

UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG
INSTITUTO DE CIÊNCIAS EXATAS - ICE_x
DEPARTAMENTO DE ESTATÍSTICA

**INFERÊNCIA EM FAMÍLIAS
ESTENDIDAS DE DISTRIBUIÇÕES NORMAIS**

Gustavo Henrique Mitraud Assis Rocha
Orientadora: Rosangela Helena Loschi
Co-orientador: Reinaldo Boris Arellano-Valle

Dissertação de mestrado

Inferência em famílias estendidas de distribuições normais

Gustavo Henrique Mitraud Assis Rocha

19/02/2010

*Dedico este trabalho
aos meus pais, Ozeres (+) e Mary,
e à minha irmã Rita e sobrinha Paloma.*

Agradecimentos

Agradeço a Deus por tudo que tem me proporcionado e por mais uma conquista.

Aos meus pais, Ozeres (*in memoriam*) e Mary, à minha irmã Rita e à minha sobrinha Paloma. Obrigado pelo apoio em relação aos estudos.

À minha família e em especial aos meus primos Alexandre, Bruno e Daniel. Obrigado por abrirem a porta e pelos momentos de descontração.

À minha orientadora, Professora Rosangela Helena Loschi, pelo apoio, paciência e auxílio durante os anos de estudo na UFMG. Ao meu co-orientador Professor Reinaldo Boris Arellano-Valle (Puc-Chile) pela confiança.

Aos professores do Departamento de Estatística da UFMG, pelos ensinamentos concedidos.

Aos membros da banca examinadora, Prof. Francisco Cribari Neto (UFPE), Prof. Márcia D'Elia Branco (IME-USP) e Prof. Denise Duarte Scarpa Magalhães Alves (UFMG), pela leitura, correções e sugestões da dissertação. Também agradeço à Prof. Lourdes Coral Contreras Montenegro, que participou da banca examinadora em meu exame de qualificação e contribuiu com sugestões para a versão final da dissertação.

À CAPES pela bolsa de mestrado, à FAPEMIG por diversos apoios financeiros prestados para participação em eventos e ao PIBIC-CNPq pela bolsa de iniciação científica, importante para o surgimento do interesse em pesquisas.

A todos os amigos e amigas cujas amizades foram feitas na UFMG - graduação e mestrado - e que se mantem até os dias de hoje. Aos amigos que estão e que já estiveram no grupo de e-mails *ggrupodosamigos*, muitas discussões e contribuições, em especial nas noites de BH. Aos amigos e amigas feitos durante o período vida. Ao Lucas, que tanto me auxilia e se faz presente.

A todos, o meu muito obrigado!

Gustavo Henrique Mitraud Assis Rocha

Resumo

Neste trabalho há revisões sobre as famílias de distribuições normais assimétricas (Azzalini, 1985), normais bimodais (Arellano-Valle *et al.*, 2008) e normais bimodais assimétricas (Elal-Oliveiro *et al.*, 2009). Serão considerados os estimadores via método dos momentos, de máxima verossimilhança, as distribuições *a posteriori* e as densidades preditivas. Em cada uma das famílias consideradas serão estabelecidas condições para a existência de estimadores de máxima verossimilhança para os parâmetros. Serão construídos os algoritmos EM para cada um deles. Estudos Monte Carlo serão feitos em dados simulados para verificar a qualidade dos estimadores de máxima verossimilhança e dos estimadores bayesianos esperança e moda *a posteriori*. Outro estudo Monte Carlo é realizado para averiguar as mudanças no comportamento dos estimadores bayesianos quando se altera a variabilidade *a priori* de cada um dos parâmetros. Por fim, serão realizadas inferências nos dados de fronteira de Azzalini.

Em relação à inferência nota-se que, tanto as distribuições *a posteriori* quanto os estimadores de máxima verossimilhança para os parâmetros de assimetria e forma da família normal bimodal assimétrica, podem ser obtidos considerando famílias mais simples.

No estudo Monte Carlo para avaliar a qualidade dos estimadores considerados verifica-se que, para amostras pequenas, o melhor estimador para o parâmetro de forma é a moda *a posteriori* quando há uma bimodalidade mais evidente e o estimador de máxima verossimilhança caso os dados sejam unimodais. Para estimar o parâmetro de assimetria o melhor estimador é a moda *a posteriori*. Para amostras grandes o melhor estimador é a moda *a posteriori* para o parâmetro de forma e o estimador de máxima verossimilhança para o parâmetro de assimetria. Para avaliar a sensibilidade dos estimadores bayesianos considerando diferentes variabilidades *a priori* nota-se que a utilização de distribuições *a priori* com alta variabilidade sempre resulta em boas estimativas quando é considerada a moda *a posteriori* como estimador dos parâmetros. Quando as distribuições *a priori* são centradas em valores mais próximos dos valores utilizados para geração das amostras os resultados se mostram bons.

Na análise dos dados de fronteira de Azzalini observa-se que os dados podem ser considerados como vindos de uma distribuição normal bimodal assimétrica padrão, uma vez que, considerando os resultados obtidos, as distribuições obtidas parecem se adequar bem aos dados e fornece resultados similares ao encontrado na literatura usando modelos menos parcimoniosos.

Sumário

1	Introdução	6
2	Classes estendidas de distribuições normais	9
2.1	A distribuição normal assimétrica padrão	10
2.1.1	Propriedades	10
2.2	A distribuição normal bimodal padrão	14
2.2.1	Propriedades	15
2.3	A distribuição normal bimodal assimétrica padrão	17
2.3.1	Propriedades	19
3	Inferência clássica e bayesiana em famílias estendidas de distribuições normais	23
3.1	Inferência na família de distribuições normais assimétricas	23
3.1.1	Inferência clássica na família normal assimétrica padrão	24
3.1.2	Construção do algoritmo EM para a família normal assimétrica padrão	30
3.1.3	Inferência bayesiana na família normal assimétrica padrão	32
3.2	Inferência na família de distribuições normais bimodais	34
3.2.1	Inferência clássica na família normal bimodal padrão	34
3.2.2	Construção do algoritmo EM para a família normal bimodal padrão	38
3.2.3	Inferência bayesiana na família normal bimodal padrão	40
3.3	Inferência na família de distribuições normais bimodais assimétricas	41
3.3.1	Inferência clássica na família normal bimodal assimétrica padrão	41
3.3.2	Inferência bayesiana na família normal bimodal assimétrica	44

4	Análise de dados simulados	47
4.1	Estudo Monte Carlo	47
4.2	Análise de sensibilidade	51
5	Análise dos dados de fronteira	57
5.1	Inferência nos dados de fronteira	58
6	Conclusão	61

Capítulo 1

Introdução

Um dos maiores desafios relacionados à modelagem de dados é encontrar classes de distribuições flexíveis o bastante para representar adequadamente diferentes comportamentos dos dados, tais como, assimetria, bimodalidade, caudas leves e pesadas, curtose, etc. Muitos dos métodos existentes para análise de dados são baseados na suposição de que as observações possuem distribuição normal. Porém, esta suposição de normalidade pode ser forte para conjuntos de dados provindos de diferentes áreas de aplicação como, por exemplo, financeira, ambiental, médica e outras. Para citar alguns exemplos, sabe-se que distribuições empíricas de retornos de mercados financeiros, principalmente de mercados emergentes, que são mais susceptíveis à atmosfera política e econômica mundial, são assimétricas e, em geral, possuem caudas mais pesadas que as da distribuição normal. Jones (1997), quando analisou a evolução da densidade da renda per capita de 121 países, verificou que tal distribuição, no período de 1960 e 1988, passou de unimodal para bimodal, dando suporte à hipótese de que, enquanto alguns países tendem a convergir em direção aos mais ricos, outro grupo de países parece ficar cada vez mais longe de tal objetivo. Pode-se pensar também na distribuição de alturas de seres humanos, onde a mistura entre as alturas de homens e mulheres geram uma bimodalidade.

Uma ideia introduzida por Azzalini (1985), que vem sendo bastante utilizada para construir classes mais flexíveis de distribuições, consiste em multiplicar uma função de densidade de probabilidade (fdp) simétrica por uma função não negativa que torna tal distribuição assimétrica. Partindo dessa ideia, Azzalini (1985) introduziu uma classe de distribuições normais assimétricas univariada. Mais tarde, essa ideia foi estendida para o caso multivariado por Azzalini e Dalla Valle (1996). A partir daí muitos trabalhos já

foram feitos sobre o assunto. Para citar alguns Arellano-Valle e Azzalini (2006), Arellano-Valle *et al.* (2006), Arellano-Valle *et al.* (2002), Arellano-Valle e Genton (2005), Azzalini (2005), Genton e Loperfido (2005), Ma e Genton (2004), Wang *et al.* (2004). Uma visão geral dessas propostas pode ser encontrada no livro editado por Genton (2004).

Numa tentativa de tornar o uso das distribuições normais assimétricas mais tratável surgiram diversas representações estocásticas para uma variável que apresenta tal distribuição. Essas representações estocásticas são úteis tanto para a parte prática quanto para obter os momentos teóricos das distribuições. Uma das mais utilizadas para a distribuição normal assimétrica foi definida por Henze (1986) e trata-se de uma combinação linear entre uma variável normal padrão e outra *half*-normal, onde os pesos são funções do parâmetro de assimetria λ .

Uma nova extensão da distribuição normal surge com o trabalho de Arellano-Valle *et al.* (2008). Nesse trabalho os autores propõem uma classe de distribuições simétricas que possuem menos curtose que a distribuição normal e que podem apresentar um comportamento bimodal - a classe de distribuições normais bimodais. Algumas de suas propriedades e sua representação estocástica também podem ser encontradas em Arellano-Valle *et al.* (2008). Elal-Olivero *et al.* (2009) introduziram uma classe de distribuições que englobam formas uni e bimodais, além de assimetria. Essa nova classe de distribuições é uma extensão das classes de Azzalini (Azzalini, 1985, 2005) e Arellano-Valle *et al.* (2008) e é chamada de classe de distribuição normal bimodal assimétrica.

