mathbf{t}}{\partial \mathbf{t}} = \text{vec}(\mathbf{A}). \quad (\text{A-2})$$

- Sejam $\mathbf{A}$ uma matriz ($n \times n$) simétrica definida positiva e t um escalar. Então,

$$\frac{\partial \mathbf{A}^{-1}}{\partial t} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial t} \mathbf{A}^{-1}, \quad (\text{A-3})$$

$$\frac{\partial \text{vec}(\mathbf{A})}{\partial t} = \text{vec}\left(\frac{\partial \mathbf{A}}{\partial t}\right), \quad (\text{A-4})$$

$$\frac{\partial \text{tr}(\mathbf{A})}{\partial t} = \text{tr}\left(\frac{\partial \mathbf{A}}{\partial t}\right), \quad (\text{A-5})$$

$$\frac{\partial |\mathbf{A}|}{\partial t} = |\mathbf{A}| \text{tr}\left\{\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial t}\right\}, \quad (\text{A-6})$$

e

$$\frac{\partial \log |\mathbf{A}|}{\partial t} = \text{tr}\left\{\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial t}\right\}. \quad (\text{A-7})$$

- Sejam $\mathbf{X}$, $\mathbf{Y}$ e $\mathbf{Z}$ matrizes de dimensões $m \times n$, $n \times v$ e $p \times q$, respectivamente. Então

$$\frac{\partial (\mathbf{X} \mathbf{Y})}{\partial \mathbf{Z}} = \frac{\partial \mathbf{X}}{\partial \mathbf{Z}} (\mathbf{I}_q \otimes \mathbf{Y}) + (\mathbf{I}_p \otimes \mathbf{X}) \frac{\partial \mathbf{Y}}{\partial \mathbf{Z}}. \quad (\text{A-8})$$

- Sejam $\mathbf{X}$, $\mathbf{Y}$ e $\mathbf{Z}$ matrizes de dimensões $m \times n$, $u \times v$ e $p \times q$, respectivamente. Então,

$$\frac{\partial(\mathbf{X} \otimes \mathbf{Y})}{\partial \mathbf{Z}} = \frac{\partial \mathbf{X}}{\partial \mathbf{Z}} \otimes \mathbf{Y} + (\mathbf{I}_p \otimes \mathbf{U}_1) \left(\frac{\partial \mathbf{Y}}{\partial \mathbf{Z}} \otimes \mathbf{X} \right) (\mathbf{I}_q \otimes \mathbf{U}_2), \quad (\text{A-9})$$

onde $\mathbf{U}_1$ é a matriz de permutações associada com uma matriz $m \times u$ e $\mathbf{U}_2$ é a matriz de permutações associada com uma matriz $n \times v$.

Os resultados seguintes estão relacionados com algumas propriedades do produto de Kronecker. Para mais detalhes, veja Magnus & Neudecker (1979) e Mardia et al. (1979)

- Sejam $\mathbf{A}$, $\mathbf{B}$ e $\mathbf{C}$ matrizes de dimensões $m \times n$, $n \times v$ e $v \times p$, respectivamente. Então,

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B}), \quad (\text{A-10})$$

$$(\text{vec}(\mathbf{A}^T))^T \text{vec}(\mathbf{B}) = \text{tr}(\mathbf{AB}). \quad (\text{A-11})$$

- Sejam $\mathbf{x}$ e $\mathbf{y}$ vetores arbitrários. Então,

$$\mathbf{x} \otimes \mathbf{y} = \text{vec}(\mathbf{yx}^T), \quad (\text{A-12})$$

$$\mathbf{x} \otimes \mathbf{y}^T = \mathbf{xy}^T = \mathbf{y}^T \otimes \mathbf{x}. \quad (\text{A-13})$$

- Se os produtos matriciais existem, então

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD}), \quad (\text{A-14})$$

$$(\mathbf{A} \otimes \mathbf{B})^T = \mathbf{A}^T \otimes \mathbf{B}^T, \quad (\text{A-15})$$

$$\mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) = (\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C}, \quad (\text{A-16})$$

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}, \quad (\text{A-17})$$

$$\text{tr}(\mathbf{A} \otimes \mathbf{B}) = \text{tr}(\mathbf{A})\text{tr}(\mathbf{B}), \quad (\text{A-18})$$

$$\text{tr}(\mathbf{ABCD}) = \text{vec}(\mathbf{D})^T (\mathbf{A} \otimes \mathbf{C}^T) \text{vec}(\mathbf{B}^T). \quad (\text{A-19})$$

O resultado seguinte está relacionado com o teorema da inversa binomial de matrizes:

- Se todas as inversas necessárias existem, então, para $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ e $\mathbf{D}$ de dimensões

$p \times p$, $p \times n$, $n \times n$ e $n \times p$, respectivamente,

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}. \quad (A-20)$$

A relação seguinte pode ser usada para encontrar os cumulantes das formas quadráticas de um vetor aleatório normal-assimétrico. (Ghazal, 1996)

