

Análise bayesiana de sensibilidade
sob distribuições *a priori* assimétricas

Luciana Graziela de Godoi

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
DOUTOR EM CIÊNCIAS

Programa: Estatística
Orientadora: Profa. Dra. Márcia D'Elia Branco

Durante o desenvolvimento deste trabalho a autora recebeu auxílio financeiro da
CAPES/CNPq

São Paulo, outubro de 2011

Análise bayesiana de sensibilidade
sob distribuições *a priori* assimétricas

Este exemplar corresponde à redação
final da tese devidamente corrigida
e defendida por Luciana Graziela de Godoi.
e aprovada pela Comissão Julgadora.

Banca Examinadora:

- Profa. Dra. Márcia D'Elia Branco - IME-USP.
- Prof. Dr. Josemar Rodrigues - UFSCar.
- Profa. Dra. Rosângela Helena Loschi - UFMG.
- Prof. Dr. Reinaldo Boris Arellano-Valle - PUCCH.
- Prof. Dr. Heleno Bolfarine - IME-USP.

*Aos meus queridos pais,
Tânia e Benedito,
pelo incondicional
apoio e carinho.*

AGRADECIMENTOS

Inicialmente, agradeço a Deus por ter me guiado até este momento e ter me permitido concluir mais esta etapa da minha jornada ao lado de pessoas tão maravilhosas.

Aos meus pais, Tânia e Benedito, e ao meu irmão Leonardo, pelos exemplos de honestidade, luta e dedicação. Simplesmente agradeço a companhia de vocês e as palavras de incentivo

À Professora Márcia Branco por ter despertado em mim a vontade de estudar estatística durante a graduação em matemática. Agradeço por sua orientação, confiança, idéias e contribuições fundamentais no progresso e elaboração deste trabalho.

Aos meus mais que amigos, Gilmar e Luana, pelas boas risadas, lembranças, conversas e momentos de distração.

Aos diversos amigos do IME-USP, em especial, Marina, Núbia, Cibele e Bet-sabé que através de palavras de apoio e incentivo e trocas de conhecimentos contribuíram significativamente durante todo este período.

Aos professores do Departamento de Estatística da Universidade de São Paulo, por terem contribuído para a minha formação acadêmica.

Ao Departamento de Estatística da Universidade Federal do Espírito Santo, pelo apoio dispensado durante a elaboração da tese. Em especial, à Professora Ana Cristina Staut Simmer Schunk pela amizade e colaboração.

E finalmente, aos membros da banca examinadora, pelas valiosas sugestões e contribuições que enriqueceram bastante este trabalho e que darão margem a futuras pesquisas.

Que Deus nos ilumine e permita continuar a compartilhar as nossas vidas.

Resumo

Neste trabalho, desenvolvemos uma análise de sensibilidade global para medir a robustez de estimadores bayesianos, com respeito a uma classe de distribuições *a priori* gerada à partir de modos de contaminação multiplicativo de uma distribuição *a priori* base, com estrutura similar ao considerado por van der Linde (2007). A esta classe denominamos classe de contaminação multiplicativa (Γ_M) e mostramos que, para particulares especificações, esta contém famílias de distribuições assimétricas conhecidas na literatura. Aqui, exploramos a classe de contaminação multiplicativa normal-assimétrica em vários contextos, a saber: como distribuição *a priori* do parâmetro de posição de um modelo normal, com variância conhecida e desconhecida, e como distribuição *a priori* do parâmetro regressor de um modelo linear normal, com a variância dos erros conhecida e desconhecida. Resultados de conjugação e expressões para medidas de distância entre as médias (variâncias) *a posteriori* fornecidas por Γ_M e a média (variância) *a posteriori* resultante da distribuição *a priori* base são apresentados. Através de um estudo de simulação, analisamos o comportamento das médias e das variâncias *a posteriori*, quando o modelo normal com variância desconhecida é considerado. Para o modelo de regressão, analisamos um conjunto de dados reais, fazendo uso da teoria desenvolvida. Por fim, mudamos o enfoque da análise de sensibilidade bayesiana, ao estudar a influência da classe de contaminação *a priori* normal-assimétrica sobre a distribuição *a posteriori* com um todo, comparando espaços de probabilidade *a posteriori* via função de concentração.

Palavras-chave: robustez bayesiana, sensibilidade global, classe de contaminação, distribuições assimétricas.

Abstract

In this work, we developed a global sensitivity analysis to measure the robustness of Bayesian estimators with respect to a class of prior distributions. This class arise from multiplicative modes of contamination of a base prior distribution, with similar structure to presented by van der Linde (2007). This class of prior is called multiplicative contamination class (Γ_M) and for some particular specifications, it contains families of skew distributions known in the literature. Here, we explored the skew-normal multiplicative contamination class on several contexts, namely as a prior distribution of the location parameter of a normal model for known and unknown variance, and as a prior distribution of the regressor parameter under the linear normal model, with known and unknown errors variance. Results of a Bayesian conjugation and expressions for some measures of distance between posterior means arise from Γ_M and the posterior mean associated to the base prior distribution are obtained. Through a simulation study we analyzed the behavior the posterior mean and the posterior variance, when the normal model with unknown variance is considered. For the regression model, we analyzed a real data set to illustrate the theory presented. Finally, we changed the focus of a Bayesian sensitivity analysis to study the posterior distribution under the skew-normal contamination class as a whole, comparing posterior probability spaces through concentration function.

Keywords: Bayesian robustness, global sensitivity, contamination class, skew distributions.

Sumário

Lista de Símbolos	iii
Lista de Figuras	v
Lista de Tabelas	ix
1 Introdução	1
1.1 Objetivos e organização da tese	5
2 Distribuições assimétricas	7
2.1 Família normal-assimétrica fechada	8
2.1.1 Definição e caracterizações	8
2.1.2 Função geradora de momentos	14
2.2 Família normal-assimétrica multivariada geral	14
3 Análise de sensibilidade sob uma classe de contaminação multipli- cativa	21
3.1 O modelo normal com variância conhecida.	23
3.1.1 Estudo de sensibilidade.	31
3.2 O modelo normal com variância desconhecida.	37
3.2.1 Estudo de sensibilidade.	40
4 Análise de sensibilidade em modelos de regressão linear normal	47
4.1 Introdução	47
4.2 Modelos de regressão normal com variância dos erros conhecida.	50
4.2.1 Modelo de regressão linear normal com um parâmetro regressor.	54
4.3 Modelos de regressão normal com variância dos erros desconhecida.	60
4.4 Análise de sensibilidade global para os dados GPM	64
4.4.1 τ^2 conhecido.	64

4.4.2	Estudo da influência de τ^2	68
4.4.3	Estudo da influência de σ^2	70
4.4.4	τ^2 desconhecido.	73
5	Sensibilidade global com o uso da função de concentração	75
5.1	A função de concentração	77
5.2	Sensibilidade global	84
5.3	Modelo normal	85
5.3.1	τ^2 conhecido	85
5.3.2	τ^2 desconhecido	87
6	Comentários finais	93
A	Demonstrações	97
A.1	Capítulo 3	97
A.2	Capítulo 4	121
A.3	Capítulo 5	128
B	Figuras	133
B.1	Situação I: $\bar{x} < 0$ e $\lambda \geq 0$	133
B.2	Situação II: $\bar{x} > 0$ e $\lambda \geq 0$	137
	Referências bibliográficas.	141

Lista de Símbolos

$N_q(\eta, \Psi)$	Distribuição normal q -variada com parâmetros de posição η e matriz de covariância Ψ .
$\phi_q(\cdot, \eta, \Psi^2)$	Função densidade de probabilidade de uma $N_q(\eta, \Psi)$.
$\Phi_q(\cdot, \eta, \Psi^2)$	Função de distribuição de acumulada de uma $N_q(\eta, \Psi)$.
$SN(\eta, \psi, \lambda)$	Distribuição normal-assimétrica univariada com parâmetros de posição η , escala ψ e assimetria λ .
$LTN_q(\theta, \eta, \Psi)$	Distribuição normal q -variada, com parâmetros de posição η e matriz de covariância Ψ , truncada à esquerda de θ .
$CSN_{p,q}(\mu, \Sigma, D, \nu, \Delta)$	Distribuição normal-assimétrica fechada, com parâmetros μ, Σ, D, ν e Δ .
$GSN_{p,q}(\mu, \Sigma, D, \nu, \Delta)$	Distribuição normal-assimétrica multivariada geral, com parâmetros μ, Σ, D, ν e Δ .
$X \perp W$	X e W são variáveis aleatórias independentes.
$X \stackrel{d}{=} W$	X e W têm a mesma distribuição.
∇_t	Operador gradiente.

Lista de Figuras

1.1	Limites da média <i>a posteriori</i>	3
2.1	Funções de densidade da distribuição normal-assimétrica fechada com $\mu = 4$, $\nu = 3$ e $D = 1$. Gráfico 1: $\Sigma = 0.5, 6$ e 9 e $\Delta = 1$. Gráfico 2: $\Sigma = 1$ e $\Delta = 0.05, 0.5$ e 5	10
2.2	Funções de densidade da distribuição normal-assimétrica fechada com $\mu = 4$, $\nu = 3$, $D = 1$ e diferentes valores para Σ e Δ	10
2.3	Funções de densidade da distribuição normal-assimétrica fechada com $\mu = 1$, $\Sigma = 9$, $D = 1$, $\Delta = 0.05$ e diferentes valores para ν	11
2.4	Funções de densidade da distribuição normal-assimétrica fechada com $\Sigma = 9$, $D = 1$, $\nu = 1$, $\Delta = 0.05$ e diferentes valores para μ	12
2.5	Funções de densidade da distribuição normal-assimétrica fechada com $\mu = 5$, $\Sigma = 9$, $\nu = 1$, $\Delta = 0.05$ e diferentes valores para D	12
2.6	Funções de densidade da distribuição normal-assimétrica multivariada geral com $\mu = 4$, $\nu = 3$ e $D = 1$. Gráfico 1: $\Sigma = 0.5, 6$ e 9 e $\Delta = 1$. Gráfico 2: $\Sigma = 1$ e $\Delta = 0.05, 0.5$ e 5	15
2.7	Funções de densidade da distribuição normal-assimétrica multivariada geral com $\mu = 4$, $\nu = 3$, $D = 1$ e diferentes valores para Σ e Δ	16
2.8	Funções de densidade da distribuição normal-assimétrica multivariada geral com $\mu = 1$, $\Sigma = 9$, $D = 1$, $\Delta = 0.05$ e diferentes valores para ν	16
2.9	Funções de densidade da distribuição normal-assimétrica multivariada geral com $\Sigma = 9$, $D = 1$, $\nu = 1$, $\Delta = 0.05$ e diferentes valores para μ	17
2.10	Funções de densidade da distribuição normal-assimétrica multivariada geral com $\mu = 3$, $\Sigma = 9$, $\nu = 1$, $\Delta = 0.05$ e diferentes valores para D	18
3.1	Gráficos de $d_M(\lambda)$ e $d_V(\lambda)$ para $n = 20$, $\tau^2 = 1$, $\mu = 0$ e $\sigma^2 = 1$	27
3.2	Distâncias extremas de d_M e d_V , quando $n = 20$ e 100 , $\sigma^2 = \tau^2 = 1$ e $\bar{x} < 0$	32

3.3	Funções de densidade da distribuição <i>a priori</i> para θ dadas por $SN(\lambda)$, com $\lambda = 0$ e $\lambda \rightarrow \infty$	34
3.4	Comparação da distribuição <i>a priori</i> , verossimilhança ($\bar{x} = -3$, $n = 20$) e distribuição <i>a posteriori</i> , para $\lambda = 20, 0$	34
3.5	Distâncias extremas de d_M e d_V , quando $n = 20$ e 100 , $\sigma^2 = \tau^2 = 1$ e $\bar{x} > 0$	35
3.6	Comparação da distribuição <i>a priori</i> , verossimilhança ($\bar{x} = 3$, $n = 20$) e distribuição <i>a posteriori</i> , para $\lambda = 20, 0$	36
3.7	Supremo das distâncias d_{M_0} , d_{V_0} , $d_{M_{\tau^2}}$ e $d_{V_{\tau^2}}$, quando $S^2 = 10$ e $n = 20, 50, 100$ e 200	42
3.8	Supremo das distâncias d_{M_0} , d_{V_0} , $d_{M_{\tau^2}}$ e $d_{V_{\tau^2}}$, quando $S^2 = 10$ e $n = 20, 50, 100$ e 200	45
4.1	Gráficos de d_M e d_V para $\sum x_i^2 = 10$, $\mu = 0$, $\sigma^2 = 1$ e $\tau^2 = 1$	57
4.2	Gráfico de dispersão para as variáveis peso do veículo (PV) e galões por milha (GPM).	65
4.3	Aproximação da função $\frac{\phi(-x)}{\Phi(-x)}$ por x e $x + \frac{1}{x}$, $\forall x > 0$	67
4.4	Comparação da distribuição <i>a priori</i> , verossimilhança e distribuição <i>a posteriori</i> , para $\lambda = 20, -20$	71
4.5	Comparação da distribuição <i>a priori</i> base, verossimilhança e distribuição <i>a posteriori</i>	72
5.1	Curva de Lorenz para a distribuição de renda na população.	76
5.2	Função de concentração de $\theta \sim SN(2)$ com respeito a $\theta \sim N(0, 1)$ e respectivas funções de densidade.	83
5.3	Função de concentração para a classe de distribuições <i>a posteriori</i> , fixado $\bar{x} = -1, 0$ e 1 . As linhas (-) e (- -) correspondem, respectivamente, ao $\hat{\varphi}(z)$ e ao $1 - \hat{\varphi}(1 - z)$	90
A.1	Curva das funções $g(k)$ e $t(k)$, para $k > 0$, no estudo da distância d_M	112
A.2	Curva das funções $g(k)$ e $t(k)$, para $k > 0$, no estudo da distância d_V	117
B.1	$S^2 = 1$	133
B.2	$S^2 = 3$	134
B.3	$S^2 = 8$	134

B.4	$S^2 = 10$	135
B.5	$S^2 = 20$	135
B.6	$S^2 = 100$	136
B.7	$n = 200$	136
B.8	$S^2 = 1$	137
B.9	$S^2 = 3$	137
B.10	$S^2 = 8$	138
B.11	$S^2 = 10$	138
B.12	$S^2 = 20$	139
B.13	$S^2 = 100$	139
B.14	$n = 100$	140
B.15	$n = 200$	140
B.16	$n = 100$	140
B.17	$n = 200$	140

Lista de Tabelas

4.1	Valores obtidos para o supremo e o ínfimo de d_M e d_V	68
4.2	Valores obtidos para o supremo e o ínfimo de d_M e d_V , $\hat{\beta}_0$ e $\hat{\zeta}_0^2$, para diferentes valores de τ^2	69
4.3	Medidas <i>a posteriori</i> , $\sup_\lambda d_M$ e $\sup_\lambda d_V$, para $\tau^2 = 100$	70
4.4	Distância entre as estimativas <i>a posteriori</i> obtidas sob a distribuição <i>a priori</i> base e a distribuição <i>a priori</i> normal-assimétrica, com $\lambda = -20, 20$	72
4.5	Valores obtidos para o supremo das distâncias <i>a posteriori</i>	74

Capítulo 1

Introdução

O termo robustez estatística tem sido amplamente difundido na literatura, entretanto é importante diferenciar as suas várias abordagens. A abordagem mais comum tem interesse em estudar o quanto os estimadores são afetados pela presença de um subconjunto de valores discrepantes na amostra. Se os estimadores são pouco sensíveis (segundo algum critério estatístico), dizemos que o estimador é robusto. Esta abordagem pode ser aplicada tanto no contexto bayesiano quanto no clássico.

Uma segunda abordagem, que será objeto de nosso estudo, está relacionada a uma possível má especificação do modelo probabilístico. Em inferência bayesiana paramétrica usualmente dois modelos probabilísticos são considerados; o primeiro associado à função de verossimilhança e o segundo à distribuição *a priori*. A escolha desses modelos é feita de forma conveniente de modo a atender certas pré-suposições do investigador. Fala-se de inferência robusta quando as estatísticas de interesse são pouco afetadas por essas escolhas. Uma maneira de verificar a robustez bayesiana é especificar uma classe de distribuições de probabilidade Γ que contenha o modelo probabilístico que estamos assumindo como de referência (base), $f_o(\theta)$, e verificar as alterações nas inferências *a posteriori* ao alterarmos a suposição de $f_o(\theta)$ para qualquer $f(\theta) \in \Gamma$. Esse estudo é também denominado análise de sensibilidade e pode ser realizado tanto para especificações *a priori* [$f(\theta)$], quanto para verossimilhança [$f(x | \theta)$] e função de perda $L(\theta, a)$.

Berger, Insua & Ruggeri (2000) fazem uma breve revisão do desenvolvimento de estudos na área de robustez bayesiana, indicando inclusive artigos que discutem aspectos filosóficos e históricos nesta área. Além disso, eles apresentam três possíveis abordagens em se tratando de robustez bayesiana, são elas:

- abordagem informal: são consideradas algumas poucas distribuições *a priori* e suas correspondentes inferências (por exemplo, a média *a posteriori*) são comparadas. Procedimento considerado simples e útil, tem como desvantagem o número limitado de distribuições *a priori*, que podem não incluir distribuições *a priori* compatíveis com a distribuição *a priori* base, mas que produzem inferências *a posteriori* muito diferentes.
- robustez global: considera-se a classe Γ de todas as distribuições *a priori* compatíveis com a distribuição *a priori* base, e medidas *a posteriori* são obtidas segundo a classe Γ . Em sua análise, compara-se o ínfimo e o supremo das quantidades de interesse *a posteriori*.

Para um melhor entendimento dessas quantidades, considere, por exemplo, a função de verossimilhança e a classe de distribuições *a priori* denotadas, respectivamente, por $f(x | \theta)$ e $\Gamma = \{f_i(\theta), i = 0, 1, \dots, n\}$. Note que a distribuição *a priori* base $f_0(\theta)$ também pertence à classe Γ . Ao associarmos a função de verossimilhança à classe de distribuições *a priori* Γ , obtemos a classe de distribuições *a posteriori* $\Gamma^* = \{f_i(\theta | x), i = 0, 1, \dots, n\}$ e, por conseguinte, um conjunto de medidas *a posteriori* (por exemplo, a média *a posteriori*) dadas por $M = \{M_i(x), i = 0, 1, \dots, n\}$, seu supremo e seu ínfimo. Suponha que $M_i(x) = E_i(\theta | x)$ para $i = 0, \dots, n$ e tenhamos obtido o gráfico da Figura 1.1.

Ao analisarmos o gráfico da Figura 1.1, notamos que, para valores de x próximos a zero, o supremo e o ínfimo das esperanças *a posteriori* estão em torno da $E_0(\theta|x)$, evidenciando a robustez da esperança *a posteriori* com relação a classe de distribuições *a priori* Γ . No entanto, o mesmo não ocorre para valores de $x > 3$, onde observa-se que as curvas contendo o supremo e o ínfimo das $\{E_i(\theta | x), i = 0, \dots, n\}$ separam-se rapidamente.

Em O'Hagan (1994), é apresentada uma análise similar que considera uma amostra aleatória com distribuição $N(\theta, 1)$, distribuição *a priori* base normal padrão e a classe de distribuições *a priori* ϵ -contaminada.

- robustez local: analisa-se o quanto mudanças infinitesimais na distribuição *a priori* influenciam na inferência *a posteriori*, através de técnicas de análise diferencial, tais como derivadas de Fréchet ou de Gâteaux. Em situações mais

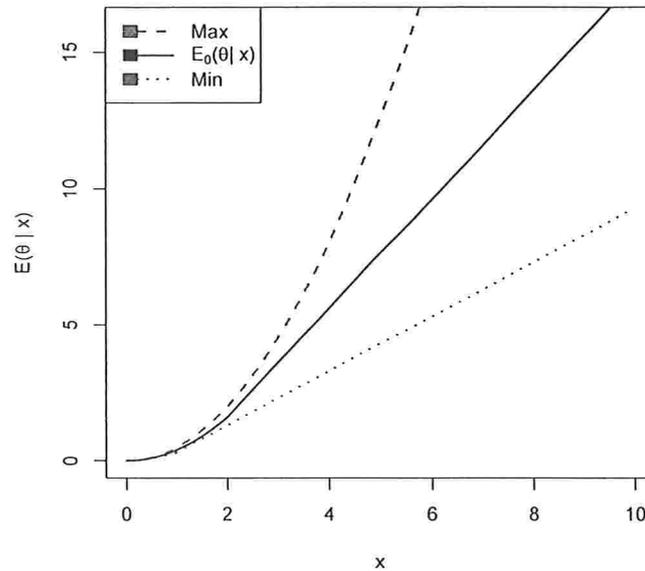


Figura 1.1: Limites da média *a posteriori*.

complicadas, medidas de robustez local são mais fáceis de calcular do que medidas globais, mas sua interpretação (e calibração) nem sempre é clara.

Maiores detalhes sobre as diferentes abordagens em robustez bayesiana podem ser encontrados em O'Hagan (1994) e Insua & Ruggeri (2000).

O'Hagan (1994) apresenta uma boa introdução a análise de sensibilidade bayesiana com respeito a variação da distribuição *a priori*, ilustrando esse tipo de análise através do uso da família ϵ -contaminada no contexto de robustez global. De maneira geral, podemos entender a família ϵ -contaminada como um processo aditivo de contaminação da distribuição *a priori* base, em que $\Gamma = \{(1-\epsilon)f_0(\theta) + \epsilon q(\theta) : q(\theta) \in \mathcal{G}\}$, onde $q(\theta)$ corresponde a função densidade de probabilidade em alguma classe \mathcal{G} que contém distribuições *a priori*. Neste contexto, ϵ corresponde a uma constante que controla o nível de contaminação da distribuição *a priori* base pela classe de distribuições \mathcal{G} .

van der Linde (2007) propõe uma estrutura simples para unificar modos de contaminação multiplicativo que implique em mudanças tanto na verossimilhança quanto

na distribuição *a priori*. Baseada nesta construção, a autora investigou quais aspectos da matriz de informação de Fisher - comumente interpretados como medida de influência local - são influenciados por diversos tipos de perturbação.

Outras especificações para classes de contaminação da distribuição *a priori* são também discutidas em Berger (1990, 1993).

Quanto a especificações de distribuições *a priori*, Mukhopadhyay & Vidakovic (1995) conduzem uma análise bayesiana robusta para explorar a eficiência da regra linear de Bayes com relação a regra exata de Bayes, nos casos em que a distribuição *a priori* considerada é assimétrica. Esses autores evidenciam a robustez do estimador linear de Bayes em relação a mudança de simetria da distribuição *a priori*, sob perda quadrática.

Segundo Azzalini & Capitanio (2003), um possível processo para assimetrizar uma distribuição simétrica em torno de zero, com função densidade de probabilidade $f(\cdot)$, contínua e d -dimensional, consiste em considerar $f_Z(z) = 2 f(z) Q(W(z))$, onde, para todo z , Q é uma função de distribuição escalar não-negativa tal que $Q(-z) = 1 - Q(z)$ e $W(\cdot)$ é uma função ímpar, de \mathbb{R}^d em \mathbb{R} . Nestas condições, Z é uma variável aleatória distribuída assimetricamente.

Distribuições assimetrizadas segundo a proposta anterior tendem a preservar algumas propriedades da distribuição simétrica f . Dessa forma, conseguimos obter uma classe de famílias de distribuições paramétricas mais flexíveis, capazes de modelar de maneira mais adequada dados oriundos de fenômenos reais.

Para obtermos a função de densidade de uma distribuição normal-assimétrica, por exemplo, basta considerar $W(z) = \lambda z$, $f = \phi$ e $Q = \Phi$, onde $\phi(\cdot)$ e $\Phi(\cdot)$ são as funções de densidade de probabilidade (f.d.p.) e de distribuição acumulada (f.d.a.) de uma distribuição normal padrão, respectivamente. Assim, a função densidade de probabilidade de uma distribuição normal assimetrizada é dada por $f_Z(z) = 2 \phi(z) \Phi(\lambda z)$.

Observe que, quando consideramos $\lambda = 0$, obtemos a distribuição normal como caso particular. Maiores detalhes sobre a distribuição normal-assimétrica univariada podem ser vistos em Azzalini (1985) e Rodríguez (2005).

Diferentes generalizações multivariadas para a distribuição normal-assimétrica têm sido propostas por vários autores, por exemplo, Azzalini & Dalla-Valle (1996), Sahu, Dey & Branco (2003), Gupta, González-Farías & Domínguez-Molina (2004).

Neste trabalho, as generalizações consideradas são as distribuições normal-assimétrica fechada e normal-assimétrica multivariada geral, propostas, respectivamente, por Domínguez-Molina, González-Farías & Gupta (2003) e Gupta, González-Farías & Domínguez-Molina (2004).

1.1 Objetivos e organização da tese

Um dos principais objetivos deste trabalho é considerar uma classe de distribuições *a priori*, gerada à partir de modos de contaminação multiplicativo de uma distribuição *a priori* base, com estrutura similar ao considerado por van der Linde (2007), para efetuar uma análise de robustez global. A esta classe denominaremos classe de contaminação multiplicativa. Para particulares especificações, mostramos que a classe contaminação multiplicativa coincide com classes de distribuições assimétricas conhecidas na literatura. Os resultados deste trabalho estão divididos em seis capítulos.

No segundo capítulo, apresentamos as distribuições normal-assimétrica fechada e normal-assimétrica multivariada geral, possíveis caracterizações e resultados. Mostramos que, sob certas condições, as distribuições normal-assimétrica fechada e normal-assimétrica multivariada geral são equivalentes.

No Capítulo 3, definimos a classe de contaminação multiplicativa *a priori* (Γ_M), associamos distribuições assimétricas a ela e detalhamos a metodologia de robustez bayesiana global. Realizamos um estudo de sensibilidade global da média e da variância da distribuição *a posteriori* de uma distribuição $N(\theta, \tau^2)$, quando a distribuição *a priori* do parâmetro de posição corresponde a classe de contaminação multiplicativa normal-assimétrica. Quando τ^2 é considerado conhecido, obtivemos resultados de conjugação bayesiana. Ainda neste caso, detalhamos quais as condições necessárias para que a robustez com respeito a escolha de distribuições *a priori* em Γ_M efetivamente ocorra. Para τ^2 desconhecido, obtivemos as distribuições condicionais completas *a posteriori*, a função de densidade marginal *a posteriori* do parâmetro de posição e mostramos que tanto a média, quanto a variância *a posteriori* marginal de τ^2 são obtidos como função dos momentos da distribuição *a posteriori* de θ . Nesta situação, promovemos um estudo de simulação, via metodologia de Gauss-Hermite, para análises de robustez global.

No quarto capítulo, generalizamos os resultados apresentados no Capítulo 3 ao discutir questões de sensibilidade bayesiana em modelos de regressão linear normal, ora supondo a variância dos erros conhecida, ora supondo desconhecida. Em ambas as situações, consideramos para a distribuição *a priori* do parâmetro regressor a classe de contaminação multiplicativa parametrizada segundo a distribuição normal-assimétrica. Resultados de conjugação são obtidos e exemplos específicos são estudados à luz da teoria desenvolvida.

De maneira geral, nos Capítulos 3 e 4 analisamos a influência da classe de distribuição multiplicativa *a priori* sobre estatísticas pontuais *a posteriori*, mais especificamente, a média e a variância *a posteriori*.

No Capítulo 5, mudamos o enfoque da análise de sensibilidade bayesiana. Nosso objetivo agora é mais global, estudamos a influência da classe de contaminação multiplicativa *a priori* sobre a distribuição *a posteriori* com um todo, comparando espaços de probabilidade *a posteriori* via função de concentração. Apresentamos uma versão mais adequada aos nossos propósitos para a definição de função de concentração, o que nos permite obter uma expressão fechada para a função de concentração *a posteriori* de modelos normais, quando a distribuição *a priori* do parâmetro de posição pertence à classe de contaminação multiplicativa *a priori*, parametrizada segundo a distribuição normal-assimétrica.

Finalmente, no sexto capítulo apresentamos conclusões dos resultados obtidos neste trabalho e perspectivas futuras de trabalho.

No Apêndice A constam as demonstrações de diversos resultados e proposições presentes nos Capítulos 3, 4 e 5. No Apêndice B apresentamos alguns gráficos das simulações discutidas no Capítulo 3.

Capítulo 2

Distribuições assimétricas

Diferentes tipos de distribuições normais-assimétricas multivariadas têm sido propostas na literatura, cada uma delas com características bastante particulares. Por exemplo, Azzalini (1985) apresenta uma extensão multivariada da distribuição normal-assimétrica, cuja as distribuições marginais não são normais-assimétricas. Já em 1996, Azzalini e Dalla-Valle apresentam uma outra versão multivariada da distribuição normal-assimétrica, construída a partir do pressuposto que a distribuição marginal do vetor de variáveis aleatórias é normal-assimétrica univariada. Neste mesmo artigo, apresentam uma construção via condicionamento.

Outras diversas extensões e formulações para a distribuição normal-assimétrica multivariada têm sido propostas na literatura, como, por exemplo, em Azzalini & Capitanio (1999), Arellano-Valle et al. (2002) e Sahu et al. (2003). Arellano-Valle & Genton (2005) apresentaram a família de distribuições normal-assimétrica fundamental que generaliza todas as definições existentes de distribuições normais-assimétricas.

Discussões sobre possíveis equivalências entre as diferentes propostas para a distribuição normal-assimétrica multivariada, propriedades interessantes e limitações de cada abordagem são apresentados em Arellano-Valle & Azzalini (2006). Ainda neste artigo, os autores apresentam uma nova generalização, a família elíptica-assimétrica.

Nas seções subsequentes, serão apresentadas versões da distribuição normal-assimétrica multivariada de relativa complexidade. De fato, esta característica se reverte em propriedades como fechamento sob conjugação, marginalização e transformações lineares (posto completo). Além disso, muitas das distribuições normais-assimétricas apresentadas na literatura podem ser vistas como casos particulares das distribuições que serão apresentadas. São elas: distribuição normal-assimétrica

fechada e distribuição normal-assimétrica multivariada geral.

Domínguez-Molina, González-Farías & Gupta (2003) apresentaram a distribuição normal-assimétrica fechada e mostraram que além de possuir as qualidades já citadas anteriormente, também a distribuição conjunta de variáveis aleatórias independentes nesta família ainda pertence a família normal-assimétrica fechada.

Parte da teoria a ser desenvolvida nos capítulos subseqüentes utilizará propriedades da distribuição normal-assimétrica multivariada geral, apresentada em Gupta, González-Farías & Domínguez-Molina (2004). Mostramos também que esta distribuição relaciona-se com a família normal-assimétrica fechada através de uma simples transformação. Dessa forma, propriedades da distribuição normal-assimétrica multivariada geral podem facilmente ser vinculados às propriedades já demonstradas da família normal-assimétrica fechada. Tais características nos permitirão obter, em algumas situações particulares, resultados analíticos para o estudo de robustez bayesiana.

Neste capítulo vamos definir, caracterizar e apresentar algumas propriedades das distribuições normal-assimétrica fechada e normal-assimétrica multivariada geral que serão úteis em estudos de robustez quando se considera particulares especificações da classe de contaminação multiplicativa *a priori*.

Durante todo este capítulo, adotaremos os símbolos $\phi_q(\cdot; \eta, \Psi)$ e $\Phi_q(\cdot; \eta, \Psi)$ para fazer referência, respectivamente, às funções de densidade e de distribuição acumulada de uma distribuição normal q -variada, com parâmetros de posição η e matriz de covariância Ψ .

2.1 Família normal-assimétrica fechada

2.1.1 Definição e caracterizações

Definição 2.1.1 *Um vetor aleatório Y p -dimensional tem distribuição normal-assimétrica fechada quando a sua função de densidade é dada por:*

$$f_{p,q}(y; \mu, \Sigma, D, \nu, \Delta) = \frac{\phi_p(y; \mu, \Sigma) \Phi_q(D(y - \mu); \nu, \Delta)}{\Phi_q(0; \nu, \Delta + D\Sigma D^T)}, \quad y \in \mathbb{R}^p, \quad (2.1)$$

onde $p \geq 1$, $q \geq 1$, $\mu \in \mathbb{R}^p$, $\nu \in \mathbb{R}^q$, $\Sigma_{p \times p}$ e $\Delta_{q \times q}$ são matrizes positivas definidas e $D_{q \times p}$ é uma matriz arbitrária. Como notação, adotamos $Y \sim CSN_{p,q}(\mu, \Sigma, D, \nu, \Delta)$.

A demonstração de que $f_{p,q}(\cdot)$ é uma genuína função densidade de probabilidade pode ser encontrada em Domínguez-Molina et al. (2003).

A distribuição normal-assimétrica fechada é um caso particular de distribuições assimétricas obtidas segundo processos de seleção baseados em distribuições condicionais. Arellano-Valle, Branco & Genton (2006) definem esta metodologia através de distribuições de seleção multivariada, discutem algumas das suas principais propriedades e mostram que muitas das distribuições assimétricas multivariadas propostas na literatura podem ser descritas como casos particulares de distribuições de seleção.

Segundo González-Farías, Domínguez-Molina & Gupta (2004b), as principais diferenças entre a distribuição normal-assimétrica fechada e as outras generalizações da distribuição normal-assimétrica univariada dizem respeito a presença dos parâmetros extras ν e Δ , que permitem, respectivamente, obter propriedades fechadas para densidades condicionais e densidades marginais. A inclusão de $\Phi_q(\cdot)$, para $q \geq 1$, nos permite obter propriedades para somas e para a distribuição conjunta de variáveis aleatórias independentes com distribuição normal-assimétrica geral.

Iversen (2010) apresenta alguns exemplos que nos permitem entender melhor a influência dos parâmetros μ , Σ , D , ν e Δ sobre a função de densidade da distribuição normal-assimétrica fechada. Considere então a distribuição normal-assimétrica fechada univariada ($p = q = 1$) e vejamos o que ocorre com o gráfico da função de densidade a medida em que alteramos os valores dos parâmetros desta distribuição. Segundo o autor,

- Quanto maior a variância Σ na função de densidade da distribuição normal, maior a assimetria considerada na distribuição normal-assimétrica fechada, como pode ser visto no primeiro gráfico da Figura 2.1.
- Como exemplificado pelo segundo gráfico da Figura 2.1, quanto menor a variância Δ presente na função de distribuição acumulada da distribuição normal, maior será a assimetria no modelo normal-assimétrico fechado.
- Na Figura 2.2, mostramos que ao associar grandes valores para o parâmetro Σ , com baixos valores para o parâmetro Δ , induzimos uma maior assimetria ao modelo normal-assimétrico fechado.

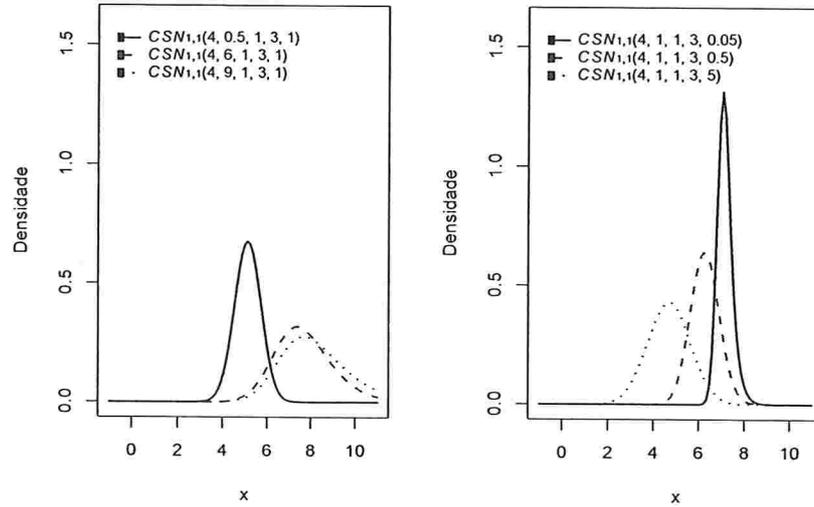


Figura 2.1: Funções de densidade da distribuição normal-assimétrica fechada com $\mu = 4, \nu = 3$ e $D = 1$. Gráfico 1: $\Sigma = 0.5, 6$ e 9 e $\Delta = 1$. Gráfico 2: $\Sigma = 1$ e $\Delta = 0.05, 0.5$ e 5 .

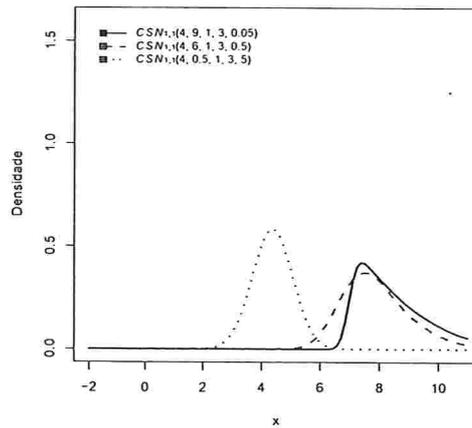


Figura 2.2: Funções de densidade da distribuição normal-assimétrica fechada com $\mu = 4, \nu = 3, D = 1$ e diferentes valores para Σ e Δ .

- Como pode ser visto na Figura 2.3, para um valor de Σ grande e de Δ pequeno, a assimetria da distribuição normal-assimétrica fechada torna-se nula, se $\nu \rightarrow -\infty$. O parâmetro ν influencia tanto a assimetria, quanto a escala e a posição da função de densidade da distribuição normal-assimétrica fechada.
- O parâmetro μ atua como parâmetro de posição no modelo normal-assimétrico fechado, quando consideramos os parâmetros Σ , D , ν e Δ fixos. Na Figura 2.4 exemplificamos esta situação.
- Já o parâmetro D também atua na assimetria do modelo. Quando $D = 0$, obtemos o modelo normal com parâmetros de posição μ e de escala Σ . Veja a Figura 2.5.

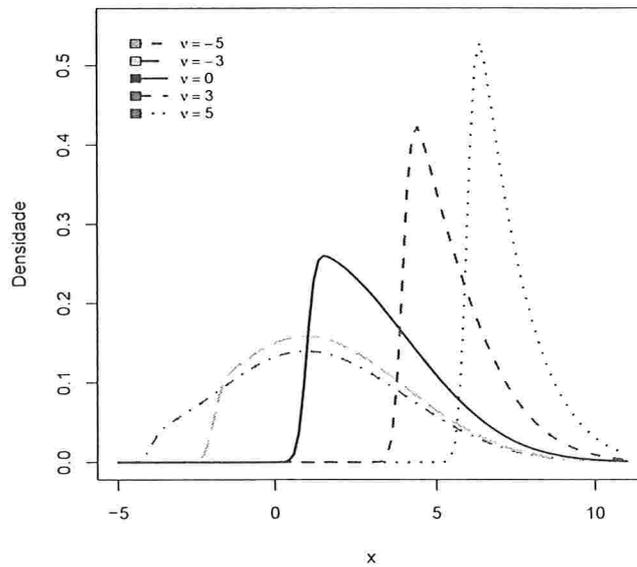


Figura 2.3: Funções de densidade da distribuição normal-assimétrica fechada com $\mu = 1$, $\Sigma = 9$, $D = 1$, $\Delta = 0.05$ e diferentes valores para ν .

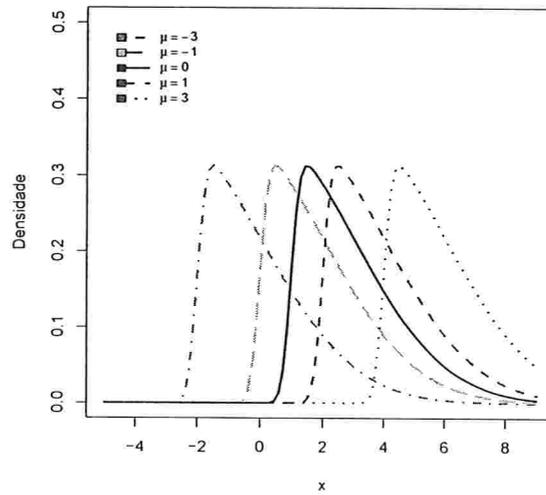


Figura 2.4: Funções de densidade da distribuição normal-assimétrica fechada com $\Sigma = 9$, $D = 1$, $\nu = 1$, $\Delta = 0.05$ e diferentes valores para μ .

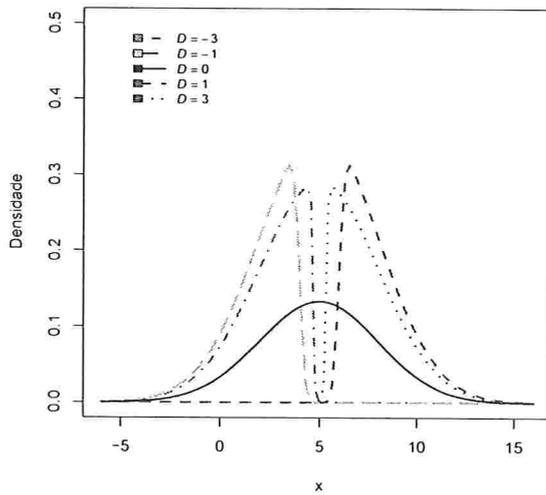


Figura 2.5: Funções de densidade da distribuição normal-assimétrica fechada com $\mu = 5$, $\Sigma = 9$, $\nu = 1$, $\Delta = 0.05$ e diferentes valores para D .

A função de distribuição acumulada associada à função densidade de probabilidade apresentada em (2.1) é dada no lema a seguir.

Lema 2.1.1 *A função de distribuição acumulada de $Y \sim CSN_{p,q}(\mu, \Sigma, D, \nu, \Delta)$ é dada por*

$$F_{p,q}(y; \mu, \Sigma, D, \nu, \Delta) = C \Phi_{p+q} \left[\begin{pmatrix} y \\ 0 \end{pmatrix}; \begin{pmatrix} \mu \\ \nu \end{pmatrix}, \begin{pmatrix} \Sigma & -\Sigma D^T \\ -D\Sigma & \Delta + D\Sigma D^T \end{pmatrix} \right], \quad (2.2)$$

onde $C^{-1} = \Phi_q(0; \nu, \Delta + D\Sigma D^T)$.

Apresentamos a seguir duas maneiras diferentes de se obter a distribuição normal-assimétrica fechada. A primeira delas, obtida via condicionamento de um vetor aleatório com distribuição normal $(p+q)$ -variado, é apresentada no Lema 2.1.2.

Lema 2.1.2 *Se $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_{p+q} \left[\begin{pmatrix} \mu \\ -\nu \end{pmatrix}, \begin{pmatrix} \Sigma & \Sigma D^T \\ D\Sigma & \Delta + D\Sigma D^T \end{pmatrix} \right]$, onde $\mu \in \mathbb{R}^p$, $\nu \in \mathbb{R}^q$, $D_{q \times p}$ é uma matriz arbitrária e $\Sigma_{p \times p}$ e $\Delta_{q \times q}$ são matrizes positivas definidas, então*

$$Y \stackrel{d}{=} X_1 \mid X_2 \geq 0 \sim CSN_{p,q}(\mu, \Sigma, D, \nu, \Delta). \quad (2.3)$$

Maiores detalhes sobre esta construção podem ser encontrados em González-Fariás et al. (2004b).

No Lema 2.1.3 apresentamos uma outra forma de construir a distribuição normal-assimétrica fechada. Proposta por Domínguez-Molina et al. (2007), consiste em obter uma variável aleatória Y como uma combinação linear de variáveis aleatórias independentes. Tal construção se mostra útil na simulação da variável Y e ficou conhecida na literatura como representação estocástica.

Lema 2.1.3 *Considere X_1 e X_2 variáveis aleatórias independentes tais que $X_1 \sim N_p(0_p, I_p)$ e $X_2 \sim LTN_q(\nu, 0_q, \Delta + D\Sigma D^T)$, então*

$$Y \stackrel{d}{=} \mu + (\Sigma^{-1} + D^T \Delta^{-1} D)^{-1/2} X_1 + \Sigma D^T (\Delta + D\Sigma D^T)^{-1} X_2 \sim CSN_{p,q}(\mu, \Sigma, D, \nu, \Delta), \quad (2.4)$$

onde o símbolo $LTN_q(\theta, \eta, \Psi)$ representa uma distribuição q -variada gerada pós truncamento à esquerda de uma distribuição normal q -variada com parâmetros de posição η e matriz de covariância Ψ , por um vetor θ , também q -variado.

2.1.2 Função geradora de momentos

Nesta seção, apresentamos a função geradora de momentos da distribuição normal-assimétrica fechada, o que nos possibilitará obter medidas como a média e a variância desta distribuição.

Proposição 2.1.1 *Se $Y \sim CSN_{p,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta)$ então a função geradora de momentos de Y é dada por:*

$$M_Y(\mathbf{t}) = \frac{\Phi_q [D\Sigma\mathbf{t}; \boldsymbol{\nu}, \Delta + D\Sigma D^T]}{\Phi_q [\mathbf{0}; \boldsymbol{\nu}, \Delta + D\Sigma D^T]} e^{\mathbf{t}^T \boldsymbol{\mu} + \frac{1}{2} \mathbf{t}^T \Sigma \mathbf{t}}, \quad \mathbf{t} \in \mathbb{R}^p. \quad (2.5)$$

Na próxima seção apresentaremos a distribuição normal-assimétrica multivariada geral, que, como veremos, pode ser obtida como uma simples transformação do modelo normal-assimétrico fechado. Como vamos obter a média e a variância do modelo normal-assimétrico multivariado geral, estaremos obtendo, implicitamente, a média e a variância da distribuição normal-assimétrica fechada, por conta disso, omitiremos a apresentação destes resultados nesta seção.

Demais propriedades da distribuição normal-assimétrica fechada tais como fechamento por transformações lineares, condicionamento e marginalização, assim como demonstrações, podem ser encontrados em González-Farías et al. (2004a,b), Domínguez-Molina et al. (2007).

2.2 Família normal-assimétrica multivariada geral

Definição 2.2.1 *Um vetor aleatório Y p -dimensional tem distribuição normal-assimétrica multivariada geral quando a sua função de densidade é dada por:*

$$f_{p,q}(\mathbf{y}; \boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta) = \frac{\phi_p(\mathbf{y}; \boldsymbol{\mu}, \Sigma) \Phi_q(D\mathbf{y}; \boldsymbol{\nu}, \Delta)}{\Phi_q(D\boldsymbol{\mu}; \boldsymbol{\nu}, \Delta + D\Sigma D^T)}, \quad \mathbf{y} \in \mathbb{R}^p, \quad (2.6)$$

onde $p \geq 1$, $q \geq 1$, $\boldsymbol{\mu} \in \mathbb{R}^p$, $\boldsymbol{\nu} \in \mathbb{R}^q$, $\Sigma_{p \times p}$ e $\Delta_{q \times q}$ são matrizes positivas definidas e $D_{q \times p}$ é uma matriz arbitrária. Como notação, adotamos $Y \sim GSN_{p,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta)$.

Como já comentado anteriormente, esta distribuição foi proposta por Gupta, González-Farías & Domínguez-Molina (2004) e possui como casos particulares as versões univariada e multivariada da distribuição normal-assimétrica apresentadas em Azzalini (1985) e Azzalini & Capitanio (1999), respectivamente.

Vejamos qual o comportamento da função de densidade da distribuição normal-assimétrica multivariada geral às mudanças de valores nos diferentes parâmetros desta distribuição. Note que

- Quanto maior a variância Σ na função de densidade da distribuição normal, maior será a assimetria e a variância da distribuição normal-assimétrica multivariada geral, como pode ser visto no primeiro gráfico da Figura 2.6.
- Como exemplificado pelo segundo gráfico da Figura 2.6, vemos que a variância Δ presente na função de distribuição acumulada da distribuição normal também influencia na assimetria do modelo normal-assimétrico multivariado geral.

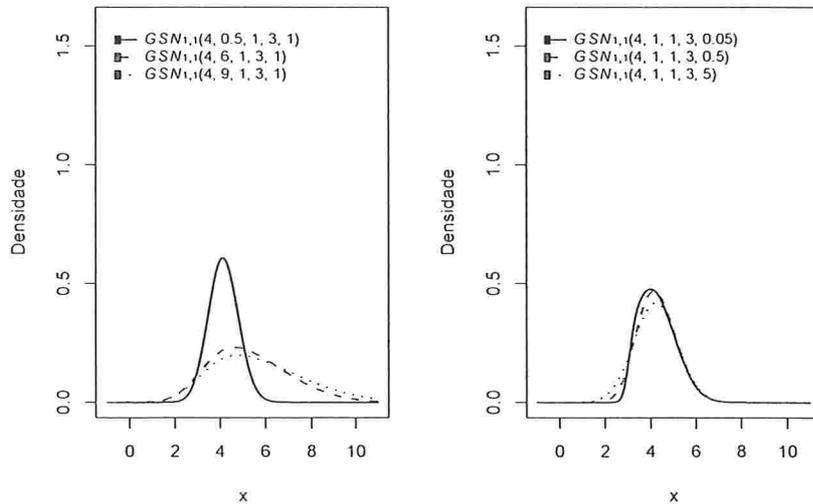


Figura 2.6: Funções de densidade da distribuição normal-assimétrica multivariada geral com $\mu = 4$, $\nu = 3$ e $D = 1$. Gráfico 1: $\Sigma = 0.5, 6$ e 9 e $\Delta = 1$. Gráfico 2: $\Sigma = 1$ e $\Delta = 0.05, 0.5$ e 5 .

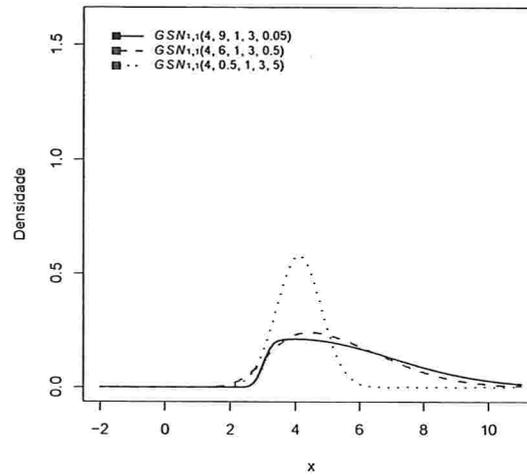


Figura 2.7: Funções de densidade da distribuição normal-assimétrica multivariada geral com $\mu = 4$, $\nu = 3$, $D = 1$ e diferentes valores para Σ e Δ .