O objetivo deste trabalho é estudar as famílias de distribuições normais assimétricas, normais bimodais e normais bimodais assimétricas. Em todos os três casos serão consideradas as famílias sem os parâmetros de posição e escala. Serão realizadas inferências sobre os parâmetros de assimetria das distribuições normais assimétricas e normais bimodais assimétricas e sobre os parâmetros de forma das distribuições normais bimodais e normais bimodais assimétricas. Será considerada tanto a abordagem clássica quanto a abordagem bayesiana fornecendo os estimadores de máxima verossimilhança (EMV), no primeiro caso, e as distribuições *a posteriori* e as distribuições preditivas *a priori* e *a posteriori* no segundo caso. Uma das principais contribuições deste trabalho é o estabelecimento de condições para a existência de estimadores de máxima verossimilhança para os parâmetros de tais famílias. Outra contribuição é averiguar se, tanto as distribuições *a posteriori*

quanto os estimadores de máxima verossimilhança para os parâmetros de assimetria e forma da família normal bimodal assimétrica podem ser obtidos considerando famílias mais simples. Para encontrar os estimadores de máxima verossimilhança é considerado o algoritmo EM (Dempster *et al.*, 1977), uma vez que os EMV não são, geralmente, obtidos analiticamente. O algoritmo EM será utilizado devido à facilidade de escrever os modelos propostos em estruturas condicionais, utilizando variáveis não observadas. Um estudo Monte Carlo será considerado para avaliar os EMV e a esperança e a moda *a posteriori*. Outro estudo Monte Carlo será feito para realizar uma análise de sensibilidade nas estimativas resultantes dos estimadores bayesianos quando é alterada a variabilidade *a priori* dos parâmetros. Uma análise nos dados de fronteira de Azzalini será realizada considerando que estes possuem distribuição normal bimodal assimétrica padrão.

Este trabalho está assim organizado. O Capítulo 2 apresenta uma revisão sobre a família de distribuições normais assimétricas, normais bimodais e normais bimodais assimétricas. O Capítulo 3 mostra os estimadores via método dos momentos para os parâmetros de cada uma das famílias consideradas no capítulo anterior e as condições de existência para os estimadores de máxima verossimilhança. São obtidas também as distribuições *a posteriori* e as densidades preditivas. O Capítulo 4 mostra resultados de análises feitas em dados simulados e, no Capítulo 5, estão os resultados das análises realizadas nos dados de fronteira de Azzalini. Finalmente, no Capítulo 6 são apresentadas algumas conclusões e sugestões de trabalhos futuros a serem desenvolvidos.

Capítulo 2

Classes estendidas de distribuições normais

Azzalini e Capitanio (2003) introduziram o seguinte resultado, que tem sido muito útil na construção de novas classes de distribuição de probabilidade. Sejam f_0 e q funções de densidade de probabilidade simétricas ao redor de zero sendo f_0 uma distribuição unimodal. Seja Q a função de distribuição acumulada (fda) obtida a partir de q . Então, para qualquer função ímpar w , segue que $f(x) = 2f_0(x)Q(w(x)), x \in \mathbb{R}$, é uma fdp. A função $Q(w(x))$ introduz assimetria na densidade simétrica $f_0(x)$.

Esse resultado generaliza o resultado introduzido em Azzalini (1985) e, partindo dele, várias distribuições assimetrizadas foram construídas. De acordo com Azzalini (1985), se f_0 e q são distribuições normais, f será a densidade da distribuição normal assimétrica, que será vista com mais detalhes na próxima seção. Gupta *et al.* (2002) consideram $q = f_0$ e f_0 sendo a distribuição de Laplace, logística ou uniforme. Nadarajah e Kotz (2003) consideram f_0 sendo uma distribuição normal padrão e Q a função de distribuição acumulada da normal, t -Student, Cauchy, Laplace, logística ou uniforme. Em Gómez *et al.* (2007) é considerada a função de distribuição acumulada da normal padrão como Q e f_0 sendo qualquer função de densidade de probabilidade simétrica.

Denote por ϕ e Φ a fdp e a fda, respectivamente, de uma distribuição normal padrão, $N(0, 1)$. Assuma $f_0 = \phi$ e $Q = \Phi$. No que segue, são apresentadas, ao longo deste capítulo, algumas extensões da distribuição normal.

2.1 A distribuição normal assimétrica padrão

A distribuição normal assimétrica padrão consiste em uma assimetização da distribuição normal padrão. Azzalini (1985) define que uma variável aleatória X tem distribuição normal assimétrica padrão, denotado por $X|\lambda \sim SN(\lambda)$, se sua função de densidade de probabilidade é dada por:

$$f(x|\lambda) = 2\phi(x)\Phi(\lambda x), \quad x \in \mathbb{R}, \lambda \in \mathbb{R}. \quad (2.1)$$

Observa-se, da expressão (2.1), que a distribuição em (2.1) é uma generalização da distribuição normal padrão, a qual é obtida quando $\lambda = 0$. Se $\lambda > 0$ observa-se assimetria à direita, ou seja, a média e a mediana estão à direita da moda, e, para $\lambda < 0$, observa-se assimetria à esquerda. A Figura 2.1 mostra as densidades para algumas distribuições normais assimétricas padrão com diferentes parâmetros de assimetria. À medida que $|\lambda|$ vai aumentando a distribuição em (2.1) vai se modificando, passando de uma normal padrão, quando $|\lambda| = 0$, para uma *half*-normal, quando $|\lambda| \rightarrow \infty$. Essas e outras propriedades são vistas na próxima seção.

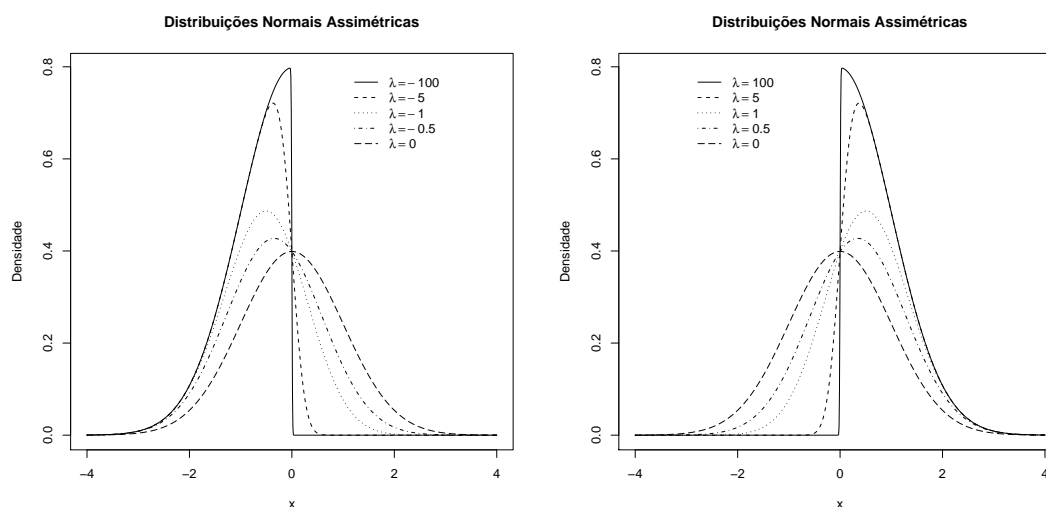


Figura 2.1: Funções de densidade da distribuição normal assimétrica padrão.

2.1.1 Propriedades

Admita que $X|\lambda \sim SN(\lambda)$. Algumas propriedades básicas da distribuição $SN(\lambda)$ (Azzalini, 1985, 1986) são:

1. a distribuição $SN(0)$ é a distribuição normal padrão;
2. quando $\lambda \rightarrow \infty$ então $f(x|\lambda) \xrightarrow{d} g(x) = 2\phi(x)\mathbf{1}_{\{x>0\}}$, ou seja, $f(x|\lambda)$ converge para uma distribuição normal positiva, onde \xrightarrow{d} denota convergência em distribuição e $\mathbf{1}_{\{G\}}$ denota a função indicadora de G ;
3. se $X|\lambda \sim SN(\lambda)$, então $-X|\lambda \sim SN(-\lambda)$;
4. se $X|\lambda \sim SN(\lambda)$, então $|X|$ tem distribuição *half*-normal (para qualquer λ);
5. se $X|\lambda \sim SN(\lambda)$, então $X^2 \sim \chi_1^2$;
6. se $X|\lambda \sim SN(\lambda)$, então a função geradora de momentos de X é $M_{X|\lambda}(t) = 2e^{\frac{t^2}{2}} \Phi\left(\frac{\lambda}{\sqrt{1+\lambda^2}}t\right)$, $t \in \mathbb{R}$.

Da função geradora de momentos segue que os momentos de ordem par e ímpar da distribuição normal assimétrica padrão são dados, respectivamente, por:

$$E(X^{2k}|\lambda) = \frac{(2k)!}{2^k k!}, \quad (2.2)$$

$$E(X^{2k-1}|\lambda) = \sqrt{\frac{2}{\pi}} \frac{\lambda}{(1+\lambda^2)^{k-1/2}} \frac{(2k-1)!}{2^{k-1}} \sum_{j=0}^{k-1} \frac{j!(2\lambda)^{2j}}{(2j+1)!(k-1-j)!}, \quad (2.3)$$

$k = 1, 2, \dots$. Dessas expressões consegue-se derivar a esperança, a variância e os coeficientes de assimetria e curtose da distribuição normal assimétrica padrão. A partir de (2.3) tem-se que o valor esperado da variável aleatória X é dado por:

$$E(X|\lambda) = \frac{\lambda}{\sqrt{1+\lambda^2}} \sqrt{\frac{2}{\pi}}. \quad (2.4)$$

Observa-se que, quanto mais próximo de zero for λ , mais próximo de zero também será $E(X|\lambda)$. Além disto, se $|\lambda|$ assumir um valor muito alto, tem-se que $E(X|\lambda)$ será próximo de $\left|\sqrt{\frac{2}{\pi}}\right|$, ou seja, $-\sqrt{\frac{2}{\pi}} < E(X|\lambda) < \sqrt{\frac{2}{\pi}}$. Utilizando resultados usuais de cálculo de probabilidade segue, a partir de (2.2) e (2.3), que a variância de X é

$$V(X|\lambda) = 1 - \frac{2}{\pi} \frac{\lambda^2}{1+\lambda^2}. \quad (2.5)$$

Pode-se observar, de (2.5), que a variância de uma distribuição normal assimétrica padrão sempre será menor ou igual à variância da distribuição normal padrão, que é igual a 1.