- Se $I - C > 0$. Então,

$$-\log |I - C| = \text{tr}C + \frac{1}{2}\text{tr}C^2 + \frac{1}{3}\text{tr}C^3 + \dots \quad (A-21)$$

```
function g=EMistorest(nj,y,x,z,C,rest)
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% ESTE PROGRAMA CALCULA OS EMV COM RESTRICOES DE IGUALDADE
%      EM UM MODELO LINEAR MISTO COMO NA SECAO 1.2.3
% nj= vetor coluna contendo os tamanhos de cada grupo
% y= vetor coluna contendo as observacoes da variavel resposta
% x= matriz de planejamento empilhada com todos os grupos
% z= matriz de planejamento do efeito aleatorio
% C=matriz de contrastes
% rest=vetor d
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
m=size(nj,1); N=size(y,1); p=size(x,2); q=size(z,2);
inicial=valinicial(y,x,q,m); % calcula as estimativas iniciais
k=0;
    for j=1:q
        for i=j:q
            k=k+1;
            d(k,1)=D(j,i);
        end
    end
theta=[beta;sigmae;d]; criterio=norm(theta); cont=0;

while criterio>0.0000001
    cont=cont+1;
```

```

for j=1:m
    y1=y(sum(nj(1:j-1))+1:sum(nj(1:j)),1);
    z1=z(sum(nj(1:j-1))+1:sum(nj(1:j)),:);
    x1=x(sum(nj(1:j-1))+1:sum(nj(1:j)),:);
    Omega((j-1)*q+1:j*q,:)=D-D*z1'*(z1*D*z1'+sigmae*eye(nj(j)))^(-1)*z1*D;
    b(:,j)=D*z1'*(z1*D*z1'+sigmae*eye(nj(j)))^(-1)*(y1-x1*beta);
end
sum1=zeros(p,p);sum2=zeros(p,1);sum3=0;sum4=zeros(q,q);sum5=zeros(p,p);

for j=1:m
    y1=y(sum(nj(1:j-1))+1:sum(nj(1:j)),1);
    z1=z(sum(nj(1:j-1))+1:sum(nj(1:j)),:);
    x1=x(sum(nj(1:j-1))+1:sum(nj(1:j)),:);
    sum1=sum1+x1'*x1;
    sum2=sum2+x1'*(y1-z1*b(:,j));
    sum3=sum3+(y1-x1*beta-z1*b(:,j))'*(y1-x1*beta-z1*b(:,j))+
        sum(diag(Omega((j-1)*q+1:j*q,:)*z1'*z1));
    sum4=sum4+b(:,j)*b(:,j)'+Omega((j-1)*q+1:j*q,:);
end
beta0=inv(sum1)*sum2;
beta=beta0+inv(sum1)*C'*(C*inv(sum1)*C')^(-1)*(rest-C*beta0);
sigmae=sum3/N; D=sum4/m; s=0;
for j=1:q
    for i=j:q
        s=s+1;
        d(s,1)=D(j,i);
    end
end
theta=[beta;sigmae;d]; param(cont,:)=theta';
if cont>2
    criterio=norm(param(cont,:)-param(cont-1,:));
end
end
g=[theta];

```

Apêndice B

Prova do Teorema 3.2

Para obter a distribuição marginal de $\mathbf{Y}_j$ eliminamos o sub-índice j , correspondente ao j -ésimo grupo (ou indivíduo), a fim de simplificar a notação. De (3.33)-(3.34) e da definição de distribuição normal-assimétrica multivariada em (2.6), segue que a densidade marginal de $\mathbf{Y}$ é obtida ao calcular a seguinte integral

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\lambda}) &= \int_{\mathbb{R}^q} f(\mathbf{y}|\mathbf{b}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}_e) f(\mathbf{b}|\boldsymbol{\alpha}, \boldsymbol{\lambda}_b) d\mathbf{b} \\ &= \int_{\mathbb{R}^q} 2^2 \phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, \boldsymbol{\psi}) \Phi_1(\boldsymbol{\lambda}_e^T \boldsymbol{\psi}^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})) \phi_q(\mathbf{b}|\mathbf{0}, \mathbf{D}) \Phi_1(\boldsymbol{\lambda}_b^T \mathbf{D}^{-1/2} \mathbf{b}) d\mathbf{b}, \end{aligned} \quad (\text{B-1})$$

cuja solução é baseada no seguinte lema:

Lema B.1. *Sob a notação considerada acima, segue que*

$$\phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, \boldsymbol{\psi}) \phi_q(\mathbf{b}|\mathbf{0}, \mathbf{D}) = \phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}) \phi_q(\mathbf{b}|\boldsymbol{\mu}_1(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \boldsymbol{\Lambda}) \quad (\text{B-2})$$

e

$$\Phi_1(\boldsymbol{\lambda}_e^T \boldsymbol{\psi}^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})) \Phi_1(\boldsymbol{\lambda}_b^T \mathbf{D}^{-1/2} \mathbf{b}) = \Phi_2(-\boldsymbol{\Gamma}\mathbf{b} | -\boldsymbol{\mu}_2(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \mathbf{I}_2), \quad (\text{B-3})$$

onde

$$\boldsymbol{\mu}_1 = \boldsymbol{\Lambda} \mathbf{Z}^T \boldsymbol{\psi}^{-1}, \quad \boldsymbol{\Sigma} = \boldsymbol{\psi} + \mathbf{Z} \mathbf{D} \mathbf{Z}^T, \quad \boldsymbol{\Lambda} = (\mathbf{D}^{-1} + \mathbf{Z}^T \boldsymbol{\psi}^{-1} \mathbf{Z})^{-1}, \quad (\text{B-4})$$

$$\boldsymbol{\mu}_2 = \begin{pmatrix} \boldsymbol{\lambda}_e^T \boldsymbol{\psi}^{-1/2} \\ 0 \end{pmatrix} \quad e \quad \boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{\lambda}_e^T \boldsymbol{\psi}^{-1/2} \mathbf{Z} \\ -\boldsymbol{\lambda}_b^T \mathbf{D}^{-1/2} \end{pmatrix}. \quad (\text{B-5})$$

Demonstração. O resultado (B-2) segue do Lema 2.2. O resultado (B-3) é provado no-

tando que se U e V são variáveis aleatórias i.i.d. $N(0, 1)$, então (B-3) pode ser escrita como

$$P(U \leq \lambda_e^T \psi^{-1/2}(\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{b}))P(V \leq \lambda_b^T \mathbf{D}^{-1/2}\mathbf{b}) = P(\mathbf{T} \leq \mu_2(\mathbf{y} - \mathbf{X}\beta)),$$

onde $\mathbf{T} = \mathbf{W} + \Gamma\mathbf{b}$, com $\mathbf{W} = (U, V)^T \sim N_2(\mathbf{0}, \mathbf{I}_2)$ e μ_2 , Γ definidos como em (B-5). Assim, uma vez que $\mathbf{T} \sim N_2(\Gamma\mathbf{b}, \mathbf{I}_2)$, temos que

$$P(\mathbf{T} \leq \mu_2(\mathbf{y} - \mathbf{X}\beta)) = \Phi_2(\mu_2(\mathbf{y} - \mathbf{X}\beta)|\Gamma\mathbf{b}, \mathbf{I}_2) = \Phi_2(-\Gamma\mathbf{b} | -\mu_2(\mathbf{y} - \mathbf{X}\beta), \mathbf{I}_2),$$

que conclui a prova. $\square$

Prova do Teorema:

Demonstração. De (B-1), (B-2) e (B-3) segue que

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\lambda}) &= \int_{\mathbb{R}^q} 2^2 \phi_n(\mathbf{y}|\mathbf{X}\beta, \Sigma) \phi_q(\mathbf{b}|\mu_1(\mathbf{y} - \mathbf{X}\beta), \Lambda) \Phi_2(-\Gamma\mathbf{b} | -\mu_2(\mathbf{y} - \mathbf{X}\beta), \mathbf{I}_2) d\mathbf{b} \\ &= 2^2 \phi_n(\mathbf{y}|\mathbf{X}\beta, \Sigma) E[\Phi_2(-\Gamma\mathbf{W} | -\mu_2(\mathbf{y} - \mathbf{X}\beta), \mathbf{I}_2)], \end{aligned}$$

onde $\mathbf{W} \sim N_q(\mu_1(\mathbf{y} - \mathbf{X}\beta), \Lambda)$. Concluí-se a prova recorrendo ao Lema 2.1. $\square$

Prova do Corolário 3.2

Demonstração. Note primeiro que

$$(\mu_2 - \Gamma\mu_1)(\mathbf{y} - \mathbf{X}\beta) = \begin{pmatrix} \lambda_e^T \psi^{-1/2}(\psi - \mathbf{Z}\Lambda\mathbf{Z}^T)\psi^{-1}(\mathbf{y} - \mathbf{X}\beta) \\ \lambda_b^T \mathbf{D}^{-1/2}\Lambda\mathbf{Z}^T\psi^{-1}(\mathbf{y} - \mathbf{X}\beta) \end{pmatrix}$$

e

$$\mathbf{I}_2 + \Gamma\Lambda\Gamma^T = \begin{pmatrix} 1 + \lambda_e^T \psi^{-1/2}\mathbf{Z}\Lambda\mathbf{Z}^T\psi^{-1/2}\lambda_e & -\lambda_e^T \psi^{-1/2}\mathbf{Z}\Lambda\mathbf{D}^{-1/2}\lambda_b \\ -\lambda_b^T \mathbf{D}^{-1/2}\Lambda\mathbf{Z}^T\psi^{-1/2}\lambda_e & 1 + \lambda_b^T \mathbf{D}^{-1/2}\Lambda\mathbf{D}^{-1/2}\lambda_b \end{pmatrix},$$

que é uma matriz diagonal quando $\lambda_e = \mathbf{0}$ ou $\lambda_b = \mathbf{0}$. Assim, para $\lambda_e = \mathbf{0}$, a parte assimétrica de (3.35) pode ser calculada como

$$\begin{aligned} \Phi_2((\mu_2 - \Gamma\mu_1)(\mathbf{y} - \mathbf{X}\beta)|\mathbf{0}, \mathbf{I}_2 + \Gamma\Lambda\Gamma^T) &= \Phi_2((\mathbf{I}_2 + \Gamma\Lambda\Gamma^T)^{-1/2}(\mu_2 - \Gamma\mu_1)(\mathbf{y} - \mathbf{X}\beta)) \\ &= \frac{1}{2} \Phi_1\left(\frac{\lambda_b^T \mathbf{D}^{-1/2}\Lambda\mathbf{Z}^T\psi^{-1}(\mathbf{y} - \mathbf{X}\beta)}{\sqrt{1 + \lambda_b^T \mathbf{D}^{-1/2}\Lambda\mathbf{D}^{-1/2}\lambda_b}}\right), \end{aligned}$$

onde algumas manipulações algébricas produzem $\Lambda Z^T = DZ^T \Sigma^{-1} \psi$. Similarmente, para $\lambda_b = 0$, temos que

$$\Phi_2(\mu_2 - \Gamma \mu_1 | 0, I_2 + \Gamma \Lambda \Gamma^T) = \frac{1}{2} \Phi_1 \left(\frac{\lambda_e^T \Psi^{-1/2} (\psi - Z \Lambda Z^T) \psi^{-1} (y - X \beta)}{\sqrt{1 + \lambda_e^T \psi^{-1/2} Z \Lambda Z^T \psi^{-1/2} \lambda_e}} \right),$$

e a prova é concluída notando que $\psi - Z \Lambda Z^T = \psi \Sigma^{-1} \psi$. □

Apêndice C

Prova do Teorema 4.1

Para obter a distribuição marginal de $\mathbf{Y}_j$, eliminamos o sub-índice j , correspondente ao j -ésimo grupo (ou indivíduo) a fim de simplificar a notação. De (4.16), (4.17) e da definição de distribuição normal-assimétrica multivariada em (2.48) com $k = n$ e $\Delta_b = \text{diag}(\delta)$, segue que a densidade marginal de $\mathbf{Y}$ é obtida calculando a seguinte integral:

$$\begin{aligned}
 f_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\lambda}) &= \int_{\mathbb{R}^q} f(\mathbf{y}|\mathbf{b}, \boldsymbol{\beta}, \sigma_e^2, \delta_e) f(\mathbf{b}|\boldsymbol{\alpha}, \delta_b) d\mathbf{b} \\
 &= \int_{\mathbb{R}^q} 2^{n+q} \phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, (\delta_e^2 + \sigma_e^2)\mathbf{I}_n) \\
 &\quad \times \phi_q(\mathbf{b}|\mathbf{0}, \Delta_b^2 + \mathbf{D}) \Phi_q(\Delta_b(\Delta_b^2 + \mathbf{D})^{-1}\mathbf{b}|\mathbf{0}, (\mathbf{I}_q + \Delta_b\mathbf{D}^{-1}\Delta_b)^{-1}) \\
 &\quad \times \Phi_n\left(\frac{\delta_e}{\sqrt{\sigma_e^2(\delta_e^2 + \sigma_e^2)}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})|\mathbf{0}, \mathbf{I}_n\right) d\mathbf{b}, \tag{C-1}
 \end{aligned}$$

para o qual a seguinte proposição deve ser usada.