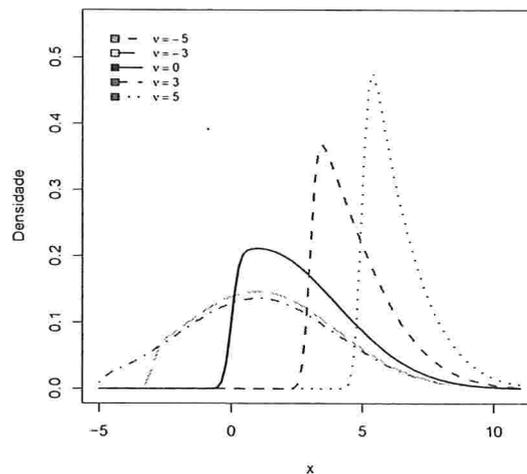


Figura 2.8: Funções de densidade da distribuição normal-assimétrica multivariada geral com $\mu = 1$, $\Sigma = 9$, $D = 1$, $\Delta = 0.05$ e diferentes valores para ν .

- Na Figura 2.7, mostramos que, assim como na distribuição normal-assimétrica fechada, ao associar grandes valores para o parâmetro Σ , com baixos valores para o parâmetro Δ , induzimos uma maior assimetria ao modelo normal-assimétrico multivariado geral.
- Como pode ser visto na Figura 2.8, o comportamento do parâmetro ν no modelo normal-assimétrico multivariado geral é similar ao discutido anteriormente para a distribuição normal-assimétrica fechada. Cabe ressaltar que o parâmetro ν gera uma maior variabilidade no modelo normal-assimétrico multivariado geral do que na distribuição normal-assimétrica fechada.
- Diferentemente do que ocorre no modelo normal-assimétrico fechado, μ não atua como parâmetro de posição no modelo normal-assimétrico fechado, quando consideramos os demais parâmetros da distribuição fixos. Na Figura 2.9 exemplificamos esta situação.
- Já o parâmetro D não atua exclusivamente na assimetria do modelo. Quando $D = 0$, obtemos o modelo normal com parâmetros de posição μ e de escala Σ . Veja a Figura 2.10.

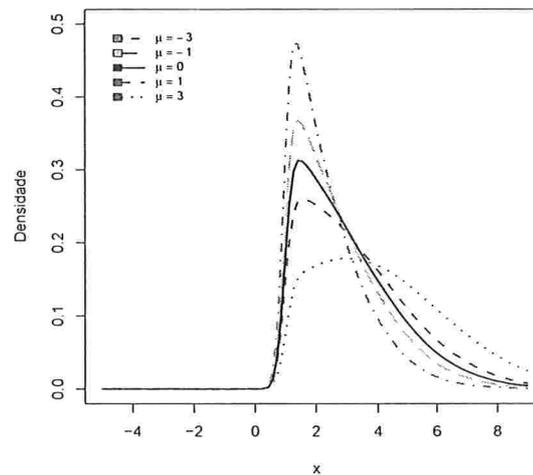


Figura 2.9: Funções de densidade da distribuição normal-assimétrica multivariada geral com $\Sigma = 9$, $D = 1$, $\nu = 1$, $\Delta = 0.05$ e diferentes valores para μ .

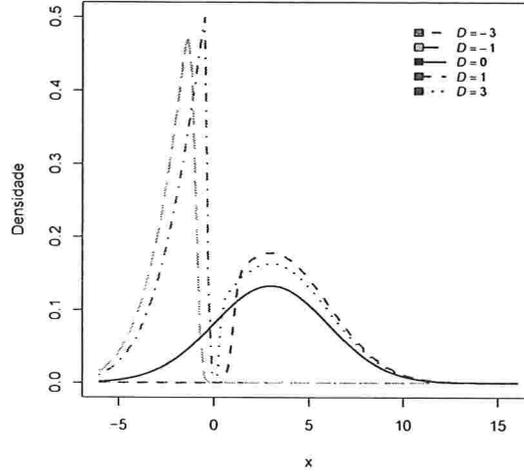


Figura 2.10: Funções de densidade da distribuição normal-assimétrica multivariada geral com $\mu = 3$, $\Sigma = 9$, $\nu = 1$, $\Delta = 0.05$ e diferentes valores para D .

Na proposição a seguir, apresentamos uma equivalência entre as distribuições normal-assimétrica fechada e normal-assimétrica multivariada geral.

Proposição 2.2.1 *Seja Y um vetor aleatório tal que $Y \sim GSN_{p,q}(\mu, \Sigma, D, \nu, \Delta)$ então*

$$Y \sim GSN_{p,q}(\mu, \Sigma, D, \nu, \Delta) \equiv CSN_{p,q}(\mu, \Sigma, D, \nu - D\mu, \Delta). \quad (2.7)$$

Prova:

Basta manipular algebricamente a função de densidade de uma variável aleatória $Y \sim GSN_{p,q}(\mu, \Sigma, D, \nu, \Delta)$. Para isso, considere $f_{p,q}(y)$ e $g_{p,q}(y)$ as funções de densidade das distribuições normal-assimétrica multivariada geral e normal-assimétrica fechada, respectivamente. Para $y \in \mathbb{R}^p$,

$$\begin{aligned} f_{p,q}(y; \mu, \Sigma, D, \nu, \Delta) &= \frac{\phi_p(y; \mu, \Sigma) \Phi_q(Dy; \nu, \Delta)}{\Phi_q(D\mu; \nu, \Delta + D\Sigma D^T)} \\ &= \frac{\phi_p(y; \mu, \Sigma) \Phi_q(Dy - D\mu; \nu - D\mu, \Delta)}{\Phi_q(D\mu - D\mu; \nu - D\mu, \Delta + D\Sigma D^T)} \\ &= \frac{\phi_p(y; \mu, \Sigma) \Phi_q[D(y - \mu); \nu - D\mu, \Delta]}{\Phi_q(0; \nu - D\mu, \Delta + D\Sigma D^T)} \\ &= g_{p,q}(y; \mu, \Sigma, D, \nu - D\mu, \Delta). \end{aligned} \quad (2.8)$$

◆

A equivalência apresentada na Proposição 2.2.1 nos permite obter importantes resultados relacionados à distribuição normal-assimétrica multivariada geral à partir de propriedades já demonstradas para a distribuição normal-assimétrica fechada.

Segundo Gupta, González-Farías & Domínguez-Molina (2004), a função geradora de momentos de um vetor aleatório $Y \sim GSN_{p,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta)$, é dada por:

$$M_Y(\mathbf{t}) = \frac{\Phi_q [D(\boldsymbol{\mu} + \Sigma \mathbf{t}); \boldsymbol{\nu}, \Delta + D\Sigma D^T]}{\Phi_q [D\boldsymbol{\mu}; \boldsymbol{\nu}, \Delta + D\Sigma D^T]} e^{\boldsymbol{\mu}^T \mathbf{t} + \frac{1}{2} \mathbf{t}^T \Sigma \mathbf{t}}, \quad \mathbf{t} \in \mathbb{R}^p. \quad (2.9)$$

Por (2.9), conseguimos obter os dois primeiros momentos de um vetor aleatório Y cuja a distribuição é normal-assimétrica multivariada geral. O primeiro momento é dado por:

$$E[Y] = \boldsymbol{\mu} + \Sigma D^T \frac{\Phi_q^*(D\boldsymbol{\mu}; \boldsymbol{\nu}, \Delta + D\Sigma D^T)}{\Phi_q(D\boldsymbol{\mu}; \boldsymbol{\nu}, \Delta + D\Sigma D^T)}, \quad (2.10)$$

tal que, para qualquer matriz Ω positiva definida,

$$\Phi_q^*(\mathbf{k}; \boldsymbol{\nu}, \Omega) = [\nabla_{\mathbf{k}} \Phi_q(\mathbf{k}; \boldsymbol{\nu}, \Omega)]^T, \quad (2.11)$$

onde $\nabla_{\mathbf{k}} = \left(\frac{\partial}{\partial k_1}, \frac{\partial}{\partial k_2}, \dots, \frac{\partial}{\partial k_q} \right)$ é o operador gradiente.

O segundo momento é:

$$\begin{aligned} E[YY^T] &= \boldsymbol{\mu} \left[\frac{\Phi_q^*(D\boldsymbol{\mu}; \boldsymbol{\nu}, \Delta + D\Sigma D^T)}{\Phi_q(D\boldsymbol{\mu}; \boldsymbol{\nu}, \Delta + D\Sigma D^T)} \right]^T D\Sigma + \Sigma D \frac{\Phi_q^*(D\boldsymbol{\mu}; \boldsymbol{\nu}, \Delta + D\Sigma D^T)}{\Phi_q(D\boldsymbol{\mu}; \boldsymbol{\nu}, \Delta + D\Sigma D^T)} \boldsymbol{\mu}^T + \\ &\quad + \Sigma + \boldsymbol{\mu} \boldsymbol{\mu}^T + \Sigma D^T \frac{\Phi_q^{**}(D\boldsymbol{\mu}; \boldsymbol{\nu}, \Delta + D\Sigma D^T)}{\Phi_q(D\boldsymbol{\mu}; \boldsymbol{\nu}, \Delta + D\Sigma D^T)} D\Sigma, \quad (2.12) \end{aligned}$$

tal que, para qualquer matriz Ω positiva definida, (2.11) ocorre e

$$\Phi_q^{**}(\mathbf{k}; \boldsymbol{\nu}, \Omega) = \nabla_{\mathbf{k}} \nabla_{\mathbf{k}}^T \Phi_q(\mathbf{k}; \boldsymbol{\nu}, \Omega). \quad (2.13)$$

Expressões que nos permitem calcular $\Phi_q^*(\mathbf{k}; \boldsymbol{\nu}, \Omega)$ e $\Phi_q^{**}(\mathbf{k}; \boldsymbol{\nu}, \Omega)$ mais facilmente são apresentadas e discutidas em Arellano-Valle, Genton & Loschi (2009) e Domínguez-Molina, González-Farías & Ramos-Quiroga (2004).

Como já comentado anteriormente, a distribuição normal-assimétrica proposta por Azzalini (1985) é um caso particular distribuição normal-assimétrica multivariada geral. Do ponto de vista univariado, seja Y uma variável aleatória tal que $Y \sim GSN_{1,1}(\mu, \sigma^2, \lambda, \lambda\mu, \sigma^2)$, então

$$Y \sim GSN_{1,1}(\mu, \sigma^2, \lambda, \lambda\mu, \sigma^2) \equiv SN(\mu, \sigma^2, \lambda), \quad (2.14)$$

cuja a notação $SN(\mu, \sigma^2, \lambda)$ corresponde à distribuição normal-assimétrica univariada, com parâmetros de posição μ , de escala σ^2 e de assimetria λ . Sua função de densidade expressa-se por:

$$f_Y(y) = \frac{2}{\sigma} \phi\left(\frac{y - \mu}{\sigma}\right) \Phi\left(\lambda \frac{y - \mu}{\sigma}\right), \quad \text{para } y \in \mathbb{R}. \quad (2.15)$$

Dizemos que uma variável aleatória Z tem distribuição normal-assimétrica padrão quando a sua função de densidade é dada por (2.15), fixados $\mu = 0$ e $\sigma^2 = 1$. Como notação, adotamos $SN(\lambda)$.

Para maiores detalhes a respeito da distribuição normal-assimétrica univariada, ver Rodríguez (2005).

Capítulo 3

Análise de sensibilidade sob uma classe de contaminação multiplicativa

Segundo muitos especialistas a década de 90 é mencionada como um período áureo em termos de estudos relacionados a análise de robustez bayesiana. Neste período muitos estatísticos trabalharam ativamente nesta área, impulsionando um rápido desenvolvimento de novas teorias. Em especial, um grande enfoque foi dado ao desenvolvimento e detalhamento de classes de contaminação, tanto para a distribuição *a priori*, quanto para a verossimilhança e funções de perda.

Berger (1990) descreve algumas características desejáveis às diferentes classes de contaminação. São elas:

- fácil eliciação,
- compatibilidade com o conhecimento *a priori*,
- cálculos simples,
- não deve conter distribuições *a priori* pouco razoáveis.

Em muitos estudos, no entanto, verificou-se que a grande maioria das classes propostas na literatura não resultam em cálculos simples para a análise de robustez bayesiana, sendo comum o uso de técnicas computacionais.

van der Linde (2007) propôs uma maneira de modificar multiplicativamente tanto a função de verossimilhança quanto a distribuição *a priori*. Neste capítulo, exploramos esta idéia de modificar multiplicativamente a distribuição *a priori*, dando origem a uma classe de contaminação multiplicativa. Mostramos que, para algumas especificações particulares desta classe, é possível determinar famílias conjugadas

a ela e obter resultados analíticos para o estudo da média e da variância *a posteriori*, quando o modelo dos dados considerado é o normal. Neste caso, conseguimos delimitar em quais situações pode ser alcançada a robustez.

Considere $f_0(\theta)$ uma função densidade de probabilidade de base e $w(\theta)$ uma função não-negativa tal que $f(\theta) = f_0(\theta)w(\theta)$ seja uma função densidade de probabilidade. A classe de contaminação multiplicativa é dada por:

$$\Gamma_M = \{ f(\theta) = f_0(\theta) w(\theta) : w \in \mathcal{G} \}, \quad (3.1)$$

onde \mathcal{G} é um determinada subclasse de funções não-negativas. Aqui, consideramos que a classe Γ_M contém a distribuição de base $f_0(\theta)$ e, portanto, qualquer que seja a subclasse \mathcal{G} escolhida, esta deve incluir a função $w(\theta) = 1, \forall \theta$.

Um caso particular desta classe ocorre quando consideramos $f_0(\theta)$ uma função de densidade de probabilidade elíptica e $w(\theta) = 2 G(\lambda \theta)$, em que G é a função de distribuição acumulada de uma variável aleatória simétrica em torno de zero, responsável pela contaminação da distribuição *a priori* base, e λ é um parâmetro que reflete o grau de contaminação da distribuição *a priori*, com $\lambda \in \mathbb{R}$.

Note que, na classe de distribuições de probabilidade Γ_M , associamos a distribuição *a priori* de base $f_0(\theta)$ a $\lambda = 0$.

Sob perda quadrática, sabe-se que um estimador de Bayes para θ é dado por $\delta(\mathbf{x}) = E[\theta | \mathbf{x}]$. Se temos interesse em estudar a sensibilidade da média da distribuição *a posteriori* com respeito às diferentes escolhas de distribuições *a priori* contidas na classe Γ_M , devemos considerar as medidas

$$s = \sup_{\Gamma_M} E[\theta | \mathbf{x}] \quad \text{e} \quad t = \inf_{\Gamma_M} E[\theta | \mathbf{x}] \quad (3.2)$$

ou, numa análise equivalente,

$$s_d = \sup_{\Gamma_M} \left| E[\theta | \mathbf{x}] - \hat{\theta}_0 \right| \quad \text{e} \quad t_d = \inf_{\Gamma_M} \left| E[\theta | \mathbf{x}] - \hat{\theta}_0 \right|, \quad (3.3)$$

em que $\hat{\theta}_0 = E_0[\theta | \mathbf{x}]$ corresponde a média da distribuição *a posteriori* obtida quando se considera a distribuição *a priori* $f_0(\theta)$.

O intervalo $\Delta = s_d - t_d$ é uma medida de robustez muito utilizada na literatura. Um valor de Δ “pequeno” indica que qualquer distribuição *a priori* na classe de

contaminação Γ_M pode ser escolhida para se proceder uma análise bayesiana, uma vez que todas elas resultam em medidas *a posteriori* com resultados próximos. No entanto, caso o valor de Δ seja considerado grande, nós dizemos que a distribuição *a posteriori* é sensível a escolha da distribuição *a priori* na classe Γ_M e, portanto, a robustez com relação à classe de contaminação da distribuição *a priori* não ocorre. Em casos como este, Berger, Insua & Ruggeri (2000) sugerem que se refaça o estudo incluindo informações adicionais na classe de contaminação considerada ou que se obtenha uma maior quantidade de dados para que se possa repetir o processo de análise de robustez.

Raciocínio análogo ao anterior, em que se apresentou uma metodologia para estudar o comportamento da média *a posteriori* com respeito a variação de especificações *a priori* contidas em Γ_M , pode ser realizado para quaisquer outras medidas *a posteriori* como, por exemplo, a variância *a posteriori*.

Neste capítulo, discutimos a robustez da média e da variância *a posteriori* no contexto do modelo normal. Sejam X_1, X_2, \dots, X_n uma sequência de variáveis aleatórias independentes e identicamente distribuídas com distribuição normal com média θ e variância τ^2 , então dizemos que $\mathbf{X} = (X_1, X_2, \dots, X_n)$ é uma amostra aleatória simples de $X \sim N(\theta, \tau^2)$. Durante todo o capítulo, os símbolos $\phi(\cdot; \mu, \sigma^2)$ e $\Phi(\cdot; \mu, \sigma^2)$ representam, respectivamente, às funções de densidade e de distribuição acumulada de uma distribuição normal com média μ e variância σ^2 . Usaremos a notação $\phi(\cdot)$ e $\Phi(\cdot)$, quando $\mu = 0$ e $\sigma^2 = 1$ em $\phi(\cdot; \mu, \sigma^2)$ e $\Phi(\cdot; \mu, \sigma^2)$, respectivamente.

3.1 O modelo normal com variância conhecida.

Seja $\mathbf{X} = (X_1, X_2, \dots, X_n)$ uma amostra aleatória simples de tamanho n da variável aleatória $X \sim N(\theta, \tau^2)$, em que τ^2 é conhecido e θ é a quantidade de interesse.

A distribuição *a priori* usualmente especificada para θ é a normal, que é conjugada para o modelo estatístico considerado. Nossa idéia é propor uma classe de distribuições *a priori* que contenha a distribuição normal, mas permita a inclusão da suposição de assimetria.

Para a distribuição *a priori* do parâmetro de posição θ , considere um caso par-

particular da classe de contaminação multiplicativa Γ_M definida em (3.1), de tal forma que, para μ e σ^2 fixados e λ variando em \mathbb{R} ,

$$f_0(\theta) = \phi(\theta; \mu, \sigma^2) \quad (3.4)$$

e

$$w(\theta) = 2 \Phi[\lambda\theta; \lambda\mu, \sigma^2]. \quad (3.5)$$

Numa interpretação da classe de contaminação multiplicativa *a priori*, podemos identificar a distribuição *a priori* de base pela $N(\mu, \sigma^2)$, a função contaminadora por $G(\lambda\theta) = \Phi_1(\lambda\theta; \lambda\mu, \sigma^2)$ e a constante normalizadora por $c = 2$. Temos então que a classe de contaminação multiplicativa *a priori* é dada por

$$\Gamma_M = \left\{ f_\lambda(\theta) = \frac{2}{\sigma} \phi\left(\frac{\theta - \mu}{\sigma}\right) \Phi\left(\lambda \frac{\theta - \mu}{\sigma}\right) : \lambda \in \mathbb{R} \right\}.$$

Note que esta específica classe de contaminação multiplicativa *a priori* é gerada à partir da variação de λ no conjunto dos reais e que, em decorrência disso, a classe Γ_M contém um número infinito de funções de densidade parametrizadas segundo a normal-assimétrica. Nosso interesse é verificar o efeito do parâmetro λ nas inferências *a posteriori*.

Neste contexto, um primeiro resultado obtido diz respeito ao fechamento sob conjugação e expressões para a média e a variância *a posteriori* do parâmetro de posição θ . Lembrando que a distribuição $SN(\mu, \sigma^2, \lambda)$ é equivalente a distribuição $GSN_{1,1}(\mu, \sigma^2, \lambda, \lambda\mu, \sigma^2)$, mostramos que tanto a classe de distribuições *a priori*, quanto a classe de distribuições *a posteriori* para o parâmetro θ decorrente dela, pertencem à família normal-assimétrica multivariada geral especificada na Seção 2.2.

Proposição 3.1.1 (Fechamento sob conjugação.) *Seja $\mathbf{X} = (X_1, X_2, \dots, X_n)$ uma amostra aleatória simples de tamanho n da variável $X \sim N(\theta, \tau^2)$, com τ^2 conhecido. Se $\theta \sim GSN_{1,1}(\mu, \sigma^2, \lambda, \lambda\mu, \sigma^2)$, então a distribuição *a posteriori*, a média e a variância *a posteriori* para o parâmetro de posição são dadas, respectivamente, por*

$$(i) \quad \theta | \mathbf{X} \sim GSN_{1,1} \left(\frac{n\bar{x}\sigma^2 + \mu\tau^2}{n\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2}, \lambda, \lambda\mu, \sigma^2 \right), \quad (3.6)$$

$$(ii) \quad E_\lambda[\theta | \mathbf{x}] = \frac{n\bar{x}\sigma^2 + \mu\tau^2}{n\sigma^2 + \tau^2} + \frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma\tau^2}{\sqrt{n\sigma^2 + \tau^2}} \times \\ \times \frac{\phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x}-\mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x}-\mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]}, \quad (3.7)$$

$$(iii) \quad Var_\lambda[\theta | \mathbf{x}] = \frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2} - \frac{\lambda^2\tau^2}{n\sigma^2 + \tau^2(1+\lambda^2)} \frac{\phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x}-\mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x}-\mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]} \times \\ \times \frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2} \left\{ \frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x}-\mu)}{\sqrt{n\sigma^2 + \tau^2}} + \frac{\phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x}-\mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x}-\mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]} \right\}. \quad (3.8)$$

A demonstração desta proposição e de outros resultados que serão apresentados neste capítulo estão no Apêndice A.

No corolário a seguir afirmamos que a variância *a posteriori* apresentada na Proposição 3.1.1, item (iii), assume valores positivos, uma vez que está limitada entre 0 e $\frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2}$.

Corolário 3.1.1 *A variância a posteriori do parâmetro de posição θ está limitada entre 0 e $\frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2}$, para todo n , \bar{x} , σ^2 , τ^2 , μ e λ .*

Note que, quando $\lambda = 0$, (3.7) e (3.8) se reduzem a

$$E_0[\theta | \mathbf{x}] = \frac{n\bar{x}\sigma^2 + \mu\tau^2}{n\sigma^2 + \tau^2} \quad \text{e} \quad Var_0[\theta | \mathbf{x}] = \frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2}, \quad (3.9)$$

em que $\theta | \mathbf{X} \sim N\left(\frac{n\bar{x}\sigma^2 + \mu\tau^2}{n\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2}\right)$. E, naturalmente, o resultado de conjugação sob normalidade é recuperado.

De posse destes resultados, obtemos na proposição a seguir medidas de distâncias relacionadas a média e a variância *a posteriori* e estudamos algumas de suas propriedades. Através destas distâncias, vamos especificar medidas de robustez para avaliar a sensibilidade da média e da variância da distribuição *a posteriori*, segundo a classe de contaminação multiplicativa *a priori* Γ_M .

Proposição 3.1.2 Considere $E_\lambda[\theta | \mathbf{x}]$, $Var_\lambda[\theta | \mathbf{x}]$, $\hat{\theta}_0 = E_0[\theta | \mathbf{x}]$ e $\hat{\tau}_0 = Var_0[\theta | \mathbf{x}]$ como apresentados em (3.7), (3.8) e (3.9), respectivamente. Então,

(i)

$$d_M = \left| E_\lambda[\theta | \mathbf{x}] - \hat{\theta}_0 \right| = \frac{\phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]} \frac{|\lambda|}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma \tau^2}{\sqrt{n\sigma^2 + \tau^2}}$$

e

(ii)

$$d_V = \left| Var_\lambda[\theta | \mathbf{x}] - \hat{\tau}_0 \right| = \frac{\lambda^2 \tau^2}{n\sigma^2 + \tau^2(1+\lambda^2)} \frac{\sigma^2 \tau^2}{n\sigma^2 + \tau^2} \frac{\phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]} \times \\ \times \left\{ \frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} + \frac{\phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]} \right\}. \quad (3.10)$$

Note que ambas as distâncias d_M e d_V são iguais a zero quando $\lambda = 0$.

Como discutido na Introdução deste trabalho, no contexto de análise de robustez global, o uso de uma classe de distribuições *a priori* associada à uma função de verossimilhança gera uma classe de distribuições *a posteriori* e, decorrente desta classe, um conjunto de medidas *a posteriori* como, por exemplo, um conjunto de médias *a posteriori*. O critério para avaliar a robustez com respeito a escolha de distribuições *a priori* na classe Γ_M será baseado na distância entre o mínimo e o máximo desvio absoluto das médias *a posteriori* obtidas segundo a distribuição *a priori* base e quaisquer outras distribuições *a priori* em Γ_M .

As próximas proposições a serem apresentadas nesta seção têm como intenção estudar classes de medidas *a posteriori*, a saber, $\{d_M(\lambda) : \lambda \in \mathbb{R}\}$ e $\{d_V(\lambda) : \lambda \in \mathbb{R}\}$, apresentar o supremo e o ínfimo de cada uma delas e fazer uma conexão destes resultados com a análise de robustez bayesiana. Para isso, faremos uma análise do comportamento das medidas d_M e d_V em função de λ , já que é a variação de λ nos reais que gera o conjunto de medidas *a posteriori*.

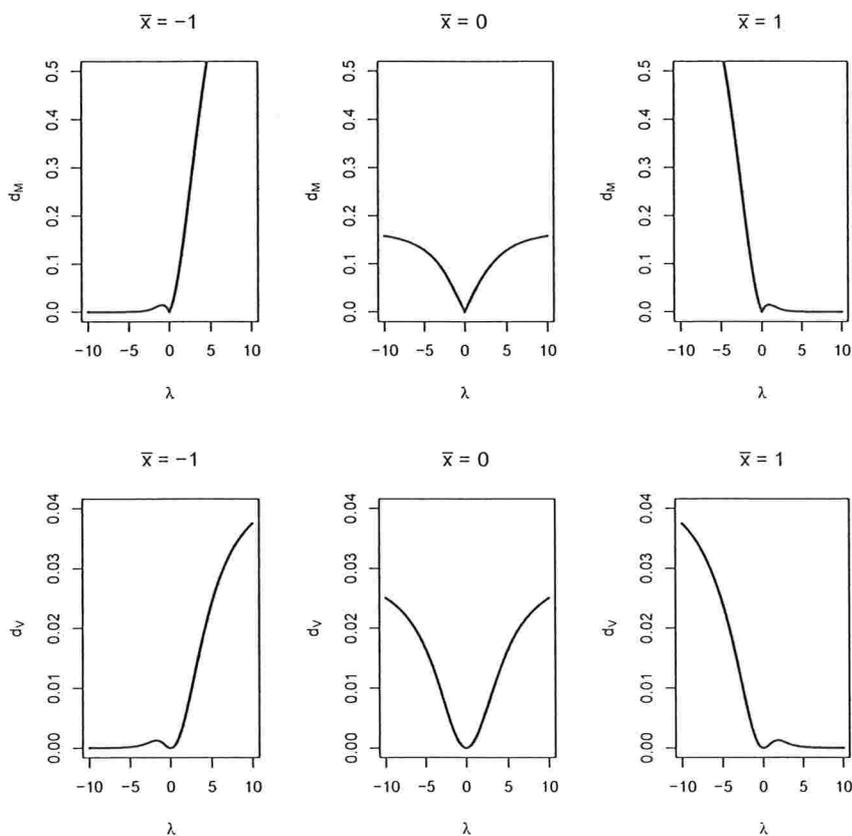


Figura 3.1: Gráficos de $d_M(\lambda)$ e $d_V(\lambda)$ para $n = 20$, $\tau^2 = 1$, $\mu = 0$ e $\sigma^2 = 1$.

Na Figura 3.1 apresentamos os gráficos das distâncias d_M e d_V em função de λ , para diferentes valores de média amostral, quando fixados os parâmetros em $n = 20$, $\tau^2 = 1$, $\mu = 0$ e $\sigma^2 = 1$. Note que, dependendo dos valores de \bar{x} considerados e dos intervalos de valores para λ analisados, as distâncias $d_M(\lambda)$ e $d_V(\lambda)$ apresentam comportamentos bem díspares, ora comportando-se monotonicamente, ora apresentando pontos de máximo local. Isto nos motiva a estudar o comportamento destas medidas sob dois aspectos, quando d_M e d_V são funções monótonas em λ ou quando não são.

Situação I: $d_M(\lambda)$ e $d_V(\lambda)$ se comportam monotonicamente.

Na Proposição 3.1.3 determinamos quais as condições para que as distâncias $d_M(\lambda)$ e $d_V(\lambda)$ se comportem monotonicamente. Veremos adiante que este conhe-

cimento facilita a determinação das medidas extremas (supremo e ínfimo) dos conjuntos de medidas *a posteriori* gerados por d_M e d_V , quando $\lambda \in \mathbb{R}$.

Proposição 3.1.3 *As distâncias $d_M(\lambda)$ e $d_V(\lambda)$ são contínuas e*

- (i) *monotônicas estritamente decrescentes, quando $\lambda \leq 0$ e $\bar{x} > \mu$,*
- (ii) *monotônicas estritamente crescentes, quando $\lambda \geq 0$ e $\bar{x} < \mu$, e*
- (iii) *simétricas em relação a $\lambda = 0$. Quando $\lambda \leq 0$ ($\lambda \geq 0$) e $\bar{x} = \mu$, $d_M(\lambda)$ e $d_V(\lambda)$ são monotônicas estritamente decrescentes (crescentes).*

Como implicação direta da Proposição 3.1.3, conseguimos determinar quais valores de λ geram valores extremos do conjunto de medidas de distâncias *a posteriori* nos intervalos onde a função d_M se comporta monotonicamente. No caso em que $\lambda \leq 0$ e $\bar{x} > \mu$, temos que $\sup_{\lambda} d_M$ e $\sup_{\lambda} d_V$ são obtidos quando λ tende para $-\infty$. No segundo caso, em que $\lambda \geq 0$ e $\bar{x} < \mu$, o supremo desses conjuntos de distâncias são obtidos quando λ tende para ∞ . No terceiro caso, em que $\bar{x} = \mu$, temos que as distâncias d_M e d_V são simétricas em torno do eixo $\lambda = 0$ e, portanto, podemos considerar λ tendendo para $-\infty$ ou ∞ para obtermos $\sup_{\lambda} d_M$ e $\sup_{\lambda} d_V$. Quanto as medidas $\inf_{\lambda} d_M$ e $\inf_{\lambda} d_V$, precisamos lembrar que, por construção, as distâncias d_M e d_V são não-negativas e, portanto, para quaisquer casos considerados na Proposição 3.1.3,

$$\inf_{\lambda} d_M = \inf_{\lambda} |E_{\lambda}[\theta | \mathbf{x}] - E_0[\theta | \mathbf{x}]| = 0 \quad \Leftrightarrow \quad \lambda = 0 \quad (3.11)$$

$$\inf_{\lambda} d_V = \inf_{\lambda} |Var_{\lambda}[\theta | \mathbf{x}] - Var_0[\theta | \mathbf{x}]| = 0 \quad \Leftrightarrow \quad \lambda = 0. \quad (3.12)$$

Através da Proposição 3.1.3 e das implicações que dela decorrem, obtemos expressões para o supremo dos conjuntos das distâncias das esperanças *a posteriori* e das variâncias *a posteriori* para três diferentes situações, apresentadas na seguinte proposição:

Proposição 3.1.4

- (i) *Quando $\lambda \leq 0$ e $\bar{x} > \mu$, temos que*

$$\sup_{\lambda} d_M(\lambda) = \lim_{\lambda \rightarrow -\infty} d_M(\lambda) = \frac{\phi \left[\frac{-\sigma n(\bar{x} - \mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[\frac{-\sigma n(\bar{x} - \mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]} \frac{\sigma \tau}{\sqrt{n\sigma^2 + \tau^2}} \quad (3.13)$$

e

$$\begin{aligned} \sup_{\lambda} d_V(\lambda) &= \lim_{\lambda \rightarrow -\infty} d_V(\lambda) = \frac{\phi \left[-\frac{\sigma n(\bar{x}-\mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[-\frac{\sigma n(\bar{x}-\mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]} \frac{\sigma^2 \tau^2}{n\sigma^2 + \tau^2} \times \\ &\quad \times \left\{ -\frac{\sigma n(\bar{x}-\mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} + \frac{\phi \left[-\frac{\sigma n(\bar{x}-\mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[-\frac{\sigma n(\bar{x}-\mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]} \right\}. \end{aligned} \quad (3.14)$$

(ii) Quando $\lambda \geq 0$ e $\bar{x} < \mu$, temos que

$$\sup_{\lambda} d_M(\lambda) = \lim_{\lambda \rightarrow \infty} d_M(\lambda) = \frac{\phi \left[\frac{\sigma n(\bar{x}-\mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[\frac{\sigma n(\bar{x}-\mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]} \frac{\sigma \tau}{\sqrt{n\sigma^2 + \tau^2}} \quad (3.15)$$

e

$$\begin{aligned} \sup_{\lambda} d_V(\lambda) &= \lim_{\lambda \rightarrow \infty} d_V(\lambda) = \frac{\phi \left[\frac{\sigma n(\bar{x}-\mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[\frac{\sigma n(\bar{x}-\mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]} \frac{\sigma^2 \tau^2}{n\sigma^2 + \tau^2} \times \\ &\quad \times \left\{ \frac{\sigma n(\bar{x}-\mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} + \frac{\phi \left[\frac{\sigma n(\bar{x}-\mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[\frac{\sigma n(\bar{x}-\mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]} \right\}. \end{aligned} \quad (3.16)$$

(iii) Quando $\bar{x} = \mu$, as expressões para $\sup_{\lambda} d_M(\lambda)$ em (3.13) e (3.15) são iguais e, portanto, qualquer uma destas expressões pode ser usada. Isto ocorre devido a simetria da função $d_M(\lambda)$ em relação ao eixo $\lambda = 0$. Conclusão análoga pode ser tomada com respeito à distância d_V .

Em geral, estamos interessados em estudar o comportamento das medidas limites das distâncias d_M e d_V , a medida em que \bar{x} varia. Considere então os resultados anteriormente apresentados e vejamos o que ocorre com essas medidas limites quando as funções d_M e d_V se comportam monotonicamente em função de λ .

Proposição 3.1.5 *Considere as medidas limites $\sup_{\lambda} d_M$ e $\sup_{\lambda} d_V$ descritas na Proposição 3.1.4. Então,*

(i) quando $\lambda \leq 0$ e $\bar{x} > \mu$,

$$\lim_{\bar{x} \rightarrow \infty} \sup_{\lambda} d_M = \infty \quad e \quad \lim_{\bar{x} \rightarrow \infty} \sup_{\lambda} d_V = \frac{\sigma^2 \tau^2}{n\sigma^2 + \tau^2},$$

(ii) quando $\lambda \geq 0$ e $\bar{x} < \mu$,

$$\lim_{\bar{x} \rightarrow -\infty} \sup_{\lambda} d_M = \infty \quad e \quad \lim_{\bar{x} \rightarrow -\infty} \sup_{\lambda} d_V = \frac{\sigma^2 \tau^2}{n\sigma^2 + \tau^2}.$$

Situação II: $d_M(\lambda)$ e $d_V(\lambda)$ não se comportam monotonicamente.

Nas regiões onde as distâncias d_M e d_V não são funções monotônicas em λ , a determinação dos valores de λ que maximizam estas funções devem ser obtidos computacionalmente, para um fixado valor de \bar{x} . Ou seja, para cada valor de \bar{x} considerado, deve-se obter o conjunto de distâncias d_M e d_V e os respectivos valores de λ que maximizam estas funções. Note que, não necessariamente o valor de λ que maximiza d_M deva ser igual ao valor de λ que maximiza d_V . Sendo assim, como notação, assumiremos que λ_M e λ_V referem-se, respectivamente, ao valor de λ que maximiza as distâncias $d_M(\lambda)$ e $d_V(\lambda)$.

Na proposição abaixo apresentamos alguns resultados que facilitam este trabalho computacional. Conseguimos mostrar que se existir este λ_M (ou λ_V) que maximiza a distância $d_M(\lambda)$ ($d_V(\lambda)$), ele será único.

Proposição 3.1.6 *Considere as distâncias d_M e d_V , $a^2 = \frac{\sigma^2 n^2 (\bar{x} - \mu)^2}{n\sigma^2 + \tau^2}$ e as constantes $K_{M_0} = 0.8399238$ e $K_{V_0} = 1.598539$. Para $\lambda > 0$ e $\bar{x} > \mu$ ($\lambda < 0$ e $\bar{x} < \mu$)*

(i) *se $a^2 > K_{M_0}^2 \tau^2$ então existe um único λ_M dado por*

$$\lambda_M = K_{M_0} \sqrt{\frac{n\sigma^2 + \tau^2}{a^2 - \tau^2 K_{M_0}^2}} \quad \left(\lambda_M = -K_{M_0} \sqrt{\frac{n\sigma^2 + \tau^2}{a^2 - \tau^2 K_{M_0}^2}} \right),$$

caso contrário não existe λ_M .

(ii) *se $a^2 > K_{V_0}^2 \tau^2$ então existe um único λ_V dado por*

$$\lambda_V = K_{V_0} \sqrt{\frac{n\sigma^2 + \tau^2}{a^2 - \tau^2 K_{V_0}^2}} \quad \left(\lambda_V = -K_{V_0} \sqrt{\frac{n\sigma^2 + \tau^2}{a^2 - \tau^2 K_{V_0}^2}} \right),$$

caso contrário não existe λ_V .

Como já mencionado na demonstração da Proposição 3.1.6 apresentada no Apêndice A, as constantes K_{M_0} e K_{V_0} foram obtidas através de métodos numéricos como solução de equações cujos os termos possuem expressões dependentes da fração $\frac{\phi(k)}{\Phi(k)}$.

Relacionados aos resultados da Proposição 3.1.6, conseguimos determinar as medidas limites de d_M e d_V , quando $\lambda > 0$ e $\bar{x} > \mu$ (ou $\lambda < 0$ e $\bar{x} < \mu$).

Nos casos em que se é possível determinar qual λ_M (λ_V) maximiza a distância d_M (d_V), a medida $\sup_{\lambda} d_M$ ($\sup_{\lambda} d_V$) é obtida substituindo-se λ_M (λ_V) na distância

d_M (d_V) definida na Proposição 3.1.2. Caso tal λ_M (λ_V) não exista, é porque em função de λ a distância d_M (d_V) se comporta monotonicamente e, nesta situação, as discussões são similares às da Proposição 3.1.3. Note que as demonstrações que permitiram a análise realizada anteriormente serão omitidas, uma vez que estas aparecem como decorrência direta da Proposição 3.1.6.

Fazendo uso da teoria discutida nesta seção, apresentamos a seguir um exemplo no qual procuramos discutir questões relativas a sensibilidade das estimativas da média e da variância *a posteriori*, quando se considera a classe de contaminação multiplicativa para a distribuição *a priori* do parâmetro de posição de um modelo normal com variância conhecida.

3.1.1 Estudo de sensibilidade.

O'Hagan (1994) apresenta um estudo de sensibilidade para a classe de contaminação ϵ , no qual avalia a influência desta classe sobre os extremos dos conjuntos de medidas *a posteriori* d_M e d_V à medida em que \bar{x} varia. Nesta seção, nós faremos um estudo similar considerando a classe de contaminação multiplicativa *a priori*.

Para isso, seja $\mathbf{X} = (X_1, X_2, \dots, X_n)$ uma amostra aleatória simples de tamanho n da variável aleatória X com distribuição $N(\theta, 1)$. Para especificação *a priori* do parâmetro de posição da distribuição normal, considere a classe de contaminação multiplicativa apresentada em (3.1) de tal forma que a distribuição *a priori* base seja $N(0, 1)$ e $w(\theta) = 2\Phi(\lambda\theta)$.

Note que $\bar{x} \in \mathbb{R}$ e que, segundo a teoria desenvolvida anteriormente, para cada valor de \bar{x} , temos relacionado um valor para $\sup_{\lambda} d_M$, $\inf_{\lambda} d_M$, $\sup_{\lambda} d_V$ e $\inf_{\lambda} d_V$.

Para avaliar a sensibilidade das distâncias d_M e d_V à variação de \bar{x} , vamos considerar apenas os casos em que $\lambda \geq 0$, pois para $\lambda < 0$ a resolução é simétrica. Consideraremos então duas situações, sendo elas $\bar{x} < 0$ (Caso 1) e $\bar{x} > 0$ (Caso 2).

Caso I: $\bar{x} < 0$.

Neste caso, para $\lambda \geq 0$, conseguimos determinar expressões explícitas para as medidas $\sup_{\lambda} d_M$ e $\sup_{\lambda} d_V$, sendo elas dadas, respectivamente, em (3.15) e (3.16).

A Figura 3.2 apresenta quatro gráficos, sendo que os dois da primeira linha são para $n = 20$, em que o primeiro refere-se a distância absoluta de $E_{\lambda}[\theta | \mathbf{x}]$ a $\hat{\theta}_0$ e o

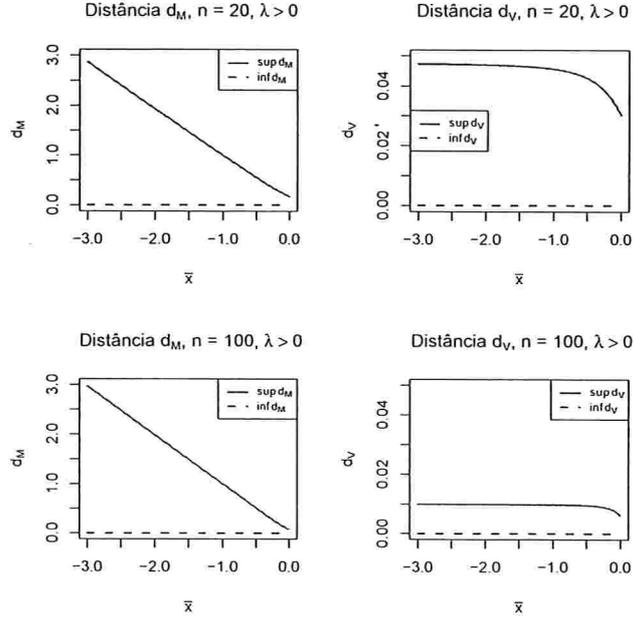


Figura 3.2: Distâncias extremas de d_M e d_V , quando $n = 20$ e 100 , $\sigma^2 = \tau^2 = 1$ e $\bar{x} < 0$.

segundo diz respeito a distância absoluta de $\text{Var}_\lambda[\theta | \mathbf{x}]$ a $\hat{\tau}_0$. Os gráficos restantes são similares aos primeiros, mas considera o tamanho da amostra igual a 100.

Note que, para $\lambda = 0$ temos que $\inf_\lambda d_M = \inf_\lambda d_V = 0$.

Por construção, fixado \bar{x} , o intervalo formado pelos valores extremos do conjunto de distâncias d_M ($\sup_\lambda d_M$ e $\inf_\lambda d_M$) corresponde a maior variação possível entre a média *a posteriori* gerada pela distribuição *a priori* base ($\hat{\theta}_0$) e qualquer outra média *a posteriori* gerada pelas demais distribuições *a priori* contidas em Γ_M . Interpretação equivalente ocorre com a distância d_V .

Analisando os gráficos da Figura 3.2, percebemos que quanto menor o valor de \bar{x} , maior a distância entre os valores extremos do conjunto de distâncias d_M . Ou seja, quanto menor o valor de \bar{x} , maior a distância entre $\sup_\lambda d_M$ e $\inf_\lambda d_M$ e, por conseguinte, maior a variação dos valores das médias *a posteriori* obtidas segundo a classe de contaminação multiplicativa *a priori* normal-assimétrica. Dessa forma, a medida em que \bar{x} se distancia de zero, a média *a posteriori* se mostrou muito sensível à escolha da distribuição *a priori* na classe de contaminação multiplicativa Γ_M , mostrando, portanto, que a média *a posteriori* não é robusta com relação a

variação de especificações *a priori* contidas em Γ_M . Mesmo quando se considera um tamanho de amostra maior, chega-se a mesma conclusão.

Para explicarmos a que se deve esta falta de robustez quando $\bar{x} < 0$ e $\lambda \geq 0$, devemos lembrar que o valor de λ que resulta na maior distância possível d_M ocorre quando $\lambda \rightarrow \infty$. Tal resultado surgiu como consequência da Proposição 3.1.3. Veja que, quando $\lambda \rightarrow \infty$, a distribuição $SN(\lambda)$ converge para uma distribuição normal truncada à esquerda de zero ($HN(0, 1)$), como pode ser visto na Figura 3.3. Dessa forma, considerar $\lambda \rightarrow \infty$ induz a idéia de que os dados amostrados devam ser positivos, o que, de fato, contradiz a informação trazida pela amostra, para a qual estamos considerando $\bar{x} < 0$.

Sendo assim, esta falta de robustez no intervalo de $\bar{x} < 0$, quando $\lambda \geq 0$, já era esperada, uma vez que nestas condições existe um conflito entre a informação sugerida pela distribuição *a priori* e a média amostral que reflete a verossimilhança.

Os gráficos da Figura 3.4 exemplificam tal situação. Consideramos aqui $n = 20$, $\bar{x} = -3$ e para representar $\lambda \rightarrow \infty$ tomamos como aproximação $\lambda = 20$. De fato, o $\sup_{\lambda} d_M$ calculado através do resultado (3.15), com $n = 20$ e $\bar{x} = -3$, resulta em 2.87. Veja que, através do primeiro gráfico da Figura 3.4, fica evidente o conflito entre a verossimilhança e a distribuição *a priori* e que quanto menor o valor de \bar{x} considerado, maior a nossa percepção da existência deste conflito. Além disso, observe o grande impacto da escolha das distribuições *a priori* sobre a distribuição *a posteriori*, principalmente quanto a sua localização.

De maneira geral, notamos como característica o conflito entre a informação fornecida pela verossimilhança e a distribuição *a priori* sempre que a média amostral tem um sinal diferente do espaço de variação do parâmetro de assimetria considerado para a distribuição *a priori*, tornando-se mais evidente na medida em que o valor absoluto de \bar{x} cresce.

Com respeito a análise da medida d_V , nota-se na Figura 3.2 que a distância entre $\text{Var}_{\lambda}[\theta | \mathbf{x}]$ e $\hat{\tau}_0$ é pequena, e torna-se menor, quanto maior o tamanho da amostra. Isto pode ser comprovado pela Propriedade 3.1.5 em que, para $n = 20$, têm-se que $\lim_{\bar{x} \rightarrow -\infty} \hat{s}_{d_V} = 1/21$ e para $n = 100$, $\lim_{\bar{x} \rightarrow -\infty} \hat{s}_{d_V} = 1/101$. Concluímos então que, para $\bar{x} < 0$ e $\lambda > 0$, ocorre a robustez da variância *a posteriori* com relação a classe de contaminação multiplicativa *a priori* Γ_M .

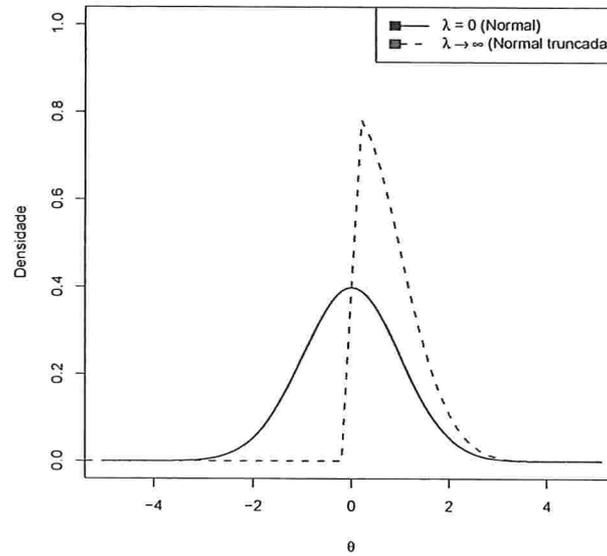


Figura 3.3: Funções de densidade da distribuição *a priori* para θ dadas por $SN(\lambda)$, com $\lambda = 0$ e $\lambda \rightarrow \infty$.

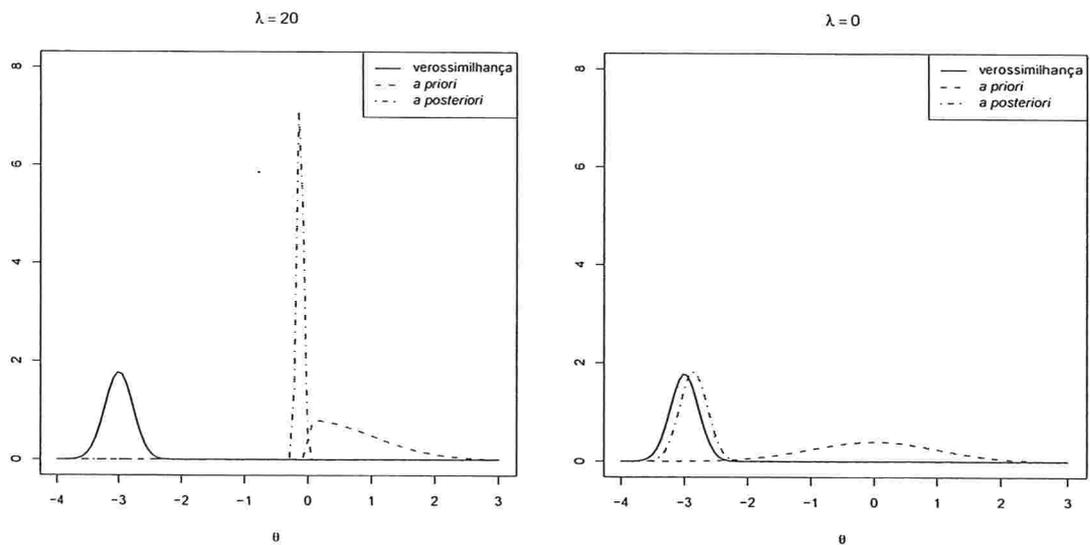


Figura 3.4: Comparação da distribuição *a priori*, verossimilhança ($\bar{x} = -3$, $n = 20$) e distribuição *a posteriori*, para $\lambda = 20, 0$.