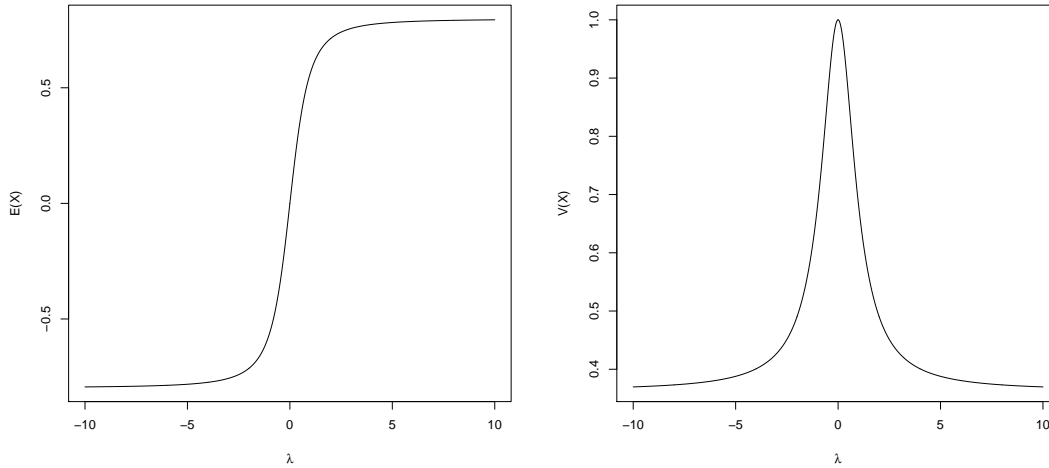


Figura 2.2: Esperança e variância da distribuição normal assimétrica padrão.

Caso $\lambda = 0$, indicando que a distribuição em questão é a normal padrão, então $V(X|\lambda) = 1$. Quando $|\lambda| \rightarrow \infty$ observa-se que $V(X|\lambda) \rightarrow 1 - \frac{2}{\pi}$. Logo, $1 - \frac{2}{\pi} < V(X|\lambda) \leq 1$. Os comportamentos de $E(X|\lambda)$ e $V(X|\lambda)$ são apresentados na Figura 2.2.

Pode-se notar que o valor esperado e a variância de uma variável aleatória com distribuição normal assimétrica padrão são limitados. Também se observa o decréscimo da variância com o aumento de $|\lambda|$.

Também são limitados os coeficientes de assimetria, $A(X|\lambda)$, e de curtose, $K(X|\lambda)$, da normal assimétrica padrão, cujas expressões são, respectivamente,

$$A(X|\lambda) = \frac{4 - \pi}{2} \left(\frac{\lambda}{\sqrt{1 + \lambda^2}} \sqrt{\frac{2}{\pi}} \right)^3 \left(1 - \frac{2}{\pi} \frac{\lambda^2}{1 + \lambda^2} \right)^{-3/2} \quad e$$

$$K(X|\lambda) = 2(\pi - 3) \left(\frac{\lambda}{\sqrt{1 + \lambda^2}} \sqrt{\frac{2}{\pi}} \right)^4 \left(1 - \frac{2}{\pi} \frac{\lambda^2}{1 + \lambda^2} \right)^{-2}.$$

Pode-se mostrar que $A(X|\lambda)$ e $K(X|\lambda)$ são tais que

$$-\frac{4 - \pi}{2} \left(\sqrt{\frac{2}{\pi}} \right)^3 \left(1 - \frac{2}{\pi} \right)^{-3/2} < A(X|\lambda) < \frac{4 - \pi}{2} \left(\sqrt{\frac{2}{\pi}} \right)^3 \left(1 - \frac{2}{\pi} \right)^{-3/2},$$

$$0 \leq K(X|\lambda) < 2(\pi - 3) \left(\sqrt{\frac{2}{\pi}} \right)^4 \left(1 - \frac{2}{\pi} \right)^{-2},$$

ou seja, aproximadamente, tem-se que $-0,9953 < A(X|\lambda) < 0,9953$ e $0 \leq K(X|\lambda) < 0,8692$. Ver também a Figura 2.3, onde percebe-se que a curtose é menor quanto mais

simétrica for a distribuição. Desta forma, uma vez que o coeficiente de assimetria é limitado, pode-se dizer que a distribuição normal assimétrica padrão não consegue modelar dados com grande assimetria.

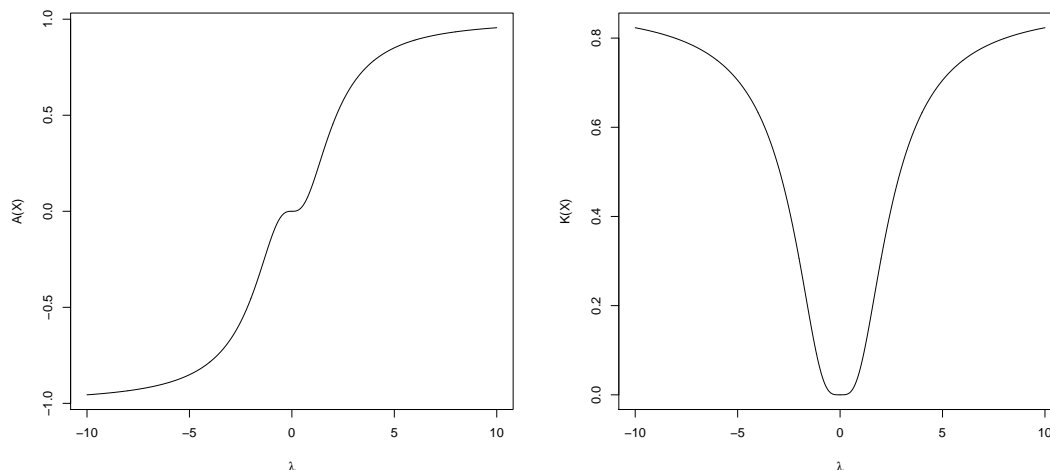


Figura 2.3: Coeficientes de assimetria e curtose da distribuição normal assimétrica padrão.

Um resultado importante, tanto do ponto de vista teórico quanto computacional, relacionado a uma variável aleatória cuja distribuição é normal assimétrica é sua representação estocástica, obtida por Henze (1986). Se $X|\lambda \sim SN(\lambda)$, então

$$X \stackrel{d}{=} \frac{\lambda}{\sqrt{1+\lambda^2}}|Z| + \frac{1}{\sqrt{1+\lambda^2}}Y, \quad (2.6)$$

onde Z e Y são variáveis aleatórias independentes e identicamente distribuídas (iid) com distribuição $N(0, 1)$ e $X \stackrel{d}{=} Y$ denota que X e Y possuem a mesma distribuição.

A representação estocástica definida em (2.6) é muito importante e útil pois fornece um modo prático de gerar amostras da família de distribuições normais assimétricas padrão e desenvolver algoritmos para a estimação.

A Figura 2.4 mostra dois histogramas de 1.000.000 de dados gerados através da representação estocástica (2.6) com parâmetros de assimetria λ iguais a, respectivamente, 5 e -2 . As curvas são feitas a partir da densidade da normal assimétrica padrão, dada em (2.1), com estes mesmos parâmetros.

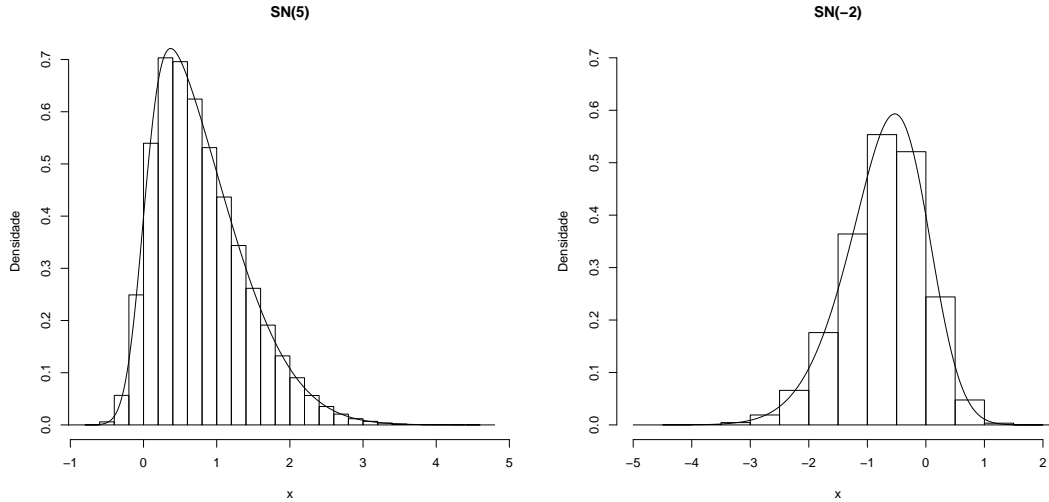


Figura 2.4: Histogramas e densidades da distribuição normal assimétrica padrão, $\lambda = 5$ e -2 .

2.2 A distribuição normal bimodal padrão

Uma nova extensão da distribuição normal é apresentada em Arellano-Valle *et al.* (2008) visando capturar comportamentos bimodais apresentados pelos dados. Tal distribuição é chamada de normal bimodal. Arellano-Valle *et al.* (2008) definem que uma variável aleatória X tem distribuição normal bimodal padrão com parâmetro de forma α , $\alpha \geq 0$, denotado por $X|\alpha \sim BN(\alpha)$, se sua função de densidade de probabilidade é dada por

$$f(x|\alpha) = \left(\frac{1 + \alpha x^2}{1 + \alpha} \right) \phi(x), x \in \mathbb{R}. \quad (2.7)$$

A Figura 2.5 mostra as densidades para algumas distribuições normais bimodais padrão com diferentes parâmetros de forma.

A distribuição introduzida por Arellano-Valle *et al.* (2008) é simétrica em torno de zero e tem a distribuição normal padrão como caso particular se $\alpha = 0$ (ver a expressão (2.7) e a Figura 2.5). Quando $\alpha \leq 0,5$ observa-se unimodalidade, enquanto que, para $\alpha > 0,5$, observa-se bimodalidade. À medida que α vai aumentando a distribuição vai se modificando, evidenciando cada vez mais a bimodalidade e as modas ficando mais distantes uma da outra. Essas e outras propriedades são vistas na próxima seção.