Proposição C.1. *Sob a notação considerada em (C-1), segue que*

$$\phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, (\delta_e^2 + \sigma_e^2)\mathbf{I}_n) \phi_q(\mathbf{b}|\mathbf{0}, \Delta_b^2 + \mathbf{D}) = \phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}) \phi_q(\mathbf{b}|\boldsymbol{\mu}_1, \boldsymbol{\Lambda}) \tag{C-2}$$

e

$$\begin{aligned}
 &\Phi_n\left(\frac{\delta_e}{\sqrt{\sigma_e^2(\delta_e^2 + \sigma_e^2)}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})|\mathbf{0}, \mathbf{I}_n\right) \Phi_q(\Delta_b(\Delta_b^2 + \mathbf{D})^{-1}\mathbf{b}|\mathbf{0}, (\mathbf{I}_q + \Delta_b\mathbf{D}^{-1}\Delta_b)^{-1}) \\
 &= \Phi_{n+q}(-\boldsymbol{\Gamma}\mathbf{b} | -\boldsymbol{\mu}_2, \mathbf{R}), \tag{C-3}
 \end{aligned}$$

onde

$$\begin{aligned}\Sigma &= (\delta_e^2 + \sigma_e^2)\mathbf{I}_n + \mathbf{Z}(\Delta_b^2 + \mathbf{D})\mathbf{Z}^T, \quad \mu_1 = \frac{\Lambda \mathbf{Z}^T(\mathbf{y} - \mathbf{X}\beta)}{\delta_e^2 + \sigma_e^2}, \quad \mu_2 = \begin{pmatrix} \frac{\delta_e}{\sqrt{\sigma_e^2(\delta_e^2 + \sigma_e^2)}}(\mathbf{y} - \mathbf{X}\beta) \\ \mathbf{0}_q \end{pmatrix}, \\ \Gamma &= \begin{pmatrix} \frac{\delta_e}{\sqrt{\sigma_e^2(\delta_e^2 + \sigma_e^2)}}\mathbf{Z} \\ -\Delta_b(\Delta_b^2 + \mathbf{D})^{-1} \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & (\mathbf{I}_q + \Delta_b \mathbf{D}^{-1} \Delta_b)^{-1} \end{pmatrix}, \\ \Lambda &= \left((\Delta_b^2 + \mathbf{D})^{-1} + \frac{\mathbf{Z}^T \mathbf{Z}}{\delta_e^2 + \sigma_e^2} \right)^{-1}. \end{aligned} \quad (\text{C-4})$$

Demonstração. O resultado (C-2) segue do Lema 2.2. O resultado (C-3) é provado notando que se $\mathbf{U} \sim N_n(\mathbf{0}, \mathbf{I}_n)$ e $\mathbf{V} \sim N_q(\mathbf{0}, (\mathbf{I}_q + \Delta_b \mathbf{D}^{-1} \Delta_b)^{-1})$ e são independentes, então (C-3) pode ser escrita como

$$P(\mathbf{U} \leq \mathbf{Y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{b})P(\mathbf{V} \leq \Delta_b(\Delta_b^2 + \mathbf{D})^{-1}\mathbf{b}) = P(\mathbf{T} \leq \mu_2),$$

onde $\mathbf{T} = \mathbf{W} + \Gamma\mathbf{b}$, com $\mathbf{W} = (\mathbf{U}, \mathbf{V})^T \sim N_{n+q}(\mathbf{0}, \mathbf{R})$, e μ_2 , Γ , $\mathbf{R}$ como definidos em (C-4). Assim, visto que $\mathbf{T} \sim N_{n+q}(\Gamma\mathbf{b}, \mathbf{R})$, temos que

$$P(\mathbf{T} \leq \mu_2) = \Phi_{n+q}(\mu_2 | \Gamma\mathbf{b}, \mathbf{R}) = \Phi_{n+q}(-\Gamma\mathbf{b} | -\mu_2, \mathbf{R}),$$

que conclui a prova. □

Prova do Teorema: De (C-1), (C-2) e (C-3) segue que

$$\begin{aligned}f_Y(\mathbf{y} | \boldsymbol{\theta}, \lambda) &= \int_{\mathbb{R}^q} 2^{n+q} \phi_n(\mathbf{y} | \mathbf{X}\beta, \Sigma) \phi_{n+q}(\mathbf{b} | \mu_1, \Lambda) \Phi_2(-\Gamma\mathbf{b} | -\mu_2, \mathbf{R}) d\mathbf{b} \\ &= 2^2 \phi_n(\mathbf{y} | \mathbf{X}\beta, \Sigma) E[\Phi_2(-\Gamma\mathbf{N} | -\mu_2, \mathbf{R})],\end{aligned}$$

onde $\mathbf{N} \sim N_q(\mu_1, \Lambda)$. Concluimos a prova usando Lema 2.1.

Código WinBUGS para ajustar o Modelo 1 aos dados de Barnett (Seção 4.2.3)

```
model {
  aux10<-sigmab0*(1+pow(lamb, 2))
  aux20<-sigmae0*(1+pow(lambe, 2))
```

```

for(i in 1:N){
  wb[i]~dnorm(0, 1) I(0,)
  a[i]<-deltab*sqrt(sigmab)*wb[i]
  b[i]~dnorm(a[i], aux10)
  for(j in 1:T){
    w[i, j] ~ dnorm(0, 1) I(0,)
    mu[i, j] <- beta[j]+b[i]+sqrt(sigmae)*deltae*w[i,j]
    y[i, j]~dnorm(mu[i, j], aux20)
  }
}

beta[1]~dnorm(0, 0.01)
beta[2]~dnorm(0, 0.01)
beta[3]~dnorm(0,0.01)
beta[4]~dnorm(0, 0.01)
sigmab0~dgamma(0.01,0.01)
sigmae0~dgamma(0.01, 0.01)
lamb~dnorm(0,0.01) I(0,)
lambe~dnorm(0, 0.01) I(0,)
sigmab<-1/sigmab0 # variancia do erro
sigmae<-1/sigmae0 #varianciado efeito aleatorio
deltab<-lamb/sqrt(1+pow(lamb, 2))
deltae<-lambe/sqrt(1+pow(lambe,2)) }

```

Código WinBUGS para ajustar o Modelo 1 aos dados de colesterol (Seção 4.5)

```

model {
  for(j in 1:N){
    for(k in 1:2){
      wb[j,k]~dnorm(0, 1) I(0,)
      a[j,k]<-deltab[k]*wb[j,k]
    }
    b[ j,1:2]~dmnorm(a[j,1:2],Omega[,])
  }
  for(i in 1:T){
    w[j,i]~ dnorm(0, 1) I(0,)
    mu[j , i]<- beta0+sex[(6*(j-1)+1)]*beta1+age[(6*(j-1)+1)]*beta2
      +b[j,1]+year[(6*(j-1)+i)]*b[j,2]+delta*w[j,i]
  }
}