Caso II: $\bar{x} > 0$.

Nesta situação, como já argumentado anteriormente, as medidas $\sup_{\lambda} d_M$ e $\sup_{\lambda} d_V$ não possuem expressões fechadas, pois não conseguimos determinar algebricamente quais valores de λ maximizam as funções $d_M(\lambda)$ e $d_V(\lambda)$. Dessa forma, métodos numéricos de maximização devem ser utilizados para se determinar qual o valor de λ que maximiza as distâncias $d_M(\lambda)$ e $d_V(\lambda)$, para um dado \bar{x} fixo. Na Proposição 3.1.6 apresentamos alguns resultados que facilitam este processo de maximização. Além disso, mostramos que se existir um valor de λ que maximiza $d_M(\lambda)$ (ou $d_V(\lambda)$), ele será único.

A Figura 3.5 apresenta os gráficos com as distâncias extremas de d_M e d_V , quando $\sigma^2 = \tau^2 = 1$ e $n = 20$ e 100.

Analisando os gráficos da distância d_M (Figura 3.5), podemos notar que a distância entre $E_{\lambda}[\theta | \mathbf{x}]$ e $\hat{\theta}_0$ tende muito rapidamente a zero, indicando a robustez da $E[\theta | \mathbf{x}]$ na classe de distribuições *a priori* multiplicativa. Mesmo para $n = 20$ e $0 < \bar{x} < 2$, onde d_M é mais significativo, nota-se que esta distância não é muito

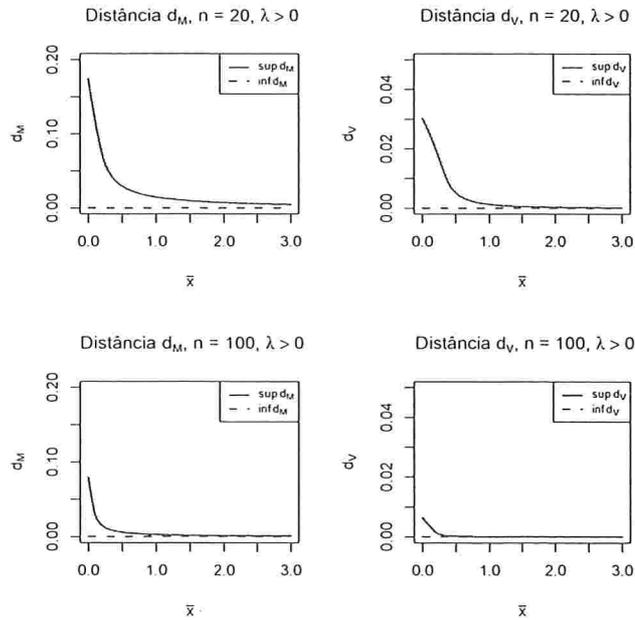


Figura 3.5: Distâncias extremas de d_M e d_V , quando $n = 20$ e 100, $\sigma^2 = \tau^2 = 1$ e $\bar{x} > 0$.

grande, sendo no máximo 0.2. Note que a indicação gráfica de robustez da medida d_V é mais evidente.

A medida em que o tamanho da amostra cresce, notamos que a robustez torna-se mais aparente, com as curvas limites indo a zero mais rapidamente, tanto nos gráficos da distância d_M quanto da distância d_V .

Na Figura 3.6 apresentamos uma comparação entre distribuições *a priori*, para $\lambda = 0.28$ e 0, e as respectivas distribuições *a posteriori* geradas. Consideramos aqui $\lambda = 0.28$ como representante do espaço $\lambda > 0$, pois é este o valor de λ que fornece o maior valor da distância d_M , quando fixados os demais parâmetros desta função em $\bar{x} = 3$, $n = 20$, $\sigma^2 = \tau^2 = 1$ e $\mu = 0$. Veja que, de fato, a distribuição *a posteriori* para o parâmetro de posição θ praticamente não é influenciada pela escolha das distribuições *a priori* na classe de contaminação multiplicativa Γ_M . Dessa maneira, podemos falar em robustez com respeito a escolha de distribuições *a priori* em Γ_M , quando $\bar{x} > 0$ e $\lambda > 0$.

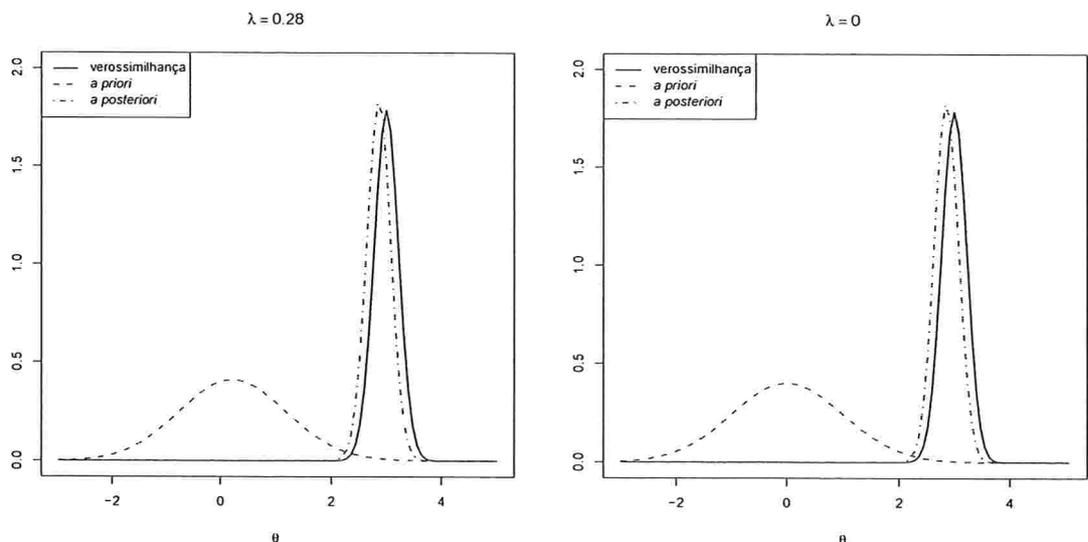


Figura 3.6: Comparação da distribuição *a priori*, verossimilhança ($\bar{x} = 3$, $n = 20$) e distribuição *a posteriori*, para $\lambda = 20, 0$.

3.2 O modelo normal com variância desconhecida.

Considere agora $\mathbf{X} = (X_1, X_2, \dots, X_n)$ uma amostra aleatória simples de tamanho n da variável aleatória X com distribuição $N(\theta, \tau^2)$, com ambos os parâmetros desconhecidos. Vamos assumir independência *a priori* entre os parâmetros θ e τ^2 de tal forma que $h(\theta, \tau^2) = h(\theta)h(\tau^2)$.

Como na Seção 3.1, vamos considerar para a distribuição *a priori* do parâmetro de posição θ distribuições contidas na classe de contaminação multiplicativa descrita em (3.1), com a parametrização dada pela distribuição normal-assimétrica.

Para a distribuição *a priori* do parâmetro de escala, vamos considerar dois tipos de distribuições *a priori*.

Priori 1. Distribuição *a priori* não-informativa, isto é, $h(\tau^2) \propto \frac{1}{\tau^2}$.

Priori 2. $\tau^2 \sim GI(a, b)$, com $a > 0$ e $b > 0$, em que $GI(a, b)$ corresponde a distribuição gama invertida, com parâmetros $a > 0$ e $b > 0$.

A seguir, apresentamos resultados associados a cada uma das distribuições *a priori* para τ^2 especificadas anteriormente. O símbolo S^2 representa a variância amostral, cuja expressão é dada por $S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$.

Priori 1:

Ao adotarmos uma distribuição *a priori* não-informativa e imprópria para o parâmetro τ^2 , uma questão natural é se a distribuição *a posteriori* conjunta resultante é uma genuína função densidade de probabilidade. Veremos adiante que a distribuição *a posteriori* resultante é própria.

Proposição 3.2.1 *Seja $\mathbf{X} = (X_1, X_2, \dots, X_n)$ uma amostra aleatória simples de tamanho n da variável $X \sim N(\theta, \tau^2)$ e considere que a distribuição *a priori* para o parâmetro de posição seja $\theta \sim SN(\mu, \sigma^2, \lambda)$. Assumindo independência *a priori* entre os parâmetros θ e τ^2 e se $h(\tau^2) \propto \frac{1}{\tau^2}$ então a distribuição *a posteriori* conjunta é própria.*

A seguir, apresentamos um resultado que diz respeito a obtenção das distribuições condicionais completas *a posteriori* para os parâmetros de posição θ e escala τ^2 .

Proposição 3.2.2 *Seja $\mathbf{X} = (X_1, X_2, \dots, X_n)$ uma amostra aleatória simples de tamanho n da variável $X \sim N(\theta, \tau^2)$ e considere que a distribuição a priori para o parâmetro de posição seja $SN(\mu, \sigma^2, \lambda)$. Assumindo independência a priori entre os parâmetros θ e τ^2 , se $h(\tau^2) \propto \frac{1}{\tau^2}$ então*

$$\theta \mid \tau^2, \mathbf{x} \sim GSN_{1,1} \left(\frac{n\bar{x}\sigma^2 + \mu\tau^2}{n\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2}, \lambda, \lambda\mu, \sigma^2 \right), \quad (3.17)$$

$$\tau^2 \mid \theta, \mathbf{x} \sim GI \left(\frac{n}{2}, \frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2} \right). \quad (3.18)$$

De posse dos resultados apresentados na proposição anterior, podemos implementar computacionalmente o método de amostragem de Gibbs, que é um caso particular dos métodos de Monte Carlo obtidos via Cadeias de Markov, para obter estimativas *a posteriori* de estimadores de interesse, tais como a média ou a variância *a posteriori*. Maiores informações a respeito destas metodologias podem ser encontradas em Paulino et al. (2003).

Na proposição a seguir, apresentamos expressões para a distribuição *a posteriori* marginal do parâmetro de posição θ . Mostramos que tanto a média, quanto a variância da distribuição *a posteriori* marginal do parâmetro de escala τ^2 , são funções da distribuição *a posteriori* marginal de θ .

Proposição 3.2.3 *Seja $\mathbf{X} = (X_1, X_2, \dots, X_n)$ uma amostra aleatória simples de tamanho n da variável $X \sim N(\theta, \tau^2)$ e considere que a distribuição a priori para o parâmetro de posição seja $\theta \sim SN(\mu, \sigma^2, \lambda)$ independente da distribuição a priori de τ^2 , se $h(\tau^2) \propto \frac{1}{\tau^2}$ então*

$$(i) \ f(\theta \mid \mathbf{x}) = \frac{1}{C} \phi \left(\frac{\theta - \mu}{\sigma} \right) \Phi \left(\lambda \frac{\theta - \mu}{\sigma} \right) \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2} \right]^{-\frac{n}{2}}, \text{ em que}$$

$$C = \int_{-\infty}^{\infty} \phi \left(\frac{\theta - \mu}{\sigma} \right) \Phi \left(\lambda \frac{\theta - \mu}{\sigma} \right) \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2} \right]^{-\frac{n}{2}} d\theta.$$

$$(ii) \ E[\tau^2 \mid \mathbf{x}] = \frac{n}{n-2} (S^2 + E[(\bar{x} - \theta)^2 \mid \mathbf{x}]).$$

$$(iii) \ Var[\tau^2 \mid \mathbf{x}] = \frac{2n^2}{(n-4)(n-2)^2} E[(S^2 + (\bar{x} - \theta)^2)^2 \mid \mathbf{x}] + \frac{n^2}{(n-2)^2} Var[(\bar{x} - \theta)^2 \mid \mathbf{x}].$$

A seguir apresentamos versões das Proposições 3.2.2, 3.2.3, ao se considerar $\tau^2 \sim GI(a, b)$.

Priori 2:

Proposição 3.2.4 *Seja $\mathbf{X} = (X_1, X_2, \dots, X_n)$ uma amostra aleatória simples de tamanho n da variável $X \sim N(\theta, \tau^2)$ e considere que a distribuição a priori para o parâmetro de posição seja $SN(\mu, \sigma^2, \lambda)$. Assumindo independência a priori entre os parâmetros θ e τ^2 , se $\tau^2 \sim GI(a, b)$ então*

$$\theta \mid \tau^2, \mathbf{x} \sim GSN_{1,1} \left(\frac{n\bar{x}\sigma^2 + \mu\tau^2}{n\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2}, \lambda, \lambda\mu, \sigma^2 \right) \quad (3.19)$$

$$e$$

$$\tau^2 \mid \theta, \mathbf{x} \sim GI \left(\frac{n}{2} + a, \frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2} + b \right). \quad (3.20)$$

Proposição 3.2.5 *Seja $\mathbf{X} = (X_1, X_2, \dots, X_n)$ uma amostra aleatória simples de tamanho n da variável $X \sim N(\theta, \tau^2)$ e considere que a distribuição a priori para o parâmetro de posição seja $\theta \sim SN(\mu, \sigma^2, \lambda)$ independente da distribuição a priori de τ^2 , se $\tau^2 \sim GI(a, b)$ então*

$$(i) \ f(\theta \mid \mathbf{x}) = \frac{1}{C} \phi \left(\frac{\theta - \mu}{\sigma} \right) \Phi \left(\lambda \frac{\theta - \mu}{\sigma} \right) \Gamma \left(\frac{n}{2} + a \right) \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2} + b \right]^{-\left(\frac{n}{2} + a\right)}, \text{ em que}$$

$$C = \int_{-\infty}^{\infty} \phi \left(\frac{\theta - \mu}{\sigma} \right) \Phi \left(\lambda \frac{\theta - \mu}{\sigma} \right) \Gamma \left(\frac{n}{2} \right) \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2} + b \right]^{-\left(\frac{n}{2} + a\right)} d\theta.$$

$$(ii) \ E[\tau^2 \mid \mathbf{x}] = \frac{1}{n+2a-2} (nS^2 + 2b + E[n(\bar{x} - \theta)^2 \mid \mathbf{x}]).$$

$$(iii) \ Var[\tau^2 \mid \mathbf{x}] = \frac{2}{(n+2a-4)(n+2a-2)^2} E[(nS^2 + 2b + n(\bar{x} - \theta)^2)^2 \mid \mathbf{x}] + \frac{n^2}{(n+2a-2)^2} \times \\ \times Var[(\bar{x} - \theta)^2 \mid \mathbf{x}].$$

A média e a variância das distribuições marginais *a posteriori* serão obtidas através de algoritmos numéricos, tais como o método de Monte Carlo ordinário ou integração numérica via método de quadratura gaussiana do tipo Hermite. De fato, estes métodos foram aplicados para calcular apenas esperanças de termos dependentes do parâmetro de posição *a posteriori*, uma vez que utilizamos os resultados (ii) e (iii) das Proposições 3.2.3 e 3.2.5 para calcular $E[\tau^2 \mid \mathbf{x}]$ e $Var[\tau^2 \mid \mathbf{x}]$.

Procuramos comparar os resultados apresentados por ambas as metodologias em algumas situações específicas e notamos que o método de quadratura gaussiana se mostrou mais eficiente, uma vez que o tempo de trabalho computacional foi muito menor do que o apresentado pelo método de Monte Carlo ordinário. Sendo assim,

neste trabalho, utilizaremos como metodologia para a obtenção das estimativas *a posteriori* o método de quadratura do tipo Gauss-Hermite, cujo os detalhes da técnica podem ser encontrados em Abramowitz & Stegun (1972) e Peixoto (2008).

De posse dos resultados de média e variância *a posteriori* e fazendo uso do método numérico anteriormente descrito para a obtenção das estimativas destas medidas, vamos considerar as distâncias d_M e d_V a fim de que possamos dar início ao processo de avaliação de sensibilidade bayesiana com respeito a escolha de distribuições *a priori* em Γ_M .

Com respeito às distâncias, adotaremos as seguintes notações:

$$\begin{aligned} d_{M_\theta} &= | E_\lambda[\theta | \mathbf{x}] - E_0[\theta | \mathbf{x}] | \quad \text{e} \quad d_{V_\theta} = | \text{Var}_\lambda[\theta | \mathbf{x}] - \text{Var}_0[\theta | \mathbf{x}] |, \\ d_{M_{\tau^2}} &= | E_\lambda[\tau^2 | \mathbf{x}] - E_0[\tau^2 | \mathbf{x}] | \quad \text{e} \quad d_{V_{\tau^2}} = | \text{Var}_\lambda[\tau^2 | \mathbf{x}] - \text{Var}_0[\tau^2 | \mathbf{x}] |. \end{aligned}$$

As nossas análises de robustez serão baseadas nas distâncias entre $\sup_\lambda d_M$ e $\inf_\lambda d_M$ e entre $\sup_\lambda d_V$ e $\inf_\lambda d_V$, tanto para $\theta | x$, quanto para $\tau^2 | x$.

Estudo similar será considerado para d_{V_θ} , $d_{M_{\tau^2}}$ e $d_{V_{\tau^2}}$.

3.2.1 Estudo de sensibilidade.

Assim como no estudo de sensibilidade apresentado na Seção 3.1.1, vamos considerar para especificação *a priori* do parâmetro de posição a classe de contaminação multiplicativa com a distribuição *a priori* base $N(0, 1)$ e $w(\theta) = 2\Phi(\lambda\theta)$. Para a distribuição *a priori* do parâmetro de escala τ^2 , consideraremos as duas distribuições *a priori* especificadas anteriormente.

Vale ressaltar que, para o caso em que a especificação *a priori* do parâmetro de escala é dada por uma distribuição gama inversa, os parâmetros $a = 1/10000$ e $b = 1/10000$ foram escolhidos de modo que a distribuição *a priori* para τ^2 fosse uma distribuição *a priori* próxima das não informativas, porém própria (distribuição *a priori* vaga).

Neste exemplo, estamos interessados em estudar qual é o comportamento do extremos das distâncias d_M e d_V para os parâmetros de posição e escala, segundo a classe de contaminação *a priori* multiplicativa, a medida em que \bar{x} varia. Aqui, d_M corresponde ao desvio absoluto entre a esperança das distribuições *a posteriori*

marginais obtidas segundo a classe de contaminação *a priori* multiplicativa e a distribuição *a priori* base. Analogamente, a distância d_V é obtida ao se considerar o desvio absoluto entre a variância das distribuições *a posteriori* e variância da distribuição *a priori* base.

Dessa forma, faremos uma análise de sensibilidade considerando os resultados apresentados na Proposição 3.2.3, com $\lambda > 0$ em duas situações: $\bar{x} \geq 0$ e $\bar{x} < 0$. Assim como discutido na Seção 3.1.1, a análise destes casos com $\lambda < 0$ será omitida, pois é similar e simétrica a situação em que $\lambda > 0$.

A média e a variância das distribuições marginais *a posteriori* serão obtidas através da aplicação da metodologia de integração numérica de Gauss-Hermite, com o número de abcissas (m) iguais a 65, nos casos em que $\bar{x} \geq 0$. Outras quantidades de abcissas foram testadas, tais como $m = 40, 80, 100, 150, 250$ e 500 , sendo que os resultados de integração foram muito pouco influenciados para valores de m superiores a 65. Além disso, a medida em que se toma o valor de m cada vez maior, o processo que permite calcular as integrais fica mais demorado e, portanto, torna-se mais viável considerar $m = 65$. Para o tamanho da amostra, consideramos $n = 20, 50, 100$ e 200 .

Note que, pelos resultados apresentados na Proposição 3.2.3, tanto a distribuição *a posteriori* marginal para θ , quanto a média e a variância dos parâmetros θ e τ^2 *a posteriori* dependem da variância amostral, a qual definimos por $S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$. Para este estudo de simulação, vamos considerar $S^2 = 1, 3, 8, 10, 20$ e 100 .

Além disso, em ambas as situações dadas a seguir, foi considerado $\lambda \in [0, 25]$. Dessa forma, para que fosse possível obter $\sup_{\lambda} d_M$ e $\sup_{\lambda} d_V$ através da metodologia de Gauss-Hermite, admitimos para cada \bar{x} fixo, $\lambda = 0.1 \times i$, $i = 0, \dots, 250$ e, então,

$$\sup_{\lambda} d_M = \max\{d_M(\bar{x}, \lambda_0), d_M(\bar{x}, \lambda_1), \dots, d_M(\bar{x}, \lambda_{250})\}. \quad (3.21)$$

Para os gráficos das situações a seguir, definimos uma grade de valores de \bar{x} e analisamos segundo o processo descrito anteriormente.

Situação I: $\bar{x} < 0$ e $\lambda > 0$.

Nesta situação, vamos considerar \bar{x} variando de -4 a 0 , através da seguinte construção: $\bar{x} = -0.01 \times j$, $j = 1, \dots, 400$.

Situação II: $\bar{x} > 0$ e $\lambda > 0$.

Vamos considerar \bar{x} variando de 0 a 4, da seguinte forma: $\bar{x} = 0.01 \times j$, $j = 0, \dots, 400$.

Vejam agora quais foram os resultados gerados ao se considerar o modelo normal e o contexto de simulação anteriormente apresentado, quando temos a classe de contaminação multiplicativa *a priori* para o parâmetro de posição θ e as especificações *a priori* para o parâmetro de escala τ^2 dadas por $h(\tau^2) \propto \frac{1}{\tau^2}$ e $\tau^2 \sim GI(1/10000, 1/10000)$.

Embora o estudo tenha sido realizado com as duas distribuições *a priori* anteriormente especificadas para τ^2 , os resultados e conclusões foram muito similares. Por isso, vamos reportar aqui somente os resultados em que $h(\tau^2) \propto \frac{1}{\tau^2}$.

Análise para $\bar{x} < 0$ e $\lambda > 0$.

Na Figura 3.7, apresentamos os extremos das distâncias d_M identificados por $\sup_{\lambda} d_M$, $\inf_{\lambda} d_M$, $\sup_{\lambda} d_V$ e $\inf_{\lambda} d_V$ tanto para a média *a posteriori*, quanto para a variância *a posteriori*, fixados $S^2 = 10$ e $n = 20, 50, 100$ e 200 . Para os demais valores de S^2 , apresentamos os gráficos no Apêndice B.

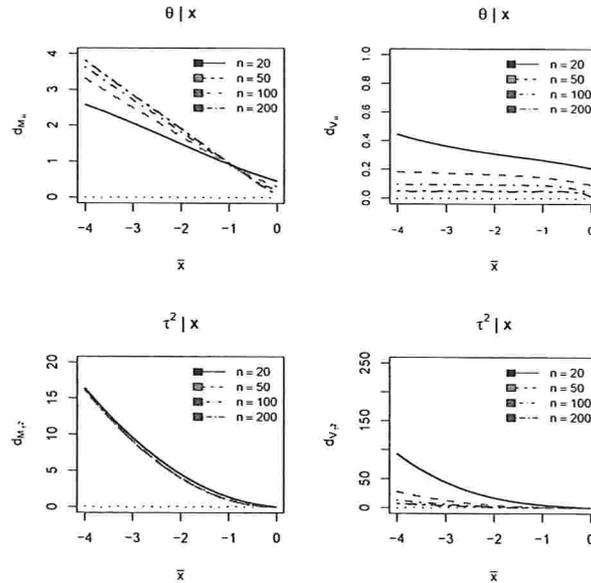


Figura 3.7: Supremo das distâncias d_{M_θ} , d_{V_θ} , $d_{M_{\tau^2}}$ e $d_{V_{\tau^2}}$, quando $S^2 = 10$ e $n = 20, 50, 100$ e 200 .

Quando comparamos os gráficos das Figuras B.1, B.2, B.3, B.4, B.5 e B.6, apresentados no Apêndice B, chegamos as seguintes conclusões:

(i) Sobre $d_{M_\theta} = | \mathbf{E}_\lambda[\theta | \mathbf{x}] - \mathbf{E}_0[\theta | \mathbf{x}] |$

Para diferentes valores de n e S^2 , o máximo das distâncias d_M apresenta-se graficamente, quando em função de \bar{x} , quase como uma reta.

Note que, para $S^2 = 1$, o máximo das distâncias d_M , em função de \bar{x} , comporta-se de maneira similar a função $f(x) = -x$. A medida em que S^2 cresce, a ‘reta’ que representa graficamente d_M tem coeficiente angular cada vez maior, principalmente para tamanhos de amostra pequenos ($n = 20$ e 50).

De maneira geral, a medida em que \bar{x} decresce, a distância entre $\sup_\lambda d_{M_\theta}$ e $\inf_\lambda d_{M_\theta}$ torna-se cada vez maior, implicando em um maior distanciamento entre os extremos das médias *a posteriori* geradas pela classe de contaminação multiplicativa *a priori*. Ou seja, a média *a posteriori* é bastante sensível a mudanças de especificações *a priori* em Γ_M , evidenciando a falta de robustez da média *a posteriori* com relação a Γ_M .

(ii) Sobre $d_{V_\theta} = | \text{Var}_\lambda[\theta | \mathbf{x}] - \text{Var}_0[\theta | \mathbf{x}] |$

Em relação a esta medidas, para quaisquer valores de S^2 , fixado \bar{x} , quanto maior o tamanho da amostra, menor é o valor atribuído ao $\sup_\lambda d_{V_\theta}$.

Para quaisquer valores de n , a medida em que S^2 cresce, maior é o intervalo de variação das medidas d_{V_θ} , exceto para o caso em $S^2 = 1$ e $\bar{x} < -2$.

Note que há pouca variação das estimativas *a posteriori* para a variância de $\theta | \mathbf{x}$ quando S^2 possui valores moderados (≤ 10) e tamanho amostral acima de 50. Podemos dizer que, nestas condições, ocorre a robustez da variância *a posteriori* com respeito a Γ_M . Nas demais situações, esta conclusão não procede.

(iii) Sobre $d_{M_{\tau^2}} = | \mathbf{E}_\lambda[\tau^2 | \mathbf{x}] - \mathbf{E}_0[\tau^2 | \mathbf{x}] |$

Note que a curva que representa $\sup d_{M_{\tau^2}}$ tem formato exponencial do tipo $f(x) = e^{-ax}$, para um $a > 0$. Para tamanhos de amostra grande ($n = 100$ e 200), as curvas de $\sup_\lambda d_{M_{\tau^2}}$ praticamente se sobrepõe, independente dos valores atribuídos a S^2 .

De maneira geral, notamos que a medida em que \bar{x} decresce, $\sup_\lambda d_{M_{\tau^2}}$ torna-se cada vez maior, mostrando que existe uma maior variação das médias *a posteriori*

para τ^2 , obtidas segundo a classe de contaminação a *a priori* Γ_M . Isto implica em dizer que a média *a posteriori* para τ^2 é muito influenciada pela escolha da distribuição *a priori* em Γ_M e, portanto, a robustez com respeito a Γ_M não ocorre.

(iv) Sobre $d_{V_{\tau^2}} = | \text{Var}_\lambda[\tau^2 | \mathbf{x}] - \text{Var}_0[\tau^2 | \mathbf{x}] |$

Fixado \bar{x} , quanto maior o tamanho da amostra, menor é a distância entre $\sup_\lambda d_{V_{\tau^2}}$ e $\inf_\lambda d_{V_{\tau^2}}$, para quaisquer valores de S^2 .

Para n fixado, quanto maior o valor de S^2 considerado, maior é o $\sup_\lambda d_{V_{\tau^2}}$. Mesmo quando o tamanho da amostra é relativamente grande, como no caso de $n = 200$, nota-se a falta de robustez da variância *a posteriori* de τ^2 , com respeito a diferentes especificações da distribuição *a priori* em Γ_M , como pode ser visto na Figura B.7, localizada no Apêndice B.

Neste contexto, de maneira geral, salvo algumas situações específicas, quando o sinal da média amostral é diferente do sinal do parâmetro de assimetria considerado para a classe de contaminação multiplicativa *a priori* Γ_M , concluímos que não ocorre a robustez da média e da variância *a posteriori* marginal dos parâmetros de posição e escala do modelo normal.

Análise para $\bar{x} > 0$ e $\lambda > 0$.

Na Figura 3.8 apresentamos gráficos similares aos da Figura 3.7, para diferentes valores de n e $S^2 = 10$, porém considerando o intervalo de variação de \bar{x} entre zero e quatro. Dessa forma, as curvas dos dois gráficos da primeira linha da Figura 3.8 correspondem, respectivamente, ao supremo das distâncias d_{M_θ} e d_{V_θ} para o parâmetro $\theta | x$. Analogamente, os dois gráficos restantes da Figura 3.8 correspondem, respectivamente, o supremo das distâncias $d_{M_{\tau^2}}$ e $d_{V_{\tau^2}}$ para o parâmetro $\tau^2 | x$.

Os demais gráficos considerando $S^2 = 1, 3, 8, 20$ e 100 para $\bar{x} > 0$ e $\lambda > 0$ são apresentados no Apêndice B, nas Figuras B.8, B.9, B.10, B.12 e B.13, respectivamente.

A fim de simplificar a notação, usaremos o símbolo $\sup d_M$ quando as análises para $\sup_\lambda d_{M_\theta}$ e $\sup_\lambda d_{M_{\tau^2}}$ forem similares. De maneira análoga, associamos as notações $\inf_\lambda d_M$, $\sup_\lambda d_V$ e $\inf_\lambda d_V$.

Como conclusões gerais podemos citar que para quaisquer valores fixados de S^2 , notamos que, quanto maior o tamanho da amostra, menor o intervalo entre $\sup d_M$

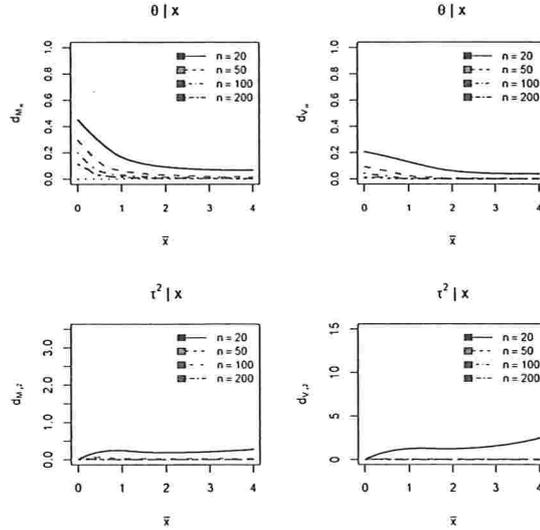


Figura 3.8: Supremo das distâncias d_{M_θ} , d_{V_θ} , $d_{M_{\tau^2}}$ e $d_{V_{\tau^2}}$, quando $S^2 = 10$ e $n = 20, 50, 100$ e 200 .

e $\inf d_M$ e entre $\sup d_V$ e $\inf d_V$, para $0 \leq \bar{x} \leq 4$. Uma vez fixado o tamanho da amostra, a medida em que S^2 cresce, maior é a distância entre $\sup d_M$ e $\inf d_M$ e entre $\sup d_V$ e $\inf d_V$.

Algumas conclusões específicas:

- (i) Sobre $d_{M_\theta} = |E_\lambda[\theta | x] - E_0[\theta | x]|$ e $d_{V_\theta} = |\text{Var}_\lambda[\theta | x] - \text{Var}_0[\theta | x]|$

Ambas as curvas de $\sup_\lambda d_{M_\theta}$ e $\sup_\lambda d_{V_\theta}$ tem um formato exponencial.

Note que, a medida em que \bar{x} cresce, a medida $\sup d_{V_\theta}$ decresce mais rapidamente para zero do que a medida $\sup d_{M_\theta}$.

Se a variância amostral S^2 possui valores moderados (≤ 10), os intervalos entre $\sup_\lambda d_{M_\theta}$ e $\inf_\lambda d_{M_\theta}$ e entre $\sup_\lambda d_{V_\theta}$ e $\inf_\lambda d_{V_\theta}$ são bem pequenos, principalmente para $n \geq 50$. Para $\bar{x} \geq 1$, estes intervalos são praticamente nulos. Concluímos então que, para valores de $S^2 \leq 10$, as estimativas da média e da variância *a posteriori* marginal de θ são robustas com respeito a diferentes especificações de distribuições *a priori* na classe de contaminação multiplicativa *a priori* Γ_M .

Mesmo quando $S^2 > 10$, podemos falar em robustez das estimativas da média e variância *a posteriori* marginal de θ com respeito a Γ_M , se o tamanho da amostra for grande ($n \geq 200$).

Conclusões similares as apresentadas no parágrafo anterior podem ser atribuídas para o caso em que $S^2 = 20$, no entanto, isto ocorre para $n \geq 100$.

Se a variância amostral é grande, representado aqui pelo valor de $S^2 = 100$, concluímos que a robustez da média e da variância a posteriori marginal para θ só ocorre para $n \geq 200$.

(ii) Sobre $d_{M_{\tau^2}} = |E_{\lambda}[\tau^2 | \mathbf{x}] - E_0[\tau^2 | \mathbf{x}]|$ e $d_{V_{\tau^2}} = |\text{Var}_{\lambda}[\tau^2 | \mathbf{x}] - \text{Var}_0[\tau^2 | \mathbf{x}]|$

Aqui, as medidas $\sup_{\lambda} d_{M_{\tau^2}}$ e $\sup_{\lambda} d_{V_{\tau^2}}$ são muito influenciadas pela variabilidade amostral.

Note que, para cada \bar{x} fixo, o valor de $\sup_{\lambda} d_{M_{\tau^2}}$ é sempre inferior ao valor de $\sup_{\lambda} d_{V_{\tau^2}}$.

Para $n \geq 100$, o intervalo entre $\sup_{\lambda} d_{V_{\tau^2}}$ e $\inf_{\lambda} d_{V_{\tau^2}}$ tem pequena magnitude, principalmente quando a variância amostral é inferior a 20. Tal fato pode ser comprovado pelas Figuras B.16 e B.17 apresentadas no Apêndice B. Concluímos então que ocorre a robustez das estimativas da média e da variância *a posteriori* marginal de τ^2 com a respeito a classe de contaminação multiplicativa *a priori* Γ_M , quando $S^2 \leq 20$ e $n \geq 100$.

Como pode ser observado nas Figuras B.14 e B.15, conclusão similar ao parágrafo anterior pode ser tomada quando analisamos o intervalo entre $\sup_{\lambda} d_{M_{\tau^2}}$ e $\inf_{\lambda} d_{M_{\tau^2}}$, para $S^2 \leq 20$.

De maneira geral, para tamanhos de amostra suficientemente grandes ($n \geq 100$) e variabilidade amostral moderada ($S^2 \leq 20$), quando o sinal da média amostral é igual ao sinal do parâmetro de assimetria considerado para a classe de contaminação multiplicativa *a priori* Γ_M , concluímos que ocorre a robustez da média e da variância *a posteriori* marginal dos parâmetros de posição e escala do modelo normal.

Capítulo 4

Análise de sensibilidade em modelos de regressão linear normal

4.1 Introdução

Uma das principais idéias que se associam a robustez em modelos de regressão tem relação com a utilização e o desenvolvimento de estratégias que permitam identificar a presença de observações atípicas e investigar o efeito destas na adequação do ajuste de um modelo (violação ou não de suas premissas) e nas estimativas dos parâmetros de interesse.

Uma boa discussão sobre metodologias de regressão robusta sob o enfoque clássico são apresentadas em Rousseeuw & Leroy (2003) e em Hampel, Ronchetti, Rousseeuw & Stahel (2005). Do ponto de vista bayesiano, Arellano-Valle, Galea-Rojas & Zuazola (2000) fazem uma revisão de medidas de influência mais comuns na literatura para modelos lineares normais com distribuição *a priori* conjugada, consideram modelos de regressão elípticos com distribuições *a priori* não-informativas para os parâmetros do modelo e investigam a influência de um determinado subconjunto dos dados na distribuição *a posteriori* dos parâmetros de posição e de escala. Além disso, os autores discutem a influência de diferentes especificações do modelo dentro da família de distribuições elípticas, ao se comparar o risco de Bayes e medidas de divergência entre as distribuições *a posteriori*.

Sempre que se considera a abordagem bayesiana, uma questão que se faz pertinente é saber o quão influenciável é a distribuição *a posteriori* em relação à escolha da distribuição *a priori*.

Como mencionado no Capítulo 1 deste trabalho, o estudo de robustez adotado

aqui é o de robustez global. Neste contexto, modelamos a incerteza sobre a distribuição *a priori* através do uso de classes de contaminação, a qual denotamos por Γ_M . Mais especificamente, utilizamos a classe de contaminação multiplicativa, tal como descrita em (3.1). Com respeito a análise, diremos que esta é robusta quando as inferências *a posteriori* pouco diferirem em relação às diferentes distribuições *a priori* escolhidas em Γ_M .

No contexto de modelos de regressão linear normal, Chib & Tiwari (1991) consideraram a classe de distribuições *a priori* ϵ -contaminada gerada à partir da mistura da distribuição normal-gama inversa (base) com a distribuição *a priori* de Jeffreys corrigida por uma constante. Neste artigo, os autores mostram que, em geral, a distribuição *a posteriori* marginal do parâmetro regressor é multimodal. Além disso, eles discutem questões de robustez bayesiana local e global. Para avaliar a robustez global, os estimadores *a posteriori* obtidos segundo *a priori* base e a classe de contaminação são comparados através de uma medida que utiliza a função de perda esperada *a posteriori*. Tal metodologia foi proposta por Berger (1984).

Problema semelhante ao descrito anteriormente é abordado por Chaturvedi (1996), utilizando outra parametrização da classe de distribuições *a priori* ϵ -contaminada e diferente técnica para a análise de robustez. Naquele trabalho, a classe corresponde a uma mistura de distribuições *a priori* do tipo g (ver Zellner 1986) e a metodologia empregada é chamada de Tipo II de máxima verossimilhança (MV-II) proposta por Good (1965).

A metodologia MV-II não avalia o intervalo de variação de medidas *a posteriori* geradas à partir da distribuição *a priori* em Γ_M . Segundo Chaturvedi (1996), o procedimento MV-II '*choosing a prior from the ϵ -contamination in a data dependent fashion*'. Neste método, a distribuição *a priori* escolhida em Γ_M é selecionada pela maximização da função de densidade preditiva e toda a análise *a posteriori* subsequente será feita com respeito a esta distribuição *a priori* selecionada. Qualquer medida *a posteriori* obtida segundo este processo aparece sempre identificado pelo termo MV-II, por exemplo, fala-se em média *a posteriori* MV-II.

Berger & Berliner (1986) reconhecem que a técnica ML-II não é infalível e pode produzir más respostas, particularmente quando a classe de distribuições *a priori* Γ_M contém distribuições pouco razoáveis. No entanto, os autores minimizam esta crítica afirmando que supostamente todas as distribuições *a priori* contidas em Γ_M devem

ser representações razoáveis de crenças *a priori* e que, dessa forma, a distribuição *a priori* escolhida em Γ_M por ML-II seria a mais plausível, à luz das opiniões *a priori* e dos dados.

Neste capítulo, também vamos estudar questões relativas à robustez bayesiana em modelos de regressão linear normal. No entanto, a abordagem utilizada fará uso da classe de contaminação *a priori* multiplicativa definida em (3.1), parametrizada de tal forma que a distribuição que a representa é a distribuição normal-assimétrica de Azzalini (1985) e a distribuição *a priori* base é a distribuição normal. Além disso, como já mencionado, estamos utilizando a abordagem de robustez bayesiana global como apresentada em O'Hagan (1994), ou seja, nossa análise será baseada na distância entre as medidas *a posteriori* geradas pela classe de contaminação multiplicativa e pela distribuição *a priori* base.

Podemos representar o modelo de regressão linear geral, com o termo dos erros normalmente distribuído, da seguinte forma:

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \dots + \beta_{p-1} X_{i,p-1} + \epsilon_i, \quad \text{para } i = 1, \dots, n, \quad (4.1)$$

onde

- Y_i é o i -ésimo valor da variável resposta ou dependente,
- $\beta_0, \beta_1, \dots, \beta_{p-1}$ são os p parâmetros regressores,
- $X_{i,j}$, para $j = 1, \dots, p - 1$, é a variável independente, explicativa ou covariável que representa constantes conhecidas,
- $\epsilon_i \sim N(0, \tau^2)$, tal que $\epsilon_i \perp \epsilon_j, \forall i \neq j$.

Equivalente a (4.1), podemos ainda utilizar a seguinte representação matricial:

$$Y = X\beta + \epsilon, \quad \text{em que} \quad (4.2)$$

- Y é o vetor de variáveis respostas ou dependentes, de dimensão $n \times 1$,
- β é o vetor de parâmetros regressores, com dimensão $p \times 1$,
- X é uma matriz com $n \times p$ covariáveis,

- ϵ é o vetor de erros aleatórios, de dimensão $n \times 1$, cuja a distribuição estamos considerando ser $N_n(\mathbf{0}, \tau^2 I_n)$, onde os símbolos I_n e $\mathbf{0}$ referem-se, respectivamente, à matriz identidade de ordem n e ao vetor $n \times 1$ de zeros.

Como consequência, obtemos que $\mathbf{Y} \sim N_n(\mathbf{X}\beta, \tau^2 I_n)$, cuja a função de densidade tem a forma

$$f(\mathbf{y}) = \left(\frac{1}{2\pi\tau^2} \right)^{\frac{n}{2}} e^{-\frac{1}{2\tau^2}(\mathbf{y}-\mathbf{X}\beta)^T(\mathbf{y}-\mathbf{X}\beta)}. \quad (4.3)$$

Para que possamos fazer uso da perspectiva bayesiana, distribuições *a priori* para os parâmetros do modelo devem ser especificadas. Duas situações serão consideradas; modelos de regressão linear normal com variância dos termos aleatórios conhecida e desconhecida. Como pressuposto para estes casos, estaremos assumindo que a matriz de covariáveis \mathbf{X} tem posto completo p . As demonstrações dos resultados deste capítulo estão no Apêndice A.

4.2 Modelos de regressão normal com variância dos erros conhecida.

Quando a matriz de covariáveis \mathbf{X} tem posto completo p , a função de verossimilhança, por expansão da forma quadrática nela contida, é dada por:

$$f(\mathbf{y} \mid \beta) = \left(\frac{1}{2\pi\tau^2} \right)^{\frac{n}{2}} e^{-\frac{1}{2\tau^2}[Ks^2 + (\beta - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\beta - \hat{\beta})]}, \quad (4.4)$$

onde $Ks^2 = (\mathbf{y} - \mathbf{X}\hat{\beta})^T(\mathbf{y} - \mathbf{X}\hat{\beta})$ é conhecido como soma dos quadrados residual, $K = n - p$ e $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ é a estimativa de mínimos quadrados (ou de máxima verossimilhança) de β .

Para a distribuição *a priori* do vetor de parâmetros regressores, vamos considerar a classe de contaminação multiplicativa dada por:

$$\Gamma_M = \{f(\beta) = f_0(\beta) \times w(\beta) : \beta \in \mathbb{R}^p\}, \quad (4.5)$$

em que $f_0(\cdot)$ e $w(\cdot)$ são como definidos em (3.1).

Segundo Arellano-Valle & Azzalini (2006), um vetor aleatório d -dimensional Y tem distribuição normal-assimétrica multivariada quando a sua função de densidade é dada por:

$$2 \phi_d(\mathbf{y} - \boldsymbol{\xi}, \Omega) \Phi[\boldsymbol{\lambda}^T \boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi})], \quad (4.6)$$

onde $\boldsymbol{\xi}$ ($\boldsymbol{\xi} \in \mathbb{R}^d$) é o parâmetro de posição, Ω é uma matriz de covariância (positiva definida), $\boldsymbol{\lambda}$ ($\boldsymbol{\lambda} \in \mathbb{R}^d$) é o parâmetro de forma e $\boldsymbol{\omega}$ é a matriz diagonal formada pelos desvios-padrões de Ω . Como notação, adotamos $Y \sim SN_d(\boldsymbol{\xi}, \Omega, \boldsymbol{\lambda})$. Note que $\Omega = \boldsymbol{\omega} \bar{\Omega} \boldsymbol{\omega}$, onde $\bar{\Omega}$ é uma matriz de correlação.

Dessa forma, para $\boldsymbol{\mu} \in \mathbb{R}^p$ e Ω com posto completo, a classe de contaminação multiplicativa *a priori* será parametrizada de tal forma que:

$$f_0(\boldsymbol{\beta}) = \phi_p(\boldsymbol{\beta}; \boldsymbol{\mu}, \Omega) \quad (4.7)$$

e

$$w(\boldsymbol{\beta}) = 2 \Phi[\boldsymbol{\lambda}^T \boldsymbol{\omega}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu})], \quad (4.8)$$

ou seja, $\Gamma_M = \{f(\boldsymbol{\beta}) = 2 \phi_p(\boldsymbol{\beta}; \boldsymbol{\mu}, \Omega) \Phi[\boldsymbol{\lambda}^T \boldsymbol{\omega}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu})] : \boldsymbol{\lambda} \in \mathbb{R}^p\}$. Quando $\boldsymbol{\lambda} = \mathbf{0}$, obtemos a distribuição *a priori* base representada pela distribuição normal multivariada.

A distribuição normal-assimétrica multivariada apresentada em (4.6) é um caso particular da distribuição normal-assimétrica multivariada geral (ver (2.6)), pois a distribuição $SN_p(\boldsymbol{\mu}, \Omega, \boldsymbol{\lambda})$ é equivalente a $GSN_{p,1}(\boldsymbol{\mu}, \Omega, \boldsymbol{\lambda}^T \boldsymbol{\omega}^{-1}, \boldsymbol{\lambda}^T \boldsymbol{\omega}^{-1} \boldsymbol{\mu}, 1)$. Neste trabalho, usamos ambas as notações conforme a nossa conveniência.

Na proposição a seguir derivamos a distribuição *a posteriori* dos parâmetros regressores e mostramos que tanto a classe de distribuições *a priori* quanto a distribuição *a posteriori* para o parâmetro $\boldsymbol{\beta}$ pertencem à família normal-assimétrica multivariada geral.

Proposição 4.2.1 (Fechamento sob conjugação) *Considere o modelo de regressão dado por (4.2) e τ^2 conhecido. Se $\boldsymbol{\beta}$ tem distribuição $GSN_{p,1}(\boldsymbol{\mu}, \Omega, \boldsymbol{\lambda}^T \boldsymbol{\omega}^{-1}, \boldsymbol{\lambda}^T \boldsymbol{\omega}^{-1} \boldsymbol{\mu}, 1)$, então a distribuição *a posteriori*, a média e a variância *a posteriori* para o vetor de parâmetros regressores são dadas, respectivamente, por*

$$(i) \quad \beta | \mathbf{Y} \sim GSN_{p,1} \left(\zeta, \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1}, \lambda^T \omega^{-1}, \lambda^T \omega^{-1} \boldsymbol{\mu}, 1 \right), \quad (4.9)$$

$$(ii) \quad E_{\lambda}[\beta | \mathbf{Y}] = \zeta + \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} (\lambda^T \omega^{-1})^T \eta \quad e \quad (4.10)$$

$$(iii) \quad Var_{\lambda}[\beta | \mathbf{Y}] = \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} - \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} (\lambda^T \omega^{-1})^T \times \\ \times \left\{ \frac{\mathcal{F} - \mathcal{G}}{\mathcal{H}} + \eta \right\} \lambda^T \omega^{-1} \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} \eta, \quad (4.11)$$

onde

$$\zeta = \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} \hat{\boldsymbol{\beta}} + \Omega^{-1} \boldsymbol{\mu} \right), \quad (4.12)$$

$$\eta = \frac{\phi(\mathcal{F}; \mathcal{G}, \mathcal{H})}{\Phi(\mathcal{F}; \mathcal{G}, \mathcal{H})}, \quad (4.13)$$

$$\mathcal{F} = \lambda^T \omega^{-1} \zeta, \quad (4.14)$$

$$\mathcal{G} = \lambda^T \omega^{-1} \boldsymbol{\mu}, \quad (4.15)$$

$$\mathcal{H} = 1 + \lambda^T \omega^{-1} \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} (\lambda^T \omega^{-1})^T. \quad (4.16)$$

Note que, quando $\lambda = 0$, resulta de (4.10) e (4.11) que

$$E_0[\beta | \mathbf{Y}] = \zeta \quad e \quad Var_0[\beta | \mathbf{Y}] = \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1}, \quad (4.17)$$

em que $\beta | \mathbf{Y} \sim N \left(\zeta, \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} \right)$, recuperando-se assim o resultado de conjugação sob normalidade.

Como discutido anteriormente, estudos de robustez conduzidos neste trabalho avaliam a distância entre medidas *a posteriori* geradas pela classe de contaminação multiplicativa *a priori* e a distribuição *a priori* base. Na proposição a seguir, apresentamos expressões para estas medidas no contexto de análise de regressão linear, com a variância dos erros conhecida.

Proposição 4.2.2 *Considere $E_\lambda[\beta | \mathbf{Y}]$, $Var_\lambda[\beta | \mathbf{Y}]$, $\hat{\beta}_0 = E_0[\beta | \mathbf{Y}]$ e $\hat{\zeta}_0^2 = Var_0[\beta | \mathbf{Y}]$ como apresentados em (4.10), (4.11) e (4.17), respectivamente. Então,*

(i)

$$\begin{aligned} d_M &= \left\| E_\lambda[\beta | \mathbf{Y}] - \hat{\beta}_0 \right\| \\ &= \eta \left\| \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} (\lambda^T \omega^{-1})^T \right\| \quad e \end{aligned} \quad (4.18)$$

(ii)

$$\begin{aligned} d_V &= \left\| Var_\lambda[\beta | \mathbf{Y}] - \hat{\zeta}_0^2 \right\| \\ &= \eta \left\{ \frac{\mathcal{F} - \mathcal{G}}{\mathcal{H}} + \eta \right\} \left\| \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} (\lambda^T \omega^{-1})^T \lambda^T \omega^{-1} \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} \right\|, \end{aligned} \quad (4.19)$$

em que o símbolo $\| \cdot \|$ corresponde a norma euclideana e η , \mathcal{F} , \mathcal{G} e \mathcal{H} são como especificados anteriormente.