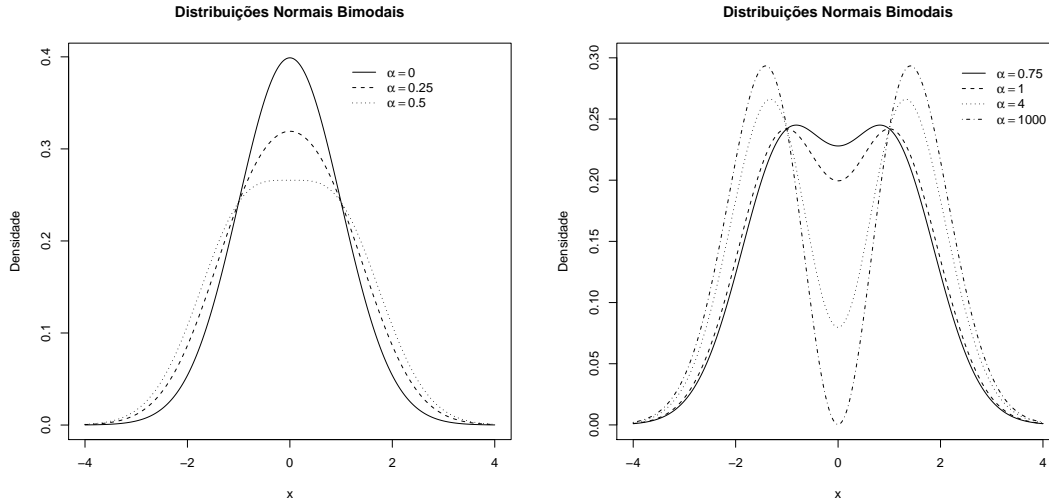


Figura 2.5: Funções de densidade de probabilidade da distribuição normal bimodal padrão.

2.2.1 Propriedades

Algumas propriedades básicas da distribuição $BN(\alpha)$ apresentadas em Arellano-Valle *et al.* (2008) são:

1. a distribuição $BN(0)$ é a distribuição normal padrão;
2. a densidade $f(x|\alpha)$ é simétrica ao redor de zero para todo valor de α ;
3. a densidade $f(x|\alpha)$ é bimodal para $\alpha > 0,5$ e $x = \pm\sqrt{\frac{2\alpha-1}{\alpha}}$ são os pontos onde ocorrem os máximos da função de densidade;
4. a densidade $f(x|\alpha)$ é unimodal para $\alpha \leq 0,5$;
5. a função de distribuição acumulada de $X|\alpha$ é $F_{X|\alpha}(x) = \Phi(x) - \frac{\alpha x}{1+\alpha}\phi(x)$;
6. se $X|\alpha \sim BN(\alpha)$, então a função geradora de momentos de X é $M_{X|\alpha}(t) = (1 + \frac{\alpha}{1+\alpha}t^2) e^{\frac{t^2}{2}}$, $t \in \mathbb{R}$.

Como a distribuição normal bimodal padrão é simétrica ao redor de zero, tem-se que todos os seus momentos de ordem ímpar são iguais a zero, ou seja,

$$E(X^{2k-1}|\alpha) = 0, \quad k = 1, 2, \dots$$

e, conseqüentemente, seu coeficiente de assimetria $A(X|\alpha)$ também é igual a zero.

Arellano-Valle *et al.* (2008) mostram que os momentos pares de uma variável aleatória X , onde $X|\alpha \sim BN(\alpha)$, são dados por

$$E(X^{2k}|\alpha) = \left[\frac{1 + \alpha(2k + 1)}{1 + \alpha} \right] \prod_{j=1}^k (2j - 1). \quad (2.8)$$

A partir de (2.8) pode-se encontrar a variância e o coeficiente de curtose da distribuição normal bimodal padrão. Suas expressões são dadas, respectivamente, por

$$V(X|\alpha) = \frac{1 + 3\alpha}{1 + \alpha} \text{ e}$$

$$K(X|\alpha) = 3 \left[\frac{(1 + 5\alpha)(1 + \alpha)}{(1 + 3\alpha)^2} - 1 \right].$$

Assim como ocorre com a distribuição normal assimétrica padrão, a distribuição normal bimodal padrão também possui variância e coeficientes de curtose limitados. Quando α é igual a zero tem-se os mesmos valores para variância e coeficiente de curtose da distribuição normal padrão, isto é, $V(X|\alpha) = 1$ e $K(X|\alpha) = 0$, respectivamente. À medida que o valor do parâmetro de forma α vai aumentando, o valor de $V(X|\alpha)$ também aumenta, enquanto que o valor de $K(X|\alpha)$ diminui. Pode-se mostrar que $1 \leq V(X|\alpha) < 3$ e $-4/3 < K(X|\alpha) \leq 0$. A Figura 2.6 mostra o comportamento de $V(X|\alpha)$ e $K(X|\alpha)$ em função de α .

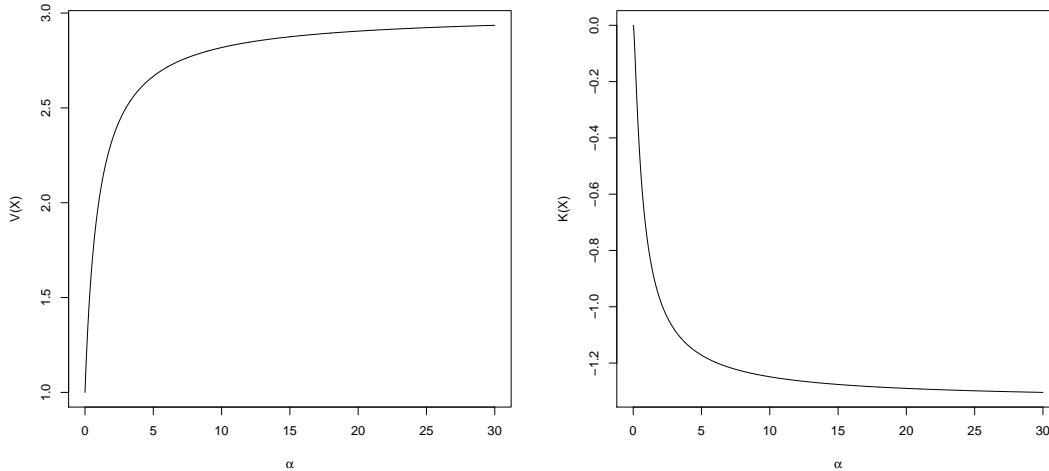


Figura 2.6: Variância e coeficiente de curtose da distribuição normal bimodal padrão.

Arellano-Valle *et al.* (2008) também obtiveram uma representação estocástica para uma variável aleatória cuja distribuição é normal bimodal padrão. Se $X|\alpha \sim BN(\alpha)$,

então

$$X \stackrel{d}{=} \sqrt{\frac{\alpha}{1+\alpha}} Z(2M-1) + \sqrt{\frac{1}{1+\alpha}} Y, \quad (2.9)$$

onde Z , M e Y são variáveis aleatórias independentes, $Z \sim \chi_3^2$, $M \sim \text{Bernoulli}(1/2)$ e $Y \sim N(0, 1)$.

A Figura 2.7 mostra os histogramas de 1.000.000 de dados gerados através da representação estocástica (2.9) com parâmetros de forma α iguais a 0,5 e 5. As curvas são desenhadas a partir da densidade da normal bimodal padrão, dada em (2.7), com estes mesmos parâmetros.

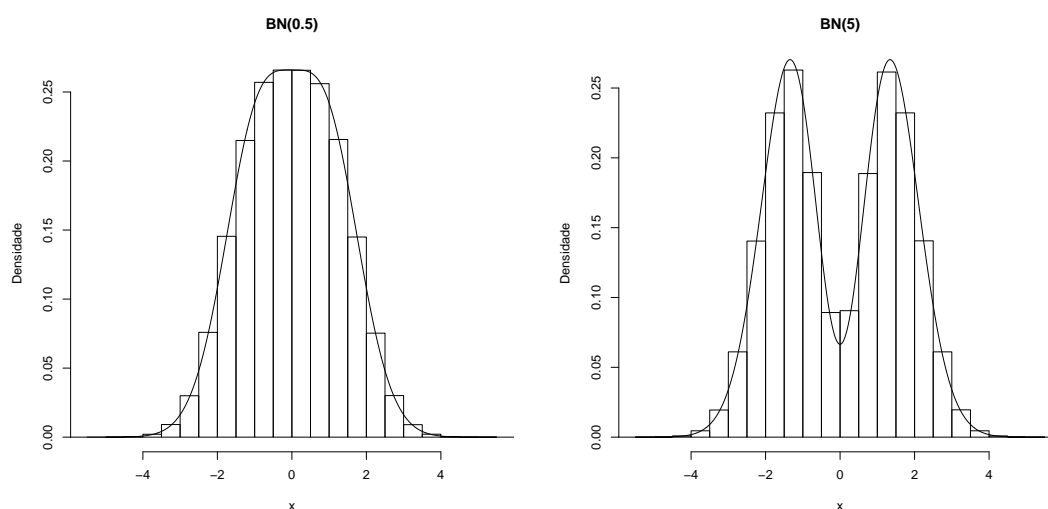


Figura 2.7: Histogramas e densidades da distribuição normal bimodal padrão.

2.3 A distribuição normal bimodal assimétrica padrão

Com o uso das distribuições assimétricas, alguns trabalhos têm sido feitos explorando a bimodalidade em tais distribuições. Para citar alguns Arellano-Valle *et al.* (2006), Ma e Genton (2004), Arnold *et al.* (2002) e Arellano-Valle *et al.* (2005). A partir do resultado apresentado em Azzalini e Capitanio (2003), Elal-Olivero *et al.* (2009) mostram que, geralmente, $f(x) = 2 \left(\frac{1+\alpha x^2}{1+\alpha k} \right) f_0(x)Q(w(x))$, $x \in \mathbb{R}$, é uma função de densidade bimodal assimétrica, onde $\alpha \geq 0$, $k = \int_{-\infty}^{\infty} x^2 f_0(x) dx < \infty$, f_0 e q são ambas simétricas e Q é a fda obtida a partir de q . Uma variável aleatória X que possui tal fdp possui uma distribuição

bimodal assimétrica com parâmetro de forma α . Neste caso, a função $Q(w(x))$ introduz assimetria na densidade simétrica $g(x) = \left(\frac{1+\alpha x^2}{1+\alpha k}\right) f_0(x)$, que pode ser uni ou bimodal.

Elal-Olivero *et al.* (2009) definem que uma variável aleatória X tem distribuição normal bimodal assimétrica padrão com parâmetro de forma α , $\alpha \geq 0$, e parâmetro de assimetria λ , $\lambda \in \mathbb{R}$, denotada por $X|\alpha, \lambda \sim SBN(\alpha, \lambda)$, se sua função de densidade de probabilidade é dada por

$$f(x|\alpha, \lambda) = 2 \left(\frac{1 + \alpha x^2}{1 + \alpha} \right) \phi(x) \Phi(\lambda x), \quad x \in \mathbb{R}. \quad (2.10)$$

Nota-se que a distribuição normal assimétrica padrão em (2.1), definida por Azzalini (1985), e a distribuição normal bimodal padrão em (2.7), introduzida por Arellano-Valle *et al.* (2008), são membros da família de distribuições em (2.10) sempre que são considerados, respectivamente, $\alpha = 0$ e $\lambda = 0$. Se α e λ são ambos iguais a zero tem-se a distribuição normal padrão como caso particular.