```



```

        y[j, i]~ dnorm(mu[j, i],tau)
      }
    }
    Omega[1:2,1:2] ~ dwish(R[ , ],2)
    v[1:2,1:2]<-inverse(Omega[,])
    beta0~dnorm(0,0.01)
    beta1~dnorm(0,0.01)
    beta2~dnorm(0,0.01)
    tau~dgamma(0.01,0.01)
    tau1<-1/tau
    deltab[1]~dnorm(0,0.01)I(0,)
    deltab[2]~dnorm(0,0.01)I(0,)
    delta~dnorm(0,0.01)I(0,) }

```

Apêndice D

Distribuições Simétricas-Assimétrica

D.1 Distribuição Simétrica-Assimétrica Univariada

Uma variável aleatória Z pertence à classe simétrica-assimétrica (SS_1) de distribuições com parâmetro de locação μ , parâmetro de escala σ^2 e parâmetro de assimetria λ , se a função de densidade é dada por

$$f_Z(z) = \frac{2}{\sigma} g\left(\frac{z - \mu}{\sigma}\right) H\left(\lambda \frac{z - \mu}{\sigma}\right), \quad Z \in \mathbb{R}, \quad (\text{D-1})$$

onde g é uma fdp padronizada, centrada em zero; H é a fda de alguma variável aleatória contínua simétrica em torno de zero. A função $\pi_k(z) = H(z)$ é chamada de função de assimetria e satisfaz $\pi(-z) = 1 - \pi(z)$. Assim, asseguramos que a integral de (D-1) seja igual a 1. Quando $\lambda = 0$, $H(0) = 1/2$ e (D-1) se reduz à função densidade simétrica g . Consideremos a situação em que H é a fda correspondente à fdp simétrica padronizada g , que é uma escolha comum na prática. Usamos a notação $Z \sim SS_1(\mu, \sigma^2, \lambda, g)$ para denotar esta classe de distribuições, que se reduz a $Z \sim SS_1(\lambda, g)$ quando $\mu = 0$ e $\sigma^2 = 1$.

Em particular, uma variável aleatória Z segue uma distribuição normal-assimétrica (SN_1 Azzalini, 1985) se a densidade de Z é dada por

$$f_Z(z) = \frac{2}{\sigma} \phi_1\left(\frac{z - \mu}{\sigma}\right) \Phi_1\left(\lambda \frac{z - \mu}{\sigma}\right), \quad Z \in \mathbb{R}. \quad (\text{D-2})$$

Similarmente, Z segue uma distribuição univariada t-assimétrica (ST_1) com ν graus

de liberdade, se a função densidade de Z é dada por

$$f_Z(z) = \frac{2}{\sigma} t_1\left(\frac{z - \mu}{\sigma}, \nu\right) T_1\left(\lambda \frac{z - \mu}{\sigma}, \nu\right), \quad Z \in \mathbb{R} \quad (\text{D-3})$$

onde $t_1(\cdot, \nu)$ e $T_1(\cdot, \nu)$, denotam a fdp e a fda, respectivamente, da distribuição t de Student padronizada com ν graus de liberdade. Usamos a notação $Z \sim ST_1(\mu, \sigma^2, \nu, \lambda)$ e $Z \sim ST_1(\lambda, \nu)$ quando $\mu = 0$ e $\sigma^2 = 1$.

D.2 Distribuição Simétrica-Assimétrica Multivariada

Um vetor aleatório n -dimensional $\mathbf{Z}$ segue uma distribuição simétrica-assimétrica com vetor de locação $\boldsymbol{\mu} \in \mathbb{R}^n$, matriz de dispersão $\boldsymbol{\Sigma}$ (uma matriz definida positiva $n \times n$) e vetor de assimetria $\boldsymbol{\lambda} \in \mathbb{R}^n$, se sua fdp é dada por

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{2}{|\boldsymbol{\Sigma}|^{1/2}} g(\boldsymbol{\Sigma}^{-1/2}(\mathbf{z} - \boldsymbol{\mu})) H(\boldsymbol{\lambda}^T \boldsymbol{\Sigma}^{-1/2}(\mathbf{z} - \boldsymbol{\mu})), \quad \mathbf{z} \in \mathbb{R}^n, \quad (\text{D-4})$$

onde g é uma fdp com simetria elíptica centrada em $\mathbf{0}$ e $H : \mathbb{R} \rightarrow [0, 1]$ é a fda de alguma variável contínua univariada simétrica em torno de zero. Na prática, H é a fda univariada correspondente à versão univariada da fdp simétrica g .

Em particular, um vetor aleatório n -dimensional $\mathbf{Z}$ segue uma distribuição normal-assimétrica com vetor de locação $\boldsymbol{\mu} \in \mathbb{R}^n$, matriz de dispersão $\boldsymbol{\Sigma}$ (uma matriz definida positiva $n \times n$) e vetor de assimetria $\boldsymbol{\lambda} \in \mathbb{R}^n$, se sua fdp é dada por

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{2}{|\boldsymbol{\Sigma}|^{1/2}} \phi_n(\boldsymbol{\Sigma}^{-1/2}(\mathbf{z} - \boldsymbol{\mu})) \Phi_1(\boldsymbol{\lambda}^T \boldsymbol{\Sigma}^{-1/2}(\mathbf{z} - \boldsymbol{\mu})), \quad \mathbf{z} \in \mathbb{R}^n. \quad (\text{D-5})$$

A distribuição t-assimétrica multivariada é definida por

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{2}{|\boldsymbol{\Sigma}|^{1/2}} t_n(\boldsymbol{\Sigma}^{-1/2}(\mathbf{z} - \boldsymbol{\mu}), \nu) T_1(\boldsymbol{\lambda}^T \boldsymbol{\Sigma}^{-1/2}(\mathbf{z} - \boldsymbol{\mu}), \nu), \quad \mathbf{z} \in \mathbb{R}^n, \quad (\text{D-6})$$

onde $t_p(\cdot, \nu)$ denota a fdp padronizada de um vetor t de Student multivariado e T_1 denota a fda de uma variável t de Student com ν graus de liberdade.