Note que, d_M representa a distância entre o estimador de Bayes fornecido pela classe Γ_M e o estimador de Bayes sob a distribuição *a priori* base $f_0(\cdot)$. Analogamente, d_V corresponde a distância entre as variâncias *a posteriori*.

Para valores fixos de μ e Ω , obtemos a classe de contaminação multiplicativa *a priori* normal-assimétrica quando variamos a distribuição $SN_p(\mu, \Omega, \lambda)$ em $\lambda \in \mathbb{R}^p$. Associada a ela e considerando o resultado apresentado na Proposição 4.2.1, a classe de distribuições *a posteriori* para o vetor de parâmetros regressores fica assim determinada:

$$\Gamma_M^* = \left\{ f(\beta | \mathbf{Y}) = \frac{\phi_p \left(\beta; \zeta, \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} \right) \Phi[\lambda^T \omega^{-1}(\beta - \mu)]}{\Phi \left[\lambda^T \omega^{-1} \zeta; \lambda^T \omega^{-1} \mu, 1 + \lambda^T \omega^{-1} \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} (\lambda^T \omega^{-1})^T \right]} : \lambda \in \mathbb{R}^p \right\}, \quad (4.20)$$

No contexto de análise de robustez global, devemos determinar um conjunto de medidas *a posteriori* e estudar o comportamento de medidas limites nesta classe.

Para isso, vamos considerar as distâncias d_M e d_V definidas na Proposição 4.2.2 e analisá-las em função de $\lambda \in \mathbb{R}^p$, nos moldes do que foi apresentado no Capítulo 3. Assim, obtemos o conjunto de medidas *a posteriori* denotados por $d_M(\lambda)$ e $d_V(\lambda)$, com $\lambda \in \mathbb{R}^p$, e suas respectivas medidas limite, dadas por:

$$\sup_{\lambda} d_M(\lambda), \quad \inf_{\lambda} d_M(\lambda), \quad \sup_{\lambda} d_V(\lambda) \quad \text{e} \quad \inf_{\lambda} d_V(\lambda). \quad (4.21)$$

Note que a classe de contaminação *a priori* contém a distribuição *a priori* base e que esta é identificada na classe Γ_M quando $\lambda = 0$. Além disso, $d_M \geq 0$ e $d_V \geq 0$. Somente quando $\lambda = 0$, temos que $d_M = d_V = 0$ e, portanto, $\inf_{\lambda} d_M(\lambda) = \inf_{\lambda} d_V(\lambda) = 0$, quando o vetor λ é identicamente nulo.

Para as expressões do supremo, algoritmos de maximização devem ser utilizados.

4.2.1 Modelo de regressão linear normal com um parâmetro regressor.

Nesta seção, estudamos com maior profundidade o modelo de regressão normal linear simples, sem intercepto, quando a distribuição *a priori* de β pertence a classe de contaminação multiplicativa normal-assimétrica. Nesta situação, conseguimos obter resultados analíticos para as medidas limites das distâncias relacionadas à média e a variância *a posteriori*. Muitos dos resultados que aqui serão apresentados têm demonstrações similares aos apresentados no Capítulo 3 e, portanto, essas demonstrações serão omitidas.

Considere o modelo de regressão normal linear simples, sem intercepto, tal que

$$Y_i = \beta x_i + \epsilon_i, \quad \epsilon_i \sim N(0, \tau^2) \quad i = 1, \dots, n, \quad \epsilon_i \text{ independente de } \epsilon_j, \quad \forall i \neq j. \quad (4.22)$$

Vamos considerar um caso particular da classe de contaminação *a priori* especificada em (4.5), em que $f_0(\beta) = \phi(\beta; \mu, \sigma^2)$ e $w(\beta) = 2 \Phi(\lambda\beta; \lambda\mu, \sigma^2)$. Dessa forma, a classe de contaminação multiplicativa *a priori* para β pode ser representada por $SN(\mu, \sigma^2, \lambda)$, com $\lambda \in \mathbb{R}$, onde λ reflete o grau de contaminação da distribuição *a priori*. Note que, quando $\lambda = 0$, $f(\beta) = f_0(\beta)$.

Um primeiro resultado está relacionado a propriedade de fechamento sob conjugação apresentado na Proposição 4.2.1. Para o modelo considerado em (4.22), temos

que a distribuição *a posteriori* para o parâmetro regressor é

$$GSN_{1,1} \left(\frac{\sigma^2 \sum_{i=1}^n x_i^2 \hat{\beta} + \mu \tau^2}{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}, \frac{\sigma^2 \tau^2}{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}, \lambda, \lambda \mu, \sigma^2 \right),$$

em que $\hat{\beta} = \frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n x_i^2}$ é o estimador de máxima verossimilhança para β no modelo de regressão linear normal simples, sem intercepto.

As expressões para a média e a variância *a posteriori* dadas em (4.10) e (4.11) se reduzem a

$$E_\lambda[\beta | \mathbf{Y}] = \frac{\sigma^2 \sum_{i=1}^n x_i^2 \hat{\beta} + \mu \tau^2}{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2} + h(\lambda) \frac{\sigma \tau^2}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \times \frac{\phi \left[h(\lambda) \frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]}{\Phi \left[h(\lambda) \frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]} \quad (4.23)$$

e

$$Var_\lambda[\beta | \mathbf{Y}] = \frac{\sigma^2 \tau^2}{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2} - \frac{\sigma^2 \tau^4}{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2} \frac{\phi \left[h(\lambda) \frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]}{\Phi \left[h(\lambda) \frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]} \times \left\{ h(\lambda) \frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} + \frac{\phi \left[h(\lambda) \frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]}{\Phi \left[h(\lambda) \frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]} \right\} \times h(\lambda)^2, \quad (4.24)$$

com $h(\lambda) = \frac{\lambda}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2 (1 + \lambda^2)}}$.

Corolário 4.2.1 A variância *a posteriori* apresentada em (4.24) está limitada entre

0 e $\frac{\sigma^2 \tau^2}{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}$.

As versões para as medidas de distância *a posteriori* apresentadas na Proposição 4.2.2, quando consideramos o modelo de regressão linear normal com um parâmetro regressor, são dadas por:

$$d_M = \left| E_\lambda[\beta | \mathbf{Y}] - \hat{\beta}_0 \right| = \frac{\phi \left[h(\lambda) \frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]}{\Phi \left[h(\lambda) \frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]} |h(\lambda)| \frac{\sigma \tau^2}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}}$$

e

$$d_V = \left| Var_\lambda[\beta | \mathbf{Y}] - \hat{\zeta}_0^2 \right| = h(\lambda)^2 \frac{\sigma^2 \tau^4}{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2} \frac{\phi \left[h(\lambda) \frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]}{\Phi \left[h(\lambda) \frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]} \times$$

$$\times \left\{ h(\lambda) \frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} + \frac{\phi \left[h(\lambda) \frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]}{\Phi \left[h(\lambda) \frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]} \right\}, \quad (4.25)$$

em que $h(\lambda) = \frac{\lambda}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2 (1 + \lambda^2)}}$.

Note que, quando $\lambda = 0$, ambas as distâncias são iguais a zero, o que implica em $\inf_\lambda d_M(\lambda) = \inf_\lambda d_V(\lambda) = 0$.

Para a análise de robustez, precisamos estudar o comportamento das distâncias d_M e d_V com respeito a mudanças nos valores do parâmetro de assimetria λ . Notamos que dependendo da região em que λ varia e da relação entre $\hat{\beta}$ e μ , as distâncias d_M e d_V , vistas como função de λ , ora se comportam monotonicamente, ora não. Na Figura 4.1, exemplificamos o comportamento das distâncias d_M e d_V ao se considerar fixos $\sum x_i^2 = 10$, $\mu = 0$, $\sigma^2 = 1$ e $\tau^2 = 1$.

Note que os gráficos da Figura 4.1 são similares aos apresentados na Figura 3.1. O estudo do comportamento destas medidas será dividido em duas situações: (I) $d_M(\lambda)$ e $d_V(\lambda)$ são monótonas e (II) $d_M(\lambda)$ e $d_V(\lambda)$ não são monótonas.

Situação I: $d_M(\lambda)$ e $d_V(\lambda)$ se comportam monotonicamente.

Vamos subdividir os estudos nos seguintes casos:

Caso 1. $\lambda \leq 0$ e $\hat{\beta} > \mu$.

Caso 2. $\lambda \geq 0$ e $\hat{\beta} < \mu$.

Caso 3. $\hat{\beta} = \mu$.

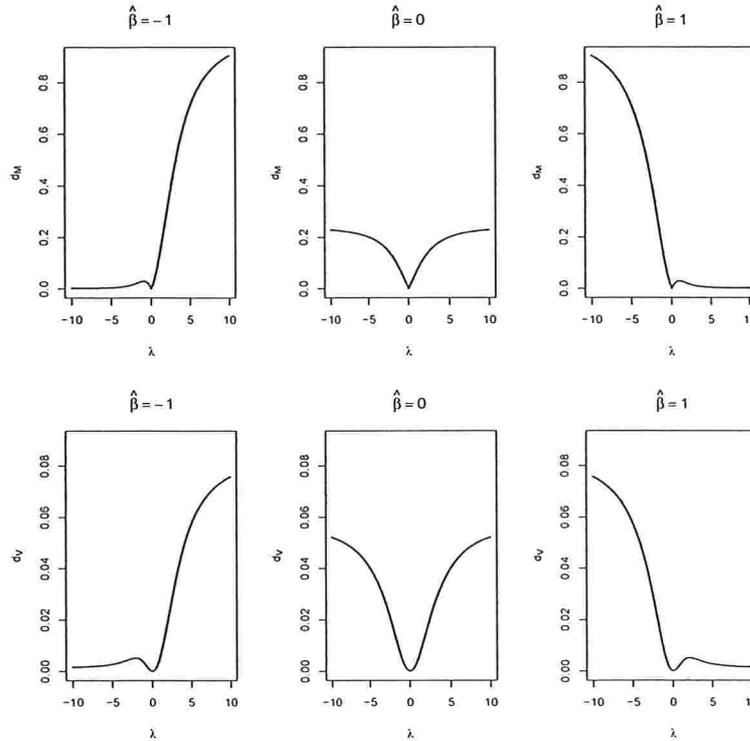


Figura 4.1: Gráficos de d_M e d_V para $\sum x_i^2 = 10$, $\mu = 0$, $\sigma^2 = 1$ e $\tau^2 = 1$.

Em cada caso, apresentamos na proposição a seguir resultados de monotonicidade para as distâncias d_M e d_V , fixado uma amostra \mathbf{x} .

Proposição 4.2.3 *As distâncias $d_M(\lambda)$ e $d_V(\lambda)$ são contínuas e*

- (i) *monotônicas estritamente decrescentes, quando $\lambda \leq 0$ e $\hat{\beta} > \mu$,*
- (ii) *monotônicas estritamente crescentes, quando $\lambda \geq 0$ e $\hat{\beta} < \mu$, e*
- (iii) *simétricas em relação a $\lambda = 0$. Quando $\lambda \leq 0$ ($\lambda \geq 0$) e $\hat{\beta} = \mu$, $d_M(\lambda)$ e $d_V(\lambda)$ são monotônicas estritamente decrescentes (crescentes).*

Como consequência direta da Proposição 4.2.3, conseguimos determinar qual valor de λ gera o supremo das distâncias *a posteriori*. Por exemplo, pela monotonicidade das funções d_M e d_V quando $\lambda \leq 0$ e $\hat{\beta} > \mu$, temos que $\sup_{\lambda} d_M$ e $\sup_{\lambda} d_V$ são obtidos quando λ tende a $-\infty$. Em (ii), analogamente, $\sup_{\lambda} d_M$ e $\sup_{\lambda} d_V$ são obtidos quando λ tende a ∞ . Em (iii), pela simetria das funções $d_M(\lambda)$ e $d_V(\lambda)$ com

relação ao eixo $\lambda = 0$, temos que tanto $\lambda \rightarrow -\infty$, quanto $\lambda \rightarrow \infty$, nos fornecem $\sup_{\lambda} d_M$ e $\sup_{\lambda} d_V$.

Dessa forma, através da Proposição 4.2.3, conseguimos obter expressões fechadas para o supremo das distâncias *a posteriori*. Apresentamos estes resultados na seguinte proposição:

Proposição 4.2.4

(i) Quando $\lambda \leq 0$ e $\hat{\beta} > \mu$, temos que

$$\sup_{\lambda} d_M(\lambda) = \frac{\phi \left[-\frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\tau \sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]}{\Phi \left[-\frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\tau \sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]} \frac{\sigma \tau}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \quad (4.26)$$

e

$$\begin{aligned} \sup_{\lambda} d_V(\lambda) &= \frac{\sigma^2 \tau^2}{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2} \frac{\phi \left[-\frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\tau \sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]}{\Phi \left[-\frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\tau \sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]} \times \\ &\times \left\{ -\frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\tau \sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} + \frac{\phi \left[-\frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\tau \sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]}{\Phi \left[-\frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\tau \sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]} \right\} \quad (4.27) \end{aligned}$$

(ii) Quando $\lambda \geq 0$ e $\hat{\beta} < \mu$, temos que

$$\sup_{\lambda} d_M(\lambda) = \frac{\phi \left[\frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\tau \sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]}{\Phi \left[\frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\tau \sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]} \frac{\sigma \tau}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \quad (4.28)$$

e

$$\begin{aligned} \sup_{\lambda} d_V(\lambda) &= \frac{\sigma^2 \tau^2}{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2} \frac{\phi \left[\frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\tau \sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]}{\Phi \left[\frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\tau \sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]} \times \\ &\times \left\{ \frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\tau \sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} + \frac{\phi \left[\frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\tau \sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]}{\Phi \left[\frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\tau \sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]} \right\} \quad (4.29) \end{aligned}$$

- (iii) Quando $\hat{\beta} = \mu$, as expressões para $\sup_{\lambda} d_M(\lambda)$ em (4.26) e (4.28) são iguais e, portanto, qualquer uma destas expressões pode ser usada. Isto ocorre devido a simetria da função $d_M(\lambda)$ em relação ao eixo $\lambda = 0$. Conclusão análoga pode ser tomada com respeito à distância d_V .

Situação II: $d_M(\lambda)$ e $d_V(\lambda)$ não se comportam monotonicamente.

Na região em que a distância $d_M(\lambda)$ ($d_V(\lambda)$) não se comporta monotonicamente, mostramos que se existir um λ_M (λ_V) que a maximiza, ele será único, obtendo uma expressão fechada para ele. Para isso, considere as distâncias $d_M(\lambda)$ e $d_V(\lambda)$, $\hat{\beta}$ o estimador de máxima verossimilhança de β e as constantes $K_{M_0} = 0.8399238$ e $K_{V_0} = 1.598539$. As constantes K_{M_0} e K_{V_0} foram obtidas através de métodos numéricos como solução de equações cujos os termos possuem expressões dependentes da fração $\frac{\phi(k)}{\Phi(k)}$. Para maiores detalhes, ver a demonstração da Proposição 3.1.6.

Proposição 4.2.5 Considere as distâncias d_M e d_V , $a^2 = \frac{[\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)]^2}{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}$ e as constantes $K_{M_0} = 0.8399238$ e $K_{V_0} = 1.598539$. Para $\lambda > 0$ e $\hat{\beta} > \mu$ ($\lambda < 0$ e $\hat{\beta} < \mu$),

- (i) se $a^2 > K_{M_0}^2 \tau^2$ então existe um único λ_M dado por

$$\lambda_M = K_{M_0} \sqrt{\frac{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}{a^2 - \tau^2 K_{M_0}^2}} \quad \left(\lambda_M = -K_{M_0} \sqrt{\frac{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}{a^2 - \tau^2 K_{M_0}^2}} \right), \quad (4.30)$$

caso contrário não existe λ_M ;

- (ii) se $a^2 > K_{V_0}^2 \tau^2$ então existe um único λ_V dado por

$$\lambda_V = K_{V_0} \sqrt{\frac{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}{a^2 - \tau^2 K_{V_0}^2}} \quad \left(\lambda_V = -K_{V_0} \sqrt{\frac{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}{a^2 - \tau^2 K_{V_0}^2}} \right), \quad (4.31)$$

caso contrário não existe λ_V .

Para obtermos o supremo das distâncias das esperanças *a posteriori*, devemos comparar a^2 com $K_{M_0}^2 \tau^2$ para estudarmos o comportamento da distância d_M em função de λ , quando $\lambda > 0$ e $\hat{\beta} > \mu$ ou $\lambda < 0$ e $\hat{\beta} < \mu$. Nestes casos, segundo a Proposição 4.2.5, se $a^2 > K_{M_0}^2 \tau^2$ então a função $d_M(\lambda)$ possui um único ponto de máximo, obtido quando λ tem o valor dado por uma das expressões de λ_M em (4.30). Se $\lambda > 0$ e $\hat{\beta} > \mu$ então $\lambda_M = K_{M_0} \sqrt{\frac{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}{a^2 - \tau^2 K_{M_0}^2}}$; se $\lambda < 0$ e $\hat{\beta} < \mu$

então $\lambda_M = -K_{M_0} \sqrt{\frac{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}{a^2 - \tau^2 K_{M_0}^2}}$. Dessa forma, para obtermos $\sup_{\lambda} d_M(\lambda)$, basta tomarmos a distância $d_M(\lambda)$ calculada em λ_M , ou seja, $d_M(\lambda_M)$.

Caso não exista tal λ_M , é porque a distância d_M em função de λ se comporta monotonicamente e, sendo assim, análise similar as das Proposições 4.2.3 e 4.2.4 deve ser considerada para que possamos determinar o supremo das medidas de distância *a posteriori*.

Raciocínio análogo deve ser considerado para obtermos o supremo da distância d_V com respeito a λ .

Aqui, assim como na Situação I, $\inf_{\lambda} d_M = \inf_{\lambda} d_V = 0$.

4.3 Modelos de regressão normal com variância dos erros desconhecida.

Considere novamente a estrutura do modelo de regressão linear normal definido em (4.2), tal que a matriz de covariáveis \mathbf{X} tenha posto completo p . A função de verossimilhança, por expansão da forma quadrática nela contida, fica dada por:

$$f(\mathbf{y} \mid \boldsymbol{\beta}, \tau^2) = \left(\frac{1}{2\pi\tau^2} \right)^{\frac{n}{2}} e^{-\frac{1}{2\tau^2} [Ks^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})]}, \quad (4.32)$$

onde $Ks^2 = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$ é conhecido como soma dos quadrados residual, $K = n - p$ e $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ é a estimativa de mínimos quadrados (ou de máxima verossimilhança) de $\boldsymbol{\beta}$.

Vamos assumir independência *a priori* entre o vetor de parâmetros regressores $\boldsymbol{\beta}$ e o parâmetro de escala τ^2 .

Assim como na Seção 4.2, vamos adotar para a distribuição *a priori* do vetor de parâmetros regressores $\boldsymbol{\beta} \sim SN_p(\boldsymbol{\mu}, \Omega, \lambda)$, com $\boldsymbol{\mu}$ e Ω fixados e $\lambda \in \mathbb{R}^p$.

Para a distribuição *a priori* do parâmetro de escala, vamos considerar a distribuição *a priori* não-informativa $h(\tau^2) \propto \frac{1}{\tau^2}$ (Priori 1) e $\tau^2 \sim GI(a, b)$, com $a > 0$ e $b > 0$ (Priori 2).

Priori 1:

Só faz sentido considerar a priori não-informativa para o parâmetro de escala, se conseguirmos mostrar que a função de densidade conjunta *a posteriori* é uma genuína função densidade de probabilidade. Este resultado é apresentado sob forma de proposição e sua demonstração consta no Apêndice A.

Proposição 4.3.1 *Considere o modelo de regressão dado por (4.2) e que a distribuição a priori para o vetor de parâmetros regressores β seja $SN_p(\mu, \Omega, \lambda)$. Assumindo independência a priori entre os parâmetros β e τ^2 e se a distribuição a priori for não-informativa tal que $h(\tau^2) \propto \frac{1}{\tau^2}$, então a distribuição a posteriori conjunta é própria.*

Num outro contexto, Arellano-Valle, Castro, Genton & Gómez (2008) mostraram que a distribuição *a posteriori* completa do modelo de regressão normal-assimétrico é própria quando uma distribuição *a priori* própria arbitrária é considerada para o parâmetro de forma e distribuições *a priori* não-informativas são assumidas para os outros parâmetros do modelo.

O resultado que apresentamos a seguir caracteriza as distribuições condicionais completas *a posteriori* do vetor de parâmetros regressores β e do parâmetro de escala τ^2 , quando se considera o modelo de regressão linear normal, a classe de contaminação *a priori* para o parâmetro β e a distribuição *a priori* não-informativa para τ^2 especificada anteriormente.

Proposição 4.3.2 *Considere o modelo de regressão dado por (4.2) e que a distribuição a priori para o vetor de parâmetros regressores seja $SN_p(\mu, \Omega, \lambda)$. Assumindo a independência a priori entre os parâmetros β e τ^2 , se $h(\tau^2) \propto \frac{1}{\tau^2}$ então*

$$\beta \mid \tau^2, Y \sim GSN_{p,1} \left(\zeta, \left(\frac{X^T X}{\tau^2} + \Omega^{-1} \right)^{-1}, \lambda^T \omega^{-1}, \lambda^T \omega^{-1} \mu, 1 \right) \quad (4.33)$$

$$\tau^2 \mid \beta, Y \sim GI \left(\frac{n}{2}, \frac{K s^2}{2} + \frac{(\beta - \hat{\beta})^T X^T X (\beta - \hat{\beta})}{2} \right), \quad (4.34)$$

onde ω é a matriz diagonal formada pelos desvios-padrões de Ω , $K = n - p$, $K s^2 = (y - X\hat{\beta})^T (y - X\hat{\beta})$ é conhecido como soma dos quadrados residual, $\hat{\beta} =$

$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ é a estimativa de mínimos quadrados (ou de máxima verossimilhança) de β e

$$\zeta = \left(\frac{\mathbf{X}^T\mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} \left(\frac{\mathbf{X}^T\mathbf{X}}{\tau^2} \hat{\beta} + \Omega^{-1} \mu \right). \quad (4.35)$$

De posse das distribuições condicionais completas apresentadas na Proposição 4.3.2, pode-se implementar o método de amostragem de Gibbs, que é um caso particular dos métodos de Monte Carlo obtidos via Cadeias de Markov, para se obter estimativas de medidas *a posteriori*, tais como a média e a variância *a posteriori*.

Na proposição a seguir, apresentamos expressões para a distribuição *a posteriori* marginal do vetor de parâmetros regressores β . Mostramos que a média e a variância da distribuição *a posteriori* marginal do parâmetro de escala τ^2 são funções da distribuição *a posteriori* marginal de β .

Proposição 4.3.3 *Considere o modelo de regressão dado por (4.2) e que a distribuição a priori para o vetor de parâmetros regressores seja $SN_p(\mu, \Omega, \lambda)$ independente da distribuição a priori de τ^2 . Se, para $Ks^2 = (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)$, onde $K = n - p$, a distribuição a priori para o parâmetro de escala τ^2 for $h(\tau^2) \propto \frac{1}{\tau^2}$ então*

$$\text{i. } f(\beta | \mathbf{Y}) = \frac{1}{C} \phi_p(\beta; \mu, \Omega) \Phi \left[\lambda^T \omega^{-1}(\beta - \mu) \right] \left[\frac{Ks^2}{2} + \frac{(\beta - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\beta - \hat{\beta})}{2} \right]^{-\frac{n}{2}},$$

em que

$$C = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi_p(\beta; \mu, \Omega) \Phi \left[\lambda^T \omega^{-1}(\beta - \mu) \right] \left[\frac{Ks^2}{2} + \frac{(\beta - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\beta - \hat{\beta})}{2} \right]^{-\frac{n}{2}} d\beta,$$

$$\text{ii. } E[\tau^2 | \mathbf{Y}] = \frac{1}{n-2} \left\{ Ks^2 + E \left[(\beta - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\beta - \hat{\beta}) \mid \mathbf{Y} \right] \right\},$$

$$\text{iii. } \text{Var}[\tau^2 | \mathbf{Y}] = \frac{2}{(n-4)(n-2)^2} E \left\{ Ks^2 + \left[(\beta - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\beta - \hat{\beta}) \right]^2 \mid \mathbf{Y} \right\} + \frac{1}{(n-2)^2} \text{Var} \left[(\beta - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\beta - \hat{\beta}) \mid \mathbf{Y} \right].$$

Observação: Cabe aqui fazer uma ressalva quanto a um abuso de notação presente na expressão da variância marginal *a posteriori* dos erros aleatórios. Veja que o termo $(\beta - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\beta - \hat{\beta})$ é unidimensional, o que justifica escrevermos o termo $[(\beta - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\beta - \hat{\beta})]^2$ na $\text{Var}[\tau^2 | \mathbf{Y}]$.

Priori 2:

A seguir apresentamos versões das Proposições 4.3.2 e 4.3.3, ao se considerar $\tau^2 \sim GI(a, b)$.

Proposição 4.3.4 *Considere o modelo de regressão dado por (4.2) e que a distribuição a priori para o vetor de parâmetros regressores seja $SN_p(\mu, \Omega, \lambda)$. Assumindo a independência a priori entre os parâmetros β e τ^2 , se $\tau^2 \sim GI(a, b)$ então*

$$\beta \mid \tau^2, \mathbf{Y} \sim GSN_{p,1} \left(\zeta, \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1}, \lambda^T \omega^{-1}, \lambda^T \omega^{-1} \mu, 1 \right) \quad (4.36)$$

$$e$$

$$\tau^2 \mid \beta, \mathbf{Y} \sim GI \left(\frac{n}{2} + a, \frac{Ks^2}{2} + \frac{(\beta - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\beta - \hat{\beta})}{2} + b \right), \quad (4.37)$$

onde ω é a matriz diagonal formada pelos desvios-padrões de Ω , $K = n - p$, $Ks^2 = (\mathbf{y} - \mathbf{X}\hat{\beta})^T (\mathbf{y} - \mathbf{X}\hat{\beta})$ é conhecido como soma dos quadrados residual, $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ é a estimativa de mínimos quadrados (ou de máxima verossimilhança) de β e

$$\zeta = \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} \hat{\beta} + \Omega^{-1} \mu \right). \quad (4.38)$$

Proposição 4.3.5 *Considere o modelo de regressão dado por (4.2) e que a distribuição a priori para o vetor de parâmetros regressores seja $SN_p(\mu, \Omega, \lambda)$ independente da distribuição a priori de τ^2 . Se, para $Ks^2 = (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta)$, onde $K = n - p$, a distribuição a priori para o parâmetro de escala τ^2 for $\tau^2 \sim GI(a, b)$ então*

$$i. f(\beta \mid \mathbf{Y}) = \frac{1}{C} \phi_p(\beta; \mu, \Omega) \Phi [\lambda^T \omega^{-1} (\beta - \mu)] \left[\frac{Ks^2}{2} + \frac{(\beta - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\beta - \hat{\beta})}{2} + b \right]^{-\left(\frac{n}{2} + a\right)},$$

em que

$$C = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi_p(\beta; \mu, \Omega) \Phi [\lambda^T \omega^{-1} (\beta - \mu)] \left[\frac{Ks^2}{2} + \frac{(\beta - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\beta - \hat{\beta})}{2} + b \right]^{-\left(\frac{n}{2} + a\right)} d\beta,$$

$$ii. E[\tau^2 \mid \mathbf{Y}] = \frac{1}{n+2a-2} \left\{ Ks^2 + E \left[(\beta - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\beta - \hat{\beta}) \mid \mathbf{Y} \right] + 2b \right\},$$

$$\text{iii. } \text{Var}[\tau^2 | \mathbf{Y}] = \frac{2}{(n+2a-4)(n+2a-2)^2} E \left\{ Ks^2 + [(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + 2b]^2 \mid \mathbf{Y} \right\} + \\ + \frac{1}{(n+2a-2)^2} \text{Var} [(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \mid \mathbf{Y}].$$

A seguir, para ilustrar os resultados apresentados nas seções anteriores, vamos utilizar um conjunto de dados apresentado em Casella (1983) para modelos de regressão simples sem intercepto.

4.4 Análise de sensibilidade global para os dados GPM

O conjunto de dados que utilizamos foi apresentado na Revista *Motor Trend* 1974 e consiste na avaliação de 10 características de 32 diferentes modelos de automóveis (1973-1974). Consideraremos aqui apenas duas variáveis: o número de galões de gasolina consumido pelo automóvel por milha percorrida, denotado por GPM, como variável dependente e o peso total do veículo, denotado por PV, como variável independente ou explicativa.

Casella (1983) argumenta que a escolha de se utilizar a variável galões por milha e não o seu inverso (milhas por galão), o que seria mais comum, se deve ao relacionamento linear observado entre as variáveis GPM e PV, o que não ocorre quando se considera as variáveis 1/GPM e PV. Além disso, ele argumenta que características físicas impõe que este relacionamento linear se dê sem se considerar o intercepto, uma vez que um veículo com peso zero deve consumir zero de gasolina.

Para análise deste conjunto de dados, vamos considerar o modelo de regressão linear normal, sem intercepto, segundo duas abordagens: variância dos erros aleatórios conhecida e desconhecida.

4.4.1 τ^2 conhecido.

Na Subseção 4.2.1, consideramos o modelo de regressão linear normal, com um parâmetro regressor e com a variância dos erros (τ^2) conhecida. Em princípio, vamos fixar $\tau^2 = 0.594$, o que corresponde à estimativa de máxima verossimilhança para a variância dos erros.

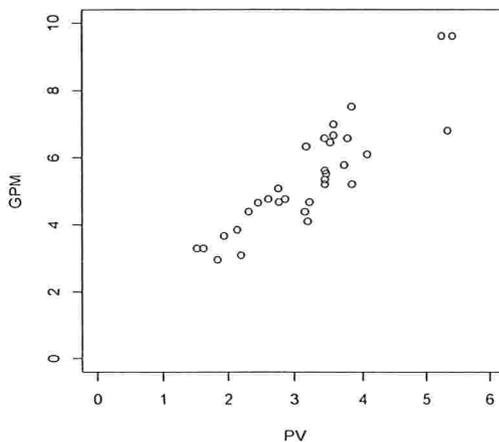


Figura 4.2: Gráfico de dispersão para as variáveis peso do veículo (PV) e galões por milha (GPM).

Para a análise de sensibilidade bayesiana, vamos considerar a classe de contaminação multiplicativa *a priori* especificada em (4.5) com a seguinte parametrização: $f_0(\beta) = \phi(\beta; 0, 10^4)$ e $w(\beta) = 2 \Phi(\lambda\beta; 0, 10^4)$. Dessa forma, a classe de contaminação multiplicativa *a priori* para β pode ser representada por $SN(0, 10^4, \lambda)$, com $\lambda \in \mathbb{R}$, onde λ reflete o grau de contaminação da distribuição *a priori*.

Observe que ao se escolher uma distribuição *a priori* base com grande variabilidade, estamos considerando uma distribuição *a priori* vaga, porém própria.

Temos então que:

$$\begin{aligned} GPM_i &= \beta PV_i + \epsilon_i, \quad \epsilon_i \sim N(0, 0.594), \quad \epsilon_i \perp \epsilon_j, \quad \forall i \neq j, \quad i = 1, \dots, 32, \\ \beta &\sim SN(0, 10^4, \lambda), \quad \text{com } \lambda \in \mathbb{R}. \end{aligned} \tag{4.39}$$

Dando início ao estudo de sensibilidade bayesiana, precisamos considerar as distâncias d_M e d_V a fim de que possamos avaliar a influência da classe de contaminação multiplicativa *a priori* sobre as estatísticas pontuais média e variância *a posteriori*. Lembre-se que a classe de contaminação multiplicativa *a priori* é obtida quando variamos $\lambda \in \mathbb{R}$ na distribuição $SN(0, 10^4, \lambda)$. Além disso, pela teoria apresentada, avaliamos as distâncias $d_M(\lambda)$ e $d_V(\lambda)$ em dois espaços distintos: $\lambda > 0$ e $\lambda < 0$.

Neste exemplo, a estimativa de máxima verossimilhança obtida para β foi 1.669, com erro padrão calculado em 0.041. Dessa forma, como $\hat{\beta}$ é maior que o valor considerado para o parâmetro de posição da distribuição *a priori* base ($\hat{\beta} > 0$), então o comportamento das distâncias $d_M(\lambda)$ e $d_V(\lambda)$ será monotônico estritamente decrescente em $\lambda \leq 0$. Para $\lambda > 0$, se existir um ponto λ_M (λ_V) que maximiza a distância $d_M(\lambda)$ ($d_V(\lambda)$), ele será único.

Dessa forma, para obtermos os valores de $\sup_{\lambda} d_M(\lambda)$ e $\sup_{\lambda} d_V(\lambda)$, quando $\lambda \leq 0$, basta utilizarmos, respectivamente, as expressões fornecidas em (4.26) e (4.27) com $\hat{\beta} = 1.669$, $\tau^2 = 0.594$, $\mu = 0$, $\tau^2 = 10^4$.

No entanto, uma dificuldade computacional surge aqui. As distâncias d_M e d_V , assim como seu supremo, dependem da razão $\frac{\phi(\cdot)}{\Phi(\cdot)}$. Neste exemplo específico

$$\frac{\phi \left[-\frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\tau \sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]}{\Phi \left[-\frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\tau \sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \right]} = \frac{\phi(-41.140)}{\Phi(-41.140)}, \quad (4.40)$$

em que $\phi(\cdot)$ e $\Phi(\cdot)$ correspondem, respectivamente, à função de densidade e de distribuição acumulada da distribuição normal padrão.

Como $\phi(-41.140)$ e $\Phi(-41.140)$ são aproximadamente iguais a zero, por uma limitação computacional $\phi(-41.140)$ e $\Phi(-41.140)$ são tomados iguais a zero, o que impossibilita o cálculo da fração em (4.40).

O que faremos aqui é determinar um limitante superior para a função $\frac{\phi(\cdot)}{\Phi(\cdot)}$ a fim de que possamos dar continuidade à análise de sensibilidade bayesiana.

Baricz (2008) apresenta diversas propriedades da razão de Mill, inclusive algumas desigualdades funcionais. Esta razão define-se como

$$r(x) = \frac{1 - \Phi(x)}{\phi(x)}, \quad \forall x \in \mathbb{R}. \quad (4.41)$$

Além disso, o autor cita que vários limites inferiores e superiores para esta razão foram propostos na literatura, sendo a mais conhecida a provada por Gordon (1941), que mostrou que

$$\frac{x}{x^2 + 1} < r(x) < \frac{1}{x}, \quad \forall x > 0. \quad (4.42)$$

À partir deste resultado, conseguimos demonstrar o seguinte lema:

Lema 4.4.1 Para todo $x > 0$, temos que $x < \frac{\phi(-x)}{\Phi(-x)} < x + \frac{1}{x}$.

Equivalentemente ao lema, se considerarmos $-x = y$, $\forall x > 0$, vemos que $-y < \frac{\phi(y)}{\Phi(y)} < -y - \frac{1}{y}$, $\forall y < 0$.

Note também que pelo lema, se x é positivo e suficientemente grande, então $\frac{\phi(-x)}{\Phi(-x)} \approx x$.

Na Figura 4.3, apresentamos uma gráfico que nos mostra esta aproximação.

De posse do resultado do Lema 4.4.1, conseguimos obter um limitante superior para as medidas $\sup_{\lambda} d_M$ e $\sup_{\lambda} d_V$ dadas em (4.26) e (4.27), pois para $\lambda \leq 0$ e $\hat{\beta} > \mu$ a expressão $\frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\tau \sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}}$ é sempre positiva. Assim,

$$\sup_{\lambda} d_M < \frac{\sigma \tau}{\sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}} \left[\mathcal{W} + \frac{1}{\mathcal{W}} \right], \quad (4.43)$$

$$\sup_{\lambda} d_V < \frac{\sigma^2 \tau^2}{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2} \left[1 + \frac{1}{\mathcal{W}^2} \right], \quad (4.44)$$

onde $\mathcal{W} = \frac{\sigma \sum_{i=1}^n x_i^2 (\hat{\beta} - \mu)}{\tau \sqrt{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}}$.

Vale ressaltar que faremos uso das cotas superiores para $\sup_{\lambda} d_M$ e $\sup_{\lambda} d_V$ somente na impossibilidade de obter um resultado exato para esses valores através das

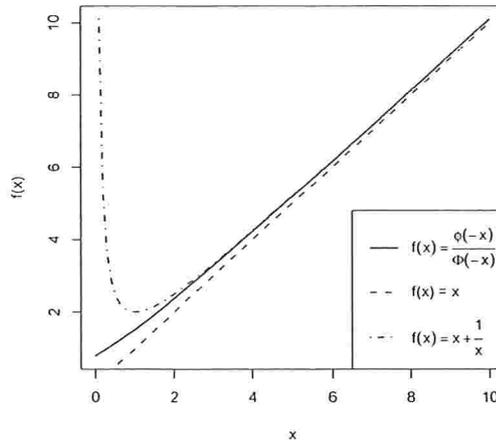


Figura 4.3: Aproximação da função $\frac{\phi(-x)}{\Phi(-x)}$ por x e $x + \frac{1}{x}$, $\forall x > 0$.

expressões (4.26) e (4.27). Isto se deve pois, como pode ser observado na Figura 4.3, valores moderados positivos de x não resultam em boas aproximações para a função $\frac{\phi(-x)}{\Phi(-x)}$, mas, no entanto, nestes casos não faz sentido usar a aproximação, já que a fração $\frac{\phi(-x)}{\Phi(-x)}$ é perfeitamente calculável.

Na Tabela 4.1, apresentamos o supremo e o ínfimo de d_M e d_V nas regiões em que $\lambda < 0$ e $\lambda \geq 0$.

Tabela 4.1: Valores obtidos para o supremo e o ínfimo de d_M e d_V .

Região	d_M		d_V	
	$\inf_{\lambda} d_M$	$\sup_{\lambda} d_M$	$\inf_{\lambda} d_V$	$\sup_{\lambda} d_V$
$\lambda < 0$	0	1.6707	0	0.0017
$\lambda \geq 0$	0	0.0003	0	0.0002

Em termos interpretativos, as medidas $\sup_{\lambda} d_M$ e $\inf_{\lambda} d_M$ correspondem, respectivamente, a maior e a menor distância entre o estimador de Bayes fornecido pela classe de contaminação multiplicativa *a priori* Γ_M e o estimador de Bayes sob a distribuição *a priori* base $f_0(\beta)$. De maneira análoga, as medidas $\sup_{\lambda} d_V$ e $\inf_{\lambda} d_V$ correspondem a maior e menor distância entre as variâncias *a posteriori* do parâmetro regressor.

Sob $f_0(\beta) = \phi(\beta; 0, 10^4)$, obtemos que a média e a variância *a posteriori* do parâmetro regressor são dados, respectivamente, por 1.669 e 0.002.

A Tabela 4.1 nos mostra que há pouca variação das estimativas da variância *a posteriori* com respeito a mudança de especificações *a priori* dentro da classe de contaminação multiplicativa *a priori* considerada. Além disso, esta pouca variação das estimativas *a posteriori* se mostra mais evidente na região em que λ tem o mesmo sinal da estimativa de máxima verossimilhança, tanto para a média, quanto para a variância *a posteriori*. No entanto, na região em que há divergência entre o sinal de λ e $\hat{\beta}$, a média *a posteriori* é bastante influenciada pela classe de contaminação *a priori*.

4.4.2 Estudo da influência de τ^2 .

Os resultados apresentados anteriormente foram obtidos considerando a variância dos erros aleatórios τ^2 igual a 0.594. Vejamos o que ocorre com as medidas $\sup_{\lambda} d_M$,

Tabela 4.2: Valores obtidos para o supremo e o ínfimo de d_M e d_V , $\hat{\beta}_0$ e $\hat{\zeta}_0^2$, para diferentes valores de τ^2 .

τ^2	$\hat{\beta}_0$	$\hat{\zeta}_0^2$	Região	$\sup_{\lambda} d_M$	$\sup_{\lambda} d_V$
3	1.669	0.008	$\lambda < 0$	1.675	0.008
			$\lambda \geq 0$	0.002	1.28×10^{-5}
5	1.669	0.014	$\lambda < 0$	1.678	0.014
			$\lambda \geq 0$	0.002	3.55×10^{-5}
100	1.669	0.277	$\lambda < 0$	1.813	0.259
			$\lambda \geq 0$	0.049	0.014
10^4	1.665	27.632	$\lambda < 0$	5.309	19.341
			$\lambda \geq 0$	3.195	15.526

$\sup_{\lambda} d_V$, quando assumimos diferentes valores para τ^2 , a saber $\tau^2 = 3, 5, 10, 100, 10^4$.

Ao analisarmos a Tabela 4.2, notamos que a medida em que se considera valores cada vez maiores para a variância dos erros aleatórios, a média *a posteriori* do parâmetro regressor obtida segundo a distribuição *a priori* base praticamente não se altera, o mesmo não podendo ser concluído para a variância *a posteriori* de β , também obtida segundo a distribuição *a priori* base.

Na Tabela 4.2 não são apresentados os valores calculados para $\inf_{\lambda} d_M$ e $\inf_{\lambda} d_V$, uma vez que já foi mostrado que $\inf_{\lambda} d_M = \inf_{\lambda} d_V = 0$, para quaisquer \mathbf{x} , \mathbf{y} , τ^2 , μ e σ^2 .

Fixado o valor de τ^2 , os valores dos supremos são menores no espaço em que $\lambda \geq 0$, justamente na região que tem o mesmo sinal da estimativa de máxima verossimilhança.

Quando $\tau^2 = 3, 5, 100$ e $\lambda \geq 0$, a magnitude dos supremos é bem pequena, indicando que neste espaço tanto a média quanto a variância *a posteriori* são robustas. Note também que, a medida em que τ^2 cresce, também cresce a magnitude dos intervalos de variação das médias e variâncias *a posteriori*, mas nada muito significativo.

Quando $\lambda < 0$ e $\tau^2 = 3, 5, 100$ e 10^4 , não verificamos as mesmas características apontadas no item anterior. De fato, a robustez com respeito a escolha de distribuições *a priori* na classe de contaminação multiplicativa Γ_M se evidencia na região em

que λ assume o mesmo sinal de $\hat{\beta}$. Neste exemplo, isso ocorre em $\lambda \geq 0$.

Numa situação extrema, consideramos $\tau^2 = 10^4$ e vimos que, mesmo neste caso, o conjunto de médias *a posteriori* pareceu não sofrer tanta influência de τ^2 , principalmente na região em que $\lambda \geq 0$. O mesmo não podemos concluir com respeito a variância *a posteriori*.

4.4.3 Estudo da influência de σ^2 .

Todas as análises realizadas até o momento consideraram as distribuições *a priori* para β com parâmetro de escala fixo $\sigma^2 = 10^4$. Vejamos o que ocorre quando consideramos uma distribuição *a priori* com menor variabilidade e cuja a variância dos erros aleatórios é relativamente grande. Assim, consideremos a estrutura do modelo de regressão similar a (4.39), substituindo $\tau^2 = 0.594$ por $\tau^2 = 100$ e $\sigma^2 = 10^4$ por $\sigma^2 = 1$.

Dessa forma, a distribuição *a priori* base será dada pela distribuição normal padrão. Na Tabela 4.3, apresentamos as estimativas da média e da variância *a posteriori* sob a distribuição *a priori* base, denotadas por $\hat{\beta}_0$ e $\hat{\zeta}_0^2$, respectivamente, e as medidas $\sup_{\lambda} d_M$ e $\sup_{\lambda} d_V$, quando $\sigma^2 = 1$ e 10^4 .

Tabela 4.3: Medidas *a posteriori*, $\sup_{\lambda} d_M$ e $\sup_{\lambda} d_V$, para $\tau^2 = 100$.

σ^2	$\hat{\beta}_0$	$\hat{\zeta}_0^2$	Região	$\sup_{\lambda} d_M$	$\sup_{\lambda} d_V$
1	1.308	0.217	$\lambda < 0$	1.446	0.200
			$\lambda \geq 0$	0.049	0.014
10^4	1.669	0.277	$\lambda < 0$	1.813	0.259
			$\lambda \geq 0$	0.049	0.014

Pela análise da Tabela 4.3, notamos que com respeito às estimativas pontuais, tanto a média quanto a variância *a posteriori* obtidas sob a distribuição *a priori* base pouco sofreram influência da mudança de $\sigma^2 = 10^4$ para $\sigma^2 = 1$. Na região em que $\lambda \geq 0$, as medidas $\sup_{\lambda} d_M$ e $\sup_{\lambda} d_V$ aparentemente mostraram-se invariantes à escolha de σ^2 . Também não houve grandes mudanças nas medidas $\sup_{\lambda} d_M$ e $\sup_{\lambda} d_V$ no espaço em que $\lambda < 0$. Vale ressaltar, no entanto, que excluindo-se a região $\lambda \geq 0$ e fixando $\sigma^2 = 1$, tanto $\hat{\beta}_0$ e $\hat{\zeta}_0^2$, quanto $\sup_{\lambda} d_M$ e $\sup_{\lambda} d_V$, apresentaram resultados inferiores às apresentadas pelas estimativas obtidas quando $\sigma^2 = 10^4$.

Também para $\sigma^2 = 1$ observamos que a robustez da média e da variância *a posteriori* do parâmetro regressor β com respeito a classe de contaminação multiplicativa *a priori* ocorre no espaço em que λ têm o mesmo sinal que o estimativa de máxima verossimilhança.

Vejam os que representa em termos de conflito entre a distribuição *a priori* e a verossimilhança o fato do espaço de variação de λ ter ou não o mesmo sinal que o estimativa de máxima verossimilhança para o parâmetro regressor.

Para um representante do espaço $\lambda > 0$, vamos tomar $\lambda = 20$. Note que aqui estamos na situação em que λ e $\hat{\beta}$ têm o mesmo sinal. Por outro lado, para um representante do espaço $\lambda < 0$, escolhemos $\lambda = -20$.

Na Figura 4.4 comparam-se a distribuição *a priori* com a função de verossimilhança e com a distribuição *a posteriori*. Para a verossimilhança, multiplicamos esta por uma constante com uma escala conveniente de tal forma que o resultado da integração da função de verossimilhança seja igual a um.

Pela Tabela 4.4, notamos que há uma pequena diferença entre as médias *a posteriori* do parâmetro β obtidas sob a distribuição *a priori* base e a distribuição *a priori* normal-assimétrica, quando $\lambda = -20$. No caso em que $\lambda = 20$, esta diferença é ainda menor. Análise semelhante pode ser feita quando consideramos a diferença entre as variâncias *a posteriori* de β .

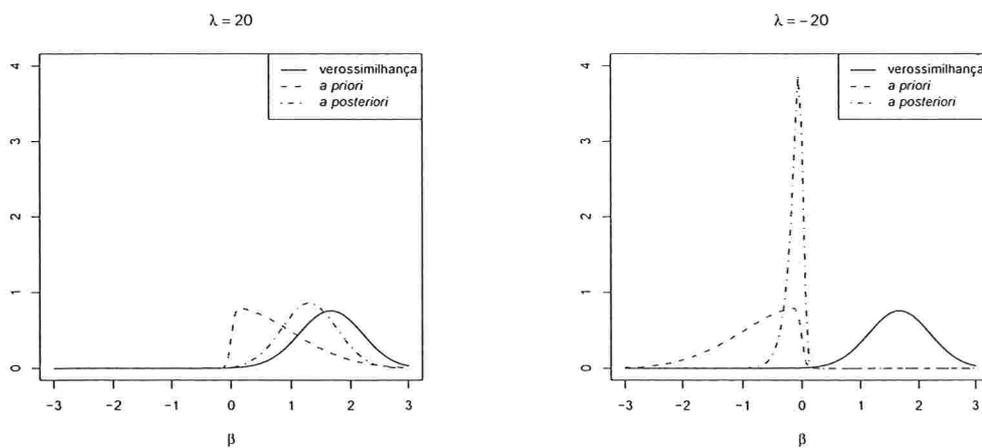


Figura 4.4: Comparação da distribuição *a priori*, verossimilhança e distribuição *a posteriori*, para $\lambda = 20, -20$.

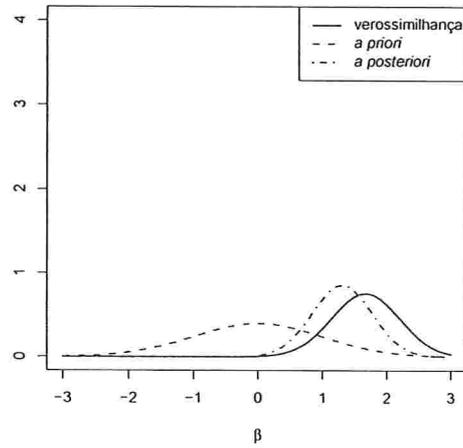


Figura 4.5: Comparação da distribuição *a priori* base, verossimilhança e distribuição *a posteriori*.

Tabela 4.4: Distância entre as estimativas *a posteriori* obtidas sob a distribuição *a priori* base e a distribuição *a priori* normal-assimétrica, com $\lambda = -20, 20$.

λ	$d_M(\lambda) = \left E_\lambda[\beta \mathbf{x}] - \hat{\beta}_0 \right $	$d_V(\lambda) = \left \text{Var}_\lambda[\beta \mathbf{x}] - \hat{\zeta}_0^2 \right $
20	0.004	0.005
-20	1.431	0.198

No entanto, apesar de pontualmente o desvio entre as médias e as variâncias *a posteriori* obtidos para β não serem muito grandes, vemos que a escolha de $\lambda = 20$ ou $\lambda = -20$ gera um grande impacto no comportamento da distribuição *a posteriori* para β (Figura 4.4). Mais do que isso, quando consideramos a distribuição normal-assimétrica com parâmetro de assimetria -20 para especificação *a priori* do parâmetro regressor, observamos um evidente conflito entre a informação fornecida pela distribuição *a priori* e a verossimilhança.