A Figura 2.8 mostra as densidades para algumas distribuições normais bimodais assimétricas padrão com diferentes parâmetros de forma e assimetria.

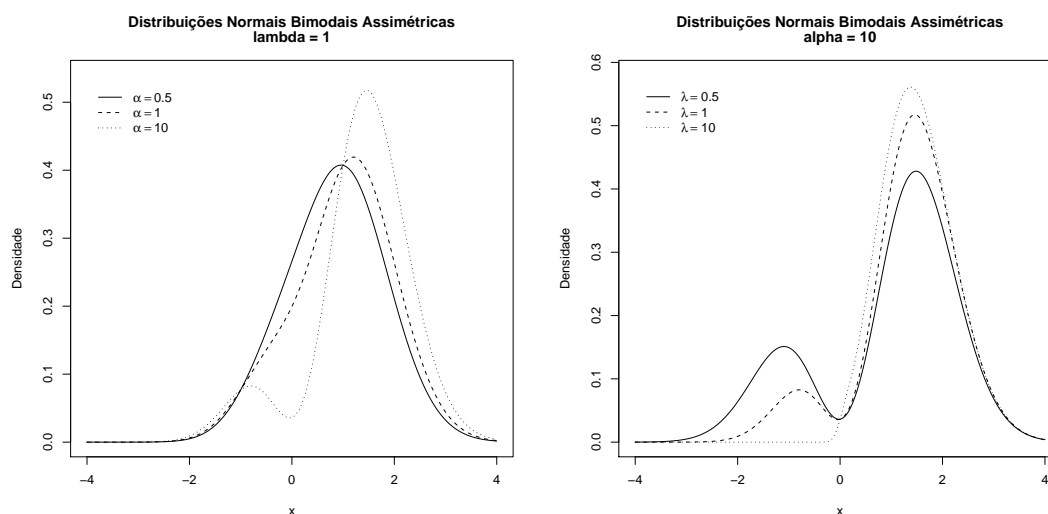


Figura 2.8: Funções de densidade de probabilidade da distribuição normal bimodal assimétrica padrão.

A partir da Figura 2.8 pode-se notar que, quando se mantém fixo o valor de λ , à medida que α aumenta, aumenta-se a evidência de bimodalidade. Quando o valor de α é fixado nota-se que, aumentando o valor de $|\lambda|$, tem-se menos evidência de bimodalidade. A assimetria é observada para valores de λ diferentes de zero, independente do valor

de α . Quando $\lambda > 0$ observa-se assimetria à direita, enquanto, para $\lambda < 0$, observa-se assimetria à esquerda. Algumas propriedades da distribuição normal bimodal assimétrica padrão serão vistas na próxima seção.

2.3.1 Propriedades

Algumas propriedades da distribuição $SBN(\alpha, \lambda)$, introduzidas por Elal-Olivero *et al.* (2009), são exibidas a seguir:

1. a distribuição $SBN(0, 0)$ é a distribuição normal padrão;
2. se $\lambda \rightarrow \pm\infty$, a densidade $f(x|\alpha, \lambda)$ não é bimodal. Nestes casos, $f(x|\alpha, \lambda)$ é positiva apenas para valores de x com mesmo sinal de λ , caso contrário, é zero;
3. se $\alpha \rightarrow \infty$, a densidade $f(x|\alpha, \lambda) \xrightarrow{d} h(x) = 2x^2\phi(x)\Phi(\lambda x)$, que é uma densidade bimodal onde a localização das modas e o valor da densidade $f(x|\alpha, \lambda)$ avaliada em cada uma delas são controladas pelo parâmetro de assimetria λ ;
4. se $\lambda > 0$, a densidade $f(x|\alpha, \lambda)$ avaliada na moda da direita é maior que quando avaliada na moda da esquerda e, avaliada na moda da esquerda, $f(x|\alpha, \lambda)$ vai se aproximando de zero à medida que λ aumenta. O oposto ocorre quando $\lambda < 0$;
5. se $X|\alpha, \lambda \sim SBN(\alpha, \lambda)$, então $|X| \sim z(x) = 2f(x|\alpha)$, onde $f(x|\alpha)$ é a mesma função definida em (2.7), ou seja, $|X|$ tem distribuição *half*-normal bimodal;
6. se $X|\alpha, \lambda \sim SBN(\alpha, \lambda)$, então os momentos de ordem par de X e $|X|$ coincidem.

Elal-Olivero *et al.* (2009) mostram que os momentos de ordem ímpar para uma variável aleatória X , onde $X|\alpha, \lambda \sim SBN(\alpha, \lambda)$, podem ser expressos como

$$E(X^{2k-1}|\alpha, \lambda) = \frac{1}{1+\alpha}E(G^{2k-1}|\lambda) + \frac{\alpha}{1+\alpha}E(G^{2k+1}|\lambda), \quad k = 1, 2, 3, \dots, \quad (2.11)$$

onde $G|\lambda \sim SN(\lambda)$.

De acordo com a propriedade 6 tem-se que os momentos de ordem par da distribuição normal bimodal assimétrica padrão são

$$E(X^{2k}|\alpha, \lambda) = E(|X|^{2k}|\alpha, \lambda) = 2 \int_0^{\infty} |x|^{2k} f(x|\alpha) dx.$$

Como a função $f(x|\alpha)$, definida em (2.7), é simétrica ao redor de zero, tem-se que $\int_0^\infty |x|^{2k} f(x|\alpha) dx = 1/2 \int_{-\infty}^\infty x^{2k} f(x|\alpha) dx$. Logo,

$$E(X^{2k}|\alpha, \lambda) = 2 \times \frac{1}{2} \int_{-\infty}^\infty x^{2k} f(x|\alpha) dx = 2 \times \frac{1}{2} E(Y^{2k}|\alpha) = E(Y^{2k}|\alpha),$$

onde $Y|\alpha \sim BN(\alpha)$. Portanto,

$$E(X^{2k}|\alpha, \lambda) = \left[\frac{1 + \alpha(2k + 1)}{1 + \alpha} \right] \prod_{j=1}^k (2j - 1), \quad k = 1, 2, 3, \dots \quad (2.12)$$

De acordo com as equações (2.11), (2.3) e (2.12) tem-se que os valores de $E(X|\alpha, \lambda)$ e $E(X^2|\alpha, \lambda)$ são dados, respectivamente, por

$$E(X|\alpha, \lambda) = \frac{1}{1 + \alpha} \sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{1 + \lambda^2}} + \frac{\alpha}{1 + \alpha} \sqrt{\frac{2}{\pi}} \frac{\lambda}{(1 + \lambda^2)^{3/2}} (3 + 2\lambda^2) \quad \text{e}$$

$$E(X^2|\alpha, \lambda) = \frac{1 + 3\alpha}{1 + \alpha}.$$

As Figuras 2.9 e 2.10 mostram, respectivamente, a esperança e variância da distribuição normal bimodal assimétrica padrão e seus coeficientes de assimetria e curtose.

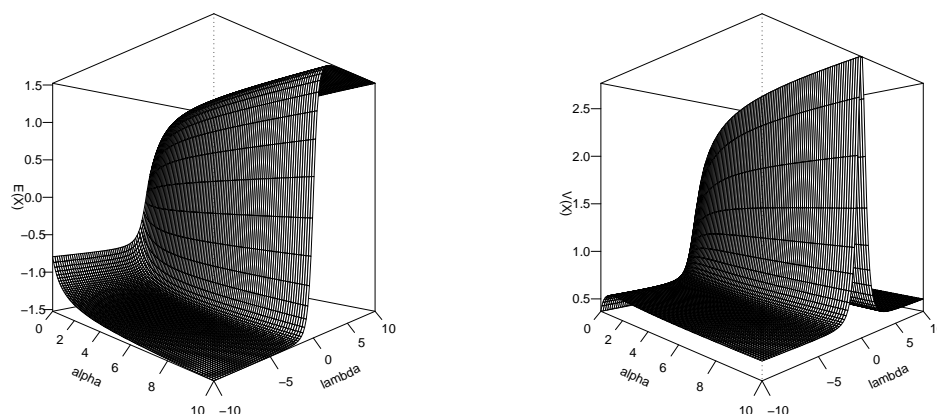


Figura 2.9: Esperança e variância da distribuição normal bimodal assimétrica padrão.

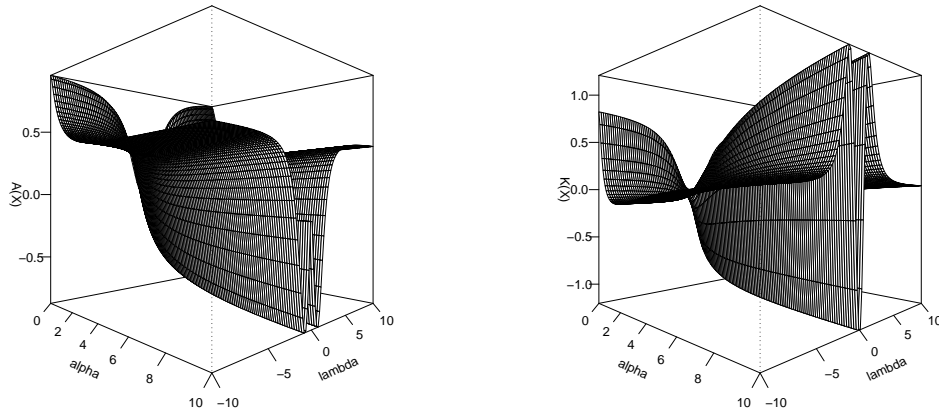


Figura 2.10: Coeficientes de assimetria e curtose da distribuição normal bimodal assimétrica padrão.

Assim como ocorre com as famílias vistas nas seções anteriores, pode-se provar que tanto a esperança quanto a variância da distribuição normal bimodal assimétrica padrão são limitadas (ver também Figura 2.9). O mesmo ocorre com os coeficientes de assimetria e curtose (ver Figura 2.10).

Se $X|\alpha, \lambda \sim SBN(\alpha, \lambda)$, então sua representação estocástica é

$$X \stackrel{d}{=} \begin{cases} Y & \text{se } Z < \lambda Y, \\ -Y & \text{se } Z \geq \lambda Y, \end{cases} \quad (2.13)$$

onde Y e Z são variáveis aleatórias independentes, $Y \sim BN(\alpha)$ e $Z \sim N(0, 1)$. A prova deste resultado pode ser encontrada em Elal-Olivero *et al.* (2009).