D.3 Distribuição Simétrica-Assimétrica Flexível Univariada

Uma variável aleatória Z pertence à classe das distribuições simétricas-assimétricas flexíveis (SSF_1) com parâmetro de locação μ , parâmetro de escala σ^2 e polinômio de ordem ímpar S_k , se a função densidade de Z é dada por

$$f_Z(z) = \frac{2}{\sigma} g\left(\frac{z-\mu}{\sigma}\right) H\left(S_k\left(\frac{z-\mu}{\sigma}\right)\right) \quad Z \in \mathbb{R}, \quad (\text{D-7})$$

onde g é uma fdp simétrica padronizada centrada em zero; H é a fda de alguma variável aleatória contínua simétrica em torno de zero; e S_k é um polinômio de grau k ímpar indexado por coeficientes $\lambda_k = (\lambda_1, \lambda_3, \dots, \lambda_k)^T$, onde k atua como um parâmetro que controla o grau de flexibilidade da distribuição. A função $\pi_k(z) = H(S_k(z))$ é chamada função de assimetria e satisfaz $\pi_k(-z) = 1 - \pi_k(z)$. Desta forma asseguramos que a integral de (D-7) é igual a um. Quando $S_k(z) \equiv 0$, então $H(0) = 1/2$, e (D-7) se reduz à fdp simétrica g . Aqui, restringimos nosso interesse em H correspondente à fdp g . Usamos $Z \sim SSFL_1(\mu, \sigma^2, S_k, g)$ para denotar esta classe de distribuições, que se reduz a $Z \sim SSFL_1(S_k, g)$ quando $\mu = 0$ e $\sigma^2 = 1$.

Em particular, uma variável aleatória Z segue uma distribuição normal-assimétrica flexível ($SNFL_1$) se a função densidade de Z é dada por

$$f_Z(z) = \frac{2}{\sigma} \phi_1\left(\frac{z-\mu}{\sigma}\right) \Phi_1\left(S_k\left(\frac{z-\mu}{\sigma}\right)\right), \quad Z \in \mathbb{R} \quad (\text{D-8})$$

Usamos a notação $Z \sim SNFL_1(\mu, \sigma^2, S_k)$, e $Z \sim SNFL_1(S_k)$ quando $\mu = 0$ e $\sigma^2 = 1$. Similarmente, Z segue uma distribuição t-assimétrica flexível ($STFL_1$), com ν graus de liberdade, se a função densidade de Z é dada por

$$f_Z(z) = \frac{2}{\sigma} t_1\left(\frac{z-\mu}{\sigma}, \nu\right) T_1\left(S_k\left(\frac{z-\mu}{\sigma}\right), \nu\right), \quad (\text{D-9})$$

Usamos a notação $Z \sim STFL_1(\mu, \sigma^2, S_k, \nu)$, e $Z \sim STFL_1(S_k, \nu)$ quando $\mu = 0$ e $\sigma^2 = 1$.

Pode ser mostrado que distribuições $SNFL_1$ de ordem $k = 3$ podem ter no máximo duas modas, enquanto que é unimodal para $k = 1$, de modo que o valor de k necessário para alcançar uma representação adequada é informativo neste caso. Como um exemplo,

a distribuição $STFL_1(S_k, \nu)$ com $k = 5$ tem a forma $2t_1(z, \nu)T_1(\lambda_1 z_1 + \lambda_3 z^3 + \lambda_5 z^5, \nu)$. Aqui, o primeiro termo controla o comportamento da cauda, enquanto que o segundo controla a forma da distribuição. Em particular, se $\lambda_3 = \lambda_5 = 0$ então a densidade se reduz à distribuição t-assimétrica (ST_1) que é unimodal; se adicionalmente $\lambda = 0$ ou equivalentemente $S_k \equiv 0$, então a densidade é simétrica (t de Student). Na Figura D.1 ilustramos a forma flexível da $STFL_1(S_5, \nu)$ para diferentes valores de λ_5 e ν . Note que a mesma pode exibir assimetria, unimodalidade, bimodalidade and trimodalidade dependendo dos valores dos parâmetros λ_1, λ_3 e λ_5 . Quando $\nu = 1$ obtemos a distribuição Cauchy-assimétrica e quando $\nu = 100$ obtemos aproximadamente a distribuição normal-assimétrica (SN_1) de Azzalini (1985). Assim, uma característica interessante desta densidade é que, problemas considerando assimetria e multimodalidade podem ser diretamente expressas em termos dos parâmetros em S_k e questões considerando caudas pesadas podem ser diretamente expressas em termos de ν . Em geral, quando o grau k do polinômio ímpar em (D-7) é grande, o número de modas possíveis na fdp cresce. A versão multivariada desta classe de distribuições pode ser encontrada em Ma & Genton (2004) e em Genton (2004).

Figura D.1: Alguns membros da distribuição $STFL_1$ com $\mu = 0$, $\sigma^2 = 1$, $k = 5$ e diferentes valores de ν (df, graus de liberdade) : (a) $\lambda_1 = 3, \lambda_3 = 0, \lambda_5 = 0$ (t-assimétrica); (b) $\lambda_1 = 2, \lambda_3 = -1, \lambda_5 = 0$; (c) $\lambda_1 = 1, \lambda_3 = -2, \lambda_5 = 0$ e (d) $\lambda_1 = 1, \lambda_3 = -5, \lambda_5 = 2$.

Bibliografia

- [1] Aoki, R. (2001). *Modelos de Regressão com Erros nas Variáveis com Intercepto Nulo*. Tese de doutorado. IME - Universidade de São Paulo.
- [2] Arellano-Valle, R. B. (1994). *Distribuições Elípticas: Propriedades, Inferência e Aplicações a Modelos de Regressão*. Tese de doutorado - IME - Universidade de São Paulo.
- [3] Arellano-Valle, R.B. and del Pino, G. (2003). From symmetric to asymmetric distributions: A unified approach. *CRC/Chapman-Hall. Ed. Vol. by M.G. Genton on Skew-distributions and their applications: A Journey Beyond Normality*.
- [4] Arellano-Valle, R.B., del Pino, G. and San Martin, E. (2002). Definition and probabilistic properties of skew distributions. *Statistics and Probability Letters* **58**, 111-121.
- [5] Arellano-Valle, R.B. and Genton, M. G. (2003). Fundamental skew distributions. *Institute of Statistics Mimeo Series #2551*, under review,
(<http://www.stat.ncsu.edu/~mrggenton/publications.html>)
- [6] Arellano-Valle R.B., Bolfarine, H. and Lachos, V.H. (2005). Skew-normal linear mixed models. *Journal of Data Science* (to appear).
- [7] Arellano-Valle R.B., Bolfarine, H. and Lachos, V.H. (2004). Bayesian inference for Skew-normal linear mixed models, RT-MAE 2004-9, IME-USP.
- [8] Arellano-Valle, R.B., Ozan S., Bolfarine, H. and Lachos, V.H. (2005). Skew normal measurement error models. *Journal of Multivariate Analysis* (to appear).
- [9] Arnold, B. C. and Beaver, R. J. (2000). Some skewed multivariate distribution. *American Journal of Mathematical and Management Sciences*, **20**, 27-38.