4.4.4 τ^2 desconhecido.

Vamos considerar o conjunto de dados referente a 10 características de 32 diferentes modelos de automóveis, publicados em 1974 na Revista *Motor Trend* e hoje divulgados no artigo de Henderson & Velleman (1981).

O modelo de regressão a ser considerado estrutura-se como

$$GPM_i = \beta PV_i + \epsilon_i, \quad \epsilon_i \sim N(0, \tau^2), \quad \epsilon_i \perp \epsilon_j, \quad \forall i \neq j, \quad i = 1, \dots, 32, \quad (4.45)$$

em que GPM é o número de galões de gasolina consumido pelo automóvel por milha percorrida e PV é o peso total do veículo.

Para uma análise bayesiana do problema, devemos especificar as distribuições *a priori* para o parâmetros β e τ^2 que são, respectivamente, o parâmetro regressor e a variância dos erros aleatórios.

Vamos considerar para a especificação *a priori* do parâmetro regressor, a classe de contaminação multiplicativa *a priori* definida em (4.5) com a seguinte parametrização: $f_0(\beta) = \phi(\beta; 0, 10^4)$ e $w(\beta) = 2 \Phi(\lambda\beta; 0, 10^4)$. Dessa forma, a classe de contaminação multiplicativa *a priori* para β pode ser representada por $SN(0, 10^4, \lambda)$, com $\lambda \in \mathbb{R}$, onde λ reflete o grau de contaminação da distribuição *a priori*.

Consideraremos as seguintes especificações *a priori* do parâmetro de escala do modelo de regressão linear normal: $h(\tau^2) \propto \frac{1}{\tau^2}$ (Priori 1) e $\tau^2 \sim GI(a, b)$ (Priori 2).

Para darmos início ao estudo de sensibilidade bayesiana, precisamos considerar as distâncias d_M e d_V a fim de que possamos avaliar a influência da classe de contaminação multiplicativa *a priori* sobre as estatísticas pontuais média e variância *a posteriori*. Estas estimativas pontuais foram obtidas por meio dos resultados apresentados nas Proposições 4.3.3 e 4.3.5 e do uso da técnica de integração numérica de Gauss-Hermite para calcular a média *a posteriori* de funções de $h(\beta)$.

Com respeito às distâncias, adotaremos as seguintes notações:

$$\begin{aligned} d_{M_\beta} &= |E_\lambda[\beta | \mathbf{x}] - E_0[\beta | \mathbf{x}]| \quad \text{e} \quad d_{V_\beta} = |Var_\lambda[\beta | \mathbf{x}] - Var_0[\beta | \mathbf{x}]|, \\ d_{M_{\tau^2}} &= |E_\lambda[\tau^2 | \mathbf{x}] - E_0[\tau^2 | \mathbf{x}]| \quad \text{e} \quad d_{V_{\tau^2}} = |Var_\lambda[\tau^2 | \mathbf{x}] - Var_0[\tau^2 | \mathbf{x}]|. \end{aligned}$$

Note que, como a classe de contaminação multiplicativa *a priori* para o parâmetro β contém o caso em que $\lambda = 0$, temos que $\inf_\lambda d_{M_\beta} = \inf_\lambda d_{V_\beta} = \inf_\lambda d_{M_{\tau^2}} = \inf_\lambda d_{V_{\tau^2}} = 0$.

Para obtermos as medidas $\sup_{\lambda} d_{M_{\beta}}$, $\sup_{\lambda} d_{V_{\beta}}$, $\sup_{\lambda} d_{M_{\tau^2}}$ e $\sup_{\lambda} d_{V_{\tau^2}}$ precisaremos recorrer a algoritmos numéricos para calcular as diversas integrais relacionadas a estas expressões. Utilizaremos aqui o método de integração do tipo Gauss-Hermite, associado a função transformadora *split-t*, para agilizar o processo de convergência. Tal metodologia está implementada no pacote *bayespack* do programa *R*. Maiores detalhes sobre este processo de integração numérica podem ser encontrados em Genz & Kass (1989) e Genz & Kass (1997).

Embora o estudo tenha sido realizado com as duas distribuições *a priori* anteriormente especificadas para τ^2 , os resultados e conclusões para $\tau^2 \sim GI(a, b)$, com $a = b = \frac{1}{10000}$, foram muito similares a $h(\tau^2) \propto \frac{1}{\tau^2}$. Por isso, apresentaremos aqui somente os resultados em que $h(\tau^2) \propto \frac{1}{\tau^2}$.

Sob a distribuição *a priori* base ($\beta \sim N(0, 10^4)$), obtivemos que as estimativas da média e da variância *a posteriori* para o parâmetro regressor foram 1.669 e 0.002, respectivamente. Para as estimativas *a posteriori* de τ^2 , obtivemos 0.626 para a média e 0.026 para a variância.

Na Tabela 4.5, apresentamos os valores dos supremos quando consideramos a distribuição *a priori* não informativa para a variância dos erros aleatórios e a classe de contaminação multiplicativa normal-assimétrica padrão para especificação da distribuição *a priori* do parâmetro regressor.

Tabela 4.5: Valores obtidos para o supremo das distâncias *a posteriori*.

Região	d_M		d_V	
	$\sup_{\lambda} d_{M_{\beta}}$	$\sup_{\lambda} d_{V_{\beta}}$	$\sup_{\lambda} d_{M_{\tau^2}}$	$\sup_{\lambda} d_{V_{\tau^2}}$
$\lambda < 0$	4.7×10^{-3}	1.2×10^{-7}	9.8×10^{-6}	4.6×10^{-6}
$\lambda \geq 0$	2.4×10^{-4}	1.0×10^{-7}	4.3×10^{-7}	6.4×10^{-8}

Note que, segundo a Tabela 4.5, há pouca variação das estimativas da média e da variância *a posteriori*, principalmente na região em que λ tem o mesmo sinal da estimativa de máxima verossimilhança de β . Dessa forma, concluímos que tanto a média quanto a variância *a posteriori* são robustas quanto a escolha de distribuições *a priori* na classe de contaminação multiplicativa Γ_M .

Capítulo 5

Sensibilidade global com o uso da função de concentração

No Capítulo 3, apresentamos um estudo de sensibilidade que nos permitiu avaliar a influência da classe de distribuições *a priori* - construída segundo a classe de contaminação multiplicativa - na média e na variância *a posteriori*. Vale salientar que tais medidas descrevem a variação da distribuição *a posteriori* em apenas um aspecto, a partir da utilização de uma estatística pontual *a posteriori*.

Com o intuito de ter uma visão mais global do comportamento da distribuição *a posteriori* segundo a classe de distribuições *a priori* multiplicativa, vamos considerar uma outra abordagem para a análise de sensibilidade *a posteriori*. Fortini & Ruggeri (1995) utilizam uma medida de divergência conhecida como função de concentração e a aplicam em um estudo de sensibilidade sob a classe de distribuições *a priori* ϵ -contaminada. Uma das vantagens do uso da função de concentração em relação a outras medidas de divergência conhecidas na literatura, é que esta nos permitiu obter uma maior gama de resultados analíticos no contexto de sensibilidade global.

Cifarelli & Regazzini (1987) definiram a função de concentração como uma generalização da curva de concentração de Lorenz, cuja representação gráfica é muito utilizada por economistas para entender a distribuição de renda entre os indivíduos/famílias de uma determinada sociedade. Marshall & Olkin (1979), por exemplo, utilizam esta contextualização para descrever a construção da curva de Lorenz empírica. Para isso, eles consideram uma população com n indivíduos e suas respectivas rendas. Ao ordenar a renda dos indivíduos de forma crescente, obtém-se $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ e $n + 1$ pontos cartesianos definidos como $(k/n, S_k/S_n)$, $k = 0, \dots, n$, onde $S_0 = 0$ e $S_k = \sum_{i=1}^k x_{(i)}$. A curva de Lorenz é obtida quando

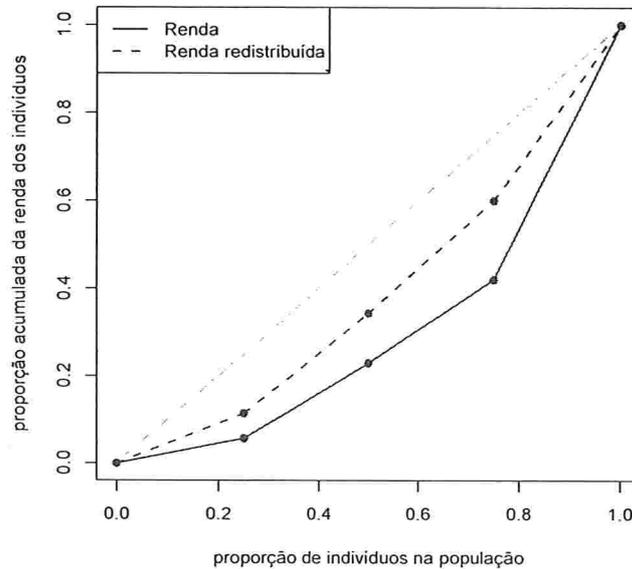


Figura 5.1: Curva de Lorenz para a distribuição de renda na população.

desenhamos os $n + 1$ pontos no plano cartesiano e os conectamos através de segmentos.

Se a renda é distribuída uniformemente entre os n indivíduos, então a curva de Lorenz corresponderá a linha reta na diagonal do quadrado unitário. Esta linha diagonal é muito utilizada como referência para interpretações, sendo que quanto mais afastada da diagonal for a curva de Lorenz, maior a desigualdade de renda entre os indivíduos.

Para exemplificar a técnica, considere uma empresa com 4 funcionários, cuja a folha de pagamento possui salários mensais dados por R\$ 1000,00, R\$ 3000,00, R\$ 3500,00 e R\$ 10000,00. Na Figura 5.1 apresentamos a curva de Lorenz para a distribuição de salários desta empresa, representada pela linha contínua. Note que, por construção, no eixo y temos a proporção acumulada da renda dos indivíduos e no eixo x a proporção de indivíduos organizados segundo a renda. Por exemplo, o quarto ponto cartesiano da curva de Lorenz corresponde a $(0.75, 0.42)$ e pode ser interpretado como segue: 42% do total da folha de pagamento desta empresa é destinado ao pagamento de 75% dos funcionários com menores salários.

Suponha agora que R\$ 3000,00 do maior salário fosse distribuído igualmente entre os demais funcionários, ou seja, os salários passassem a R\$ 2000,00, R\$ 4000,00, R\$ 4500,00 e R\$ 7000,00. Na Figura 5.1 utilizamos a linha tracejada para representar a curva de Lorenz nesta situação. Veja que quanto melhor distribuído o pagamento entre os funcionários, mais próximo da diagonal do quadrado unitário tende a estar a curva de Lorenz.

Como propriedades da curva de Lorenz podemos citar que está é não-decrescente, contínua e convexa. Tais propriedades também são compartilhadas pela função de concentração.

Segundo Fortini & Ruggeri (2000), ‘Cifarelli & Regazzini (1987) defined the concentration function of a probability measure P with respect to another one, say P_0 , extending the classical notion of the Lorenz-Gini curve. By the concentration function, the discrepancy between two measures defined on the same probability space is studied, comparing the different concentrations of probability determined by the measures.’

Assim, temos por objetivo utilizar a teoria de função de concentração para podermos comparar diferentes medidas de probabilidade *a posteriori* geradas à partir da variação da distribuição *a priori*, segundo a classe de contaminação multiplicativa, ao se considerar a função de verossimilhança fixa.

No Apêndice A apresentamos as demonstrações das proposições presentes neste capítulo.

5.1 A função de concentração

Antes de definirmos a função de concentração, vamos apresentar alguns conceitos iniciais que serão necessários para o entendimento desta seção. Para isso, considere Π e Π_0 duas medidas de probabilidade em um mesmo espaço mensurável (Θ, \mathcal{F}) .

Uma medida Π é absolutamente contínua com respeito a uma medida Π_0 , se para cada $A \in \mathcal{F}$, $\Pi_0(A) = 0$ implicar em $\Pi(A) = 0$. Neste caso, também dizemos que Π é dominada por Π_0 , denotando esta relação por $\Pi \ll \Pi_0$. Por outro lado, dizemos que as medidas Π e Π_0 são mutuamente singulares se têm suportes disjuntos, isto é, se existem conjuntos S_Π e S_{Π_0} tal que $\Pi(\Theta - S_\Pi) = 0$, $\Pi_0(\Theta - S_{\Pi_0}) = 0$ e $S_\Pi \cap S_{\Pi_0} = \emptyset$.

De posse destas definições, apresentaremos agora a função de concentração como

definida em Fortini & Ruggeri (2000). Para isso, considere Π e Π_0 duas medidas de probabilidade em um mesmo espaço mensurável (Θ, \mathcal{F}) . De acordo com o teorema de Radon-Nikodym, existe uma única partição $\{N, N^C\} \subset \mathcal{F}$ de Θ e uma função não-negativa h em N^C tal que

$$\Pi(E) = \Pi_a(E \cap N^C) + \Pi_s(E \cap N), \quad \forall E \in \mathcal{F}, \quad (5.1)$$

onde Π_a e Π_s são, respectivamente, as partes absolutamente contínua e singular de Π com respeito a Π_0 , em que

$$\Pi_a(E \cap N^C) = \int_{E \cap N^C} h(\theta) \Pi_0(d\theta), \quad \Pi_0(N) = 0, \quad \Pi_s(N) = \Pi_s(\Theta). \quad (5.2)$$

Considere $h(\theta) = \infty$, $\forall N$ e defina

$$H(y) = \Pi_0(\{\theta \in \Theta : h(\theta) \leq y\}), \quad c_z = \inf\{y \in \mathbb{R} : H(y) \geq z\} \text{ e } c_z^- = \lim_{t \rightarrow z^-} c_t. \quad (5.3)$$

Finalmente, seja

$$L_z = \{\theta \in \Theta : h(\theta) \leq c_z\} \text{ e } L_z^- = \{\theta \in \Theta : h(\theta) < c_z\}. \quad (5.4)$$

Definição 5.1.1 A função $\varphi_\Pi: [0, 1] \rightarrow [0, 1]$ é a função de concentração de Π com respeito a Π_0 se $\varphi_\Pi(z) = \Pi(L_z^-) + c_z[z - H(c_z^-)]$, para $z \in (0, 1)$, $\varphi_\Pi(0) = 0$ e $\varphi_\Pi(1) = \Pi_a(\Theta)$.

Neste trabalho, por considerarmos medidas de probabilidade com características mais específicas, vamos apresentar a definição da função de concentração com uma linguagem mais simples do que a anteriormente apresentada.

Para isso, considere (Θ, \mathcal{F}, P) um espaço de probabilidade tal que $\Theta \equiv \mathbb{R}$ e $\mathcal{F} \equiv \mathcal{B}$, onde \mathcal{B} é a σ -álgebra de Borel nos subconjuntos de Θ . Além disso, seja Π e Π_0 duas medidas de probabilidade absolutamente contínuas com respeito a medida de Lebesgue, em um mesmo espaço mensurável $(\mathbb{R}, \mathcal{B})$.

Segundo o teorema de Radon-Nikodym (Billingsley (1986), Capiński & Kopp (2004)), dado duas medidas σ -finitas ν e μ tal que $\nu \ll \mu$, então existe uma função h não-negativa, tal que $\nu(A) = \int_A h(\theta) \mu(d\theta)$, para todo $A \in \mathcal{F}$. Além disso, para duas funções h e g não-negativas, $\mu[h \neq g] = 0$, o que nos garante a unicidade da função h com respeito a medida μ .

A função h é conhecida como derivada de Radon-Nikodym de ν com respeito a μ e é denotada por $\frac{d\nu}{d\mu}$.

Sabemos que uma medida μ em $(\mathbb{R}, \mathcal{B})$ tem uma função densidade de probabilidade associada a ela se e somente se μ é absolutamente contínua com respeito a medida de Lebesgue λ . Neste caso, a derivada de Radon-Nikodym $\frac{d\mu}{d\lambda} \equiv f$ é a função de densidade.

Nesse contexto, sejam f e f_0 as respectivas funções de densidade associadas as medidas Π e Π_0 absolutamente contínuas com respeito a medida de Lebesgue.

Note que a construção da função de concentração como apresentada na Definição 5.1.1 depende da determinação da derivada de Radon-Nikodym de uma medida Π com respeito a uma medida Π_0 . Assim,

$$\Pi(A) = \int_A f(\theta) d\theta = \int_A \frac{f(\theta)}{f_0(\theta)} f_0(\theta) d\theta = \int_A \frac{f(\theta)}{f_0(\theta)} \Pi_0(d\theta), \quad (5.5)$$

ou seja, a derivada de Radon-Nikodym da medida Π com respeito a medida Π_0 é a razão entre funções de densidades associadas a Π e a Π_0 . Daqui por diante, consideraremos $h(\theta) = \frac{f(\theta)}{f_0(\theta)}$, $\forall \theta$. Note que $E_{\Pi_0}[h(\theta)] = 1$.

Neste contexto, vejamos agora como determinar uma expressão mais concisa da função de concentração.

À partir da definição de função de concentração em (5.1.1), Cifarelli & Regazzini (1987) e Fortini & Ruggeri (1994) argumentam que

$$\varphi_{\Pi}(z) = \begin{cases} \Pi_a(\{\theta \in \Theta : h(\theta) < c_z\}), & \text{se } z = H(c_z^-) = \Pi_0(L_z^-) \\ \Pi_a(\{\theta \in \Theta : h(\theta) \leq c_z\}), & \text{se } z = H(c_z) = \Pi_0(L_z). \end{cases} \quad (5.6)$$

Se, por outro lado, $H(c_z^-) < z < H(c_z)$, então o termo complementar $c_z[z - H(c_z^-)]$ deve ser considerado, a fim de que propriedades como continuidade, convexidade, etc. possam ser atribuídas à função de concentração, como ocorre com a curva de Lorenz-Gini.

Note que, como $\Pi \ll \Pi_0$, então $\Pi_a = \Pi$. Além disso, como f e f_0 são funções de densidade, temos então que f e f_0 são funções contínuas não-negativas e, dessa forma, $h(\theta) = \frac{f(\theta)}{f_0(\theta)}$ também é uma função contínua não-negativa.

Vejamos se no nosso contexto, existe $z \in [0, 1]$ tal que $H(c_z^-) \leq z \leq H(c_z)$. Note

que, por (5.6),

$$\begin{aligned} H(c_z^-) \leq z \leq H(c_z) &\iff \Pi_0(L_z^-) \leq z \leq \Pi_0(L_z) \\ &\iff \Pi_0(\{\theta \in \Theta : h(\theta) < c_z\}) \leq z \leq \Pi_0(\{\theta \in \Theta : h(\theta) \leq c_z\}). \end{aligned} \quad (5.7)$$

Como $c_z \in \mathbb{R}$, pois $c_z = \inf\{y \in \mathbb{R} : H(y) \geq z\}$, e h é uma função contínua não-negativa então, pelo teorema do confronto, conseguimos mostrar que $z = \Pi_0(\{\theta \in \Theta : h(\theta) \leq c_z\}) = \Pi_0(L_z)$. Além disso,

$$\begin{aligned} c_z &= \inf\{y \in \mathbb{R} : H(y) \geq z\} \\ &= \inf\{y \in \mathbb{R} : \Pi_0(\{\theta \in \Theta : h(\theta) \leq y\}) \geq z\} \\ &= \inf\{y \in \mathbb{R} : \Pi_0(\{\theta \in \Theta : h(\theta) \leq y\}) \geq \Pi_0(\{\theta \in \Theta : h(\theta) \leq c_z\})\} \\ &= \inf\{y \in \mathbb{R} : y \geq c_z\}. \end{aligned} \quad (5.8)$$

Assim, $c_z = \inf\{y \in \mathbb{R} : y \geq c_z\} \Rightarrow y = c_z$.

Dessa forma, conseguimos mostrar que a Definição 5.1.1 para função de concentração pode ser reescrita como segue.

Seja $L_y = \{\theta \in \Theta : h(\theta) \leq y\}$. A função $\varphi_\Pi : [0, 1] \rightarrow [0, 1]$ tal que, para $z \in (0, 1)$,

$$\begin{aligned} \varphi_\Pi(z) &= \Pi(L_y), \quad \text{se } z = \Pi_0(L_y), \\ \varphi_\Pi(0) &= 0, \\ \varphi_\Pi(1) &= \Pi(\Theta) = 1, \end{aligned}$$

é conhecida como função de concentração de Π com respeito a Π_0 .

Note que $\Pi(L_y) \equiv \int_{L_y} f(\theta) d\theta$ e $\Pi_0(L_y) \equiv \int_{L_y} f_0(\theta) d\theta$.

Scarsini (1990) obteve a mesma definição reduzida da função de concentração ao considerar duas medidas de probabilidade no conjunto das partes de um espaço finito X . Como uma possível interpretação para esta definição reduzida, ele argumenta que, para cada conjunto L obtido em termos da razão de funções de densidade h , φ_Π relaciona a massa de probabilidade concentrada em L por Π_0 a massa de probabilidade concentrada em L por Π .

Fortini & Ruggeri (1995) apresentam diversos exemplos e usos da função de concentração em vários contextos. Eles descrevem, por exemplo, como determinar

numericamente a função de concentração de $\theta \sim \text{Gama}(2, 2)$ com respeito a $\theta_0 \sim \text{Exp}(1)$ através de uma estrutura similar a apresentada na definição reduzida da função de concentração.

Como discutido em Fortini & Ruggeri (2000), a função $\varphi_{\Pi}(z)$ é não-decrescente, contínua e convexa. Se $\Pi = \Pi_0$ então $\varphi_{\Pi}(z) = z, \forall z \in [0, 1]$. Além disso, quando $\varphi_{\Pi}(z) = 0, 0 \leq z \leq \alpha, 0 < \alpha < 1$, podemos dizer que a medida Π não tem massa no subconjunto $A \in \mathcal{F}$ tal que $\Pi_0(A) = \alpha$.

A seguir, apresentaremos uma importante propriedade da função de concentração destacada por Fortini & Ruggeri (1995) e que será útil no estudo da análise de sensibilidade global.

Propriedade 5.1.1 *Para qualquer $z \in [0, 1]$, a probabilidade sob Π , de todos os subconjuntos A com medida Π_0 dada por $z, \Pi_0(A) = z$, satisfaz que*

$$\varphi_{\Pi}(z) \leq \Pi(A) \leq 1 - \varphi_{\Pi}(1 - z). \quad (5.9)$$

Vejamos agora um exemplo que pretende apenas discutir a função de concentração como uma ferramenta de comparação entre duas medidas de probabilidade. Para isso, vamos determinar a função de concentração de $\theta \sim \text{SN}(\mu, \sigma^2, \lambda)$ com respeito a $\theta_0 \sim N(\mu, \sigma^2)$. O suporte de ambas as distribuições está em \mathbb{R} e

$$h(\theta) = \frac{f(\theta)}{f_0(\theta)} = 2\Phi\left(\lambda \frac{\theta - \mu}{\sigma}\right). \quad (5.10)$$

Segundo a teoria desenvolvida, nós calculamos a função de concentração para qualquer $z \in [0, 1]$ encontrando o valor de y tal que $z = \Pi_0(\{\theta \in \Theta : h(\theta) \leq y\})$. Dessa forma, para $\lambda > 0$,

$$\begin{aligned} 0 < h(\theta) \leq y &\iff \\ 0 < 2\Phi\left(\lambda \frac{\theta - \mu}{\sigma}\right) \leq y &\iff \\ -\infty < \theta \leq \frac{\sigma\Phi^{-1}(y/2)}{\lambda} + \mu. & \end{aligned} \quad (5.11)$$

Assim, seja $q = \frac{\sigma\Phi^{-1}(y/2)}{\lambda} + \mu$, então

$$\begin{aligned} z &= \Pi_0(\{\theta \in \Theta : h(\theta) \leq y\}) \\ &= \Pi_0(\{\theta \in \Theta : -\infty < \theta \leq q\}) \\ &= \Phi\left(\frac{q - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{\Phi^{-1}(y/2)}{\lambda}\right). \end{aligned} \quad (5.12)$$

Portanto, o y para o qual $z = \Phi\left(\frac{\Phi^{-1}(y/2)}{\lambda}\right)$ é dado por $2 \Phi[\lambda \Phi^{-1}(z)]$. Então, como $\varphi_{\Pi}(z) = \Pi(\{\theta \in \Theta : h(\theta) \leq y\}) = \Pi(\{\theta \in \Theta : h(\theta) \leq 2 \Phi[\lambda \Phi^{-1}(z)]\})$, temos que

$$\begin{aligned} 0 < h(\theta) \leq 2 \Phi[\lambda \Phi^{-1}(z)] &\iff \\ -\infty < \theta \leq \sigma\Phi^{-1}(z) + \mu. & \end{aligned} \quad (5.13)$$

Portanto, para $\lambda > 0$,

$$\begin{aligned} \varphi_{\Pi}(z) &= \int_{-\infty}^{\sigma\Phi^{-1}(z) + \mu} \frac{2}{\sigma} \phi\left(\frac{\theta - \mu}{\sigma}\right) \Phi\left(\lambda \frac{\theta - \mu}{\sigma}\right) d\theta \\ &= \int_{-\infty}^{\Phi^{-1}(z)} 2 \phi(t) \Phi(\lambda t) dt. \end{aligned} \quad (5.14)$$

Segundo Rodríguez (2005), podemos reescrever $\varphi_{\Pi}(z)$ como

$$\varphi_{\Pi}(z) = 2 \Phi_2 \left[\begin{pmatrix} \Phi^{-1}(z) \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Omega \right], \quad (5.15)$$

onde Φ_2 corresponde a função de distribuição acumulada da distribuição normal bivariada e $\Omega = \begin{bmatrix} 1 & -\delta \\ -\delta & 1 \end{bmatrix}$ é a matriz de correlação, com $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

Analogamente, para $\lambda < 0$,

$$\varphi_{\Pi}(z) = 1 - 2 \Phi_2 \left[\begin{pmatrix} \Phi^{-1}(1-z) \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Omega \right], \quad (5.16)$$

com a matriz de correlação Ω como especificado anteriormente.

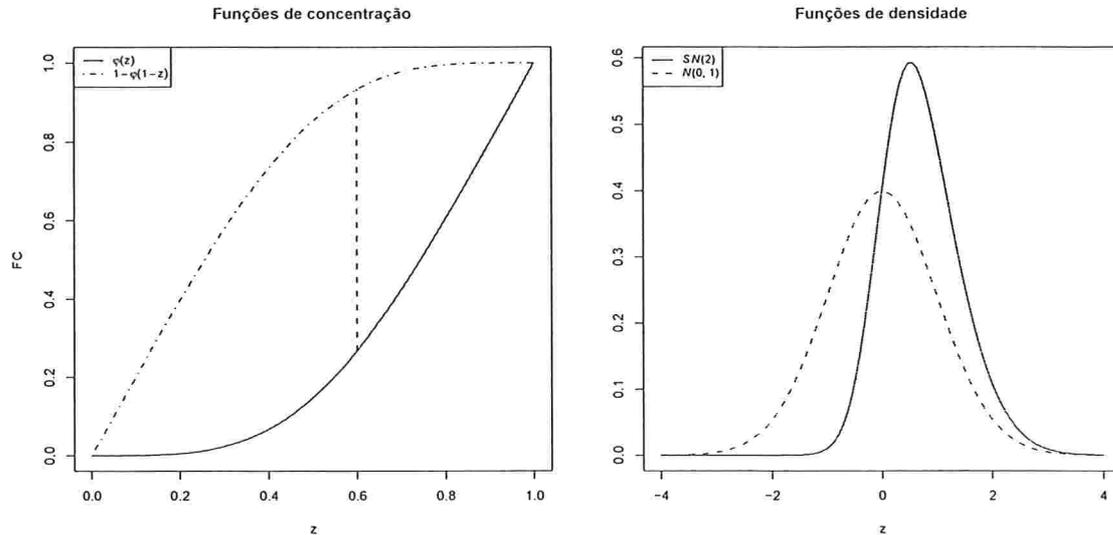


Figura 5.2: Função de concentração de $\theta \sim SN(2)$ com respeito a $\theta \sim N(0, 1)$ e respectivas funções de densidade.

Note que, pelas expressões dadas em (5.15) e (5.16), a função de concentração é invariante a mudanças de posição e de escala, quando $\theta \sim SN(\mu, \sigma^2, \lambda)$ e $\theta_0 \sim N(\mu, \sigma^2)$.

Na Figura 5.2, apresentamos o gráfico da função de concentração no caso específico em que $\theta \sim SN(2)$ e $\theta_0 \sim N(0, 1)$ e suas respectivas curvas da função de densidade.

No primeiro gráfico, vemos um segmento tracejado que intercepta a função de concentração em dois pontos, para $x = 0.6$. O intervalo entre as duas curvas está variando de $[0.267, 0.933]$, ou seja, $0.267 \leq \Pi(A) \leq 0.933$, onde A é formado por todos os possíveis conjuntos tais que $\Pi_0(A) = 0.6$.

De fato, quanto mais próximo $\varphi_\Pi(z)$ for de $1 - \varphi_\Pi(1 - z)$, mais próximas estarão as curvas em relação a diagonal do quadrado unitário e, por consequência, mais “próximas” estarão as medidas de probabilidade.

Fortini & Ruggeri (2000) ressaltam que ‘*the c.f.* [função de concentração], far from substituting other usual distribution summaries, e.g. the mean, furnishes different information about probability measures. As an example, consider two measures

concentrated on disjoint, very close sets in \mathbb{R} : their means are very close, their variances might be the same but their c.f. is 0 in $[0, 1]$.’

5.2 Sensibilidade global

Para termos uma visão mais global do comportamento da distribuição *a posteriori*, quando variamos a distribuição *a priori* segundo a classe de contaminação multiplicativa, vamos considerar uma outra abordagem para a análise de sensibilidade *a posteriori*, que leva em conta a teoria das funções de concentração apresentada na seção anterior.

No caso da análise de sensibilidade global, vamos considerar a classe de contaminação multiplicativa para a distribuição *a priori* especificada nos termos de (5.17), em que $f_0(\cdot)$ é a distribuição *a priori* base e $f(\cdot)$ são as demais distribuições *a priori* pertencentes a Γ_M , tal que

$$\Gamma_M = \{ f(\theta) = f_0(\theta) w(\theta) : w \in \mathcal{G} \}, \text{ onde} \quad (5.17)$$

\mathcal{G} é um determinada subclasse de funções não-negativas.

Ao relacionarmos a função de verossimilhança à classe de distribuições *a priori* Γ_M , obtemos a classe de distribuições *a posteriori* Γ_M^* . Seja f_0^* a função de densidade *a posteriori* associada à distribuição *a priori* base e f^* as demais funções de densidade *a posteriori* associadas as distribuições *a priori* em Γ_M .

Conseqüentemente, vamos obter uma classe de funções de concentração (f.c.), dada por:

$$\Psi = \{ \varphi_f(z) : \varphi_f(z) \text{ é a f. c. de } f^* \text{ com respeito a } f_0^*, f^* \in \Gamma_M^* \}. \quad (5.18)$$

Segundo Fortini & Ruggeri (2000), para qualquer $f \in \Gamma$ e $A \in \mathcal{F}$ com $f_0^*(A) = z$, temos que

$$\hat{\varphi}(z) \leq f^*(A) \leq 1 - \hat{\varphi}(1 - z), \quad (5.19)$$

onde $\hat{\varphi}(z) = \inf_{f \in \Gamma} \varphi_f(z)$, para qualquer $x \in [0, 1]$.

Em termos de robustez bayesiana, Fortini & Ruggeri (1995) nos dizem que quanto mais próximos forem as medidas $\hat{\varphi}(z)$ e $1 - \hat{\varphi}(1 - z)$, para todo $z \in [0, 1]$, mais próximas as medidas *a posteriori* serão. É então possível fazer julgamentos sobre a

robustez observando quão próximos estão no gráfico as curvas de $\hat{\varphi}(z)$ e $y = z$, que corresponde a medidas de probabilidade iguais.

5.3 Modelo normal

Nesta seção, faremos um estudo de sensibilidade via função de concentração no mesmo contexto do Capítulo 3, ou seja, modelo normal com a distribuição *a priori* do parâmetro de posição dada pela classe de contaminação multiplicativa normal-assimétrica, em duas situações: τ^2 conhecido e desconhecido.

5.3.1 τ^2 conhecido

Nesta situação, como mostrado na Proposição 3.1.1, temos que se $X_i \sim N(\theta, \tau^2)$, em que τ^2 é conhecido, X_i independente de $X_j, \forall i \neq j$, então

$$\begin{aligned} \theta \sim SN(\mu, \sigma^2, \lambda) &\implies \theta^* \sim GSN_{1,1}(a, b^2, \lambda, \lambda\mu, \sigma^2) \\ \theta_0 \sim N(\mu, \sigma^2) &\implies \theta_0^* \sim GSN_{1,1}(a, b^2, 0, 0, \sigma^2), \end{aligned} \quad (5.20)$$

onde $a = \frac{n\bar{x}\sigma^2 + \mu\tau^2}{n\sigma^2 + \tau^2}$ e $b^2 = \frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2}$.

Com o objetivo de fazer uma estudo de robustez bayesiana nos moldes do que foi descrito na Seção 5.2, vamos apresentar a função de concentração de f^* com respeito a f_0^* , com θ^* e θ_0^* como definidos em (5.20).

Proposição 5.3.1 *A função de concentração de $\theta^* \sim GSN_{1,1}(a, b^2, \lambda, \lambda\mu, \sigma^2)$ com respeito a $\theta_0^* \sim GSN_{1,1}(a, b^2, 0, 0, \sigma^2)$ é dada por,*

I.

$$\varphi_{\Pi}(z) = \frac{\Phi_2 \left[\left(\begin{array}{c} \frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \\ \Phi^{-1}(z) \end{array} \right); \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \rho = -\frac{\lambda\tau}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \right]}{\Phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]}, \quad (5.21)$$

se $\lambda > 0$,

II.

$$\varphi_{\Pi}(z) = 1 - \frac{\Phi_2 \left[\left(\begin{array}{c} \frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \quad \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \\ \Phi^{-1}(1-z) \end{array} \right); \quad \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \rho = -\frac{\lambda\tau}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \right]}{\Phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \quad \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]}, \quad (5.22)$$

se $\lambda < 0$,

onde $\Phi_2(\cdot)$ corresponde a função de distribuição acumulada da distribuição normal bivariada e ρ é o coeficiente de correlação da distribuição normal bivariada.

Como em (5.18), obtivemos então uma classe de funções de concentração, em que

$$\Psi = \{\varphi_f(z) : \varphi_f(z) \text{ é a f. c. de } f^* \text{ com respeito a } f_0^*, f^* \in \Gamma_M^*\} \quad (5.23)$$

No nosso contexto, sabemos que a variação das distribuições *a priori* na classe de contaminação multiplicativa, representada pela distribuição normal-assimétrica, é dada pela variação do parâmetro de assimetria, com $\lambda \in \mathbb{R}$. Assim, para cada z fixado, temos uma classe Ψ de funções de concentração, variando segundo λ , ou seja,

$$\Psi_z = \{\varphi_\lambda(z) : \varphi_\lambda(z) \text{ é a f. c. de } f_\lambda^* \text{ com respeito a } f_0^*, \lambda \in \mathbb{R}\}, \quad (5.24)$$

com $\varphi_{\Pi}(z)$ definido pela Proposição 5.3.1.

Para a análise de robustez bayesiana, assim como em (5.19), devemos obter a medida $\hat{\varphi}(z) = \inf_{\lambda} \varphi_\lambda(z)$ para cada $z \in [0, 1]$ fixado. Dessa forma, para $A \in \mathbb{R}$, tal que $f_0^*(A) = z$, teremos que $\hat{\varphi}(z) \leq f_\lambda^*(A) \leq 1 - \hat{\varphi}(1 - z)$ ocorre $\forall \lambda$ e, por conseqüência, dizemos que quanto mais próximos forem as medidas $\hat{\varphi}(z)$ e $1 - \hat{\varphi}(1 - z)$, para todo $z \in [0, 1]$, mais próximas serão as distribuições *a posteriori* em Γ_M^* .

Fixado $z \in [0, 1]$, a obtenção de $\hat{\varphi}(z) = \inf_{\lambda} \varphi_\lambda(z)$ se dará computacionalmente, uma vez que não existem expressões analíticas para a minimização da função de distribuição acumulada de uma distribuição normal bivariada, da qual decorre Ψ_z em (5.24).

5.3.2 τ^2 desconhecido

Para a especificação *a priori* do parâmetro de escala, vamos considerar como antes duas situações. A primeira *a priori* não-informativa $h(\tau^2) \propto \frac{1}{\tau^2}$ e a segunda $\tau^2 \sim GI(a, b)$, com $a > 0$ e $b > 0$, em que $GI(a, b)$ corresponde a distribuição gama invertida, com parâmetros $a > 0$ e $b > 0$. Aqui, estaremos assumindo independência *a priori* entre θ e τ^2 .

$$\text{Priori 1: } h(\tau^2) \propto \frac{1}{\tau^2}$$

Nesta subseção, temos o interesse em estudar a influência da classe de contaminação multiplicativa *a priori* sobre a distribuição *a posteriori* do parâmetro de posição do modelo normal. Para isso, vamos considerar a distribuição *a posteriori* marginal de θ apresentada na Proposição 3.2.3 para obtermos o conjunto de funções de concentração dependentes em λ , como em 5.24.

Para isso, seja $X_i \sim N(\theta, \tau^2)$, X_i independente de X_j , $\forall i \neq j$. Considere $h(\tau^2) \propto \frac{1}{\tau^2}$ e, quando

$$\begin{aligned} \theta \sim SN(\mu, \sigma^2, \lambda) &\implies f^*(\theta | x) = \frac{1}{C_\lambda} \phi\left(\frac{\theta - \mu}{\sigma}\right) \Phi\left(\lambda \frac{\theta - \mu}{\sigma}\right) \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2}\right]^{-\frac{n}{2}} \\ \theta_0 \sim N(\mu, \sigma^2) &\implies f_0^*(\theta | x) = \frac{1}{C_0} \phi\left(\frac{\theta - \mu}{\sigma}\right) \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2}\right]^{-\frac{n}{2}}, \end{aligned} \quad (5.25)$$

onde C_0 e C_λ são as constantes normalizadoras das respectivas distribuições *a posteriori*.

A partir das distribuições *a posteriori* marginais para θ , obtemos a seguinte expressão para a função de concentração de f^* com respeito a f_0^* , com f^* e f_0^* como especificados em (5.25).

Proposição 5.3.2 *A função de concentração de f^* com respeito a f_0^* , com f^* e f_0^* como definidos em (5.25), é dada por,*

I. se $\lambda > 0$,

$$\varphi_{\Pi}(z) = \int_{-\infty}^{\frac{\sigma}{\lambda} \Phi^{-1}\left\{\frac{C_\lambda}{C_0} z\right\} + \mu} \frac{1}{C_\lambda} \phi\left(\frac{\theta - \mu}{\sigma}\right) \Phi\left(\lambda \frac{\theta - \mu}{\sigma}\right) \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2}\right]^{-\frac{n}{2}} d\theta, \quad (5.26)$$

onde y_z^* corresponde ao valor de y que valida a igualdade abaixo

$$z = \int_{-\infty}^{\frac{\sigma}{\lambda} \Phi^{-1} \left\{ \frac{C_\lambda}{C_0} y \right\} + \mu} \frac{1}{C_0} \phi \left(\frac{\theta - \mu}{\sigma} \right) \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2} \right]^{-\frac{n}{2}} d\theta, \quad (5.27)$$

II. se $\lambda < 0$,

$$\varphi_{\Pi}(z) = \int_{\frac{\sigma}{\lambda} \Phi^{-1} \left\{ \frac{C_\lambda}{C_0} y_z^* \right\} + \mu}^{\infty} \frac{1}{C_\lambda} \phi \left(\frac{\theta - \mu}{\sigma} \right) \Phi \left(\lambda \frac{\theta - \mu}{\sigma} \right) \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2} \right]^{-\frac{n}{2}} d\theta, \quad (5.28)$$

onde y_z^* corresponde ao valor de y que valida a igualdade abaixo

$$z = \int_{\frac{\sigma}{\lambda} \Phi^{-1} \left\{ \frac{C_\lambda}{C_0} y \right\} + \mu}^{\infty} \frac{1}{C_0} \phi \left(\frac{\theta - \mu}{\sigma} \right) \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2} \right]^{-\frac{n}{2}} d\theta. \quad (5.29)$$

Priori 2: $\tau^2 \sim GI(a, b)$, com $a > 0$ e $b > 0$

Para $X_i \sim N(\theta, \tau^2)$, X_i independente de X_j , $\forall i \neq j$, considere $\tau^2 \sim GI(a, b)$, com $a > 0$ e $b > 0$ e, quando

$$\begin{aligned} \theta \sim SN(\mu, \sigma^2, \lambda) &\implies f^*(\theta | x) = \frac{1}{C_\lambda} \phi \left(\frac{\theta - \mu}{\sigma} \right) \Phi \left(\lambda \frac{\theta - \mu}{\sigma} \right) \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2} + b \right]^{-\left(\frac{n}{2} + a\right)} \\ \theta_0 \sim N(\mu, \sigma^2) &\implies f_0^*(\theta | x) = \frac{1}{C_0} \phi \left(\frac{\theta - \mu}{\sigma} \right) \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2} + b \right]^{-\left(\frac{n}{2} + a\right)}, \end{aligned} \quad (5.30)$$

onde C_0 e C_λ são as constantes normalizadoras das respectivas distribuições a posteriori.

Por consequência, definimos abaixo uma expressão para a função de concentração.

Proposição 5.3.3 A função de concentração de f^* com respeito a f_0^* , com f^* e f_0^* como definidos em (5.30), é dada por,

I. se $\lambda > 0$,

$$\varphi_{\Pi}(z) = \int_{-\infty}^{\frac{\sigma}{\lambda} \Phi^{-1} \left\{ \frac{C_\lambda}{C_0} y_z^* \right\} + \mu} \frac{1}{C_\lambda} \phi \left(\frac{\theta - \mu}{\sigma} \right) \Phi \left(\lambda \frac{\theta - \mu}{\sigma} \right) \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2} + b \right]^{-\left(\frac{n}{2} + a\right)} d\theta, \quad (5.31)$$

onde y_z^* corresponde ao valor de y que valida a igualdade abaixo

$$z = \int_{-\infty}^{\frac{\sigma}{\lambda} \Phi^{-1} \left\{ \frac{C_\lambda}{C_0} y \right\} + \mu} \frac{1}{C_0} \phi \left(\frac{\theta - \mu}{\sigma} \right) \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2} + b \right]^{-\left(\frac{n}{2} + a\right)} d\theta, \quad (5.32)$$

II. se $\lambda < 0$,

$$\varphi_{\Pi}(z) = \int_{\frac{\sigma}{\lambda} \Phi^{-1}\left\{\frac{c_{\lambda}}{c_0} y_i^*\right\} + \mu}^{\infty} \frac{1}{C_{\lambda}} \phi\left(\frac{\theta - \mu}{\sigma}\right) \Phi\left(\lambda \frac{\theta - \mu}{\sigma}\right) \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2} + b\right]^{-\left(\frac{n}{2} + a\right)} d\theta, \quad (5.33)$$

onde y_i^* corresponde ao valor de y que valida a igualdade abaixo

$$z = \int_{\frac{\sigma}{\lambda} \Phi^{-1}\left\{\frac{c_{\lambda}}{c_0} y\right\} + \mu}^{\infty} \frac{1}{C_0} \phi\left(\frac{\theta - \mu}{\sigma}\right) \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2} + b\right]^{-\left(\frac{n}{2} + a\right)} d\theta. \quad (5.34)$$

Em ambas as escolhas da distribuição *a priori* para τ^2 , devemos considerar, assim como na Seção 5.3.1, as medidas $\hat{\varphi}(z)$ e $1 - \hat{\varphi}(1 - z)$, em que $\hat{\varphi}(z) = \inf_{\lambda \in \Gamma_M} \varphi_{\Pi}(z)$, para todo $z \in [0, 1]$, para avaliar a robustez bayesiana das medidas de probabilidade *a posteriori*.

Estudo de sensibilidade

Seja $X_i \sim N(\theta, 1)$, $i = 1, \dots, n$, $X_i \perp X_j$, $\forall i \neq j$. Considere $f(\theta) \in \Gamma_M$, tal que a classe de contaminação multiplicativa *a priori* seja formada por $c = 2$, $G(\lambda\theta) = \Phi(\lambda\theta)$ e contenha a distribuição *a priori* base $N(0, 1)$. Ou seja, a distribuição que representa a classe de contaminação multiplicativa *a priori* é $\theta \sim SN(\lambda)$.

Com este exemplo, vamos procurar saber o quão díspares são as distribuições *a posteriori* geradas à partir da variação das distribuições *a priori* na classe de contaminação multiplicativa. Para isso, vamos considerar a teoria relativa a função de concentração.

Para cada \bar{x} fixado, teremos um conjunto de funções de concentração indexados por λ . Dessa forma, assim como nos Capítulos 3 e 4 analisaremos as curvas da função de concentração em duas situações: $\lambda > 0$ e $\lambda < 0$.

Na Figura 5.3, apresentamos os gráficos para $\hat{\varphi}(z) = \inf_{\lambda \in \Gamma_M} \varphi_{\Pi}(z)$ e $1 - \hat{\varphi}(1 - z)$, quando $\bar{x} = -1, 0$ e 1 . Para a construção destes, assumimos $\sigma^2 = \tau^2 = 1$, $\mu = 0$ e, para cada z fixo, admitimos $\lambda \in (0, 250]$, tal que $\lambda = 0.1 \times i$, $i = 1, \dots, 250$, se $\lambda > 0$, ou $\lambda = -0.1 \times i$, $i = 1, \dots, 250$, se $\lambda < 0$ então,

$$\hat{\varphi}(z) = \inf_{\lambda} \varphi_{\lambda}(z) = \min\{\varphi_{\lambda_1}(z), \dots, \varphi_{\lambda_{250}}(z)\}. \quad (5.35)$$

Pelo mesmo processo obtemos $1 - \hat{\varphi}(1 - z)$.

Ao analisarmos os gráficos da Figura 5.3, notamos que, quando a média amostral e os valores considerados para o parâmetro de assimetria têm o mesmo sinal, há fortes evidências de robustez com respeito à escolha de diferentes distribuições *a priori* em Γ_M . Nesta situação, $\hat{\varphi}(z)$ e $1 - \hat{\varphi}(1 - z)$ estão muito próximos. Destacamos também

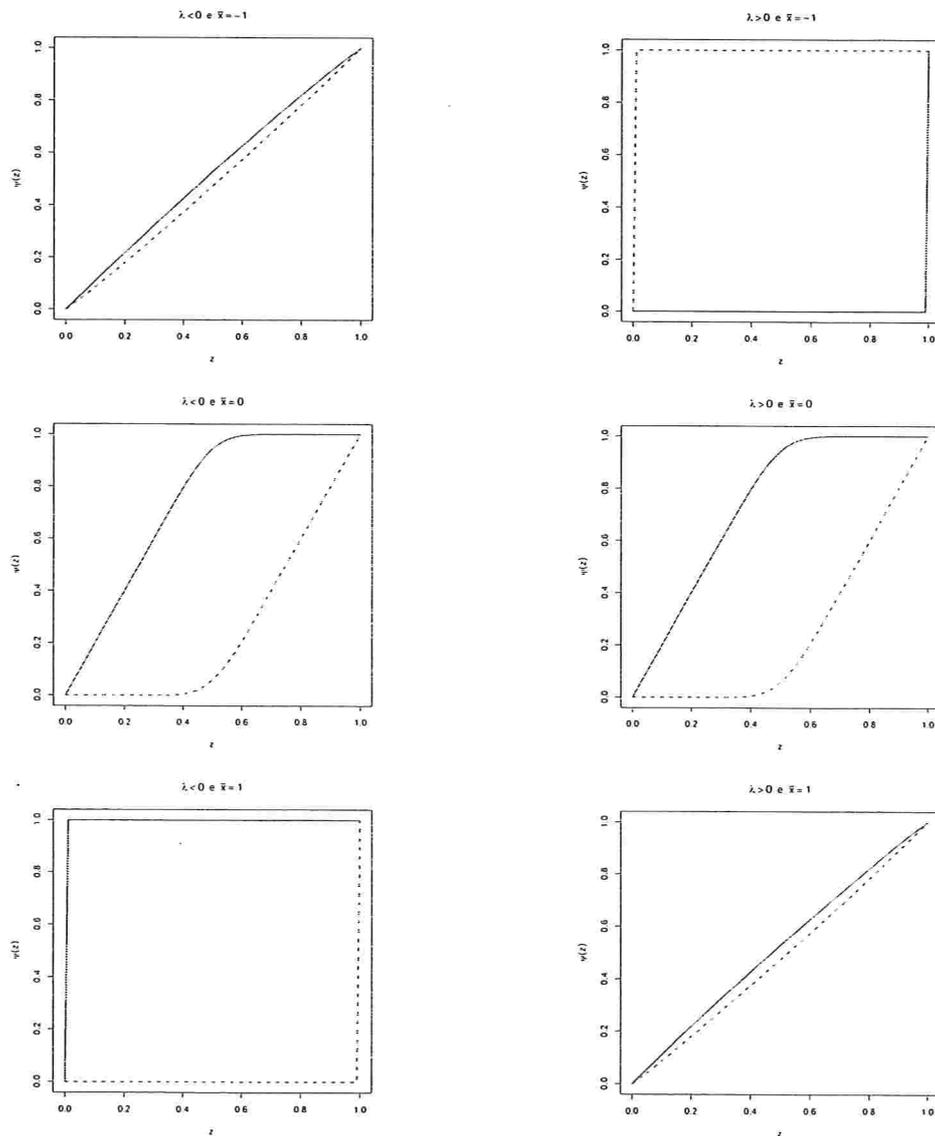


Figura 5.3: Função de concentração para a classe de distribuições *a posteriori*, fixado $\bar{x} = -1, 0$ e 1 . As linhas (—) e (---) correspondem, respectivamente, ao $\hat{\varphi}(z)$ e ao $1 - \hat{\varphi}(1 - z)$.

que, se \bar{x} e λ têm o mesmo sinal, quanto maior a magnitude de \bar{x} , mais próximo será $\hat{\varphi}(z)$ de $1 - \hat{\varphi}(1 - z)$.