A Figura 2.11 mostra dois histogramas de 1.000.000 de dados gerados através da representação estocástica (2.13) com parâmetros de forma e assimetria (α, λ) iguais a, respectivamente, $(10; 0, 5)$ e $(5.000, -2)$. As curvas são desenhadas a partir da densidade da normal bimodal assimétrica padrão, dada em (2.10), com estes mesmos parâmetros.

O Capítulo 3 será destinado ao problema de inferência nas famílias estendidas de distribuições normais apresentadas neste capítulo.

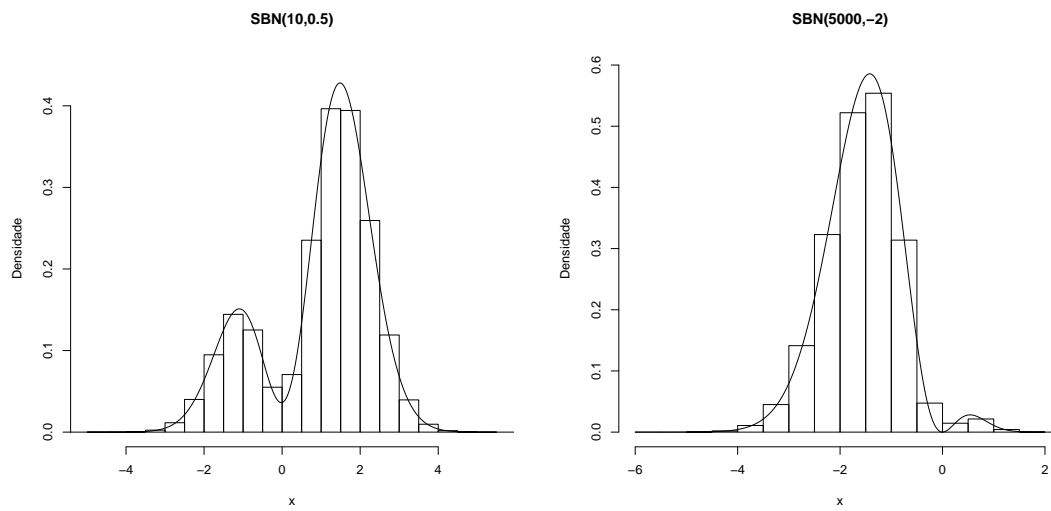


Figura 2.11: Histogramas e densidades da distribuição normal bimodal assimétrica padrão.

Capítulo 3

Inferência clássica e bayesiana em famílias estendidas de distribuições normais

Neste capítulo será tratado o problema de inferência, tanto do ponto de vista clássico quanto bayesiano, para as extensões da família de distribuições normais apresentadas no Capítulo anterior.

Do ponto de vista da inferência clássica são encontrados os estimadores via método dos momentos e estabelecidas as condições de existência dos estimadores de máxima verossimilhança. Nestas famílias os EMV dos parâmetros, geralmente, não têm formas analíticas. Então, são construídos algoritmos EM para encontrá-los.

Do ponto de vista bayesiano serão encontradas as distribuições *a posteriori* para cada um dos parâmetros das famílias de distribuições consideradas e também as funções de densidade preditivas *a priori* e *a posteriori*.

Na família da distribuição normal bimodal assimétrica padrão observa-se que, tanto considerando inferência clássica quanto inferência bayesiana, o estimador dos parâmetros se reduz aos estimadores encontrados para as famílias da distribuição normal assimétrica padrão e normal bimodal padrão.

3.1 Inferência na família de distribuições normais assimétricas

O trabalho de Azzalini (1985) mostrou que a estimação do parâmetro de assimetria na distribuição normal assimétrica definida em (2.1) não é fácil. Para estimar λ tanto o

método dos momentos quanto o método via estimador de máxima verossimilhança podem apresentar problemas (Sartori, 2006). Na parametrização exibida em (2.1) pode haver máximos locais, por exemplo. Métodos alternativos foram estudados por Sartori (2006) na abordagem clássica, e, na abordagem bayesiana, por Liseo e Loperfido (2006). Estes últimos propõem a utilização da distribuição *a priori* de Jeffreys para λ . Entretanto, essa distribuição tem uma expressão difícil de trabalhar, o que levou Bayes e Branco (2007) a proporem uma aproximação para ela por uma certa distribuição *t*-Student. Além da aproximação para a distribuição *a priori* de Jeffreys, Bayes e Branco (2007) também propuseram uma aproximação para o fator de correção de viés que é utilizado no estimador para λ proposto por Sartori (2006).

3.1.1 Inferência clássica na família normal assimétrica padrão

Suponha que, fixado λ , as variáveis aleatórias X_1, \dots, X_n são independentes e identicamente distribuídas (iid) com distribuição normal assimétrica padrão dada em (2.1). Assim, para cada amostra observada $\mathbf{x} = (x_1, \dots, x_n)^t$ do vetor aleatório $\mathbf{X} = (X_1, \dots, X_n)^t$, tem-se que um possível estimador de λ via método dos momentos é dado por

$$\hat{\lambda} = \frac{\bar{x}}{\sqrt{2/\pi - \bar{x}^2}}, \quad \text{se } \bar{x}^2 < 2/\pi, \quad (3.1)$$

onde $\bar{x} = 1/n \sum_{i=1}^n x_i$ é a média amostral observada. A expressão (3.1) é encontrada ao resolver, para λ , a igualdade $\bar{x} = E(X|\lambda)$, onde $E(X|\lambda)$ é o valor esperado de uma variável cuja distribuição é normal assimétrica padrão com parâmetro de assimetria λ , dado em (2.4). De acordo com Sartori (2006) o estimador via método dos momentos pode apresentar estimativas muito ruins para λ . Assim, neste trabalho será considerado o EMV para λ . O EMV tem a vantagem de não violar o princípio da verossimilhança (Berger e Wolpert, 1984).

Para uma amostra de tamanho n , selecionada independentemente, tem-se que a função de verossimilhança é dada por

$$f(\mathbf{x}|\lambda) = 2^n \left[\prod_{i=1}^n \phi(x_i) \Phi(\lambda x_i) \right]. \quad (3.2)$$

A Figura 3.1 apresenta exemplos da função de verossimilhança dada em (3.2) quando todas as observações são positivas, negativas e quando há observações tanto positivas

quanto negativas, respectivamente. Para construção da Figura 3.1 foram feitas modificações nos 50 valores amostrais que formam os dados de fronteira, que serão analisados no Capítulo 5, para avaliar o comportamento da função de verossimilhança dada em (3.2). Nota-se que, quando todas as observações são positivas (negativas), a função de verossimilhança é monótona crescente (decrecente) em λ . Logo, a estimativa via máxima verossimilhança para o parâmetro de assimetria λ é infinito (menos infinito). Tal fato ocorre com probabilidade positiva (Liseo, 1990). No entanto, essa probabilidade decresce rapidamente com o aumento do tamanho da amostra n . Como o EMV para λ pode ser infinito ou menos infinito, Sartori (2006) utilizou uma função escore modificada como uma equação para estimar o parâmetro de assimetria. O estimador alternativo ao EMV, proposto por Sartori (2006), é baseado numa correção de viés apresentada em Firth (1993) e é sempre finito.

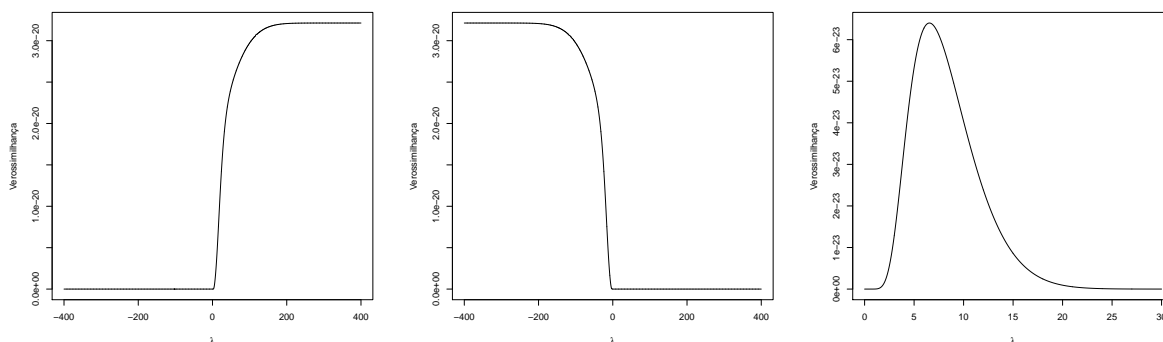


Figura 3.1: Funções de verossimilhança da distribuição normal assimétrica padrão quando todas as observações são positivas, negativas e quando há observações positivas e negativas, respectivamente.

No caso onde há observações tanto positivas quanto negativas é de interesse encontrar o valor do parâmetro de assimetria λ que maximiza a função de verossimilhança em (3.2). Apesar de não haver provas de que o EMV para λ é único na família normal assimétrica padrão, resultados empíricos dão evidências que isso ocorre. Nesses casos, porém, não há soluções analíticas para o estimador de máxima verossimilhança para λ . No entanto há como saber quando o EMV para λ existe e quais os possíveis valores que ele pode assumir. Isso é mostrado na proposição abaixo.