- [10] Arnold, B. C. and Beaver, R. J. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Test*, **11**,1, 7-54.
- [11] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171-178.
- [12] Azzalini, A. and Capitanio, A. (1999). Statistical Applications of the Multivariate Skew Normal Distributions. *Journal Royal Statistics Society*, **61**, 579-602.
- [13] Azzalini, A. and Dalla-Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**, 715-726.
- [14] Azzalini, A. and Kotz, S. (2002). Log-skew-normal and log-skew-t distributions as models for family income data. *Technical Report*, Available at (<http://azzalini.stat.unipd.it/SN/income-distr.pdf>.)
- [15] Barnett, V. D. (1969). Simultaneous pairwise linear structural relationships. *Biometrics* **25**, 129-142.
- [16] Bolfarine, H. and Arellano-Valle, R. (1999). Weak nondifferential measurement error models. *Statistics and Probability Letters*, **40**, 279-287.
- [17] Bolfarine, H. e Galea-Rojas, M. (1995). Functional comparative calibration using on EM algorithm - comment. *Biometrics* **51**, 4, 1579 - 1580.
- [18] Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*(3rd ed.) San Fransisco: Holden-Day.
- [19] Branco, M. and Dey, D. (2001). A general class of multivariate skew-elliptical distribution. *Journal of Multivariate Analysis*, **79**, 93-113.
- [20] Butler, S.M. and Louis, T. A. (1992). Random effects models with nonparametric priors. *Statistics in Medicine*, **11**, 1981-2000.
- [21] Cambanis, S., Huang, S. and Simons, G. (1981). On the theory of elliptical contoured distributions. *J. Multivariate Anal.*, **11**, 368-385.
- [22] Capitanio, A., Azzalini, A., and Stanghellini. (2002). Graphical models for skew-normal variates. *Scandinavian Journal of statistics*, **29**.

- [23] Carrol, R. J., Spiegelman, C.H., Gordon-Lan, K.K., Bailey, K.T., and Abbott, R.D. (1984). On errors in variables for binary regression models. *Biometrika*, **71**, 19-25.
- [24] Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*. Chapman and Hall: London.
- [25] Cheng C. L. and Van Ness, J. (1999). *Statistical Regression with Measurement Error*. Arnold Publishers, London.
- [26] Chibs, S. and Greenberg, E. (1995). Understanding the Metropolis-Hansting Algorithm. *The American Statistician*, **49**, 327-335.
- [27] Cook, J. and Stefanski, L.A. (1994). A simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, **89**, 1314-1328.
- [28] Cysneiros, F. J. A. (2004). *Métodos Restritos e Validação de Modelos Simétricos de Regressão*. Tese de doutorado - IME - Universidade de São Paulo.
- [29] Davidian, M, and Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*. London: Chapman and Hall.
- [30] Dempster, A.P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, ser. **B**, **39**, 1-22.
- [31] Diggle, P. J., Liang, K. Y. and Zeger, S. L. (1994). *Analysis of Longitudinal Data*. Clarendon Press, Oxford.
- [32] Fang, K. -T., Kotz, S. and Ng, K.-W. (1990). *Symmetric multivariate and related distributions*. Chapman and Hall, London.
- [33] Fuller, W. A. (1987). *Measurement error models*. John Wiley. New York.
- [34] Galea M. (1995). *Calibração comparativa estrutural e funcional*. Tese de doutorado, IME - Universidade de São Paulo.
- [35] Gallant, A.R. and Nychka, D.W. (1987). Semiparametric maximum likelihood estimation. *Econometrica*, **55**, 363-390.

- [36] Gamermann, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. London, Chapman & Hill.
- [37] Ghazal, G. H. (1996). Recurrence formula for spectations of products of quadratic forms. *Statistics and Probability Letters*, **27**, 101-109.
- [38] Gelman A.(2004). Running WingBUGS from R. Available for download at (<http://cran.us.r-project.org/src/contrib/Descriptions/R2WinBUGS.html>)
- [39] Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457-511.
- [40] Genton, M. G. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Edited Volume, Chapman & Hall / CRC, Boca Raton, FL, 416 pp.
- [41] Genton, M. G., He, L. and Liu, X. (2001). Moments of skew-normal vectors and their quadratic forms. *Statistics and Probability Letters*, **51**, 319-325.
- [42] Genton, M. G. and Loperfido, N. (2001). Generalized skew-elliptical distributions and their quadratic forms. *Institute of Statistics Mimeo Series #2541*, under review.
- [43] Gimenez, P. and Bolfarine H. (1997). Corrected score functins in classical error-in-variables and incidental parameter models. *Australian Journal Statistics*, **39**, 325-344.
- [44] Graham, A. (1981). *Kronecker Products and Matrix Calculus with Applications*. Ellis Horwood Limited, Chichester.
- [45] Hartley, H.O. and Rao, J.N.K. (1967). Maximum likelihood stimation for the analysis of variance model. *Biometrika*, **54**, 93-108.
- [46] Henderson, C.R. (1984). Applications of linear models in animal breeding. University of Guelph Press, Guelph.
- [47] Henze, N., (1986). A probabilistic representation of the "skew normal" distribution. *Scand. J. Statist.*, **13**, 271-275.
- [48] Horn, R.A. and Johnson C. R. (1999). *Matrix Analysis*. Cambridge University Press, Cambridge.