As conclusões aqui apresentadas confirmam a teoria e as análises discutidas nos Capítulos 3 e 4.

O estudo de sensibilidade será realizado apenas para o caso em que τ^2 é conhecido. Espera-se obter conclusões similares para o caso em que τ^2 é desconhecido.

Capítulo 6

Comentários finais

O objetivo principal deste trabalho foi realizar uma análise de sensibilidade bayesiana que nos permita entender o quanto a distribuição *a posteriori* é influenciada pela escolha da distribuição *a priori*. Para isso, discutimos uma nova maneira de contaminar a distribuição *a priori*, a qual nos referimos como contaminação multiplicativa.

Para particulares especificações, vimos que a classe de contaminação multiplicativa coincide com classes de distribuições assimétricas conhecidas na literatura.

No Capítulo 3, apresentamos alguns resultados teóricos obtidos quando consideramos a classe de contaminação multiplicativa normal-assimétrica para a distribuição *a priori* do parâmetro de posição. A estrutura desta classe assume $w(\theta) = c G(\lambda \theta)$, em que a distribuição *a priori* base é $N(\mu, \sigma^2)$, a função contaminadora é dada por $G(\lambda \theta) = \Phi_1(\lambda \theta; \lambda \mu, \sigma^2)$ e $c = 2$. Ou seja, temos que a classe de distribuições *a priori* para θ contém um número infinito de diferentes distribuições normais-assimétricas e a distribuição normal. Nesse contexto, os seguintes casos foram estudados:

- I. Amostra aleatória de uma distribuição $N(\theta, \tau^2)$, com τ^2 conhecido.
- II. Amostra aleatória de uma distribuição $N(\theta, \tau^2)$, com ambos os parâmetros desconhecidos. Para a distribuição *a priori* do parâmetro de escala, utilizamos duas propostas, sendo elas: $h(\tau^2) \propto \frac{1}{\tau^2}$ e $\tau^2 \sim GI(a, b)$, com $a > 0$ e $b > 0$.

Em ambas as situações, procuramos realizar um estudo de sensibilidade analisando as estimativas pontuais da distribuição *a posteriori*, tais como a média e a variância *a posteriori*.

Na situação I, obtivemos resultados de conjugação para o parâmetro de posição do modelo normal, expressões para a média e a variância *a posteriori*. O estudo de sensibilidade foi baseado na análise de medidas de distância *a posteriori* entre as médias *a posteriori* geradas por Γ_M e a média *a posteriori* obtida quando se considera a distribuição *a priori* base. Análise similar foi realizada para as variâncias *a posteriori*. Em ambos os casos, conseguimos delimitar em quais situações pode ser alcançada a robustez. Vimos que, em geral, ocorre a robustez da média e da variância *a posteriori* com respeito a variação de distribuições *a priori* em Γ_M , quando λ e $\bar{x} - \mu$ têm o mesmo sinal.

Na situação II, determinamos as distribuições condicionais completas, a distribuição *a posteriori* marginal para θ e expressões para a média e variância *a posteriori* marginal de τ^2 . Para obter estimativas da média e da variância *a posteriori* das distribuições *a posteriori* marginais de θ e τ^2 , fizemos uso de metodologias de integração numérica via método de quadratura gaussiana do tipo Gauss-Hermite.

Aqui, exemplificamos a situação II através de um estudo de simulação em que consideramos amostras aleatórias da distribuição normal com tamanhos $n = 20, 50, 100$ e 200 , distribuição *a priori* do parâmetro de escala $h(\tau^2) \propto \frac{1}{\tau^2}$ e para a distribuição *a priori* do parâmetro de posição a classe de contaminação multiplicativa, com a distribuição *a priori* base normal padrão e $w(\theta) = 2\Phi(\lambda\theta)$. Vimos que, de maneira geral, para tamanhos de amostra suficientemente grandes ($n \geq 100$) e variabilidade amostral moderada ($S^2 \leq 20$), quando o sinal da média amostral é igual ao sinal do parâmetro de assimetria considerado para a classe de contaminação multiplicativa *a priori* Γ_M , ocorre a robustez da média e da variância *a posteriori* marginal dos parâmetros de posição e escala do modelo normal, com respeito a diferentes escolhas da distribuição *a priori* em Γ_M .

No Capítulo 4, generalizamos os resultados apresentados no Capítulo 3 ao discutir questões de sensibilidade bayesiana no contexto de modelos de regressão linear normal com a variância dos erros aleatórios (τ^2) conhecida e desconhecida. Resultados fechados similares aos do Capítulo 3 foram obtidos quando o modelo linear normal possui apenas um parâmetro regressor. Exploramos esta situação através de uma aplicação, em que a classe de contaminação multiplicativa *a priori* definida em (4.5) é dada por $f_0(\beta) = \phi(\beta; 0, 10^4)$ e $w(\beta) = 2 \Phi(\lambda\beta; 0, 10^4)$. Vimos que, se τ^2 é conhecido, a robustez das estimativas de média e variância *a posteriori*

ocorrem no espaço em que λ tem o mesmo sinal que o da estimativa de máxima verossimilhança do parâmetro regressor. Se τ^2 é desconhecido, conclusões similares são obtidas quando a especificação *a priori* do parâmetro de escala do modelo de regressão linear normal é $h(\tau^2) \propto \frac{1}{\tau^2}$ ou $\tau^2 \sim GI(1/10000, 1/10000)$.

De maneira geral, nos Capítulos 3 e 4, analisamos o intervalo de variação de algumas medidas *a posteriori*, quando se considera a classe de contaminação multiplicativa *a priori*. Tais medidas descrevem a variação da distribuição *a posteriori* em apenas um aspecto, a partir da utilização de uma estatística pontual *a posteriori*.

No Capítulo 5, mudamos este enfoque ao estudar a sensibilidade da distribuição *a posteriori* através de uma medida de divergência global (função de concentração) que nos permite avaliar discrepâncias entre a distribuição *a posteriori* decorrente da distribuição *a priori* base ($f_0(\theta)$) e as distribuições *a posteriori* geradas pelas distribuições *a priori* contidas em Γ_M . Neste contexto, apresentamos uma expressão mais concisa para a definição de função de concentração ao considerar medidas de probabilidade e suas respectivas funções de densidade. Obtivemos uma expressão fechada para a função de concentração entre $N(\mu, \sigma^2)$ e $SN(\mu, \sigma^2, \lambda)$ e mostramos que, neste caso, a função de concentração é invariante a mudanças de posição e escala. Do ponto de vista de análise de robustez, obtivemos expressões para a função de concentração com respeito a medidas de probabilidade *a posteriori* do parâmetro de posição, quando os dados têm distribuição $N(\theta, \tau^2)$, com τ^2 ora conhecido, ora desconhecido. Aqui, a distribuição *a priori* para θ pertence a Γ_M e, se τ^2 é desconhecido, consideramos dois contextos: $h(\tau^2) \propto \frac{1}{\tau^2}$ ou $\tau^2 \sim GI(a, b)$, com $a, b > 0$. Através da função de concentração, vimos que, quando τ^2 é conhecido, não só a média e a variância *a posteriori*, como também as medidas de probabilidade *a posteriori*, são robustas com respeito a Γ_M , quando $\bar{x} - \mu$ e λ têm o mesmo sinal.

A seguir, apresentamos algumas possibilidades de futuras pesquisas relacionadas ao estudo de robustez bayesiana:

- Fazer um estudo de simulação para a função de concentração, quando o modelo normal possui a variância desconhecida, através das Proposições 5.3.2 e 5.3.3.
- Explorar questões de robustez bayesiana com respeito a funções de concentração no contexto de modelos de regressão linear normal.
- Estender os resultados obtidos neste trabalho para modelos amostrais mais

complexos, como, por exemplo, o modelo normal-assimétrico.

- Estudar a influência da classe de contaminação multiplicativa normal-assimétrica para a distribuição *a priori* do parâmetro regressor de modelos mais gerais, como, por exemplo, o modelo de regressão linear simples t-Student.
- Considerar especificações mais gerais para a função contaminadora $w(\cdot)$ como, por exemplo, $w(\theta) = c Q$, onde c e Q corresponderiam, respectivamente, a constante normalizadora da distribuição *a priori* e a função de distribuição acumulada de distribuições simétricas.
- Estudar o comportamento das medidas de distância *a posteriori* relacionadas a classe de contaminação multiplicativa *a priori*, quando consideramos dados amostrais que contém valores aberrantes.
- Aplicar a metodologia de Tipo II de máxima-verossimilhança (MV-II) no contexto considerado nos Capítulos 3 e 4.
- Refazer o estudo realizado neste trabalho considerando outras funções de perda.

Apêndice A

Demonstrações

A.1 Capítulo 3

Prova da Proposição 3.1.1:

Para a prova do item (i), notamos que

$$f(\mathbf{x} | \theta) = \left(\frac{1}{\tau\sqrt{2\pi}} \right)^n e^{-\frac{\sum_{i=1}^n (x_i - \theta)^2}{2\tau^2}}. \quad (\text{A.1})$$

Ao considerar $\theta \sim SN(\mu, \sigma^2, \lambda)$ como distribuição *a priori* do parâmetro de posição e a função de verossimilhança especificada em (A.1), temos como distribuição *a posteriori* para o parâmetro θ

$$\begin{aligned} f(\theta | \mathbf{x}) &\propto f(\mathbf{x} | \theta) f(\theta) \\ &= \left(\frac{1}{\tau\sqrt{2\pi}} \right)^n e^{-\frac{\sum_{i=1}^n (x_i - \theta)^2}{2\tau^2}} \frac{2}{\sigma} \phi\left(\frac{\theta - \mu}{\sigma}\right) \Phi\left(\lambda \frac{\theta - \mu}{\sigma}\right) \\ &\propto e^{-\frac{\sum_{i=1}^n (x_i - \theta)^2}{2\tau^2} - \frac{(\theta - \mu)^2}{2\sigma^2}} \Phi\left(\lambda \frac{\theta - \mu}{\sigma}\right). \end{aligned} \quad (\text{A.2})$$

Ao assumir $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$, note que

$$\begin{aligned}
e^{-\frac{\sum_{i=1}^n (x_i - \theta)^2}{2\tau^2} - \frac{(\theta - \mu)^2}{2\sigma^2}} &= e^{-\frac{1}{2} \left\{ \frac{\sum_{i=1}^n (x_i^2 - 2x_i\theta + \theta^2)}{\tau^2} + \frac{(\theta^2 - 2\theta\mu + \mu^2)}{\sigma^2} \right\}} \\
&= e^{-\frac{1}{2} \left\{ \frac{\sum_{i=1}^n x_i^2}{\tau^2} + \frac{\mu^2}{\sigma^2} \right\}} e^{-\frac{1}{2} \left\{ -\frac{2n\bar{x}\theta + n\theta^2}{\tau^2} + \frac{\theta^2 - 2\theta\mu}{\sigma^2} \right\}} \\
&= e^{-\frac{1}{2} \left\{ \frac{\sum_{i=1}^n x_i^2}{\tau^2} + \frac{\mu^2}{\sigma^2} \right\}} e^{-\frac{1}{2} \left\{ \theta^2 \left[\frac{n}{\tau^2} + \frac{1}{\sigma^2} \right] - 2\theta \left[\frac{n\bar{x}}{\tau^2} + \frac{\mu}{\sigma^2} \right] \right\}} \\
&= e^{-\frac{1}{2} \left\{ \frac{\sum_{i=1}^n x_i^2}{\tau^2} + \frac{\mu^2}{\sigma^2} \right\}} e^{-\frac{1}{2} \left\{ \left[\theta \sqrt{\frac{n}{\tau^2} + \frac{1}{\sigma^2}} - \frac{\left(\frac{n\bar{x}}{\tau^2} + \frac{\mu}{\sigma^2} \right)}{\sqrt{\frac{n}{\tau^2} + \frac{1}{\sigma^2}}} \right]^2 - \frac{\left(\frac{n\bar{x}}{\tau^2} + \frac{\mu}{\sigma^2} \right)^2}{\frac{n}{\tau^2} + \frac{1}{\sigma^2}} \right\}} \\
&= e^{-\frac{1}{2} \left\{ \frac{\sum_{i=1}^n x_i^2}{\tau^2} + \frac{\mu^2}{\sigma^2} \right\}} e^{\frac{1}{2} \frac{\left(\frac{n\bar{x}}{\tau^2} + \frac{\mu}{\sigma^2} \right)^2}{\frac{n}{\tau^2} + \frac{1}{\sigma^2}}} e^{-\frac{1}{2} \left\{ \sqrt{\frac{n}{\tau^2} + \frac{1}{\sigma^2}} \left[\theta - \frac{\frac{n\bar{x}}{\tau^2} + \frac{\mu}{\sigma^2}}{\sqrt{\frac{n}{\tau^2} + \frac{1}{\sigma^2}}} \right] \right\}^2} \\
&= e^{-\frac{1}{2} \left\{ \frac{\sum_{i=1}^n x_i^2}{\tau^2} + \frac{\mu^2}{\sigma^2} \right\}} e^{\frac{1}{2} \frac{\left(\frac{n\bar{x}}{\tau^2} + \frac{\mu}{\sigma^2} \right)^2}{\frac{n}{\tau^2} + \frac{1}{\sigma^2}}} e^{-\frac{1}{2} \frac{1}{\frac{n}{\sigma^2} + \tau^2} \left[\theta - \frac{n\bar{x}\sigma^2 + \mu\tau^2}{n\sigma^2 + \tau^2} \right]^2}. \quad (A.3)
\end{aligned}$$

Portanto,

$$\begin{aligned}
f(\theta | \mathbf{x}) &\propto e^{-\frac{1}{2} \frac{1}{\frac{n}{\sigma^2} + \tau^2} \left[\theta - \frac{n\bar{x}\sigma^2 + \mu\tau^2}{n\sigma^2 + \tau^2} \right]^2} \Phi \left(\lambda \frac{\theta - \mu}{\sigma} \right) \\
&\propto \phi \left(\theta; \frac{n\bar{x}\sigma^2 + \mu\tau^2}{n\sigma^2 + \tau^2}, \frac{\tau^2 \sigma^2}{n\sigma^2 + \tau^2} \right) \Phi \left(\lambda \frac{\theta - \mu}{\sigma} \right), \quad (A.4)
\end{aligned}$$

que é o núcleo de uma distribuição normal-assimétrica multivariada geral. Dessa forma, conseguimos mostrar que a distribuição *a posteriori* para o parâmetro de posição θ é $GSN_{1,1} \left(\frac{n\bar{x}\sigma^2 + \mu\tau^2}{n\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2}, \lambda, \lambda\mu, \sigma^2 \right)$.

Para demonstrarmos os itens (ii) e (iii), vamos utilizar a função geradora de momentos da distribuição normal-assimétrica multivariada geral expressa em (2.9). Para reduzirmos a notação, seja $K = \Phi[\lambda a; \lambda\mu, \sigma^2 + \lambda^2 \Sigma]$, onde $a = \frac{n\bar{x}\sigma^2 + \mu\tau^2}{n\sigma^2 + \tau^2}$ e $\Sigma = \frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2}$, e, assim, a função geradora de momentos de $\theta | \mathbf{X}$ é dada por:

$$\begin{aligned}
M_{\theta|\mathbf{x}}(t) &= \frac{\Phi[\lambda(a + \Sigma t); \lambda\mu, \sigma^2 + \lambda^2 \Sigma]}{K} e^{a t + \frac{1}{2} \Sigma t^2} \\
&= \frac{1}{K} \Phi \left[\frac{\lambda(a + \Sigma t) - \lambda\mu}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right] e^{a t + \frac{1}{2} \Sigma t^2}. \quad (A.5)
\end{aligned}$$

Sabemos que $E_\lambda[\theta | \mathbf{x}] = \frac{\partial}{\partial t} M_{\theta|\mathbf{x}}(t) \Big|_{t=0}$ e $E_\lambda[\theta^2 | \mathbf{x}] = \frac{\partial^2}{\partial t^2} M_{\theta|\mathbf{x}}(t) \Big|_{t=0}$ então

$$\begin{aligned}
 E_\lambda[\theta | \mathbf{x}] &= \frac{1}{K} \left\{ e^{a t + \frac{1}{2} \Sigma t^2} (a + \Sigma t) \Phi \left[\frac{\lambda(a + \Sigma t) - \lambda \mu}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right] + e^{a t + \frac{1}{2} \Sigma t^2} \times \right. \\
 &\quad \left. \times \phi \left[\frac{\lambda(a + \Sigma t) - \lambda \mu}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right] \frac{\lambda}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \Sigma \right\} \Big|_{t=0} \\
 &= \frac{1}{K} \left\{ a \Phi \left[\frac{\lambda(a - \mu)}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right] + \phi \left[\frac{\lambda(a - \mu)}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right] \frac{\lambda}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \Sigma \right\} \quad (\text{A.6})
 \end{aligned}$$

e

$$\begin{aligned}
 E_\lambda[\theta^2 | \mathbf{x}] &= \frac{1}{K} \frac{\partial}{\partial t} \left\{ e^{a t + \frac{1}{2} \Sigma t^2} (a + \Sigma t) \Phi \left[\frac{\lambda(a + \Sigma t) - \lambda \mu}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right] + e^{a t + \frac{1}{2} \Sigma t^2} \times \right. \\
 &\quad \left. \times \phi \left[\frac{\lambda(a + \Sigma t) - \lambda \mu}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right] \frac{\lambda}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \Sigma \right\} \Big|_{t=0} \\
 &= \frac{1}{K} \left\{ \left[e^{a t + \frac{1}{2} \Sigma t^2} (a + \Sigma t)^2 + e^{a t + \frac{1}{2} \Sigma t^2} \Sigma \right] \Phi \left[\frac{\lambda(a + \Sigma t) - \lambda \mu}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right] + \right. \\
 &\quad \left. + 2 e^{a t + \frac{1}{2} \Sigma t^2} (a + \Sigma t) \phi \left[\frac{\lambda(a + \Sigma t) - \lambda \mu}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right] \frac{\lambda \Sigma}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} + \right. \\
 &\quad \left. + e^{a t + \frac{1}{2} \Sigma t^2} \frac{\lambda \Sigma}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \frac{\partial}{\partial t} \phi \left[\frac{\lambda(a + \Sigma t) - \lambda \mu}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right] \right\} \Big|_{t=0}. \quad (\text{A.7})
 \end{aligned}$$

Note que

$$\frac{\partial}{\partial t} \phi \left[\frac{\lambda(a + \Sigma t) - \lambda \mu}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right] = \phi \left[\frac{\lambda(a + \Sigma t) - \lambda \mu}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right] \left\{ -\frac{\lambda^2 \Sigma [a + \Sigma t - \mu]}{\sigma^2 + \lambda^2 \Sigma} \right\}. \quad (\text{A.8})$$

Substituindo (A.8) em (A.7) e fazendo $t = 0$, temos que

$$\begin{aligned}
 E_\lambda[\theta^2 | \mathbf{x}] &= \frac{1}{K} \left\{ (a^2 + \Sigma) \Phi \left[\frac{\lambda(a - \mu)}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right] + 2 a \phi \left[\frac{\lambda(a - \mu)}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right] \frac{\lambda \Sigma}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} + \right. \\
 &\quad \left. + \phi \left[\frac{\lambda(a - \mu)}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right] \left[-\frac{\lambda^2 \Sigma (a - \mu)}{\sigma^2 + \lambda^2 \Sigma} \right] \frac{\lambda \Sigma}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right\}. \quad (\text{A.9})
 \end{aligned}$$

Como $K = \Phi[\lambda a; \lambda \mu, \sigma^2 + \lambda^2 \Sigma]$, então temos em (A.6) e (A.9) que

$$E_\lambda[\theta | \mathbf{x}] = a + \frac{\phi \left[\frac{\lambda(a - \mu)}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right]}{\Phi \left[\frac{\lambda(a - \mu)}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right]} \frac{\lambda}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \Sigma \quad \text{e} \quad (\text{A.10})$$

$$E_\lambda[\theta^2 | \mathbf{x}] = a^2 + \Sigma + \frac{\phi \left[\frac{\lambda(a - \mu)}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right]}{\Phi \left[\frac{\lambda(a - \mu)}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right]} \frac{\lambda \Sigma}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \left[2 a - \frac{\lambda^2 \Sigma (a - \mu)}{\sigma^2 + \lambda^2 \Sigma} \right], \quad (\text{A.11})$$

implicando em

$$\begin{aligned} \text{Var}_\lambda[\theta | \mathbf{x}] &= E_\lambda[\theta^2 | \mathbf{x}] - E_\lambda[\theta | \mathbf{x}]^2 \\ &= \Sigma - \frac{\lambda^2 \Sigma^2}{\sigma^2 + \lambda^2 \Sigma} \frac{\phi \left[\frac{\lambda(a-\mu)}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right]}{\Phi \left[\frac{\lambda(a-\mu)}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right]} \left\{ \frac{\lambda(a-\mu)}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} + \frac{\phi \left[\frac{\lambda(a-\mu)}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right]}{\Phi \left[\frac{\lambda(a-\mu)}{\sqrt{\sigma^2 + \lambda^2 \Sigma}} \right]} \right\} \end{aligned} \quad (\text{A.12})$$

Sabemos que $a = \frac{n\bar{x}\sigma^2 + \mu\tau^2}{n\sigma^2 + \tau^2}$ e $\Sigma = \frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2}$ então

$$\begin{aligned} \frac{\lambda}{\sqrt{\sigma^2 + \lambda^2 \Sigma}}(a - \mu) &= \frac{\lambda}{\sqrt{\sigma^2 + \lambda^2 \frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2}}} \left(\frac{n\bar{x}\sigma^2 + \mu\tau^2}{n\sigma^2 + \tau^2} - \mu \right) \\ &= \frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1 + \lambda^2)}} \left[\frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \right] \end{aligned} \quad (\text{A.13})$$

e

$$\frac{\Sigma^2}{\sigma^2 + \lambda^2 \Sigma} = \frac{\left(\frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2} \right)^2}{\sigma^2 + \lambda^2 \frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2}} = \frac{\tau^4\sigma^2}{[n\sigma^2 + \tau^2][n\sigma^2 + \tau^2(1 + \lambda^2)]}. \quad (\text{A.14})$$

Substituindo a , Σ , (A.13) e (A.14) em (A.10) e (A.12), obtemos as expressões fechadas mencionadas anteriormente para a esperança e a variância da distribuição *a posteriori* do parâmetro de posição θ . ◆

A seguir apresentamos os Lemas A.1.1, A.1.2 e A.1.3 que serão fundamentais nas demonstrações de diversas proposições.

Lema A.1.1 A função $t + \frac{\phi(t)}{\Phi(t)}$ é positiva em $t \in \mathbb{R}$.

Prova:

Sabemos que $\phi(t) > 0$ e $\Phi(t) > 0$, $\forall t \in \mathbb{R}$ e, dessa forma, quando $t \geq 0$, temos que $\left[t + \frac{\phi(t)}{\Phi(t)} \right] > 0$.

Devemos estudar agora o caso em que $t < 0$. Sampford (1953) cita que, $\forall u > 0$, a razão¹ $\frac{1-\Phi(u)}{\phi(u)} < \frac{1}{u}$. Esta desigualdade é equivalente a

$$\frac{\phi(u)}{1 - \Phi(u)} > u \iff \frac{\phi(u)}{1 - \Phi(u)} - u > 0, \quad \forall u > 0. \quad (\text{A.15})$$

¹Esta razão é conhecida na literatura como razão de Mill (*Mill's ratio*).

Seja $u = -t$, onde $t < 0$. Então, considerando $\phi(t) = \phi(-t) = \phi(u)$, $\Phi(t) = 1 - \Phi(-t) = 1 - \Phi(u)$ e a desigualdade dada em (A.15), temos que, $\forall t < 0$,

$$t + \frac{\phi(t)}{\Phi(t)} = t + \frac{\phi(-t)}{\Phi(t)} = -u + \frac{\phi(u)}{1 - \Phi(u)} > 0. \quad (\text{A.16})$$

Logo, mostramos que $\forall t \in \mathbb{R}$, $t + \frac{\phi(t)}{\Phi(t)} > 0$. ♦

Lema A.1.2 A função $h(\lambda) = \frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1 + \lambda^2)}}$ é estritamente crescente em λ , limitada entre $-\frac{1}{\tau}$ e $\frac{1}{\tau}$.

Prova:

Ao derivar a função $h(\lambda)$ com respeito a λ , obtemos

$$\begin{aligned} \frac{\partial}{\partial \lambda} h(\lambda) &= \frac{1}{\sqrt{n\sigma^2 + \tau^2(1 + \lambda^2)}} - \frac{\lambda^2 \tau^2}{\sqrt{[n\sigma^2 + \tau^2(1 + \lambda^2)]^3}} \\ &= \frac{n\sigma^2 + \tau^2}{[n\sigma^2 + \tau^2(1 + \lambda^2)]^{\frac{3}{2}}} > 0, \quad \forall \lambda \in \mathbb{R}. \end{aligned} \quad (\text{A.17})$$

Portanto, a função $h(\lambda) = \frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1 + \lambda^2)}}$ é estritamente crescente em λ .

Além disso,

$$\lim_{\lambda \rightarrow \infty} h(\lambda) = \lim_{\lambda \rightarrow \infty} \sqrt{\frac{\lambda^2}{n\sigma^2 + \tau^2(1 + \lambda^2)}} = \lim_{\lambda \rightarrow \infty} \sqrt{\frac{1}{\frac{n\sigma^2 + \tau^2}{\lambda^2} + \tau^2}} = \frac{1}{\tau}. \quad (\text{A.18})$$

Analogamente, $\lim_{\lambda \rightarrow -\infty} h(\lambda) = -\frac{1}{\tau}$. ♦

Lema A.1.3 A função $Z(t) = \frac{\phi(t)}{\Phi(t)}$ é estritamente decrescente em $t \in \mathbb{R}$.

Prova:

Considere

$$\frac{\partial}{\partial t} Z(t) = \frac{-t \phi(t) \Phi(t) - \phi(t)^2}{\Phi(t)^2} = -\frac{\phi(t)}{\Phi(t)} \left[t + \frac{\phi(t)}{\Phi(t)} \right]. \quad (\text{A.19})$$

Sabemos que $\forall t \in \mathbb{R}$, $\phi(t) > 0$ e $\Phi(t) > 0$. Além disso, pelo Lema A.1.1, temos que $t + \frac{\phi(t)}{\Phi(t)} > 0$, $\forall t$ e, então, $\frac{\partial}{\partial t} Z(t) < 0$, $\forall t \in \mathbb{R}$.

Portanto, $\forall t \in \mathbb{R}$, temos que $\frac{\partial}{\partial t} Z(t) < 0$, o que implica que a função $Z(t)$ é estritamente decrescente em \mathbb{R} . ♦

Prova do Corolário 3.1.1:

Considere $y = \frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1 + \lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}}$, tal que $-\infty < y < \infty$. Reescrevendo a variância *a posteriori* de θ em função de y temos que

$$\begin{aligned} \text{Var}[\theta \mid \mathbf{x}] &= \frac{\sigma^2 \tau^2}{n\sigma^2 + \tau^2} - \frac{\lambda^2 \tau^2}{n\sigma^2 + \tau^2(1 + \lambda^2)} \frac{\sigma^2 \tau^2}{n\sigma^2 + \tau^2} \frac{\phi(y)}{\Phi(y)} \left[y + \frac{\phi(y)}{\Phi(y)} \right] \\ &= \frac{\sigma^2 \tau^2}{n\sigma^2 + \tau^2} \left\{ 1 - \frac{\lambda^2 \tau^2}{n\sigma^2 + \tau^2(1 + \lambda^2)} \frac{\phi(y)}{\Phi(y)} \left[y + \frac{\phi(y)}{\Phi(y)} \right] \right\}. \end{aligned} \quad (\text{A.20})$$

Note que a expressão $\frac{\lambda^2 \tau^2}{n\sigma^2 + \tau^2(1 + \lambda^2)}$ gera valores entre zero e um, pois $0 \leq \lambda^2 \tau^2 < n\sigma^2 + \tau^2(1 + \lambda^2)$, $\forall n \geq 1$. Então, para mostrar que $0 < \text{Var}[\theta \mid \mathbf{x}] \leq \frac{\sigma^2 \tau^2}{n\sigma^2 + \tau^2}$, basta mostrar que $0 < \frac{\phi(y)}{\Phi(y)} \left[y + \frac{\phi(y)}{\Phi(y)} \right] < 1$, $\forall y$.

Dessa forma, vamos estudar a função

$$G(y) = \frac{\phi(y)}{\Phi(y)} \left[y + \frac{\phi(y)}{\Phi(y)} \right] = \frac{\phi(y)}{\Phi(y)} y + \frac{\phi(y)^2}{\Phi(y)^2}. \quad (\text{A.21})$$

Seja U uma variável aleatória com distribuição normal padrão, truncada de tal forma que $-\infty < u \leq y$. Então, segundo Johnson, Kotz & Balakrishnan (1994), a função de densidade de U é dada por

$$f_U(u) = \frac{\phi(u)}{\Phi(y)} \quad (\text{A.22})$$

e a expressão de sua variância é

$$\text{Var}(U) = 1 - \frac{\phi(y)}{\Phi(y)} y - \frac{\phi(y)^2}{\Phi(y)^2}. \quad (\text{A.23})$$

Sabemos que $\text{Var}(U) > 0$, então

$$1 - \frac{\phi(y)}{\Phi(y)} y - \frac{\phi(y)^2}{\Phi(y)^2} > 0 \implies \frac{\phi(y)}{\Phi(y)} y + \frac{\phi(y)^2}{\Phi(y)^2} < 1. \quad (\text{A.24})$$

Além disso, $\forall t \in \mathbb{R}$, $\phi(y) > 0$, $\Phi(y) > 0$ e, pelo Lema A.1.1, $y + \frac{\phi(y)}{\Phi(y)} > 0$ então

$$\frac{\phi(y)}{\Phi(y)} y + \frac{\phi(y)^2}{\Phi(y)^2} > 0, \quad \forall y. \quad (\text{A.25})$$

Portanto, por (A.24) e (A.25) temos que $0 < G(y) < 1$. Concluimos então que $0 < Var[\theta | \mathbf{x}] \leq \frac{\sigma^2 \tau^2}{n\sigma^2 + \tau^2}$. \blacklozenge

Prova da Proposição 3.1.2:

Pelos resultados dados em (3.9), sabemos que $E_0[\theta | \mathbf{x}] = \frac{n\bar{x}\sigma^2 + \mu\tau^2}{n\sigma^2 + \tau^2}$ e $Var_0[\theta | \mathbf{x}] = \frac{\sigma^2 \tau^2}{n\sigma^2 + \tau^2}$. Além disso, considere as medidas pontuais $E_\lambda[\theta | \mathbf{x}]$ e $Var_\lambda[\theta | \mathbf{x}]$ como definidas em (3.7) e (3.8).

Manipulando as expressões de $E_\lambda[\theta | \mathbf{x}]$, $E_0[\theta | \mathbf{x}]$, $Var_\lambda[\theta | \mathbf{x}]$ e $Var_0[\theta | \mathbf{x}]$ segundo as expressões de d_M e d_V , fica fácil ver que

$$d_M = \left| E_\lambda[\theta | \mathbf{x}] - \hat{\theta}_0 \right| = \frac{\phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]} \frac{|\lambda|}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma \tau^2}{\sqrt{n\sigma^2 + \tau^2}}$$

e

$$d_V = \left| Var_\lambda[\theta | \mathbf{x}] - \hat{\tau}_0 \right| = \frac{\lambda^2 \tau^2}{n\sigma^2 + \tau^2(1+\lambda^2)} \frac{\sigma^2 \tau^2}{n\sigma^2 + \tau^2} \frac{\phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]} \times$$

$$\times \left| \frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} + \frac{\phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]} \right|. \quad (\text{A.26})$$

Para obtermos a expressão da distância d_V como apresentada na Proposição 3.1.2, basta considerar $y = \frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}}$ e lembrar que, pelo Lema A.1.1, $K(y) = y + \frac{\phi(y)}{\Phi(y)} > 0$, $\forall y \in \mathbb{R}$. Dessa forma, $|K(y)| = K(y)$, $\forall y \in \mathbb{R}$ e fica demonstrada a expressão da distância d_V .

É imediato que quando $\lambda = 0$ ambas as distâncias d_M e d_V serão iguais a zero. \blacklozenge

Prova da Proposição 3.1.3:

Parte (i):

Vamos estudar a distância d_M em função de λ no caso em que $\lambda \leq 0$ e $\bar{x} > \mu$ através da função

$$G(\lambda) = \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} [-h(\lambda)] b, \quad (\text{A.27})$$

onde $h(\lambda) = \frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1+\lambda^2)}}$, $a = \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} > 0$ e $b = \frac{\sigma\tau^2}{\sqrt{n\sigma^2 + \tau^2}} > 0$ pois $n, \sigma, \tau > 0$ e $\bar{x} > \mu$.

Como $\lambda \leq 0$, considere então $\lambda_1 < \lambda_2 \leq 0$. Segundo o Lema A.1.2, a função $h(\lambda)$ é estritamente crescente em λ e, como $a > 0$ e $b > 0$, então

$$h(\lambda_1) a < h(\lambda_2) a \leq 0 \quad (\text{A.28})$$

e

$$h(\lambda_1) b < h(\lambda_2) b \leq 0 \implies 0 \leq -h(\lambda_2) b < -h(\lambda_1) b. \quad (\text{A.29})$$

No Lema A.1.3 foi mostrado que $\frac{\phi(y)}{\Phi(y)}$ é uma função estritamente decrescente, que gera apenas valores positivos, $\forall y \in \mathbb{R}$. Logo, por (A.28),

$$0 < \frac{\phi(h(\lambda_2) a)}{\Phi(h(\lambda_2) a)} < \frac{\phi(h(\lambda_1) a)}{\Phi(h(\lambda_1) a)}. \quad (\text{A.30})$$

Então

$$\begin{aligned} G(\lambda_2) = \frac{\phi[h(\lambda_2) a]}{\Phi[h(\lambda_2) a]} [-h(\lambda_2)] b &< \frac{\phi[h(\lambda_1) a]}{\Phi[h(\lambda_1) a]} [-h(\lambda_2)] b \\ &< \frac{\phi[h(\lambda_1) a]}{\Phi[h(\lambda_1) a]} [-h(\lambda_1)] b = G(\lambda_1). \end{aligned} \quad (\text{A.31})$$

Portanto, $G(\lambda)$ é uma função estritamente decrescente em λ , quando $\lambda \leq 0$ e $\bar{x} > \mu$.

Note que a função $G(\lambda)$ é equivalente a distância d_M , quando substituimos $h(\lambda)$, a e b por suas respectivas expressões. Concluímos então que d_M é uma função estritamente decrescente em λ , quando $\lambda \leq 0$ e $\bar{x} > \mu$.

Para o estudo da distância d_V , considere inicialmente a função auxiliar

$$T(z) = \frac{\phi(z)}{\Phi(z)} z^2 \left[z + \frac{\phi(z)}{\Phi(z)} \right] = \frac{\phi(z)}{\Phi(z)} z^3 + \frac{\phi(z)^2}{\Phi(z)^2} z^2 \quad (\text{A.32})$$

e mostremos que ela é uma função estritamente decrescente em $z \leq 0$.

Segue então que

$$\begin{aligned}
 \frac{\partial}{\partial z} T(z) &= \left[-z \frac{\phi(z)}{\Phi(z)} - \frac{\phi(z)^2}{\Phi(z)^2} \right] z^3 + 3 z^2 \frac{\phi(z)}{\Phi(z)} + 2 z \frac{\phi(z)^2}{\Phi(z)^2} + \\
 &\quad + 2 \frac{\phi(z)}{\Phi(z)} \left[-z \frac{\phi(z)}{\Phi(z)} - \frac{\phi(z)^2}{\Phi(z)^2} \right] z^2 \\
 &= \frac{\phi(z)}{\Phi(z)} \left[-z^4 - 3 z^3 \frac{\phi(z)}{\Phi(z)} + 3 z^2 + 2 z \frac{\phi(z)}{\Phi(z)} - 2 z^2 \frac{\phi(z)^2}{\Phi(z)^2} \right] \\
 &= M(z).
 \end{aligned} \tag{A.33}$$

Segundo Sampford (1953), $\forall x$ finito,

$$\nu(x) \{ [\nu(x) - x] [2 \nu(x) - x] - 1 \} > 0, \tag{A.34}$$

em que

$$\nu(x) = \frac{1}{R_x} \quad \text{e} \quad R_x = \frac{1 - \Phi(x)}{\phi(x)} = \frac{\Phi(-x)}{\phi(x)} = \frac{\Phi(-x)}{\phi(-x)}. \tag{A.35}$$

Seja $-x = z$. Então por (A.35)

$$R_{-z} = \frac{\Phi(z)}{\phi(z)} \quad \text{e} \quad \nu(-z) = \frac{\phi(z)}{\Phi(z)}. \tag{A.36}$$

Substituindo (A.36) em (A.34), temos que

$$\begin{aligned}
 \frac{\phi(z)}{\Phi(z)} \left\{ \left[\frac{\phi(z)}{\Phi(z)} + z \right] \left[2 \frac{\phi(z)}{\Phi(z)} + z \right] - 1 \right\} &> 0 \iff \\
 \frac{\phi(z)}{\Phi(z)} \left\{ 2 \frac{\phi(z)^2}{\Phi(z)^2} + 3 z \frac{\phi(z)}{\Phi(z)} + z^2 - 1 \right\} &> 0 \iff \\
 \frac{\phi(z)}{\Phi(z)} \left\{ -2 \frac{\phi(z)^2}{\Phi(z)^2} - 3 z \frac{\phi(z)}{\Phi(z)} - z^2 \right\} &< -\frac{\phi(z)}{\Phi(z)} \quad (\times z^2) \\
 \frac{\phi(z)}{\Phi(z)} \left\{ -2 z^2 \frac{\phi(z)^2}{\Phi(z)^2} - 3 z^3 \frac{\phi(z)}{\Phi(z)} - z^4 \right\} &< -z^2 \frac{\phi(z)}{\Phi(z)}.
 \end{aligned} \tag{A.37}$$

Somando $3 z^2 \frac{\phi(z)}{\Phi(z)} + 2 z \frac{\phi(z)^2}{\Phi(z)^2}$ em ambos os lados da desigualdade (A.37), obtemos

$$\begin{aligned}
 \frac{\phi(z)}{\Phi(z)} \left\{ -2 z^2 \frac{\phi(z)^2}{\Phi(z)^2} - 3 z^3 \frac{\phi(z)}{\Phi(z)} - z^4 + 3 z^2 + 2 z \frac{\phi(z)}{\Phi(z)} \right\} &< \\
 &< -z^2 \frac{\phi(z)}{\Phi(z)} + 3 z^2 \frac{\phi(z)}{\Phi(z)} + 2 z \frac{\phi(z)^2}{\Phi(z)^2} \iff
 \end{aligned}$$

$$\begin{aligned}
M(z) &< 2 z^2 \frac{\phi(z)}{\Phi(z)} + 2 z \frac{\phi(z)^2}{\Phi(z)^2} && \iff \\
M(z) &< 2 z \frac{\phi(z)}{\Phi(z)} \left\{ z + \frac{\phi(z)}{\Phi(z)} \right\}. && \text{(A.38)}
\end{aligned}$$

Sabemos que $\frac{\phi(z)}{\Phi(z)} > 0$ e $\left\{ z + \frac{\phi(z)}{\Phi(z)} \right\} > 0$, $\forall z$. Além disso, note que, quando $z \leq 0$, temos

$$2 z \frac{\phi(z)}{\Phi(z)} \left\{ z + \frac{\phi(z)}{\Phi(z)} \right\} < 0. \quad \text{(A.39)}$$

Então, por (A.38), $M(z) < 0$, $\forall z < 0$. Como definimos $\frac{\partial}{\partial z} T(z) = M(z)$, concluímos que a função $T(z)$ considerada em (A.32) é estritamente decrescente em $z \leq 0$.

Uma vez mostrada a monotocidade da função $T(z)$ no intervalo em que $z \leq 0$, vamos estudar a monotocidade da função $d_V(\lambda)$, quando $\lambda \leq 0$ e $\bar{x} > \mu$.

Como visto em (A.28), quando $\lambda_1 < \lambda_2 \leq 0$, temos que $h(\lambda_1) a < h(\lambda_2) a \leq 0$. Já foi mostrado que $T(z)$ é uma função estritamente decrescente em $z \leq 0$, então $\lambda_1 < \lambda_2 \leq 0$ implica que

$$\begin{aligned}
\frac{\phi[h(\lambda_2) a]}{\Phi[h(\lambda_2) a]} [h(\lambda_2) a]^2 \left\{ h(\lambda_2) a + \frac{\phi[h(\lambda_2) a]}{\Phi[h(\lambda_2) a]} \right\} &< \\
&< \frac{\phi[h(\lambda_1) a]}{\Phi[h(\lambda_1) a]} [h(\lambda_1) a]^2 \left\{ h(\lambda_1) a + \frac{\phi[h(\lambda_1) a]}{\Phi[h(\lambda_1) a]} \right\}. && \text{(A.40)}
\end{aligned}$$

Multiplicando por $\frac{b^2}{a^2}$ em ambos os lados da desigualdade (A.40), obtemos que

$$\begin{aligned}
\frac{\phi[h(\lambda_2) a]}{\Phi[h(\lambda_2) a]} h(\lambda_2)^2 b^2 \left\{ h(\lambda_2) a + \frac{\phi[h(\lambda_2) a]}{\Phi[h(\lambda_2) a]} \right\} &< \\
&< \frac{\phi[h(\lambda_1) a]}{\Phi[h(\lambda_1) a]} h(\lambda_1)^2 b^2 \left\{ h(\lambda_1) a + \frac{\phi[h(\lambda_1) a]}{\Phi[h(\lambda_1) a]} \right\}, && \text{(A.41)}
\end{aligned}$$

implicando em $0 \leq d_V(\lambda_2) < d_V(\lambda_1)$, quando $\lambda_1 < \lambda_2 \leq 0$ e $\bar{x} > \mu$. Ou seja, quando $\lambda \leq 0$ e $\bar{x} > \mu$, temos que a distância d_V é uma função estritamente decrescente em $\lambda \leq 0$.

Parte (ii): Omitiremos a demonstração deste caso, pois esta prova é similar a demonstração apresentada para o item (i).

Parte (iii):: Mostrando que, quando $\bar{x} = \mu$, as distâncias $d_M(\lambda)$ e $d_V(\lambda)$ são simétricas, monotônicas estritamente decrescentes em $\lambda \leq 0$ e monotônicas estritamente crescentes em $\lambda \geq 0$.

Ao considerar $\bar{x} = \mu$, temos que as distâncias d_M e d_V reduzem-se a

$$\begin{aligned} d_M &= \left| E_\lambda[\theta | \mathbf{x}] - \hat{\theta}_0 \right| = \frac{\phi[0]}{\Phi[0]} \frac{|\lambda|}{\sqrt{n\sigma^2 + \tau^2(1 + \lambda^2)}} \frac{\sigma\tau^2}{\sqrt{n\sigma^2 + \tau^2}} \\ &= \sqrt{\frac{2}{\pi}} \frac{|\lambda|}{\sqrt{n\sigma^2 + \tau^2(1 + \lambda^2)}} \frac{\sigma\tau^2}{\sqrt{n\sigma^2 + \tau^2}}. \end{aligned} \quad (\text{A.42})$$

e

$$\begin{aligned} d_V &= \left| \text{Var}_\lambda[\theta | \mathbf{x}] - \hat{\tau}_0 \right| = \frac{\lambda^2\tau^2}{n\sigma^2 + \tau^2(1 + \lambda^2)} \frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2} \frac{\phi(0)^2}{\Phi(0)^2} \\ &= \frac{\lambda^2\tau^2}{n\sigma^2 + \tau^2(1 + \lambda^2)} \frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2} \frac{2}{\pi}. \end{aligned} \quad (\text{A.43})$$

Uma vez fixados n , τ^2 e σ^2 em (A.42) e (A.43), ao analisarmos as distâncias d_M e d_V como funções de λ , temos que $d_M(\lambda) = d_M(-\lambda)$ e $d_V(\lambda) = d_V(-\lambda)$, implicando que ambas as distâncias são simétricas em relação a $\lambda = 0$.

Vamos estudar inicialmente a monotocidade da distância d_M como função de λ , quando $\lambda \geq 0$. Neste intervalo, a expressão de d_M em (A.42) reduz-se a

$$d_M = \sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1 + \lambda^2)}} \frac{\sigma\tau^2}{\sqrt{n\sigma^2 + \tau^2}}. \quad (\text{A.44})$$

Pelo Lema A.1.3, já foi mostrado que a função $\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1 + \lambda^2)}}$ é estritamente crescente em $\lambda \in \mathbb{R}$, em particular, no intervalo em que $\lambda \geq 0$, este resultado também se aplica. Como $\sqrt{\frac{2}{\pi}} \frac{\sigma\tau^2}{\sqrt{n\sigma^2 + \tau^2}} > 0$, então a distância d_M em $\lambda \geq 0$, também é uma função estritamente crescente em λ .

Pela simetria de $d_M(\lambda)$ nos intervalos $\lambda < 0$ e $\lambda \geq 0$, temos que a distância d_M em $\lambda < 0$, é uma função estritamente decrescente em λ .

Para o estudo da monotocidade da distância d_V como função de λ , iremos considerar o intervalo $\lambda \in \mathbb{R}$ particionado em $\lambda < 0$ e $\lambda \geq 0$.

Note que, para $\lambda < 0$, a função $w(\lambda) = \frac{\lambda^2\tau^2}{n\sigma^2 + \tau^2(1 + \lambda^2)}$ é uma função estritamente

decrecente pois

$$\begin{aligned} \frac{\partial}{\partial \lambda} w(\lambda) &= \frac{2 \lambda \tau^2}{n\sigma^2 + \tau^2(1 + \lambda^2)} - \frac{2 \lambda^3 \tau^4}{[n\sigma^2 + \tau^2(1 + \lambda^2)]^2} \\ &= \frac{2 \tau^2 \lambda (n\sigma^2 + \tau^2)}{[n\sigma^2 + \tau^2(1 + \lambda^2)]^2} < 0, \quad \forall \lambda < 0. \end{aligned} \quad (\text{A.45})$$

Como $\frac{2}{\pi} \frac{\sigma^2 \tau^2}{n\sigma^2 + \tau^2} > 0$, então a distância d_V em $\lambda < 0$ é uma função estritamente decrescente em λ . Por simetria da função $d_V(\lambda)$ nos intervalos $\lambda < 0$ e $\lambda \geq 0$, temos que a distância d_V com relação a λ é estritamente crescente em $\lambda \geq 0$. ♦

Prova da Proposição 3.1.4:

Apresentaremos agora a demonstração da medida $\sup_{\lambda} d_M(\lambda)$ definida no item (i). Pelo Lema A.1.2, vimos que $\lim_{\lambda \rightarrow -\infty} h(\lambda) = -\frac{1}{\tau}$, onde $h(\lambda) = \frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1 + \lambda^2)}}$.