Proposição 1. *Seja (x_1, \dots, x_n) uma amostra iid da distribuição normal assimétrica padrão. Denote por I_+ o conjunto de índices tais que $x_i > 0$ e por I_- o conjunto de*

índices tais que $x_i < 0$. Seja C_λ o conjunto de valores distintos de $\lambda \in \mathbb{R}$ tais que

$$\sum_{i \in I_+} \frac{\phi(\lambda x_i)}{\Phi(\lambda x_i)} x_i = - \sum_{i \in I_-} \frac{\phi(\lambda x_i)}{\Phi(\lambda x_i)} x_i.$$

Então,

1. se $I_+ \neq \emptyset$ e $I_- = \emptyset$ o EMV para λ é infinito;
2. se $I_+ = \emptyset$ e $I_- \neq \emptyset$ o EMV para λ é menos infinito;
3. se $I_+ \neq \emptyset$, $I_- \neq \emptyset$ e $C_\lambda = \{0\}$, então o EMV para λ é zero;
4. se $I_+ \neq \emptyset$, $I_- \neq \emptyset$, $C_\lambda \neq \emptyset$ e $C_\lambda \neq \{0\}$ então o EMV para λ poderá não ser único e segue que:

(a) os EMV para λ serão os valores $\hat{\lambda}_1, \dots, \hat{\lambda}_k \in C_\lambda$, $k \geq 1$, que satisfaçam a uma das seguintes condições:

i.

$$\left| \sum_{i \in I_-} H(x_i, \hat{\lambda}_j) \hat{\lambda}_j x_i \Phi(\hat{\lambda}_j x_i) \right| > \left| \sum_{i \in I_-} H(x_i, \hat{\lambda}_j) \phi(\hat{\lambda}_j x_i) \right| \quad e$$

$$\left| \sum_{i \in I_+} H(x_i, \hat{\lambda}_j) [\hat{\lambda}_j x_i \Phi(\hat{\lambda}_j x_i) + \phi(\hat{\lambda}_j x_i)] \right| >$$

$$\left| \sum_{i \in I_-} H(x_i, \hat{\lambda}_j) [\hat{\lambda}_j x_i \Phi(\hat{\lambda}_j x_i) + \phi(\hat{\lambda}_j x_i)] \right| \quad \text{ou}$$

ii.

$$\left| \sum_{i \in I_-} H(x_i, \hat{\lambda}_j) \hat{\lambda}_j x_i \Phi(\hat{\lambda}_j x_i) \right| < \left| \sum_{i \in I_-} H(x_i, \hat{\lambda}_j) \phi(\hat{\lambda}_j x_i) \right|,$$

tal que $f(\mathbf{x}|\hat{\lambda}_j) = \max_{\lambda \in C_\lambda} f(\mathbf{x}|\lambda)$, $j = 1, \dots, k$, onde $H(x, \lambda) = \frac{\phi(\lambda x)}{\Phi^2(\lambda x)} x^2 > 0$ para todo $x \in \mathbb{R}$ e $\lambda \in \mathbb{R}$. Nesse caso, os EMV para λ são positivos;

(b) os EMV para λ serão os valores $\hat{\lambda}_1, \dots, \hat{\lambda}_k \in C_\lambda$, $k \geq 1$, que satisfaçam a uma das seguintes condições:

i.

$$\left| \sum_{i \in I_+} H(x_i, \hat{\lambda}_j) \hat{\lambda}_j x_i \Phi(\hat{\lambda}_j x_i) \right| > \left| \sum_{i \in I_-} H(x_i, \hat{\lambda}_j) \phi(\hat{\lambda}_j x_i) \right| \quad e$$

$$\left| \sum_{i \in I_+} H(x_i, \hat{\lambda}_j) [\hat{\lambda}_j x_i \Phi(\hat{\lambda}_j x_i) + \phi(\hat{\lambda}_j x_i)] \right| < \\ \left| \sum_{i \in I_-} H(x_i, \hat{\lambda}_j) [\hat{\lambda}_j x_i \Phi(\hat{\lambda}_j x_i) + \phi(\hat{\lambda}_j x_i)] \right| \quad \text{ou}$$

ii.

$$\left| \sum_{i \in I_+} H(x_i, \hat{\lambda}_j) \hat{\lambda}_j x_i \Phi(\hat{\lambda}_j x_i) \right| < \left| \sum_{i \in I_+} H(x_i, \hat{\lambda}_j) \phi(\hat{\lambda}_j x_i) \right|,$$

tal que $f(\mathbf{x}|\hat{\lambda}_j) = \max_{\lambda \in C_\lambda} f(\mathbf{x}|\lambda)$, $j = 1, \dots, k$, onde $H(x, \lambda)$ é como definidos em 4a. Nesse caso, os EMV para λ são negativos.

Prova: Seja (x_1, \dots, x_n) uma amostra iid da distribuição normal assimétrica padrão. Denote por I_+ o conjunto de índices das observações tais que $x_i > 0$ e por I_- o conjunto de índices das observações tais que $x_i < 0$.

Para encontrar os possíveis pontos de máximo e, conseqüentemente, os estimadores de máxima verossimilhança para λ , deve-se realizar um estudo em $\frac{d}{d\lambda} f(\mathbf{x}|\lambda)$ ou, equivalentemente, em $\frac{d}{d\lambda} l(\lambda)$, onde $l(\lambda) = \ln[f(\mathbf{x}|\lambda)]$ é a log-verossimilhança da função de verossimilhança em (3.2). Seja

$$\begin{aligned} \frac{d}{d\lambda} l(\lambda) &= \sum_{i=1}^n \frac{\phi(\lambda x_i)}{\Phi(\lambda x_i)} x_i \\ &= \sum_{i \in I_+} \frac{\phi(\lambda x_i)}{\Phi(\lambda x_i)} x_i + \sum_{i \in I_-} \frac{\phi(\lambda x_i)}{\Phi(\lambda x_i)} x_i \end{aligned} \quad (3.3)$$

a primeira derivada de $l(\lambda)$ em relação a λ . Portanto:

1. se $I_+ \neq \emptyset$ e $I_- = \emptyset$ tem-se $\frac{d}{d\lambda} l(\lambda) > 0$ para todo $\lambda \in \mathbb{R}$. Logo, $l(\lambda)$ e, conseqüentemente, $f(\mathbf{x}|\lambda)$ são funções estritamente crescentes em λ e, portanto, o estimador de máxima verossimilhança para λ é infinito;
2. se $I_+ = \emptyset$ e $I_- \neq \emptyset$ tem-se $\frac{d}{d\lambda} l(\lambda) < 0$ para todo $\lambda \in \mathbb{R}$. Logo, $l(\lambda)$ e, conseqüentemente, $f(\mathbf{x}|\lambda)$ são funções estritamente decrescentes em λ e, portanto, o estimador de máxima verossimilhança para λ é menos infinito;
3. se $I_+ \neq \emptyset$ e $I_- \neq \emptyset$ então deve-se encontrar, primeiramente, os valores de λ tais que $\frac{d}{d\lambda} f(x|\lambda) = 0$ ou, equivalentemente, $\frac{d}{d\lambda} l(\lambda) = 0$. Logo, os valores de λ tais que

$\frac{d}{d\lambda}l(\lambda) = 0$ são os valores de $\lambda \in \mathbb{R}$ tais que

$$\sum_{i \in I_+} \frac{\phi(\lambda x_i)}{\Phi(\lambda x_i)} x_i = - \sum_{i \in I_-} \frac{\phi(\lambda x_i)}{\Phi(\lambda x_i)} x_i. \quad (3.4)$$

Portanto, os valores de λ que satisfazem (3.4) são pontos críticos em $f(\mathbf{x}|\lambda)$ e formam o conjunto C_λ .

Para que esses valores de λ sejam pontos de máximo deve-se ter que $\frac{d^2}{d\lambda^2}l(\lambda)$ avaliada nesses valores seja negativa. Seja

$$\begin{aligned} \frac{d^2}{d\lambda^2}l(\lambda) &= - \sum_{i=1}^n H(x_i, \lambda) [\lambda x_i \Phi(\lambda x_i) + \phi(\lambda x_i)] \\ &= - \sum_{i \in I_+} H(x_i, \lambda) [\lambda x_i \Phi(\lambda x_i) + \phi(\lambda x_i)] - \sum_{i \in I_-} H(x_i, \lambda) [\lambda x_i \Phi(\lambda x_i) + \phi(\lambda x_i)] \end{aligned}$$

a segunda derivada de $l(\lambda)$ em relação a λ , onde $H(x, \lambda) = \frac{\phi(\lambda x)}{\Phi^2(\lambda x)} x^2 > 0$ para todo $x \in \mathbb{R}$ e $\lambda \in \mathbb{R}$.

Se o único valor que satisfaz (3.4) é zero, ou seja, $C_\lambda = \{0\}$, então tem-se que $\frac{d^2}{d\lambda^2}l(0)$ é negativa, o que garante que 0 é ponto de máximo global e, portanto, o estimador de máxima verossimilhança para λ é zero;

4. suponha $I_+ \neq \emptyset$, $I_- \neq \emptyset$, $C_\lambda \neq \emptyset$ e $C_\lambda \neq \{0\}$:

(a) sejam $\hat{\lambda}_1, \dots, \hat{\lambda}_k \in C_\lambda$, $k \geq 1$, os pontos positivos de máximo global, ou seja, os valores positivos de C_λ tais que

$$f(\mathbf{x}|\hat{\lambda}_j) = \max_{\lambda \in C_\lambda} f(\mathbf{x}|\lambda) \quad j = 1, \dots, k.$$

Suponha que $\lambda > 0$. Então, tem-se $\sum_{i \in I_+} H(x_i, \lambda) [\lambda x_i \Phi(\lambda x_i) + \phi(\lambda x_i)] > 0$, uma vez que $x_i > 0$ para todo $i \in I_+$, e $\sum_{i \in I_-} H(x_i, \lambda) \lambda x_i \Phi(\lambda x_i) < 0$, uma vez que $x_i < 0$ para todo $i \in I_-$. Logo, para que $\frac{d^2}{d\lambda^2}l(\lambda) < 0$ deve-se ter

i.

$$\left| \sum_{i \in I_-} H(x_i, \lambda) \lambda x_i \Phi(\lambda x_i) \right| > \left| \sum_{i \in I_+} H(x_i, \lambda) \phi(\lambda x_i) \right| \quad \text{e}$$

$$\left| \sum_{i \in I_+} H(x_i, \lambda) [\lambda x_i \Phi(\lambda x_i) + \phi(\lambda x_i)] \right| >$$

$$\left| \sum_{i \in I_-} H(x_i, \lambda) [\lambda x_i \Phi(\lambda x_i) + \phi(\lambda x_i)] \right| \quad \text{ou}$$

ii.

$$\left| \sum_{i \in I_-} H(x_i, \lambda) \lambda x_i \Phi(\lambda x_i) \right| < \left| \sum_{i \in I_-} H(x_i, \lambda) \phi(\lambda x_i) \right|.$$

Logo, para $\hat{\lambda}_1, \dots, \hat{\lambda}_k \in C_\lambda$, $k \geq 1$, que são valores positivos pertencentes a C_λ , tem-se que, se alguma das condições acima é satisfeita, $\frac{d^2}{d\lambda^2} l(\lambda)$ avaliada em $\hat{\lambda}_j$, $j = 1, \dots, k$, $k \geq 1$, será negativa. Portanto, nessas situações, $\hat{\lambda}_1, \dots, \hat{\lambda}_k \in C_\lambda$ são pontos de máximo globais, ou seja, estimadores de máxima verossimilhança para λ ;