- [49] Hu, P., Tsiatis, A.A. and Davidian, M. (1988). Estimating the parameters in the cox models when covariate variables are measured with error. *Biometrics*, **54**, 1407-1419.
- [50] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous univariate distributions*, Vol. 1. Wiley, New York.
- [51] Kelker (1970). Distributional theory of spherical distributions and a location-scale parameter. *Sankhya A*, **32**, 419-430.
- [52] Lachos, V.H. (2002). *Inferência e diagnóstico em modelos de Grubbs*. Dissertação de mestrado. IMECC-UNICAMP.
- [53] Lachos, V.H., Heleno, B., Labra, F.V. and Galea M.(2004). Estimation and influence diagnostics for structural calibration comparative model under the skew-normal distribution. Under Review.
- [54] Laird, N.M. and Ware, J.H. (1982). Random effects models for longitudinal data. *Biometrics*, **38**, 963-974.
- [55] Lange, K. and Sinsheimer, J. S. (1993). Normal/independent distribution and their applications in robust regression. *Journal of the American Statistics Association*, **2**, 175-198.
- [56] Lin, X. and Carroll, R.J. (1999). SIMEX variance component test in generalized linear mixed measurement error models. *Biometrics*, **55**, 613-619.
- [57] Lindstrom, M. J., and Bates, D. M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated measures data. *Journal of the American Statistical Association*, **83**, 1014-1022.
- [58] Liseo, B. (1990). LA classe delle densità normali sghembe: aspetti inferenziali da punto di vista bayesiano. *Statistica*, **50**, 59-70.
- [59] Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. (1996). *SAS system for mixed models*. Cary, North Carolina: SAS Institute.
- [60] Liu, C. (1996). Bayesian robust multivariate linear regression with incomplete data. *Journal of the American Statistical Association*, **91**, 1219-1227.
- [61] Loperfido, n. (2002). Statistical implications of selectively reported inferential results. *Statistics and Probability Letters*, **56**, 13-22.

- [62] Ma, Y., Genton, M. G. (2004). A flexible class of skew-symmetric distributions, Institute of Statistics Mimeo Series #2541, to appear in *Scandinavian Journal of Statistics*
- [63] Magnus, J. R. and Neudecker, H. (1979). The commutation matrix: some properties and applications. *Ann. Statist.*, **7**, 381-394.
- [64] Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press: London.
- [65] Meng, X. L. and Rubin, D.B.(1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, **80**, 267-278.
- [66] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.
- [67] Nakamura, T. (1990). Corrected score function for errors-in-variables models: Methology and application to generalized linear models. *Biometrika*, **77**, 127-137.
- [68] Nyquist, H. (1991). Restricted estimation of generalized linear models. *Applied Statistics*, **40**, 133-141.
- [69] Odell, P.L. and Fieveson, A.H. (1966). A numerical procedure to generate a sample covariance matrix. *Journal of the American Statistical Association*, **61**, 199-209.
- [70] O'Hagan, A. and Leonard, T. (1976). Bayes estimation subject to uncertainty about constraints. *Biometrika*, **63** 201-203.
- [71] Palta, M. and Lin, C.Y. (1996). Latent variables, measurement error and methods for analyzing longitudinal ordinal data, it Proceedings of American Statistical Association in Biometrics, 340-345.
- [72] Pinheiro, J. C. and Bates, D.M.(2000). *Mixed-Effects Models in S and S-plus*. New York: Springer-Verlag.
- [73] Pinheiro, J. C., Liu, C. H. and 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.
- [74] Rosa, G.J.M., Padovani, C.R., and Gianola D. (2003). Robust linear mixed models with normal/independent distributions and Bayesian MCMC implementation. *Biometrical Journal*, **45**, 573-590.

- [75] Sahu, S.K. and Dey, D.K. (2001). On multivariate survival models with a skewed frailty and a correlated baseline hazard process. *Technical Report*, Available at (<http://www.maths.soton.ac.uk/staff/Sahu/research/papers/logskew.html>.)
- [76] Sahu, S.K., Dey, D.K., and Branco, M. (2003). A new class of multivariate distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics*, **29**, to appear.
- [77] Searle, S.R., Cassela, G. and McCulloch, C.E. (1992). *Variance Components*. New York: John Wiley and Sons.
- [78] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Series B*, **64**, 583-639.
- [79] Stefanski, L. A. and Carroll, R. J. (1987). Conditional scores and optimal scores for generalised linear measurement error models. *Biometrika*, **74**, 703-716.
- [80] Stefanski, L. A. and Cook, J. (1995). Simulação-extrapolação: the measurement error jackknife. *Journal of the American Statistical Association*, **90**, 1247-1256.
- [81] Tanner, M. A. (1996). *Tools for Statistical Inference Methods for Exploration of posterior Distribution and Likelihood Functions* (3rd. edition). New York: Springer Verlag.
- [82] Tao, H., Palta, M., Yandell, B.S. and Newton, M.A. (1999). An estimation method for the semiparametric mixed effects model. *Biometrics*, **55**, 102-110.
- [83] Valença, D.M. (2003). *Teste de homogeneidade e estimação para dados de sobrevivência agrupados e com erros de medida*. Tese de doutorado - IME - Universidade de São Paulo.
- [84] Verbeke, G. and Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, **91**, 217-221.
- [85] Verbeke, G. and Lesaffre, E. (1997). The effect of misspecifying the random effects distribution in linear mixed models for longitudinal data. *Computational Statistics and Data Analysis*, **23**, 541-556.

- [86] Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.
- [87] Wang, N., Lin, X., Gutierrez, R. G. and Carroll, R. J. (1998). Bias analysis and SIMEX approach in generalized mixed measurement error models. *Journal of the American Statistical Association*, **93**, 249-261.
- [88] Wang, J., Boyer, J., and Genton, M. G. (1996). A skew-symmetric representation of multivariate distributions. *Statist. Sinica*. To appear.
- [89] Wolfinger, R., Tobias, R., and Sall, J. (1991). "Mixed Models: A future direction". *SUGI*, 16, 1380-1388.
- [90] Xu-Ping, Z., Wing-Kam, F. and Bo-Cheng, W. (2002). Estimation in linear models with random effects and errors-in- variables. *Ann. Ins. Statist. Math.*, Vol. **54**, 3, 595-606.
- [91] Zhang, D. and Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*, **57**, 795-802.