Utilizando conhecidas propriedades de limite de uma função, temos que a medida $\sup_{\lambda} d_M$, quando $\lambda \leq 0$ e $\bar{x} > \mu$, fica dada por:

$$\begin{aligned} \sup_{\lambda} d_M &= \lim_{\lambda \rightarrow -\infty} d_M = \lim_{\lambda \rightarrow -\infty} \frac{\phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1 + \lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[\frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1 + \lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]} \frac{|\lambda|}{\sqrt{n\sigma^2 + \tau^2(1 + \lambda^2)}} \times \\ &\hspace{20em} \times \frac{\sigma \tau^2}{\sqrt{n\sigma^2 + \tau^2}} \\ &= \frac{\phi \left[\lim_{\lambda \rightarrow -\infty} \frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1 + \lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[\lim_{\lambda \rightarrow -\infty} \frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1 + \lambda^2)}} \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} \right]} \lim_{\lambda \rightarrow -\infty} \left| \frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1 + \lambda^2)}} \right| \\ &\hspace{20em} \times \frac{\sigma \tau^2}{\sqrt{n\sigma^2 + \tau^2}} \\ &= \frac{\phi \left[-\frac{\sigma n(\bar{x} - \mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[-\frac{\sigma n(\bar{x} - \mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]} \frac{\sigma \tau}{\sqrt{n\sigma^2 + \tau^2}}. \end{aligned} \quad (\text{A.46})$$

De maneira análoga a demonstração de A.46, conseguimos obter as expressões das medidas limites dadas em (3.14), (3.15) e (3.16). ♦

Prova da Proposição 3.1.5:

Sabemos que, quando $\lambda \leq 0$ e $\bar{x} > \mu$,

$$\lim_{\bar{x} \rightarrow \infty} \sup_{\lambda} d_M = \frac{\sigma\tau}{\sqrt{n\sigma^2 + \tau^2}} \lim_{\bar{x} \rightarrow \infty} \frac{\phi \left[-\frac{\sigma n(\bar{x} - \mu)}{\tau\sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[-\frac{\sigma n(\bar{x} - \mu)}{\tau\sqrt{n\sigma^2 + \tau^2}} \right]}. \quad (\text{A.47})$$

Como

$$\frac{\partial}{\partial \bar{x}} \phi \left[-\frac{\sigma n(\bar{x} - \mu)}{\tau\sqrt{n\sigma^2 + \tau^2}} \right] = -\phi \left[-\frac{\sigma n(\bar{x} - \mu)}{\tau\sqrt{n\sigma^2 + \tau^2}} \right] \frac{\sigma^2 n^2 (\bar{x} - \mu)}{\tau^2 (n\sigma^2 + \tau^2)} \quad (\text{A.48})$$

e

$$\frac{\partial}{\partial \bar{x}} \Phi \left[-\frac{\sigma n(\bar{x} - \mu)}{\tau\sqrt{n\sigma^2 + \tau^2}} \right] = -\phi \left[-\frac{\sigma n(\bar{x} - \mu)}{\tau\sqrt{n\sigma^2 + \tau^2}} \right] \frac{\sigma n}{\tau\sqrt{n\sigma^2 + \tau^2}}, \quad (\text{A.49})$$

temos, pela regra de L'Hospital, que

$$\lim_{\bar{x} \rightarrow \infty} \sup_{\lambda} d_M = \frac{\sigma\tau}{\sqrt{n\sigma^2 + \tau^2}} \lim_{\bar{x} \rightarrow \infty} \frac{\sigma n}{\tau\sqrt{n\sigma^2 + \tau^2}} (\bar{x} - \mu) = \infty. \quad (\text{A.50})$$

Para mostramos que $\lim_{\bar{x} \rightarrow \infty} \sup_{\lambda} d_V(\lambda) = \frac{\sigma^2 \tau^2}{n\sigma^2 + \tau^2}$, precisamos recorrer a um resultado apresentado na prova do Corolário 3.1.1 que nos diz que

$$0 < \frac{\phi(y)}{\Phi(y)} y + \frac{\phi(y)^2}{\Phi(y)^2} < 1, \quad \forall y, \quad (\text{A.51})$$

e mostrar que a função $L(y) = \frac{\phi(y)}{\Phi(y)} \left\{ y + \frac{\phi(y)}{\Phi(y)} \right\}$ é estritamente decrescente. Demonstra-se este resultado ao notar que

$$\begin{aligned} \frac{\partial}{\partial y} L(y) &= \left\{ \frac{\partial}{\partial y} \frac{\phi(y)}{\Phi(y)} \right\} \left\{ y + \frac{\phi(y)}{\Phi(y)} \right\} + \frac{\phi(y)}{\Phi(y)} \left\{ \frac{\partial}{\partial y} \left[y + \frac{\phi(y)}{\Phi(y)} \right] \right\} \\ &= -\frac{\phi(y)}{\Phi(y)} \left\{ -1 + y^2 + 3y \frac{\phi(y)}{\Phi(y)} + 2 \frac{\phi(y)^2}{\Phi(y)^2} \right\} < 0, \end{aligned} \quad (\text{A.52})$$

como pode ser verificado pela segunda equivalência em (A.37).

Como consequência, a função $L(-y)$ é uma função estritamente crescente e $0 <$

$\frac{\phi(-y)}{\Phi(-y)} \left\{ -y + \frac{\phi(-y)}{\Phi(-y)} \right\} < 1$. Portanto,

$$\begin{aligned}
\lim_{\bar{x} \rightarrow \infty} \sup_{\lambda} d_V &= \lim_{\bar{x} \rightarrow \infty} \frac{\phi \left[-\frac{\sigma n(\bar{x} - \mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[-\frac{\sigma n(\bar{x} - \mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]} \frac{\sigma^2 \tau^2}{n\sigma^2 + \tau^2} \times \\
&\quad \times \left\{ -\frac{\sigma n(\bar{x} - \mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} + \frac{\phi \left[-\frac{\sigma n(\bar{x} - \mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[-\frac{\sigma n(\bar{x} - \mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]} \right\} \\
&= \frac{\sigma^2 \tau^2}{n\sigma^2 + \tau^2} \lim_{\bar{x} \rightarrow \infty} \frac{\phi \left[-\frac{\sigma n(\bar{x} - \mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[-\frac{\sigma n(\bar{x} - \mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]} \times \\
&\quad \times \left\{ -\frac{\sigma n(\bar{x} - \mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} + \frac{\phi \left[-\frac{\sigma n(\bar{x} - \mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]}{\Phi \left[-\frac{\sigma n(\bar{x} - \mu)}{\tau \sqrt{n\sigma^2 + \tau^2}} \right]} \right\} \\
&= \frac{\sigma^2 \tau^2}{n\sigma^2 + \tau^2}. \tag{A.53}
\end{aligned}$$

◆

Prova da Proposição 3.1.6:

item (i): Mostrando que se existir um ponto de máximo para a distância d_M em função de λ , no caso em que $\lambda > 0$ e $\bar{x} > \mu$, ele será único.

Considere

$$P(\lambda) = \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} h(\lambda) b, \tag{A.54}$$

onde $h(\lambda) = \frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1 + \lambda^2)}}$, $a = \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}}$ e $b = \frac{\sigma \tau^2}{\sqrt{n\sigma^2 + \tau^2}}$.

Derivando a função $P(\lambda)$ com respeito a λ e igualando a zero obtemos

$$\frac{\partial}{\partial \lambda} P(\lambda) = b \left\{ h(\lambda) \frac{\partial}{\partial \lambda} \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} + \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} \frac{\partial}{\partial \lambda} h(\lambda) \right\} = 0. \tag{A.55}$$

Mas

$$\frac{\partial}{\partial \lambda} \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} = \left\{ -h(\lambda) a^2 \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} - a \frac{\phi[h(\lambda) a]^2}{\Phi[h(\lambda) a]^2} \right\} \frac{\partial}{\partial \lambda} h(\lambda). \tag{A.56}$$

Como $b > 0$, vamos dividir ambos os lados da segunda igualdade em (A.55) por

b e substituir nela a expressão (A.56). Assim,

$$h(\lambda) \left\{ -h(\lambda) a^2 \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} - a \frac{\phi[h(\lambda) a]^2}{\Phi[h(\lambda) a]^2} \right\} \frac{\partial}{\partial \lambda} h(\lambda) + \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} \frac{\partial}{\partial \lambda} h(\lambda) = 0 \iff \left[\frac{\partial}{\partial \lambda} h(\lambda) \right] \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} \left\{ -h(\lambda)^2 a^2 - a h(\lambda) \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} + 1 \right\} = 0. \quad (\text{A.57})$$

Dessa forma,

$$\frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} = 0, \quad \frac{\partial}{\partial \lambda} h(\lambda) = \frac{n\sigma^2 + \tau^2}{[n\sigma^2 + \tau^2(1 + \lambda^2)]^{\frac{3}{2}}} = 0 \quad (\text{A.58})$$

e

$$-h(\lambda)^2 a^2 - a h(\lambda) \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} + 1 = 0. \quad (\text{A.59})$$

Note que não existe λ para o qual as igualdades em (A.58) ocorram, pois $\forall x \in \mathbb{R}$, $\frac{\phi[x]}{\Phi[x]} > 0$ e $\forall \lambda \in \mathbb{R}$, $\frac{\partial}{\partial \lambda} h(\lambda) > 0$.

Estamos estudando o caso em que $\lambda > 0$ e $\bar{x} > \mu$, o que implica em $a = \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}} > 0$ e $h(\lambda) > 0$. Assim, por (A.59) temos que

$$a h(\lambda) \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} = 1 - h(\lambda)^2 a^2 \iff \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} = \frac{1}{a h(\lambda)} - h(\lambda) a. \quad (\text{A.60})$$

Seja $k = h(\lambda) a$, em que $k > 0$. Substituindo k em (A.60), obtemos

$$\underbrace{\frac{\phi[k]}{\Phi[k]}}_{g(k)} = \underbrace{\frac{1}{k} - k}_{t(k)}. \quad (\text{A.61})$$

Vamos analisar agora as funções $g(k)$ e $t(k)$, para todo $k > 0$.

- *Análise de* $g(k) = \frac{\phi(k)}{\Phi(k)}$, $\forall k > 0$.

Pelo Lema A.1.3, mostramos que a função $\frac{\phi[k]}{\Phi[k]}$ é estritamente decrescente, $\forall k \in \mathbb{R}$. Além disso,

$$\lim_{k \rightarrow 0^+} g(k) = \frac{\phi[0]}{\Phi[0]} = \sqrt{\frac{2}{\pi}} \quad \text{e} \quad \lim_{k \rightarrow \infty} g(k) = 0. \quad (\text{A.62})$$

- Análise de $t(k) = \frac{1}{k} - k, \forall k > 0$.

Note que $\frac{\partial}{\partial k} t(k) = -\frac{1}{k^2} - 1 < 0, \forall k$. Logo, em particular, para $k > 0$, a função $t(k)$ é estritamente decrescente. Além disso,

$$\begin{aligned} \lim_{k \rightarrow 0^+} t(k) &= \frac{1}{k} - k = -\infty \quad \text{e} \\ \lim_{k \rightarrow \infty} t(k) &= \lim_{k \rightarrow \infty} \frac{1 - k^2}{k} = \lim_{k \rightarrow \infty} -2k = -\infty \quad (\text{aplicando L'Hospital}). \end{aligned} \quad (\text{A.63})$$

Na Figura A.1, apresentamos as curvas das funções $g(k)$ e $t(k)$, para $k > 0$. Note que, pela análise algébrica das funções, mostramos que existe um único ponto K_{M_0} tal que a igualdade (A.61) ocorre. Na Figura A.1, K_{M_0} é o ponto de intersecção entre as curvas $g(k)$ e $t(k)$. Através de métodos numéricos obtivemos que $K_{M_0} = 0.8399238$.

Agora, como $k = h(\lambda)$, vamos determinar qual deve ser o valor de λ_M que resulte em $K_{M_0} = 0.8399238$. Assim, para $a > 0$ e $\lambda > 0$,

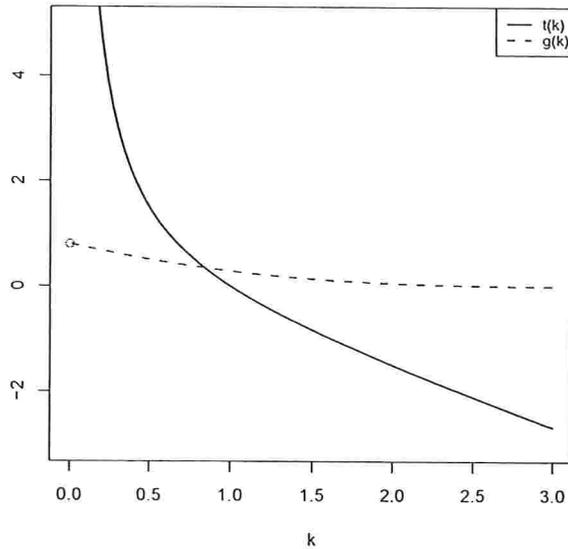


Figura A.1: Curva das funções $g(k)$ e $t(k)$, para $k > 0$, no estudo da distância d_M .

$$\begin{aligned}
 h(\lambda_M) a = K_{M_0} &\iff \frac{\lambda_M}{\sqrt{n\sigma^2 + \tau^2(1 + \lambda_M^2)}} a = K_{M_0} \iff \\
 \sqrt{\frac{1}{\frac{n\sigma^2 + \tau^2}{\lambda_M^2} + \tau^2}} = \frac{K_{M_0}}{a} &\iff \left(\frac{a}{K_{M_0}}\right)^2 = \frac{n\sigma^2 + \tau^2}{\lambda_M^2} + \tau^2 \iff \\
 \frac{n\sigma^2 + \tau^2}{\lambda_M^2} = \frac{a^2 - K_{M_0}^2 \tau^2}{K_{M_0}^2} &\iff \lambda_M^2 = \frac{K_{M_0}^2 (n\sigma^2 + \tau^2)}{a^2 - K_{M_0}^2 \tau^2}. \tag{A.64}
 \end{aligned}$$

Ao resolvermos a última igualdade em (A.64), obtemos que $\lambda_M = K_{M_0} \sqrt{\frac{n\sigma^2 + \tau^2}{a^2 - K_{M_0}^2 \tau^2}}$ e $\lambda_M = -K_{M_0} \sqrt{\frac{n\sigma^2 + \tau^2}{a^2 - K_{M_0}^2 \tau^2}}$. Como estamos analisando o caso em que $\lambda > 0$ e $\bar{x} > \mu$, obtemos uma única solução para a equação (A.64) que é dada por

$$\lambda_M = K_{M_0} \sqrt{\frac{n\sigma^2 + \tau^2}{a^2 - K_{M_0}^2 \tau^2}}. \tag{A.65}$$

Note que, o valor de λ_M dado em (A.65) existirá apenas quando em $a^2 - K_{M_0}^2 \tau^2 > 0$. Isto ocorre devido a condição de existência presente no denominador da fração $\sqrt{\frac{n\sigma^2 + \tau^2}{a^2 - K_{M_0}^2 \tau^2}}$, que deve ser sempre maior que zero.

A prova para $\lambda < 0$ e $\bar{x} < \mu$ é equivalente a apresentada para o caso $\lambda > 0$ e $\bar{x} > \mu$ e, portanto, será omitida.

item (ii): Mostrando que se existir um ponto de máximo para a distância d_V em função de λ , no caso em que $\lambda > 0$ e $\bar{x} > \mu$, ele será único.

Reescrevendo a distância d_V , temos que

$$d_V(\lambda) = \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} h(\lambda)^3 a b^2 + \frac{\phi[h(\lambda) a]^2}{\Phi[h(\lambda) a]^2} h(\lambda)^2 b^2, \tag{A.66}$$

onde $h(\lambda) = \frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1 + \lambda^2)}}$, $a = \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}}$ e $b = \frac{\sigma \tau^2}{\sqrt{n\sigma^2 + \tau^2}}$.

Derivando a função $d_V(\lambda)$ com respeito a λ e igualando a zero obtemos

$$\begin{aligned}
 \frac{\partial}{\partial \lambda} d_V(\lambda) &= b^2 a \left\{ 3 h(\lambda)^2 \left[\frac{\partial}{\partial \lambda} h(\lambda) \right] \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} + h(\lambda)^3 \frac{\partial}{\partial \lambda} \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} \right\} + \\
 &+ b^2 \left\{ 2 h(\lambda) \left[\frac{\partial}{\partial \lambda} h(\lambda) \right] \frac{\phi[h(\lambda) a]^2}{\Phi[h(\lambda) a]^2} + h(\lambda)^2 2 \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} \frac{\partial}{\partial \lambda} \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} \right\} = 0. \tag{A.67}
 \end{aligned}$$

Como $b > 0$, vamos dividir ambos lados da segunda igualdade em (A.67) por b^2 e substituir na igualdade resultante a expressão de $\frac{\partial}{\partial \lambda} \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]}$ dada em (A.56). Assim,

$$\begin{aligned}
& 3 a h(\lambda)^2 \left[\frac{\partial}{\partial \lambda} h(\lambda) \right] \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} + 2 h(\lambda) \left[\frac{\partial}{\partial \lambda} h(\lambda) \right] \frac{\phi[h(\lambda) a]^2}{\Phi[h(\lambda) a]^2} + \left[\frac{\partial}{\partial \lambda} h(\lambda) \right] \times \\
& \times \left\{ -h(\lambda) a^2 \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} - a \frac{\phi[h(\lambda) a]^2}{\Phi[h(\lambda) a]^2} \right\} \left\{ a h(\lambda)^3 + h(\lambda)^2 2 \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} \right\} = 0 \\
& \iff \\
& h(\lambda) \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} \left[\frac{\partial}{\partial \lambda} h(\lambda) \right] \left\{ 3a h(\lambda) + 2 \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} + \left[-h(\lambda) a^2 - a \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} \right] \times \right. \\
& \left. \times \left[h(\lambda)^2 a + 2 h(\lambda) \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} \right] \right\} = 0. \quad (\text{A.68})
\end{aligned}$$

Dessa forma, para $\lambda > 0$, temos que resolver as equações dadas em (A.58),

$$h(\lambda) = 0 \iff \frac{\lambda}{\sqrt{n\sigma^2 + \tau^2(1 + \lambda^2)}} = 0 \quad (\text{A.69})$$

e

$$3ah(\lambda) + 2 \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} - \left[h(\lambda) a^2 + a \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} \right] \left[h(\lambda)^2 a + 2h(\lambda) \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} \right] = 0, \quad (\text{A.70})$$

em λ .

Na demonstração da Proposição 3.1.6, item i , já foi discutido que não existe $\lambda \in \mathbb{R}$ para os quais as equações dadas em (A.58) tenham solução. Note também que em (A.69), não existe $\lambda > 0$ para o qual $h(\lambda) = 0$.

Vamos estudar agora para qual λ a equação em (A.70) tem solução. Para isso, desenvolvendo a equação (A.70), temos que

$$3ah(\lambda) + 2 \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} - h(\lambda)^3 a^3 - 3h(\lambda)^2 a^2 \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} - 2h(\lambda) a \frac{\phi[h(\lambda) a]^2}{\Phi[h(\lambda) a]^2} = 0.$$

\iff

$$h(\lambda)^3 a^3 + 3h(\lambda)^2 a^2 \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]} + 2h(\lambda) a \frac{\phi[h(\lambda) a]^2}{\Phi[h(\lambda) a]^2} = 3ah(\lambda) + 2 \frac{\phi[h(\lambda) a]}{\Phi[h(\lambda) a]}. \quad (\text{A.71})$$

Vamos estudar o caso em que $\lambda > 0$ e $\bar{x} > \mu$ implicam em $h(\lambda) > 0$ e $a = \frac{\sigma n(\bar{x}-\mu)}{\sqrt{n\sigma^2+\tau^2}} > 0$. Ao considerar $k = h(\lambda)$ a em (A.71), obtemos que

$$\underbrace{k^3 + 3k^2 \frac{\phi[k]}{\Phi[k]} + 2k \frac{\phi[k]^2}{\Phi[k]^2}}_{g(k)} = \underbrace{3k + 2 \frac{\phi[k]}{\Phi[k]}}_{t(k)}. \quad (\text{A.72})$$

Vamos analisar agora as funções $g(k)$ e $t(k)$, para todo $k > 0$.

- *Análise de $g(k) = k^3 + 3k^2 \frac{\phi(k)}{\Phi(k)} + 2k \frac{\phi(k)^2}{\Phi(k)^2}$, $\forall k > 0$.*

Vamos mostrar inicialmente que a função $g(k)$ é uma função estritamente crescente em $k > 0$. Então:

$$\begin{aligned} \frac{\partial}{\partial k} g(k) &= \underbrace{k \frac{\phi[k]}{\Phi[k]} \left\{ 6 - 4k \frac{\phi[k]}{\Phi[k]} - 4 \frac{\phi[k]^2}{\Phi[k]^2} \right\}}_{P(k)} + 2 \frac{\phi[k]^2}{\Phi[k]^2} + \\ &\quad + \underbrace{3k^2 \left\{ 1 - k \frac{\phi[k]}{\Phi[k]} - \frac{\phi[k]^2}{\Phi[k]^2} \right\}}_{Q(k)}. \end{aligned} \quad (\text{A.73})$$

Considere a função $P(k)$, $\forall k$. Sabemos que $\forall k$, por (A.24) e (A.25),

$$\begin{aligned} 0 &< k \frac{\phi[k]}{\Phi[k]} + \frac{\phi[k]^2}{\Phi[k]^2} < 1 && \iff \\ 2 &< 6 - 4 \left\{ k \frac{\phi[k]}{\Phi[k]} + \frac{\phi[k]^2}{\Phi[k]^2} \right\} < 6, \forall k && \iff \\ 2k \frac{\phi[k]}{\Phi[k]} &< P(k) < 6k \frac{\phi[k]}{\Phi[k]}, \forall k > 0. \end{aligned} \quad (\text{A.74})$$

De maneira análoga a (A.74), podemos mostrar que $Q(k) > 0$, $\forall k$ finito. Dessa forma,

$$\frac{\partial}{\partial k} g(k) = P(k) + Q(k) + 2 \frac{\phi[k]^2}{\Phi[k]^2} > 0, \forall k > 0 \quad (\text{A.75})$$

e concluímos que a função $g(k)$ apresentada em (A.72) é uma função estritamente crescente em k , quando $k > 0$. Além disso,

$$\lim_{k \rightarrow 0^+} g(k) = 0 \quad \text{e} \quad \lim_{k \rightarrow \infty} g(k) = \infty, \quad (\text{A.76})$$

pois $\lim_{k \rightarrow \infty} k^2 \frac{\phi[k]}{\Phi[k]} = 0$ e $\lim_{k \rightarrow \infty} k \frac{\phi[k]^2}{\Phi[k]^2} = 0$.

- *Análise de $t(k) = 3k + 2 \frac{\phi(k)}{\Phi(k)}$, $\forall k > 0$.*

Vamos mostrar agora que a função $t(k)$ é uma função estritamente crescente em $k > 0$. Então:

$$\frac{\partial}{\partial k} t(k) = 3 - 2 \left\{ k \frac{\phi[k]}{\Phi[k]} + \frac{\phi[k]^2}{\Phi[k]^2} \right\}. \quad (\text{A.77})$$

Mas, por (A.24),

$$2 - 2 \left\{ k \frac{\phi[k]}{\Phi[k]} + \frac{\phi[k]^2}{\Phi[k]^2} \right\} > 0 \implies 3 - 2 \left\{ k \frac{\phi[k]}{\Phi[k]} + \frac{\phi[k]^2}{\Phi[k]^2} \right\} > 1. \quad (\text{A.78})$$

Dessa forma, $\frac{\partial}{\partial k} t(k) > 0$, $\forall k$. Portanto, função $t(k)$ apresentada em (A.72) é uma função estritamente crescente em k , $\forall k$. Além disso,

$$\lim_{k \rightarrow 0^+} t(k) = 2 \frac{\phi[0]}{\Phi[0]} = 2 \sqrt{\frac{2}{\pi}} \quad \text{e} \quad \lim_{k \rightarrow \infty} t(k) = \infty. \quad (\text{A.79})$$

Sabemos também que, $\forall k \geq \sqrt{3}$, as desigualdades $k^3 \geq 3k$ e $3k^2 > 2 \implies 3k^2 \frac{\phi[k]}{\Phi[k]} > 2 \frac{\phi[k]}{\Phi[k]}$. Esta última desigualdade implica em $\frac{\phi[k]}{\Phi[k]}$ ocorrem. Logo, $\forall k \geq \sqrt{3}$,

$$k^3 + 3k^2 \frac{\phi[k]}{\Phi[k]} > 3k + 2 \frac{\phi[k]}{\Phi[k]}. \quad (\text{A.80})$$

Para $k > 0$, temos que a parcela $2k \frac{\phi[k]^2}{\Phi[k]^2} > 0$, logo

$$k^3 + 3k^2 \frac{\phi[k]}{\Phi[k]} + 2k \frac{\phi[k]^2}{\Phi[k]^2} > k^3 + 3k^2 \frac{\phi[k]}{\Phi[k]}. \quad (\text{A.81})$$

Por (A.80) e (A.81), mostramos que, $\forall k \geq \sqrt{3}$,

$$k^3 + 3k^2 \frac{\phi[k]}{\Phi[k]} + 2k \frac{\phi[k]^2}{\Phi[k]^2} > 3k + 2 \frac{\phi[k]}{\Phi[k]} \implies g(k) > t(k). \quad (\text{A.82})$$

Considerando as análises das funções $g(k)$ e $t(k)$, para todo $k > 0$, complementado pela desigualdade em (A.82), para $k \geq \sqrt{3}$, conseguimos mostrar que existe um único ponto K_{V_0} tal que a igualdade (A.72) ocorre. Na Figura A.2, K_{V_0} é o ponto

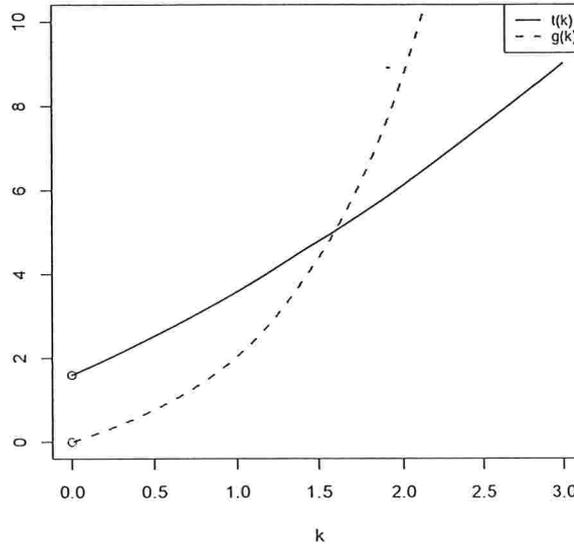


Figura A.2: Curva das funções $g(k)$ e $t(k)$, para $k > 0$, no estudo da distância d_V .

de intersecção entre as curvas $g(k)$ e $t(k)$. Através de métodos numéricos obtivemos que $K_{V_0} = 1.598539$.

Sabemos que $a > 0$ e $\lambda > 0$, pois $a = \frac{\sigma n(\bar{x} - \mu)}{\sqrt{n\sigma^2 + \tau^2}}$ e $\bar{x} > \mu$. Para determinar qual deve ser o valor de λ_V que resulte em $K_{V_0} = 1.598539$ na equação $K_V = h(\lambda_V) a$, basta considerar os cálculos apresentados em (A.64) e as conclusões posteriores a eles, utilizando K_{V_0} no lugar de K_{M_0} .

A prova para $\lambda < 0$ e $\bar{x} < \mu$ é equivalente a apresentada para o caso $\lambda > 0$ e $\bar{x} > \mu$ e, portanto, será omitida. ♦

Prova da Proposição 3.2.1:

Considere $m(\mathbf{x})$ a distribuição marginal ou preditiva de \mathbf{X} , ou seja,

$$m(\mathbf{x}) = \int_{-\infty}^{\infty} \int_0^{\infty} f(\mathbf{x} | \theta, \tau^2) f(\theta) f(\tau^2) d\tau^2 d\theta. \quad (\text{A.83})$$

É sabido que a distribuição *a posteriori* é própria se e somente se $0 < m(\mathbf{x}) < \infty$. Dessa forma, vamos mostrar que no nosso contexto a distribuição preditiva de

\mathbf{X} tem tal característica. Assim, por (A.83),

$$m(\mathbf{x}) = \int_{-\infty}^{\infty} \int_0^{\infty} \left(\frac{1}{\sqrt{2\pi\tau^2}} \right)^n e^{-\frac{\sum_{i=1}^n (x_i - \theta)^2}{2\tau^2}} \frac{2}{\sigma} \phi\left(\frac{\theta - \mu}{\sigma}\right) \times \\ \times \Phi\left(\lambda \frac{\theta - \mu}{\sigma}\right) \frac{1}{\tau^2} d\tau^2 d\theta. \quad (\text{A.84})$$

Lembrando que $0 < \phi\left(\frac{\theta - \mu}{\sigma}\right) \leq \frac{1}{\sqrt{2\pi}}$ e $0 < \Phi\left(\lambda \frac{\theta - \mu}{\sigma}\right) < 1$, temos que

$$m(\mathbf{x}) \leq \frac{2}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} \int_0^{\infty} \left(\frac{1}{\tau}\right)^{n+2} \prod_{i=1}^n \phi\left(\frac{x_i - \theta}{\tau}\right) d\tau^2 d\theta \\ = \frac{2}{\sigma \sqrt{2\pi}} \underbrace{\int_{-\infty}^{\infty} \int_0^{\infty} \left(\frac{1}{\tau}\right)^{n+2} \phi_n\left[\frac{\mathbf{x} - \boldsymbol{\theta}}{\tau}\right] d\tau^2 d\theta}_A \\ < \infty, \quad (\text{A.85})$$

pois a integral expressa em A é finita, uma vez que corresponde a distribuição preditiva de \mathbf{X} quando se considera o modelo de regressão linear normal dado por $X_1, \dots, X_n \mid \theta, \tau^2 \sim N_n(\theta, \tau^2 I_n)$ e a distribuição *a priori* não-informativa usual $h(\theta, \tau^2) \propto \frac{1}{\tau^2}$, em que $\boldsymbol{\theta}^T = \theta \times \{1, 1, \dots, 1\}_n$ e I_n corresponde a matriz identidade de ordem n .

◆

Prova da Proposição 3.2.2:

A função densidade de probabilidade conjunta *a posteriori* é dada por:

$$f(\theta, \tau^2 \mid \mathbf{x}) \propto f(\mathbf{x} \mid \theta, \tau^2) f(\theta) f(\tau^2) \\ = \left(\frac{1}{\sqrt{2\pi\tau^2}} \right)^n e^{-\frac{\sum_{i=1}^n (x_i - \theta)^2}{2\tau^2}} \frac{2}{\sigma} \phi\left(\frac{\theta - \mu}{\sigma}\right) \Phi\left(\lambda \frac{\theta - \mu}{\sigma}\right) \frac{1}{\tau^2} \\ \propto \left(\frac{1}{\tau^2}\right)^{\frac{n}{2}+1} e^{-\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{2\tau^2} - \frac{n(\bar{x} - \theta)^2}{2\tau^2} - \frac{(\theta - \mu)^2}{2\sigma^2}} \Phi\left(\lambda \frac{\theta - \mu}{\sigma}\right). \quad (\text{A.86})$$

Mas, assumindo que $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$, obtemos

$$\begin{aligned}
 e^{-n \frac{(\bar{x}-\theta)^2}{2\tau^2} - \frac{(\theta-\mu)^2}{2\sigma^2}} &= e^{-\frac{1}{2} \left\{ \frac{n\bar{x}^2 - 2n\bar{x}\theta + n\theta^2}{\tau^2} + \frac{\theta^2 - 2\theta\mu + \mu^2}{\sigma^2} \right\}} \\
 &= e^{-\frac{1}{2} \left\{ \theta^2 \left[\frac{n}{\tau^2} + \frac{1}{\sigma^2} \right] - 2\theta \left[\frac{n\bar{x}}{\tau^2} + \frac{\mu}{\sigma^2} \right] + \left[\frac{n\bar{x}^2}{\tau^2} + \frac{\mu^2}{\sigma^2} \right] \right\}} \\
 &= e^{-\frac{1}{2} \left\{ \left[\theta \sqrt{\frac{n}{\tau^2} + \frac{1}{\sigma^2}} - \frac{\left(\frac{n\bar{x}}{\tau^2} + \frac{\mu}{\sigma^2} \right)}{\sqrt{\frac{n}{\tau^2} + \frac{1}{\sigma^2}}} \right]^2 - \frac{\left(\frac{n\bar{x}}{\tau^2} + \frac{\mu}{\sigma^2} \right)^2}{\frac{n}{\tau^2} + \frac{1}{\sigma^2}} \right\}} e^{-\frac{1}{2} \left[\frac{n\bar{x}^2}{\tau^2} + \frac{\mu^2}{\sigma^2} \right]} \\
 &= e^{-\frac{1}{2} \left\{ \sqrt{\frac{n}{\tau^2} + \frac{1}{\sigma^2}} \left[\theta - \frac{\frac{n\bar{x}}{\tau^2} + \frac{\mu}{\sigma^2}}{\sqrt{\frac{n}{\tau^2} + \frac{1}{\sigma^2}}} \right] \right\}^2} e^{\frac{1}{2} \frac{\left(\frac{n\bar{x}}{\tau^2} + \frac{\mu}{\sigma^2} \right)^2}{\frac{n}{\tau^2} + \frac{1}{\sigma^2}}} e^{-\frac{1}{2} \left[\frac{n\bar{x}^2}{\tau^2} + \frac{\mu^2}{\sigma^2} \right]} \\
 &= e^{-\frac{1}{2} \frac{\left(\frac{n\bar{x}}{\tau^2} + \frac{\mu}{\sigma^2} \right)^2}{\frac{n}{\tau^2} + \frac{1}{\sigma^2}}} e^{-\frac{1}{2} \left[\frac{n\bar{x}^2}{\tau^2} + \frac{\mu^2}{\sigma^2} \right]}. \quad (\text{A.87})
 \end{aligned}$$

Dessa forma, substituindo (A.87) em (A.86), temos que

$$\begin{aligned}
 f(\theta, \tau^2 | \mathbf{x}) &\propto \left(\frac{1}{\tau^2} \right)^{\frac{n}{2}+1} e^{-\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{2\tau^2} - \frac{1}{2} \frac{\sigma^2}{n\sigma^2 + \tau^2} \left[\theta - \frac{n\bar{x}\sigma^2 + \mu\tau^2}{n\sigma^2 + \tau^2} \right]^2} \times \\
 &\quad \times e^{\frac{1}{2} \frac{\left(\frac{n\bar{x}}{\tau^2} + \frac{\mu}{\sigma^2} \right)^2}{\frac{n}{\tau^2} + \frac{1}{\sigma^2}}} e^{-\frac{1}{2} \left[\frac{n\bar{x}^2}{\tau^2} + \frac{\mu^2}{\sigma^2} \right]} \Phi \left(\lambda \frac{\theta - \mu}{\sigma} \right). \quad (\text{A.88})
 \end{aligned}$$

Observe que

$$f(\theta | \tau^2, \mathbf{x}) \propto \phi \left(\theta; \frac{n\bar{x}\sigma^2 + \mu\tau^2}{n\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2} \right) \Phi(\lambda\theta; \lambda\mu, \sigma^2), \quad (\text{A.89})$$

que é o núcleo de uma função densidade de probabilidade de uma distribuição normal-assimétrica multivariada geral. Dessa forma, temos que a distribuição condicional completa *a posteriori* para o parâmetro de posição é dada por $\theta | \tau^2, \mathbf{x} \sim GSN_{1,1} \left(\frac{n\bar{x}\sigma^2 + \mu\tau^2}{n\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2}, \lambda, \lambda\mu, \sigma^2 \right)$.

Para obtermos a distribuição condicional *a posteriori* de τ^2 dado θ , basta notarmos que,

$$f(\tau^2 | \theta, \mathbf{x}) \propto \left(\frac{1}{\tau^2} \right)^{\frac{n}{2}+1} e^{-\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{2\tau^2} - \frac{n(\bar{x}-\theta)^2}{2\tau^2}}. \quad (\text{A.90})$$

Considerando $S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$, temos que

$$f(\tau^2 | \theta, \mathbf{x}) \propto \left(\frac{1}{\tau^2} \right)^{\frac{n}{2}+1} e^{-\frac{1}{\tau^2} \left[\frac{nS^2}{2} + \frac{n(\bar{x}-\theta)^2}{2} \right]}. \quad (\text{A.91})$$

que é o núcleo de uma função densidade de probabilidade de uma distribuição gama invertida com parâmetros $\frac{n}{2}$ e $\frac{nS^2}{2} + \frac{n(\bar{x}-\theta)^2}{2}$.

◆

Prova da Proposição 3.2.3:

Sabemos que a distribuição marginal *a posteriori* do parâmetro de posição é dada por:

$$f(\theta | \mathbf{x}) = \int_0^\infty f(\theta, \tau^2 | \mathbf{x}) d\tau^2. \quad (\text{A.92})$$

Utilizando uma expressão equivalente a apresentada em (A.86) para a função de densidade conjunta *a posteriori* de θ e τ^2 , temos que

$$\begin{aligned} f(\theta | \mathbf{x}) &\propto \int_0^\infty \left(\frac{1}{\tau^2}\right)^{\frac{n}{2}+1} e^{-\frac{\sum_{i=1}^n (x_i - \theta)^2}{2\tau^2}} \frac{2}{\sigma} \phi\left(\frac{\theta - \mu}{\sigma}\right) \Phi\left(\lambda \frac{\theta - \mu}{\sigma}\right) d\tau^2 \\ &\propto \phi\left(\frac{\theta - \mu}{\sigma}\right) \Phi\left(\lambda \frac{\theta - \mu}{\sigma}\right) \int_0^\infty \left(\frac{1}{\tau^2}\right)^{\frac{n}{2}+1} e^{-\frac{1}{\tau^2} \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2}\right]} d\tau^2 \\ &\propto \phi\left(\frac{\theta - \mu}{\sigma}\right) \Phi\left(\lambda \frac{\theta - \mu}{\sigma}\right) \Gamma\left(\frac{n}{2}\right) \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2}\right]^{-\frac{n}{2}} \times \\ &\times \int_0^\infty \frac{1}{\Gamma\left(\frac{n}{2}\right)} \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2}\right]^{\frac{n}{2}} \left(\frac{1}{\tau^2}\right)^{\frac{n}{2}+1} e^{-\frac{1}{\tau^2} \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2}\right]} d\tau^2. \end{aligned}$$

Note que o integrando presente na última expressão é a função densidade de probabilidade de uma distribuição gama com parâmetros $\frac{n}{2}$ e $\frac{nS^2}{2} + \frac{n(\bar{x}-\theta)^2}{2}$. Portanto, a integral em τ^2 é igual a um. Logo,

$$f(\theta | \mathbf{x}) \propto \phi\left(\frac{\theta - \mu}{\sigma}\right) \Phi\left(\lambda \frac{\theta - \mu}{\sigma}\right) \Gamma\left(\frac{n}{2}\right) \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2}\right]^{-\frac{n}{2}}. \quad (\text{A.93})$$

Para obtermos a esperança e a variância da distribuição marginal *a posteriori* de τ^2 , precisamos utilizar o resultado apresentado na Proposição 3.2.2 em que obtivemos $\tau^2 | \theta, \mathbf{x} \sim GI\left(\frac{n}{2}, \frac{nS^2}{2} + \frac{n(\bar{x}-\theta)^2}{2}\right)$. Assim,

$$E[\tau^2 | \theta, \mathbf{x}] = \frac{nS^2 + n(\bar{x} - \theta)^2}{n - 2}, \text{ para } n > 2, \quad (\text{A.94})$$

$$Var[\tau^2 | \theta, \mathbf{x}] = 2n^2 \frac{[S^2 + (\bar{x} - \theta)^2]^2}{(n - 2)^2(n - 4)}, \text{ para } n > 4. \quad (\text{A.95})$$

Lembrando que

$$E[\tau^2 | \mathbf{x}] = E[E[\tau^2 | \theta, \mathbf{x}] | \mathbf{x}] \quad (\text{A.96})$$

e

$$Var[\tau^2 | \mathbf{x}] = E[Var[\tau^2 | \theta, \mathbf{x}] | \mathbf{x}] + Var[E[\tau^2 | \theta, \mathbf{x}] | \mathbf{x}], \quad (\text{A.97})$$

as expressões apresentadas para $E[\tau^2 | \mathbf{x}]$ e $Var[\tau^2 | \mathbf{x}]$ na Proposição 3.2.3 surgem com simples manipulações algébricas de (A.96) e (A.97). ♦

Prova da Proposição 3.2.4: Será omitida, uma vez que é similar a demonstração da Proposição 3.2.2.

Prova da Proposição 3.2.5: Será omitida, uma vez que é similar a demonstração da Proposição 3.2.3.

A.2 Capítulo 4

Prova da Proposição 4.2.1:

A distribuição *a posteriori* para β será dada por:

$$\begin{aligned} f(\beta | \mathbf{y}) &\propto f(\mathbf{y} | \beta) f(\beta) \\ &\propto \left(\frac{1}{2\pi\tau^2} \right)^{\frac{n}{2}} e^{-\frac{1}{2\tau^2} [ks^2 + (\beta - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\beta - \hat{\beta})]} \phi_p(\beta; \mu, \Omega) \Phi[\lambda^T \omega^{-1}(\beta - \mu)] \\ &\propto e^{-\frac{1}{2\tau^2} [ks^2 + (\beta - \hat{\beta})^T \mathbf{X}^T \mathbf{X} (\beta - \hat{\beta})]} \Phi[\lambda^T \omega^{-1}(\beta - \mu)] \times \\ &\quad \times e^{-\frac{1}{2} [(\beta - \mu)^T \Omega^{-1} (\beta - \mu)]}, \end{aligned} \quad (\text{A.98})$$

onde $ks^2 = (\mathbf{y} - \mathbf{X}\hat{\beta})^T (\mathbf{y} - \mathbf{X}\hat{\beta})$ é conhecido como soma dos quadrados residual, $k = n - p$ e $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ é a estimativa de mínimos quadrados (ou de máxima verossimilhança) de β .

Para desenvolvermos o termo

$$e^{-\frac{1}{2} \left\{ (\beta - \hat{\beta})^T \frac{\mathbf{X}^T \mathbf{X}}{\tau^2} (\beta - \hat{\beta}) + (\beta - \mu)^T \Omega^{-1} (\beta - \mu) \right\}}, \quad (\text{A.99})$$

faremos uso do seguinte lema apresentado em Paulino et al. (2003), cuja a demonstração encontra-se em Box & Tiao (1973).

Lema A.2.1 *Sendo x , a e b vetores $p \times 1$ e A e B matrizes $p \times p$ simétricas definidas positivas tais que $A + B$ é não singular, tem-se*

$$(x-a)^T A(x-a) + (x-b)^T B(x-b) = (x-c)^T (A+B)(x-c) + (a-b)^T C(a-b), \quad (\text{A.100})$$

onde $c = (A + B)^{-1}(Aa + Bb)$ e $C = A(A + B)^{-1}B = (A^{-1} + B^{-1})^{-1}$.

Vamos utilizar o Lema A.2.1 com $x = \beta$, $a = \hat{\beta}$, $b = \mu$, $A = \frac{\mathbf{X}^T \mathbf{X}}{\tau^2}$ e $B = \Omega^{-1}$. Além disso, seguem algumas observações:

- 1) $\mathbf{X}^T \mathbf{X}$ é positiva-definida, pois assumimos que X tem posto completo.
- 2) $\frac{\mathbf{X}^T \mathbf{X}}{\tau^2}$ também é positiva definida pois, por propriedade, se r é uma constante estritamente positiva e M é uma matriz positiva-definida então $r \times M$ é positiva-definida.
- 3) Ω é positiva-definida, portanto Ω invertível e Ω^{-1} também é positiva-definida.
- 4) $\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1}$ é positiva-definida pois, por propriedade, se A e B são matrizes positiva-definidas de mesma dimensão, então $A+B$ também é positiva-definida.
- 5) Pela observação 4, $\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1}$ é invertível e, portanto, não-singular.

Então, pelo Lema A.2.1,

$$\begin{aligned} & (\beta - \hat{\beta})^T \frac{\mathbf{X}^T \mathbf{X}}{\tau^2} (\beta - \hat{\beta}) + (\beta - \mu)^T \Omega^{-1} (\beta - \mu) = \\ & = (\beta - c)^T \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right) (\beta - c) + (\hat{\beta} - \mu)^T C (\hat{\beta} - \mu), \end{aligned} \quad (\text{A.101})$$

onde

$$\zeta = c = \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} \hat{\beta} + \Omega^{-1} \mu \right), \quad (\text{A.102})$$

$$C = \left[\tau^2 (\mathbf{X}^T \mathbf{X})^{-1} + \Omega \right]^{-1}. \quad (\text{A.103})$$

Sendo assim, considerando (A.101) em (A.98), obtemos que

$$\begin{aligned} f(\beta | y) & \propto e^{-\frac{1}{2} \{ (\beta - \hat{\beta})^T \frac{\mathbf{X}^T \mathbf{X}}{\tau^2} (\beta - \hat{\beta}) + (\beta - \mu)^T \Omega^{-1} (\beta - \mu) \}} \Phi[\lambda^T \omega^{-1} (\beta - \mu)] \\ & \propto e^{-\frac{1}{2} \{ (\beta - c)^T \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right) (\beta - c) + (\hat{\beta} - \mu)^T C (\hat{\beta} - \mu) \}} \Phi[\lambda^T \omega^{-1} (\beta - \mu)] \\ & \propto e^{-\frac{1}{2} \{ (\beta - c)^T \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right) (\beta - c) \}} \Phi[\lambda^T \omega^{-1} (\beta - \mu)] \\ & \propto \phi_p \left(\zeta, \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} \right) \Phi[\lambda^T \omega^{-1} (\beta - \mu)], \end{aligned} \quad (\text{A.104})$$

que é o núcleo de uma distribuição normal-assimétrica multivariada geral. Dessa forma, conseguimos mostrar que a distribuição *a posteriori* para β é

$$GSN_{p,1} \left(\zeta, \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1}, \lambda^T \omega^{-1}, \lambda^T \omega^{-1} \mu, 1 \right),$$

onde ζ é como especificado em (A.102) e ω é uma matriz diagonal formada pelos desvios-padrões de Ω .

Para obtermos a média e variância *a posteriori* para o vetor de parâmetros regressores β , basta considerar os resultados apresentados em (2.10) e (2.12), assim como saber que \mathcal{F} , \mathcal{G} e \mathcal{H} são unidimensionais e que

$$\eta = \frac{\Phi_1^*(\mathcal{F}; \mathcal{G}, \mathcal{H})}{\Phi_1(\mathcal{F}; \mathcal{G}, \mathcal{H})} = \frac{\phi_1(\mathcal{F}; \mathcal{G}, \mathcal{H})}{\Phi_1(\mathcal{F}; \mathcal{G}, \mathcal{H})} \quad (\text{A.105})$$

e

$$\Lambda = \frac{\Phi_1^{**}(\mathcal{F}; \mathcal{G}, \mathcal{H})}{\Phi_1(\mathcal{F}; \mathcal{G}, \mathcal{H})} = -\frac{\phi_1(\mathcal{F}; \mathcal{G}, \mathcal{H})}{\Phi_1(\mathcal{F}; \mathcal{G}, \mathcal{H})} \left[\frac{\mathcal{F} - \mathcal{G}}{\mathcal{H}} \right]. \quad (\text{A.106})$$

◆

Prova da Proposição 4.2.2:

Para demonstrarmos a parte (i), devemos considerar $\lambda = \mathbf{0}$ em (4.10) para mostrarmos que $\widehat{\beta}_0 = \zeta$ e lembrarmos que η é estritamente positivo, assim

$$\begin{aligned} d_M &= \left\| E_\lambda[\beta | \mathbf{Y}] - \widehat{\beta}_0 \right\| \\ &= \eta \left\| \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} (\lambda^T \omega^{-1})^T \right\|. \end{aligned} \quad (\text{A.107})$$

Para a parte (ii), precisamos notar que, quando $\lambda = \mathbf{0}$ em 4.11, $\widehat{\tau}_0 = \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1}$, assim

$$\begin{aligned} d_V &= \left\| \text{Var}_\lambda[\beta | \mathbf{Y}] - \widehat{\tau}_0 \right\| \quad (\text{A.108}) \\ &= \left\| - \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} (\lambda^T \omega^{-1})^T \left\{ \frac{\mathcal{F} - \mathcal{G}}{\mathcal{H}} + \eta \right\} \lambda^T \omega^{-1} \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} \eta \right\|. \end{aligned}$$

Sabemos, por propriedade, que se α é um número real e \mathbf{x} é um vetor então a norma euclidiana $\|\alpha \mathbf{x}\|$ é igual a $|\alpha| \|\mathbf{x}\|$. Como η , \mathcal{F} , \mathcal{G} e \mathcal{H} são unidimensionais, então

$$d_V = \left| \eta \left\{ \frac{\mathcal{F} - \mathcal{G}}{\mathcal{H}} + \eta \right\} \right| \left\| \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} (\boldsymbol{\lambda}^T \boldsymbol{\omega}^{-1})^T \boldsymbol{\lambda}^T \boldsymbol{\omega}^{-1} \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} \right\|.$$

Além disso,

$$\begin{aligned} \left| \eta \left\{ \frac{\mathcal{F} - \mathcal{G}}{\mathcal{H}} + \eta \right\} \right| &= \left| \frac{\phi(\mathcal{F}; \mathcal{G}, \mathcal{H}) \left\{ \frac{\mathcal{F} - \mathcal{G}}{\mathcal{H}} + \frac{\phi(\mathcal{F}; \mathcal{G}, \mathcal{H})}{\Phi(\mathcal{F}; \mathcal{G}, \mathcal{H})} \right\}}{\Phi(\mathcal{F}; \mathcal{G}, \mathcal{H}) \left\{ \frac{\mathcal{F} - \mathcal{G}}{\mathcal{H}} + \frac{\phi(\mathcal{F}; \mathcal{G}, \mathcal{H})}{\Phi(\mathcal{F}; \mathcal{G}, \mathcal{H})} \right\}} \right| \\ &= \left| \frac{1}{\sqrt{\mathcal{H}}} \frac{\phi\left(\frac{\mathcal{F} - \mathcal{G}}{\sqrt{\mathcal{H}}}\right)}{\Phi\left(\frac{\mathcal{F} - \mathcal{G}}{\sqrt{\mathcal{H}}}\right)} \left\{ \frac{\mathcal{F} - \mathcal{G}}{\mathcal{H}} + \frac{1}{\sqrt{\mathcal{H}}} \frac{\phi\left(\frac{\mathcal{F} - \mathcal{G}}{\sqrt{\mathcal{H}}}\right)}{\Phi\left(\frac{\mathcal{F} - \mathcal{G}}{\sqrt{\mathcal{H}}}\right)} \right\} \right| \\ &= \frac{1}{\mathcal{H}} \left| \frac{\phi\left(\frac{\mathcal{F} - \mathcal{G}}{\sqrt{\mathcal{H}}}\right)}{\Phi\left(\frac{\mathcal{F} - \mathcal{G}}{\sqrt{\mathcal{H}}}\right)} \left\{ \frac{\mathcal{F} - \mathcal{G}}{\sqrt{\mathcal{H}}} + \frac{\phi\left(\frac{\mathcal{F} - \mathcal{G}}{\sqrt{\mathcal{H}}}\right)}{\Phi\left(\frac{\mathcal{F} - \mathcal{G}}{\sqrt{\mathcal{H}}}\right)} \right\} \right|. \end{aligned} \quad (\text{A.109})$$

Mas, por (A.25), sabemos que

$$\frac{\phi(y)}{\Phi(y)} y + \frac{\phi(y)^2}{\Phi(y)^2} > 0, \quad \forall y, \quad (\text{A.110})$$

então

$$\begin{aligned} \left| \eta \left\{ \frac{\mathcal{F} - \mathcal{G}}{\mathcal{H}} + \eta \right\} \right| &= \frac{1}{\mathcal{H}} \frac{\phi\left(\frac{\mathcal{F} - \mathcal{G}}{\sqrt{\mathcal{H}}}\right)}{\Phi\left(\frac{\mathcal{F} - \mathcal{G}}{\sqrt{\mathcal{H}}}\right)} \left\{ \frac{\mathcal{F} - \mathcal{G}}{\sqrt{\mathcal{H}}} + \frac{\phi\left(\frac{\mathcal{F} - \mathcal{G}}{\sqrt{\mathcal{H}}}\right)}{\Phi\left(\frac{\mathcal{F} - \mathcal{G}}{\sqrt{\mathcal{H}}}\right)} \right\} \\ &= \frac{1}{\sqrt{\mathcal{H}}} \frac{\phi\left(\frac{\mathcal{F} - \mathcal{G}}{\sqrt{\mathcal{H}}}\right)}{\Phi\left(\frac{\mathcal{F} - \mathcal{G}}{\sqrt{\mathcal{H}}}\right)} \left\{ \frac{\mathcal{F} - \mathcal{G}}{\mathcal{H}} + \frac{1}{\sqrt{\mathcal{H}}} \frac{\phi\left(\frac{\mathcal{F} - \mathcal{G}}{\sqrt{\mathcal{H}}}\right)}{\Phi\left(\frac{\mathcal{F} - \mathcal{G}}{\sqrt{\mathcal{H}}}\right)} \right\} \\ &= \eta \left\{ \frac{\mathcal{F} - \mathcal{G}}{\mathcal{H}} + \eta \right\}. \end{aligned} \quad (\text{A.111})$$

Vale ressaltar que $\mathcal{H} = 1 + \boldsymbol{\lambda}^T \boldsymbol{\omega}^{-1} \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1} (\boldsymbol{\lambda}^T \boldsymbol{\omega}^{-1})^T$ é um número escalar estritamente positivo, pois \mathcal{H} é unidimensional e sabemos, pelas observações 3 e 4 apresentadas na demonstração da Proposição 4.2.1, que $\mathcal{B} = \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1} \right)^{-1}$ é positiva-definida.

Se uma matriz \mathcal{B} de dimensão $p \times p$ é positiva-definida, então para todo vetor x de dimensão $p \times 1$ temos que $\mathbf{x}^T \mathcal{B} \mathbf{x} > 0$, exceto quando o vetor \mathbf{x} for identicamente

nulo. Em particular, esta propriedade vale quando $\mathbf{x}^T = \boldsymbol{\lambda}^T \boldsymbol{\omega}^{-1}$, logo \mathcal{H} é um número estritamente positivo.

◆

Prova do Corolário 4.2.1:

$$\text{Var}_\lambda[\beta | \mathbf{Y}] = \frac{\sigma^2 \tau^2}{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2} \left\{ 1 - \frac{\lambda^2 \tau^2}{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2 (1 + \lambda^2)} \frac{\phi(y)}{\Phi(y)} \times \left[y + \frac{\phi(y)}{\Phi(y)} \right] \right\}. \quad (\text{A.112})$$

Como $0 \leq \lambda^2 \tau^2 \leq \sigma^2 \sum_{i=1}^n x_i^2 + \tau^2 (1 + \lambda^2)$, então $0 \leq \frac{\lambda^2 \tau^2}{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2 (1 + \lambda^2)} \leq 1$.

Por (A.24) e (A.25), temos que

$$0 < \frac{\phi(y)}{\Phi(y)} \left[y + \frac{\phi(y)}{\Phi(y)} \right] < 1. \quad (\text{A.113})$$

Logo, $0 \leq \text{Var}_\lambda[\beta | \mathbf{Y}] \leq \frac{\sigma^2 \tau^2}{\sigma^2 \sum_{i=1}^n x_i^2 + \tau^2}$.

◆

Prova da Proposição 4.3.1:

Considere $m(\mathbf{y})$ a distribuição marginal ou preditiva de \mathbf{Y} , ou seja,

$$m(\mathbf{y}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_0^{\infty} f(\mathbf{y} | \boldsymbol{\beta}, \tau^2) f(\boldsymbol{\beta}) f(\tau^2) d\tau^2 d\boldsymbol{\beta}. \quad (\text{A.114})$$

É sabido que a distribuição *a posteriori* é própria se e somente se $0 < m(\mathbf{y}) < \infty$. Dessa forma, vamos mostrar que no nosso contexto a distribuição preditiva de \mathbf{Y} tem tal característica. Assim, por (A.114),

$$m(\mathbf{y}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_0^{\infty} \left(\frac{1}{2\pi\tau^2} \right)^{\frac{n}{2}} e^{-\frac{1}{2\tau^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} 2 \phi_p(\boldsymbol{\beta}; \boldsymbol{\mu}, \Omega) \times \Phi[\boldsymbol{\lambda}^T \boldsymbol{\omega}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu})] \frac{1}{\tau^2} d\tau^2 d\boldsymbol{\beta}. \quad (\text{A.115})$$

Note que $0 < \phi_p(\boldsymbol{\beta}; \boldsymbol{\mu}, \Omega) \leq \frac{1}{(2\pi)^{p/2} |\Omega|^{1/2}}$ e $0 < \Phi[\boldsymbol{\lambda}^T \boldsymbol{\omega}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu})] < 1$. Além disso, como definimos Ω uma matriz positiva-definida, temos que $|\Omega|^{\frac{1}{2}} > 0$ e

$$\begin{aligned}
m(\mathbf{y}) &\leq \frac{2}{(2\pi)^{p/2} |\Omega|^{1/2}} \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_0^{\infty} \left(\frac{1}{\tau}\right)^{n+2} \phi_n \left[\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}}{\tau} \right] d\tau^2 d\boldsymbol{\beta}}_A \\
&< \infty,
\end{aligned} \tag{A.116}$$

pois a integral expressa em A é finita, uma vez que corresponde a distribuição preditiva de \mathbf{X} quando se considera o modelo de regressão linear normal dado por $Y_1, \dots, Y_n \mid \boldsymbol{\beta}, \tau^2 \sim N_n(\mathbf{X}\boldsymbol{\beta}, \tau^2 I_n)$ e a distribuição *a priori* não-informativa usual $h(\boldsymbol{\beta}, \tau^2) \propto \frac{1}{\tau^2}$, em que I_n corresponde a matriz identidade de ordem n .

◆

Prova da Proposição 4.3.2:

A função densidade de probabilidade *a posteriori* conjunta é dada por

$$\begin{aligned}
f(\boldsymbol{\beta}, \tau^2 \mid \mathbf{y}) &\propto f(\mathbf{y} \mid \boldsymbol{\beta}, \tau^2) f(\boldsymbol{\beta}) f(\tau^2) \\
&\propto \left(\frac{1}{2\pi\tau^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\tau^2} [ks^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})]} \phi_p(\boldsymbol{\beta}; \boldsymbol{\mu}, \Omega) \times \\
&\quad \times \Phi[\boldsymbol{\lambda}^T \boldsymbol{\omega}^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu})] \frac{1}{\tau^2} \\
&\propto \left(\frac{1}{\tau^2}\right)^{\frac{n}{2}+1} e^{-\frac{1}{2\tau^2} [ks^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})]} \Phi[\boldsymbol{\lambda}^T \boldsymbol{\omega}^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu})] \times \\
&\quad \times e^{-\frac{1}{2} [(\boldsymbol{\beta} - \boldsymbol{\mu})^T \boldsymbol{\Omega}^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu})]}. \tag{A.117}
\end{aligned}$$

Observe que

$$f(\tau^2 \mid \mathbf{y}, \boldsymbol{\beta}) \propto \left(\frac{1}{\tau^2}\right)^{\frac{n}{2}+1} e^{-\frac{1}{2\tau^2} [ks^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})]}, \tag{A.118}$$

que é o núcleo de uma distribuição $GI\left(\frac{n}{2}, \frac{ks^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}{2}\right)$, onde $K = n - p$, $ks^2 = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$ é conhecido como soma dos quadrados residual, $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ é a estimativa de mínimos quadrados (ou de máxima verossimilhança) de $\boldsymbol{\beta}$.

Reescrevendo a função de densidade conjunta em A.117 através do resultado A.101, temos que

$$\begin{aligned}
f(\boldsymbol{\beta}, \tau^2 \mid \mathbf{y}) &\propto \left(\frac{1}{\tau^2}\right)^{\frac{n}{2}+1} e^{-\frac{1}{2} \{(\boldsymbol{\beta} - \boldsymbol{\zeta})^T \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \boldsymbol{\Omega}^{-1}\right) (\boldsymbol{\beta} - \boldsymbol{\zeta}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\mu})^T \mathbf{C} (\hat{\boldsymbol{\beta}} - \boldsymbol{\mu})\}} \times \\
&\quad \times e^{-\frac{1}{2\tau^2} ks^2} \Phi[\boldsymbol{\lambda}^T \boldsymbol{\omega}^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu})]. \tag{A.119}
\end{aligned}$$

e, portanto, a distribuição condicional completa *a posteriori* para β é

$$f(\beta | y, \tau^2) \propto e^{-\frac{1}{2}\{(\beta-\zeta)^T \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1}\right) (\beta-\zeta)\}} \Phi[\lambda^T \omega^{-1}(\beta - \mu)], \quad (\text{A.120})$$

que é o núcleo de uma distribuição $GSN_{p,1} \left(\zeta, \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \Omega^{-1}\right)^{-1}, \lambda^T \omega^{-1}, \lambda^T \omega^{-1} \mu, 1 \right)$, com ζ como especificado no enunciado da proposição.

◆

Prova da Proposição 4.3.3:

Sabemos que a distribuição marginal *a posteriori* do vetor de parâmetros regressores é dada por:

$$f(\beta | \mathbf{Y}) = \int_0^\infty f(\beta, \tau^2 | \mathbf{Y}) d\tau^2. \quad (\text{A.121})$$

Utilizando uma expressão equivalente a apresentada em (A.117) para a função de densidade conjunta *a posteriori* de β e τ^2 , temos que

$$\begin{aligned} f(\beta | \mathbf{Y}) &\propto \int_0^\infty \left(\frac{1}{\tau^2}\right)^{\frac{n}{2}+1} e^{-\frac{1}{2\tau^2}[Ks^2+(\beta-\hat{\beta})^T \mathbf{X}^T \mathbf{X}(\beta-\hat{\beta})]} \phi_p(\beta; \mu, \Omega) \times \\ &\quad \times \Phi[\lambda^T \omega^{-1}(\beta - \mu)] d\tau^2 \\ &\propto \phi_p(\beta; \mu, \Omega) \Phi[\lambda^T \omega^{-1}(\beta - \mu)] \times \\ &\quad \times \int_0^\infty \left(\frac{1}{\tau^2}\right)^{\frac{n}{2}+1} e^{-\frac{1}{2\tau^2}[Ks^2+(\beta-\hat{\beta})^T \mathbf{X}^T \mathbf{X}(\beta-\hat{\beta})]} d\tau^2 \\ &\propto \phi_p(\beta; \mu, \Omega) \Phi[\lambda^T \omega^{-1}(\beta - \mu)] \Gamma\left(\frac{n}{2}\right) \times \\ &\quad \left[\frac{Ks^2}{2} + \frac{(\beta - \hat{\beta})^T \mathbf{X}^T \mathbf{X}(\beta - \hat{\beta})}{2}\right]^{-\frac{n}{2}} \times \\ &\quad \times \int_0^\infty \frac{1}{\Gamma\left(\frac{n}{2}\right)} \left[\frac{Ks^2}{2} + \frac{(\beta - \hat{\beta})^T \mathbf{X}^T \mathbf{X}(\beta - \hat{\beta})}{2}\right]^{\frac{n}{2}} \left(\frac{1}{\tau^2}\right)^{\frac{n}{2}+1} \\ &\quad \times e^{-\frac{1}{2\tau^2}[Ks^2+(\beta-\hat{\beta})^T \mathbf{X}^T \mathbf{X}(\beta-\hat{\beta})]} d\tau^2. \quad (\text{A.122}) \end{aligned}$$

Note que o integrando presente na última expressão é a função densidade de probabilidade de uma distribuição gama com parâmetros $\frac{n}{2}$ e $\frac{Ks^2}{2} + \frac{(\beta-\hat{\beta})^T \mathbf{X}^T \mathbf{X}(\beta-\hat{\beta})}{2}$.

Portanto, a integral em τ^2 é igual a um. Logo,

$$f(\boldsymbol{\beta} | \mathbf{Y}) \propto \phi_p(\boldsymbol{\beta}, \boldsymbol{\mu}, \Omega) \Phi[\boldsymbol{\lambda}^T \boldsymbol{\omega}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu})] \Gamma\left(\frac{n}{2}\right) \left[\frac{ks^2}{2} + \frac{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}{2} \right]^{-\frac{n}{2}},$$

e o resultado (i) fica demonstrado.

Para obtermos a esperança e a variância da distribuição marginal *a posteriori* de τ^2 , precisamos utilizar o resultado apresentado na Proposição 4.3.2, em que obtivemos $\tau^2 | \boldsymbol{\beta}, \mathbf{Y} \sim GI\left(\frac{n}{2}, \frac{Ks^2}{2} + \frac{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}{2}\right)$. Assim,

$$E[\tau^2 | \boldsymbol{\beta}, \mathbf{Y}] = \frac{Ks^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}{n - 2}, \text{ para } n > 2, \quad (\text{A.123})$$

$$\text{Var}[\tau^2 | \boldsymbol{\beta}, \mathbf{Y}] = 2 \frac{[(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})]^2}{(n - 2)^2(n - 4)}, \text{ para } n > 4. \quad (\text{A.124})$$

Note que, apesar do abuso de notação, podemos escrever $[(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})]^2$ pois a expressão contida nos colchetes é unidimensional.

Lembrando que

$$E[\tau^2 | \mathbf{Y}] = E[E[\tau^2 | \boldsymbol{\beta}, \mathbf{Y}] | \mathbf{Y}] \quad (\text{A.125})$$

e

$$\text{Var}[\tau^2 | \mathbf{Y}] = E[\text{Var}[\tau^2 | \boldsymbol{\beta}, \mathbf{Y}] | \mathbf{Y}] + \text{Var}[E[\tau^2 | \boldsymbol{\beta}, \mathbf{Y}] | \mathbf{Y}], \quad (\text{A.126})$$

as expressões apresentadas para $E[\tau^2 | \mathbf{x}]$ e $\text{Var}[\tau^2 | \mathbf{x}]$ são obtidas através de simples manipulações algébricas de (A.125) e (A.126).

◆

A.3 Capítulo 5

Prova da Proposição 5.3.1:

Vamos determinar a função de concentração de $\theta^* \sim GSN_{1,1}(a, b^2, \lambda, \lambda\mu, \sigma^2)$ com respeito a $\theta_0^* \sim GSN_{1,1}(a, b^2, 0, 0, \sigma^2)$, onde $a = \frac{n\bar{x}\sigma^2 + \mu\tau^2}{n\sigma^2 + \tau^2}$ e $b^2 = \frac{\sigma^2\tau^2}{n\sigma^2 + \tau^2}$. Veja que o suporte de ambas as distribuições estão em \mathbb{R} e

$$h^*(\theta) = \frac{f^*(\theta)}{f_0^*(\theta)} = \frac{\Phi \left[\lambda \left(\frac{\theta - \mu}{\sigma} \right) \right]}{\Phi \left[\lambda \left(\frac{a - \mu}{\sqrt{\sigma^2 + \lambda^2 b^2}} \right) \right]}. \quad (\text{A.127})$$

Segundo a teoria desenvolvida, nós calculamos a função de concentração para qualquer $z \in [0, 1]$ encontrando o valor de y tal que $z = \Pi_0^*(\{\theta \in \Theta : h^*(\theta) \leq y\})$. Dessa forma, para $\lambda > 0$,

$$\begin{aligned} 0 < h^*(\theta) \leq y &\iff \\ 0 < \frac{\Phi \left[\lambda \left(\frac{\theta - \mu}{\sigma} \right) \right]}{\Phi \left[\lambda \left(\frac{a - \mu}{\sqrt{\sigma^2 + \lambda^2 b^2}} \right) \right]} \leq y &\iff \\ -\infty < \theta \leq \frac{\sigma}{\lambda} \Phi^{-1} \left\{ y \Phi \left[\frac{\lambda(a - \mu)}{\sqrt{\sigma^2 + \lambda^2 b^2}} \right] \right\} + \mu. & \quad (\text{A.128}) \end{aligned}$$

Assim, seja $d = \frac{\sigma}{\lambda} \Phi^{-1} \left\{ y \Phi \left[\frac{\lambda(a - \mu)}{\sqrt{\sigma^2 + \lambda^2 b^2}} \right] \right\} + \mu$, então

$$\begin{aligned} z &= \Pi_0^*(\{\theta \in \Theta : h^*(\theta) \leq y\}) \\ &= \Pi_0^*(\{\theta \in \Theta : -\infty < \theta \leq d\}) \\ &= \Phi \left(\frac{d - a}{b} \right) \\ &= \Phi \left[\frac{\sigma}{\lambda b} \Phi^{-1} \left\{ y \Phi \left[\frac{\lambda(a - \mu)}{\sqrt{\sigma^2 + \lambda^2 b^2}} \right] \right\} + \frac{1}{b}(\mu - a) \right]. \quad (\text{A.129}) \end{aligned}$$

Portanto, o y para o qual $z = \Phi \left[\frac{\sigma}{\lambda b} \Phi^{-1} \left\{ y \Phi \left[\frac{\lambda(a - \mu)}{\sqrt{\sigma^2 + \lambda^2 b^2}} \right] \right\} + \frac{1}{b}(\mu - a) \right]$ é dado por $\frac{\Phi \left[\frac{\lambda}{\sigma} \{ b \Phi^{-1}(z) - (\mu - a) \} \right]}{\Phi \left[\lambda \left(\frac{a - \mu}{\sqrt{\sigma^2 + \lambda^2 b^2}} \right) \right]}$. Então

$$\begin{aligned} \varphi_{\Pi}(z) &= \Pi^*(\{\theta \in \Theta : h^*(\theta) \leq y\}) \\ &= \Pi^* \left(\left\{ \theta \in \Theta : h^*(\theta) \leq \frac{\Phi \left[\frac{\lambda}{\sigma} \{ b \Phi^{-1}(z) - (\mu - a) \} \right]}{\Phi \left[\lambda \left(\frac{a - \mu}{\sqrt{\sigma^2 + \lambda^2 b^2}} \right) \right]} \right\} \right) \quad (\text{A.130}) \end{aligned}$$

e, dessa forma,

$$\begin{aligned} 0 < h^*(\theta) \leq \frac{\Phi \left[\frac{\lambda}{\sigma} \{ b \Phi^{-1}(z) - (\mu - a) \} \right]}{\Phi \left[\lambda \left(\frac{a - \mu}{\sqrt{\sigma^2 + \lambda^2 b^2}} \right) \right]} &\iff \\ -\infty < \theta \leq b \Phi^{-1}(z) + a. & \quad (\text{A.131}) \end{aligned}$$

Portanto, para $\lambda > 0$,

$$\varphi_{\Pi}(z) = \int_{-\infty}^{b\Phi^{-1}(z)+a} \frac{\phi(\theta; a, b^2)\Phi(\lambda\theta; \lambda\mu, \sigma^2)}{\Phi(\lambda a; \lambda\mu, \sigma^2 + \lambda^2 b^2)} d\theta \quad (\text{A.132})$$

Através da Proposição 2.2.1 conseguimos obter a função de distribuição acumulada de uma variável aleatória com distribuição normal-assimétrica multivariada geral a partir do Lema 2.1.1. Assim,

$$\varphi_{\Pi}(z) = \frac{1}{\Phi(\lambda a; \lambda\mu, \sigma^2 + \lambda^2 b^2)} \Phi_2 \left[\begin{pmatrix} b\Phi^{-1}(z) + a \\ \lambda a \end{pmatrix}; \begin{pmatrix} a \\ \lambda\mu \end{pmatrix}, \Omega \right], \quad (\text{A.133})$$

onde Φ_2 corresponde a função de distribuição acumulada da distribuição normal bivariada e $\Omega = \begin{bmatrix} b^2 & -\lambda b^2 \\ -\lambda b^2 & \sigma^2 + \lambda^2 b^2 \end{bmatrix}$ é a matriz de variância. Note que a matriz Ω é positiva-definida.

Fazendo uso de conhecidas propriedades da distribuição normal bivariada conseguimos mostrar o item I da proposição. A demonstração do item II é análoga a do item I e, portanto, será omitida.

◆

Prova da Proposição 5.3.2:

Note que

$$h^*(\theta) = \frac{f^*(\theta | x)}{f_0^*(\theta | x)} = \frac{C_0}{C_\lambda} \Phi \left[\lambda \left(\frac{\theta - \mu}{\sigma} \right) \right]. \quad (\text{A.134})$$

Para qualquer valor de $z \in [0, 1]$, nós calculamos a função de concentração encontrando o valor de y tal que $z = \Pi_0^*(\{\theta \in \Theta : h^*(\theta) \leq y\})$. Dessa forma, para $\lambda > 0$. Assim,

$$\begin{aligned} 0 < h^*(\theta) &\leq y && \iff \\ 0 < \frac{C_0}{C_\lambda} \Phi \left[\lambda \left(\frac{\theta - \mu}{\sigma} \right) \right] &\leq y && \iff \\ -\infty < \theta &\leq \frac{\sigma}{\lambda} \Phi^{-1} \left\{ \frac{C_\lambda}{C_0} y \right\} + \mu. \end{aligned} \quad (\text{A.135})$$

Assim, seja $d = \frac{\sigma}{\lambda} \Phi^{-1} \left\{ \frac{C_\lambda}{C_0} y \right\} + \mu$. Dessa forma,

$$\begin{aligned}
 z &= \Pi_0^*(\{\theta \in \Theta : h^*(\theta) \leq y\}) \\
 &= \Pi_0^*(\{\theta \in \Theta : -\infty < \theta \leq d\}) \\
 &= \int_{-\infty}^{\frac{\sigma}{\lambda} \Phi^{-1}\left\{\frac{C_\lambda}{C_0} y\right\} + \mu} \frac{1}{C_0} \phi\left(\frac{\theta - \mu}{\sigma}\right) \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2}\right]^{-\frac{n}{2}} d\theta. \quad (\text{A.136})
 \end{aligned}$$

Portanto, fixado z , preciso conhecer qual é o y para o qual a igualdade anterior ocorre. Chamemos este y de y_z^* , então

$$\begin{aligned}
 0 < h^*(\theta) \leq y_z^* &\iff \\
 -\infty < \theta \leq \frac{\sigma}{\lambda} \Phi^{-1}\left\{\frac{C_\lambda}{C_0} y_z^*\right\} + \mu. &\quad (\text{A.137})
 \end{aligned}$$

Logo,

$$\varphi_{\Pi}(z) = \int_{-\infty}^{\frac{\sigma}{\lambda} \Phi^{-1}\left\{\frac{C_\lambda}{C_0} y_z^*\right\} + \mu} \frac{1}{C_\lambda} \phi\left(\frac{\theta - \mu}{\sigma}\right) \Phi\left(\lambda \frac{\theta - \mu}{\sigma}\right) \left[\frac{nS^2}{2} + \frac{n(\bar{x} - \theta)^2}{2}\right]^{-\frac{n}{2}} d\theta. \quad (\text{A.138})$$

A demonstração do item II é análoga a do item I e, portanto, será omitida.

◆

Prova da Proposição 5.3.3: Será omitida, uma vez que é similar a demonstração da Proposição 5.3.2.

Apêndice B

Figuras

B.1 Situação I: $\bar{x} < 0$ e $\lambda \geq 0$.

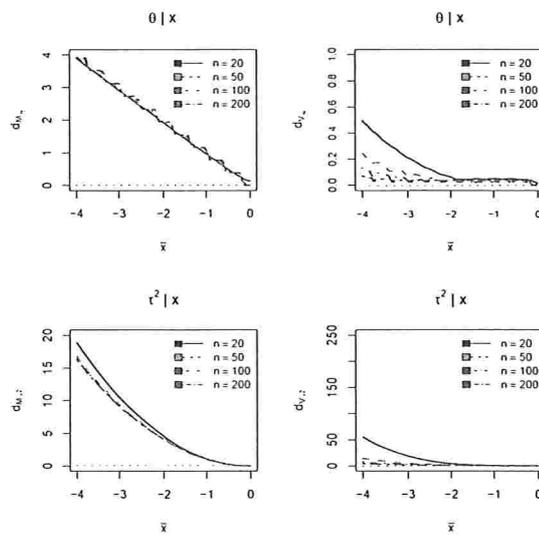
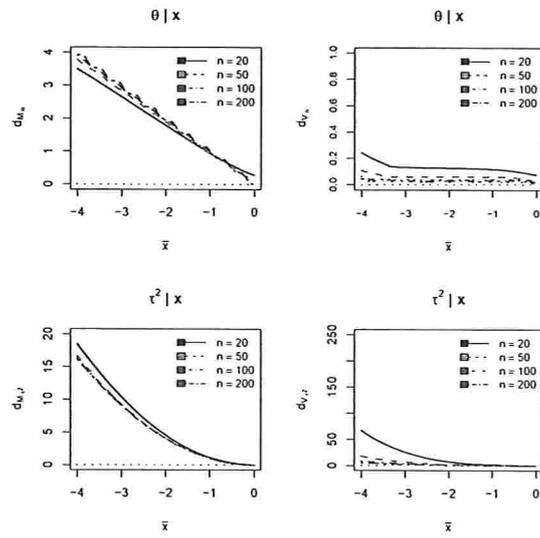
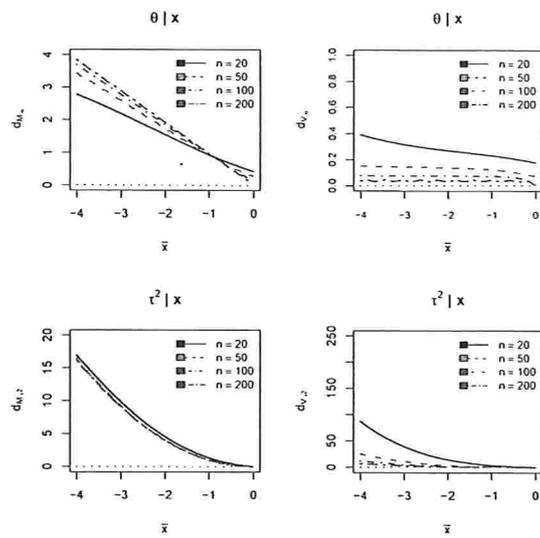


Figura B.1: $S^2 = 1$

Figura B.2: $S^2 = 3$ Figura B.3: $S^2 = 8$

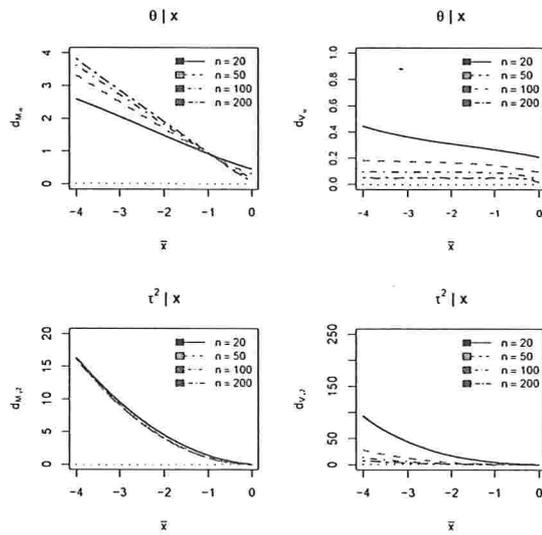


Figura B.4: $S^2 = 10$

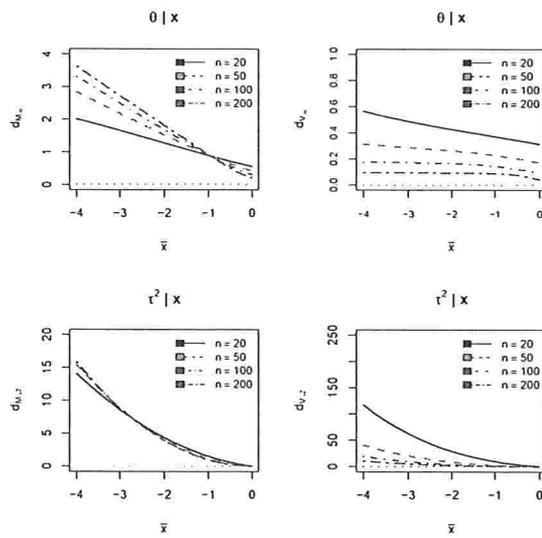


Figura B.5: $S^2 = 20$

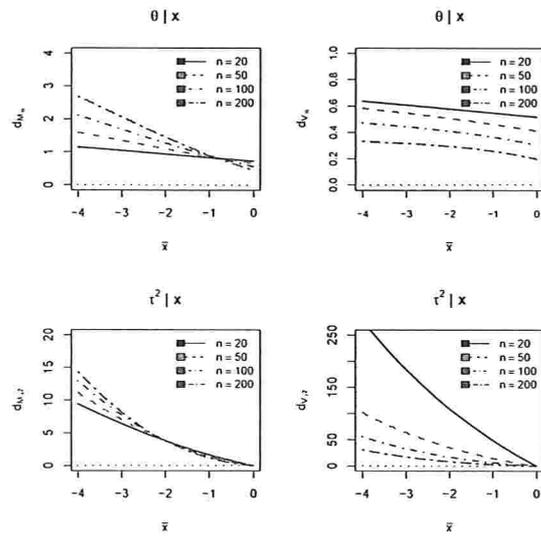


Figura B.6: $S^2 = 100$

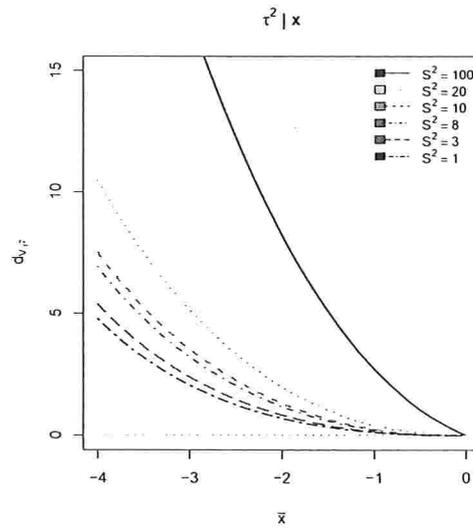


Figura B.7: $n = 200$

B.2 Situação II: $\bar{x} > 0$ e $\lambda \geq 0$.

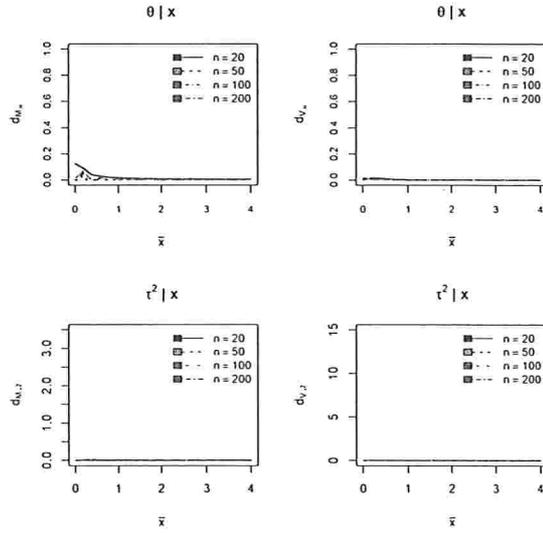


Figura B.8: $S^2 = 1$

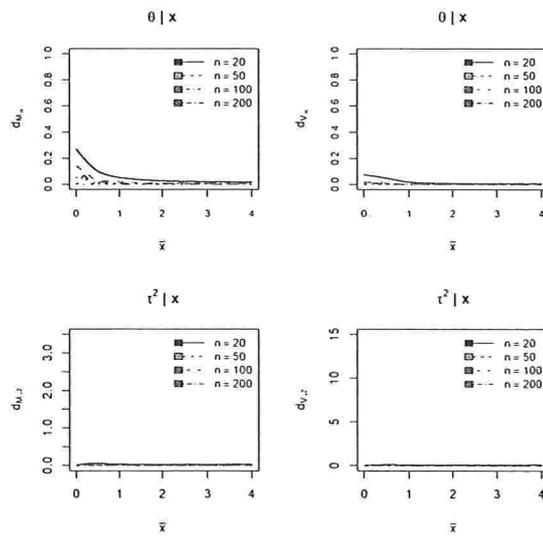


Figura B.9: $S^2 = 3$

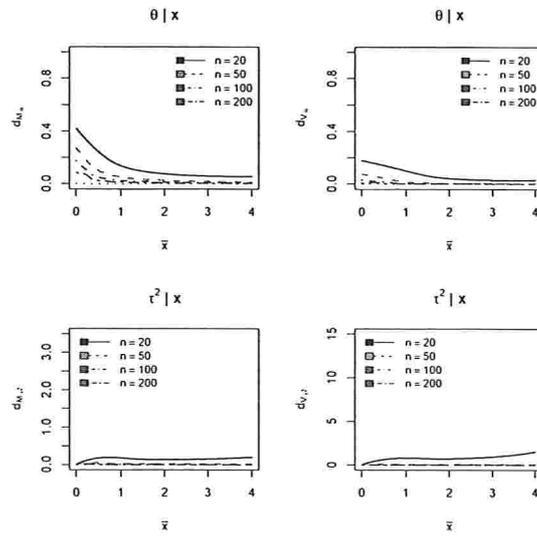


Figura B.10: $S^2 = 8$

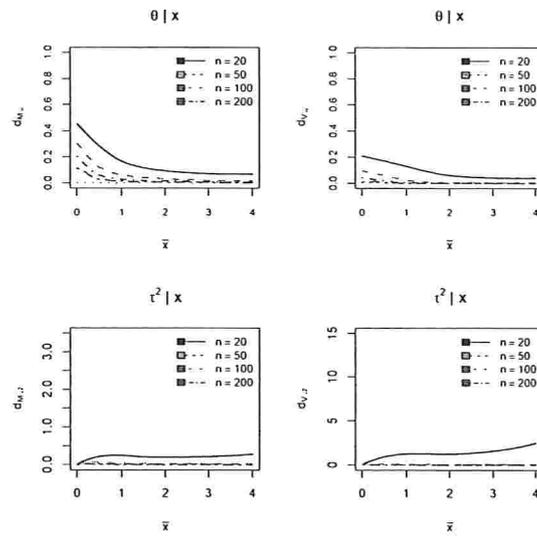


Figura B.11: $S^2 = 10$

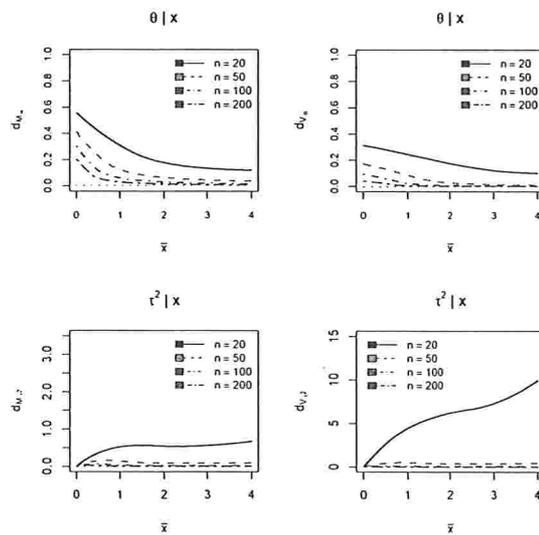


Figura B.12: $S^2 = 20$

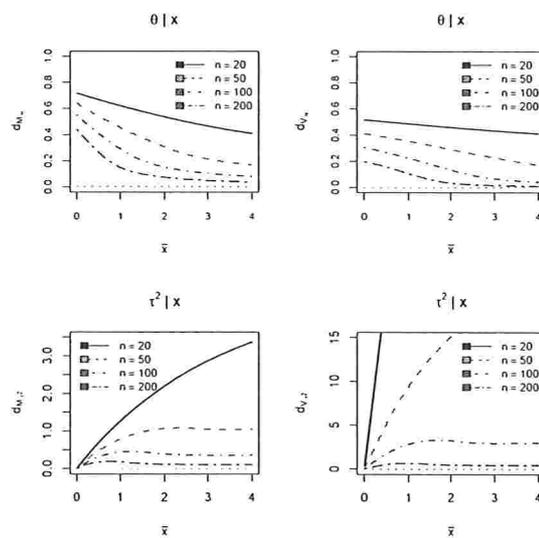


Figura B.13: $S^2 = 100$

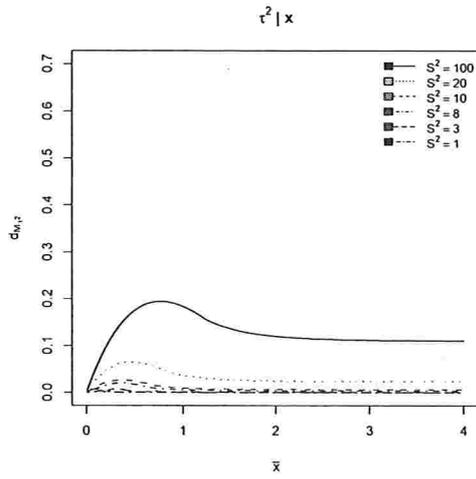


Figura B.14: $n = 100$

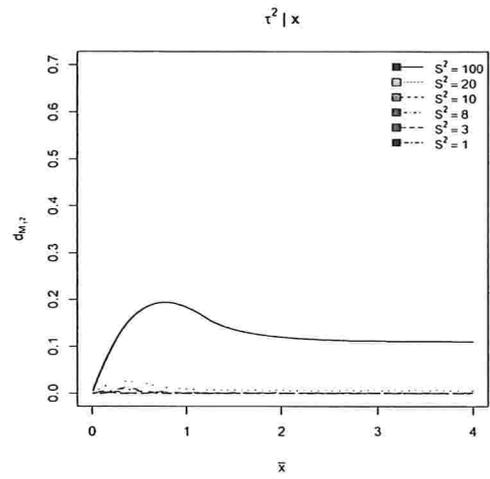


Figura B.15: $n = 200$.

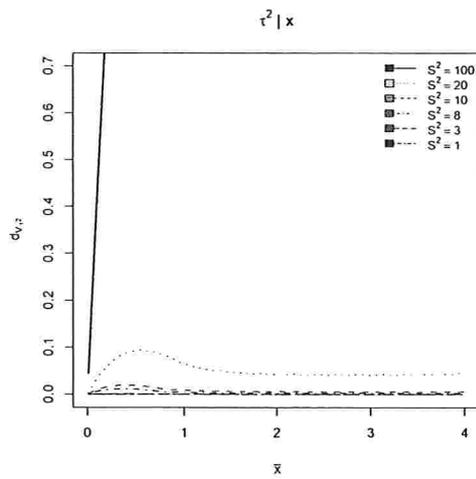


Figura B.16: $n = 100$

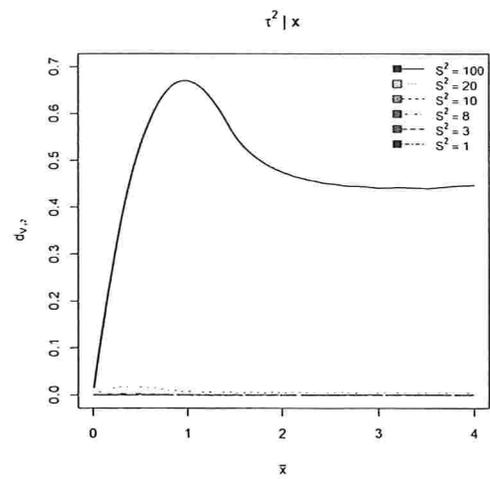


Figura B.17: $n = 200$

Referências Bibliográficas

- Abramowitz, M. & Stegun, I. A., eds (1972), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Washington, D. C.: U.S. Government Printing Office. 40
- Arellano-Valle, R. B. & Azzalini, A. (2006), 'On the unification of families of skew-normal distributions', *Scandinavian Journal of Statistics* **33**, 561–574. 7, 51
- Arellano-Valle, R. B., Branco, M. D. & Genton, M. G. (2006), 'A unified view on skewed distributions arising from selections', *The Canadian Journal of Statistics* **34**(4), 581–601. 9
- Arellano-Valle, R. B., Castro, L. M., Genton, M. G. & Gómez, H. W. (2008), 'Bayesian inference for shape mixtures of skewed distributions, with application to regression analysis', *Bayesian Analysis* **3**(3), 513–540. 61
- Arellano-Valle, R. B., del Pino, G. & Martín, E. S. (2002), 'Definition and probabilistic properties of skew-distributions', *Statistics and Probability Letters* **58**, 111–121. 7
- Arellano-Valle, R. B., Galea-Rojas, M. & Zuazola, P. I. (2000), 'Bayesian sensitivity analysis in elliptical linear regression models', *Journal of Statistical Planning and Inference* **86**, 175–199. 47
- Arellano-Valle, R. B. & Genton, M. G. (2005), 'On fundamental skew distributions', *Journal of Multivariate Analysis* **96**, 93–116. 7
- Arellano-Valle, R. B., Genton, M. G. & Loschi, R. H. (2009), 'Shape mixtures of multivariate skew-normal distributions', *Journal of Multivariate Analysis* **100**, 91–101. 19

- Azzalini, A. (1985), 'A class of distributions which includes the normal ones', *Scandinavian Journal of Statistics* **12**, 171–178. 4, 7, 15, 20, 49
- Azzalini, A. & Capitanio, A. (1999), 'Statistical applications of the multivariate skew-normal distribution', *Journal of the Royal Statistical Society: Series B* **61**(3), 579–602. 7, 15
- Azzalini, A. & Capitanio, A. (2003), 'Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution', *Journal of the Royal Statistical Society: Series B* **65**(2), 367–389. 4
- Azzalini, A. & Dalla-Valle, A. (1996), 'The multivariate skew-normal distribution', *Biometrika* **83**(4), 715–726. 4, 7
- Baricz, A. (2008), 'Mill's ratio: monotonicity patterns and functional inequalities', *Journal of Mathematical Analysis and Applications* **340**, 1362–1370. 66
- Berger, J. (1990), 'Robust Bayesian analysis: sensitivity to the prior', *Journal of Statistical Planning and Inference* **25**, 303–328. 4, 21
- Berger, J. (1993), An overview of robust Bayesian analysis, Technical report, Purdue University, Indiana, USA.
URL: <http://www.isds.duke.edu/~berger/papers/overview.ps> 4
- Berger, J. & Berliner, L. M. (1986), 'Robust Bayes and empirical Bayes analysis with ϵ -contaminated priors', *The Annals of Statistics* **14**(2), 461–486. 48
- Berger, J. O. (1984), The robust Bayesian viewpoint (with discussion), in J. B. Kadane, ed., 'Robustness of Bayesian Analysis', Amsterdam: North-Holland, pp. 63–124. 48
- Berger, J. O., Insua, D. R. & Ruggeri, F. (2000), Bayesian robustness, in D. R. Insua & F. Ruggeri, eds, 'Robust Bayesian Analysis', Nova York: Springer-Verlag. 1, 23
- Billingsley, P. (1986), *Probability and Measure*, Nova York: John Wiley & Sons. 78
- Box, G. E. P. & Tiao, G. C. (1973), *Bayesian Inference in Statistical Analysis*, Reading : Addison-Wesley, 1973. 121

- Capiński, M. & Kopp, E. (2004), *Measure, Integral and Probability*, Nova York: Springer-Verlag. 78
- Casella, G. (1983), 'Leverage and regression through the origin', *The American Statistician* 37(2), 147–152. 64
- Chaturvedi, A. (1996), 'Robust Bayesian analysis of the linear regression model', *Journal of Statistical Planning and Inference* 50, 175–186. 48
- Chib, S. & Tiwari, R. C. (1991), 'Robust Bayes analysis in normal linear regression with an improper mixture prior', *Communications in Statistics - Theory and Methods* 20(3), 807–829. 48
- Cifarelli, D. M. & Regazzini, E. (1987), 'On a general definition of concentration function', *Sankhya: The Indian Journal of Statistics* 49(Series B), 307–319. 75, 77, 79
- Domínguez-Molina, J. A., González-Farías, G. & Gupta, A. K. (2003), The multivariate closed skew normal distribution, Technical report, Bowling Green State University, Ohio, USA. 5, 8, 9
- Domínguez-Molina, J. A., González-Farías, G. & Ramos-Quiroga, R. (2004), Skew-normality in stochastic frontier analysis, in M. G. Genton, ed., 'Skew-Elliptical Distributions and Their Applications : A Journey Beyond Normality', Boca Raton: Chapman & Hall / CRC, pp. 223–241. 19
- Domínguez-Molina, J. A., González-Farías, G., Ramos-Quiroga, R. & Gupta, A. K. (2007), 'A matrix variate closed skew-normal distribution with applications to stochastic frontier analysis', *Communications in Statistics - Theory and Methods* 36, 1691–1703. 13, 14
- Fortini, S. & Ruggeri, F. (1994), 'On defining neighbourhoods of measures through the concentration function', *Sankhya: The Indian Journal of Statistics* 56(Series A), 444–457. 79
- Fortini, S. & Ruggeri, F. (1995), 'Concentration function and sensitivity to the prior', *Journal of the Italian Statistical Society* 3, 283–297. 75, 80, 81, 84

- Fortini, S. & Ruggeri, F. (2000), On the use of the concentration function in Bayesian robustness, *in* D. R. Insua & F. Ruggeri, eds, 'Robust Bayesian Analysis', Nova York: Springer-Verlag, pp. 109–126. 77, 78, 81, 83, 84
- Genz, A. & Kass, R. E. (1989), 'Bayespack: A collection of numerical integration software for bayesian analysis', *Versão de 09 de junho de 2011* .
URL: <http://www.math.wsu.edu/faculty/genz/papers/bayespack.pdf>
74
- Genz, A. & Kass, R. E. (1997), 'Subregion-adaptive integration of functions having a dominant peak', *Journal of Computational and Graphical Statistics* 6(1), 92–111.
74
- González-Farías, G., Domínguez-Molina, J. A. & Gupta, A. K. (2004a), 'Additive properties of skew normal random vectors', *Journal of Statistical Planning and Inference* 126, 521–534. 14
- González-Farías, G., Domínguez-Molina, J. A. & Gupta, A. K. (2004b), The closed skew-normal distribution, *in* M. G. Genton, ed., 'Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality', Boca Raton: Chapman & Hall / CRC, pp. 25–42. 9, 13, 14
- Good, I. J. (1965), *The Estimation of Probabilities*, Cambridge: M.I.T. Press. 48
- Gordon, R. D. (1941), 'Values of Mill's ratio of area bounding ordinate and of the normal probability integral for large values of the argument', *The Annals of Mathematical Statistics* 12(3), 364–366. 66
- Gupta, A. K., González-Farías, G. & Domínguez-Molina, J. A. (2004), 'A multivariate skew normal distribution', *Journal of Multivariate Analysis* 89, 181–190. 4, 5, 8, 15, 19
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. & Stahel, W. A. (2005), *Robust Statistics: The approach based on influence functions*, New Jersey: John Wiley & Sons. 47
- Henderson, H. V. & Velleman, P. F. (1981), 'Building multiple regression models interactively', *Biometrics* 37, 391–411. 73

- Insua, D. R. & Ruggeri, F., eds (2000), *Robust Bayesian Analysis*, Nova York: Springer-Verlag. 3
- Iversen, D. H. (2010), Closed skew distributions: simulation, inversion and parameter estimation, Master's thesis, NTNU, Trondheim, Noruega. 9
- Johnson, N., Kotz, S. & Balakrishnan, N. (1994), *Continuous univariate distributions*, Vol. 1, Nova York: John Wiley & Sons. 102
- Marshall, A. W. & Olkin, I. (1979), *Inequalities: theory of majorizations and its applications*, New York: Academic Press. 75
- Mukhopadhyay, S. & Vidakovic, B. (1995), 'Efficiency of linear Bayes rules for a normal mean: skewed priors class', *The Statistician* 44, 389–397. 4
- O'Hagan, A. (1994), *Bayesian Inference - Kendall's Advanced Theory of Statistics*, Vol. 2B, Inglaterra: Edward Arnold. 2, 3, 31, 49
- Paulino, C. D., Turkman, M. A. A. & Murteira, B. (2003), *Estatística Bayesiana*, Lisboa: Fundação Calouste Gulbenkian. 38, 121
- Peixoto, L. L. (2008), Quadratura de gauss iterativa com base nos polinômios ortogonais clássicos, Master's thesis, CEFET-MG, Minas Gerais, Brasil. 40
- Rodríguez, C. L. B. (2005), Inferência bayesiana no modelo normal assimétrico, Master's thesis, IME-USP, São Paulo, Brasil. 4, 20, 82
- Rousseeuw, P. J. & Leroy, A. M. (2003), *Robust Regression and Outlier Detection*, New Jersey: John Wiley & Sons. 47
- Sahu, S. K., Dey, D. K. & Branco, M. D. (2003), 'A new class of multivariate skew distributions with applications to Bayesian regression models', *The Canadian Journal of Statistics* 31(2), 129–150. 4, 7
- Sampford, M. R. (1953), 'Some inequalities on Mill's ratio and related functions', *The Annals of Mathematical Statistics* 24(1), 130–132. 100, 105
- Scarsini, M. (1990), An ordering of dependence, in H. W. Block, A. R. Sampson & T. H. Savits, eds, 'Topics in Statistical Independence', CA: Institute of Mathematical Statistics, pp. 403–414. 80

- van der Linde, A. (2007), 'Local influence on posterior distributions under multiplicative modes of perturbation', *Bayesian Analysis* 2(2), 319–332. 9, 11, 3, 5, 21
- Zellner, A. (1986), On assessing prior distributions and Bayesian regression analysis with g -prior distribution, *in* P. K. Goel & A. Zellner, eds, 'Bayesian Inference and Decision Techniques', Amsterdam: North-Holland. 48