(b) sejam $\hat{\lambda}_1, \dots, \hat{\lambda}_k \in C_\lambda$, $k \geq 1$, os pontos negativos de máximo global, ou seja, os valores negativos de C_λ tais que

$$f(\mathbf{x}|\hat{\lambda}_j) = \max_{\lambda \in C_\lambda} f(\mathbf{x}|\lambda) \quad j = 1, \dots, k.$$

Suponha $\lambda < 0$. Então, tem-se $\sum_{i \in I_-} H(x_i, \lambda) [\lambda x_i \Phi(\lambda x_i) + \phi(\lambda x_i)] > 0$, uma vez que $x_i < 0$ para todo $i \in I_-$, e $\sum_{i \in I_+} H(x_i, \lambda) \lambda x_i \Phi(\lambda x_i) < 0$, uma vez que $x_i > 0$ para todo $i \in I_+$. Logo, para que $\frac{d^2}{d\lambda^2} l(\lambda) < 0$ deve-se ter

i.

$$\left| \sum_{i \in I_+} H(x_i, \lambda) \lambda x_i \Phi(\lambda x_i) \right| > \left| \sum_{i \in I_-} H(x_i, \lambda) \phi(\lambda x_i) \right| \quad \text{e}$$

$$\left| \sum_{i \in I_+} H(x_i, \lambda) [\lambda x_i \Phi(\lambda x_i) + \phi(\lambda x_i)] \right| < \left| \sum_{i \in I_-} H(x_i, \lambda) [\lambda x_i \Phi(\lambda x_i) + \phi(\lambda x_i)] \right| \quad \text{ou}$$

ii.

$$\left| \sum_{i \in I_+} H(x_i, \lambda) \lambda x_i \Phi(\lambda x_i) \right| < \left| \sum_{i \in I_+} H(x_i, \lambda) \phi(\lambda x_i) \right|.$$

Logo, para $\hat{\lambda}_1, \dots, \hat{\lambda}_k \in C_\lambda$, $k \geq 1$, que são valores negativos pertencentes a C_λ , tem-se que, se alguma das condições acima é satisfeita, $\frac{d^2}{d\lambda^2} l(\lambda)$ avaliada em $\hat{\lambda}_j$,

$j = 1, \dots, k$, $k \geq 1$, será negativa. Portanto, nessas situações, $\hat{\lambda}_1, \dots, \hat{\lambda}_k \in C_\lambda$ seriam pontos de máximo globais, ou seja, estimadores de máxima verossimilhança para λ .

□

Apesar da Proposição 1 fornecer condições para a existência do EMV para λ , sua forma analítica permanece desconhecida. Além disso, não fica garantida a unicidade do EMV. Será considerado o algoritmo EM (Dempster *et al.*, 1977) para obter um valor aproximado para o EMV para λ .

3.1.2 Construção do algoritmo EM para a família normal assimétrica padrão

O algoritmo EM foi introduzido por Dempster *et al.* (1977) e é uma técnica computacional utilizada para encontrar as estimativas de máxima verossimilhança dos parâmetros nos modelos probabilísticos, onde o modelo depende das chamadas variáveis latentes, que são as variáveis não observadas. É constituído de dois passos: o passo E (*expectation*) e o passo M (*maximization*).

O passo E do algoritmo EM calcula os valores esperados da log-verossimilhança completa com respeito às variáveis latentes, dadas as variáveis observadas e a estimativa atualizada dos parâmetros. O passo M calcula os parâmetros que maximizam o valor esperado da log-verossimilhança completa encontrada no passo E.

Uma vantagem do algoritmo EM é a obtenção, através da maximização da log-verossimilhança completa, de expressões fechadas dos estimadores dos parâmetros.

No entanto, algumas desvantagens devem ser consideradas. A escolha dos valores iniciais tecnicamente não importa, mas, na prática, uma escolha ruim pode levar a uma má estimação. Além disso, a convergência, embora garantida, pode demorar muito para ocorrer. Na prática, se os valores dos parâmetros ou da log-verossimilhança avaliada nos parâmetros não mudam significativamente entre duas iterações, então o algoritmo termina.

Para construir o algoritmo EM para obter aproximações para as estimativas de máxima verossimilhança para λ utiliza-se a representação estocástica de Henze (1986) dada

na expressão (2.6), ou seja, considera-se que, se $X \sim SN(\lambda)$, então

$$X \stackrel{d}{=} \frac{\lambda}{\sqrt{1+\lambda^2}}|Z| + \frac{1}{\sqrt{1+\lambda^2}}Y,$$

onde Z e Y são variáveis aleatórias iid com distribuição normal padrão $N(0, 1)$ e $\stackrel{d}{=}$ significa igualdade em distribuição. Seja $H = \frac{\lambda}{\sqrt{1+\lambda^2}}|Z|$. Considerando essa definição, o modelo normal assimétrico padrão pode ser hierarquicamente obtido assumindo-se

$$\begin{aligned} X|H = h, \lambda &\sim N(h, (1+\lambda^2)^{-1}) \\ H|\lambda &\sim f_{H|\lambda}(h) = 2\frac{\sqrt{1+\lambda^2}}{|\lambda|}\phi\left(h\frac{\sqrt{1+\lambda^2}}{\lambda}\right), \quad \text{se } \frac{h}{\lambda} \geq 0. \end{aligned} \quad (3.5)$$

Utilizando o Teorema de Bayes pode-se mostrar que

$$H|X = x, \lambda \sim \begin{cases} NT\left(x\frac{\lambda^2}{1+\lambda^2}, \frac{\lambda^2}{(1+\lambda^2)^2}; 0, \infty\right) & \text{se } \lambda > 0, \\ NT\left(x\frac{\lambda^2}{1+\lambda^2}, \frac{\lambda^2}{(1+\lambda^2)^2}; -\infty, 0\right) & \text{se } \lambda < 0, \end{cases} \quad (3.6)$$

onde $NT(\mu, \sigma^2; a, b)$ denota a distribuição normal truncada no intervalo (a, b) com parâmetros de posição $\mu \in \mathbb{R}$ e de escala $\sigma > 0$.

Denote por \mathbf{H} o vetor das variáveis latentes, não observadas, $(H_1, \dots, H_n)^t$. De (3.5) tem-se que a função de log-verossimilhança aumentada associada a (\mathbf{X}, \mathbf{H}) é dada por

$$\begin{aligned} \ln f(\mathbf{X}, \mathbf{H}|\lambda) &= n \ln(1+\lambda^2) - n \ln \pi - n \ln |\lambda| - \\ &\frac{1+\lambda^2}{2} \left[\sum_{i=1}^n X_i^2 - 2 \sum_{i=1}^n X_i H_i + \frac{1+\lambda^2}{\lambda^2} \sum_{i=1}^n H_i^2 \right]. \end{aligned} \quad (3.7)$$

Para o passo E do algoritmo EM deve-se calcular a esperança condicional da expressão em (3.7) dados a amostra observada \mathbf{x} e o parâmetro estimado $\hat{\lambda}$. Tal esperança é dada por

$$\begin{aligned} Q(\lambda, \hat{\lambda}) &= E \left[\ln f(\mathbf{X}, \mathbf{H}|\lambda) | \mathbf{X} = \mathbf{x}, \lambda = \hat{\lambda} \right] \\ &= n \ln(1+\lambda^2) - n \ln \pi - n \ln |\lambda| - \\ &\frac{1+\lambda^2}{2} \left\{ \sum_{i=1}^n x_i^2 - 2 \sum_{i=1}^n x_i E \left[H_i | x_i, \hat{\lambda} \right] + \frac{1+\lambda^2}{\lambda^2} \sum_{i=1}^n E \left[H_i^2 | x_i, \hat{\lambda} \right] \right\}. \end{aligned} \quad (3.8)$$

Para obter-se (3.8) faz-se necessário calcular $E \left[H_i | x_i, \hat{\lambda} \right]$ e $E \left[H_i^2 | x_i, \hat{\lambda} \right]$. De (3.6) segue que, condicional em $X_i = x_i$ e $\lambda = \hat{\lambda}$, H_i tem distribuição normal truncada. Se

$V \sim NT(\mu, \sigma^2; a, b)$, então $E(V)$ e $E(V^2)$ são dados, respectivamente, por

$$E(V) = \mu + \frac{\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}\sigma \quad \text{e} \quad (3.9)$$

$$E(V^2) = \sigma^2 \left\{ 1 + \frac{\frac{a-\mu}{\sigma}\phi\left(\frac{a-\mu}{\sigma}\right) - \frac{b-\mu}{\sigma}\phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} - \left[\frac{E(V) - \mu}{\sigma}\right]^2 \right\} + [E(V)]^2. \quad (3.10)$$

Sendo assim, para obter-se $E[H_i|x_i, \hat{\lambda}]$ e $E[H_i^2|x_i, \hat{\lambda}]$ basta substituir os valores para μ , σ^2 , a e b dados em (3.6) nas equações (3.9) e (3.10), respectivamente.

No passo M do algoritmo EM deve-se maximizar a expressão (3.8) em relação a λ para que se possa atualizar o valor de $\hat{\lambda}$. Para isso, deve-se encontrar os valores de λ que são as raízes da seguinte expressão:

$$\begin{aligned} \frac{d}{d\lambda}Q(\lambda, \hat{\lambda}) &= 2n\frac{\lambda}{1+\lambda^2} - \frac{n}{\lambda} - \\ &\lambda \left[\sum_{i=1}^n x_i^2 - 2 \sum_{i=1}^n x_i E[H_i|x_i, \hat{\lambda}] + \frac{1+\lambda^2}{\lambda^2} \sum_{i=1}^n E[H_i^2|x_i, \hat{\lambda}] \right] - \\ &\frac{1+\lambda^2}{2} \left[\frac{2}{\lambda} \sum_{i=1}^n E[H_i^2|x_i, \hat{\lambda}] - 2\frac{1+\lambda^2}{\lambda^3} \sum_{i=1}^n E[H_i^2|x_i, \hat{\lambda}] \right]. \end{aligned}$$

A implementação do algoritmo EM se resume a:

- **Passo E:** Dado $\lambda = \hat{\lambda}$, calcular $E[H_i|x_i, \hat{\lambda}]$ e $E[H_i^2|x_i, \hat{\lambda}]$;
- **Passo M:** Atualizar $\hat{\lambda}$ maximizando $Q(\lambda, \hat{\lambda})$ com respeito a λ .

3.1.3 Inferência bayesiana na família normal assimétrica padrão

Considere que, condicional em λ , as variáveis aleatórias X_1, \dots, X_n e Y_1, \dots, Y_m sejam iid com distribuição normal assimétrica padrão dada em (2.1). Como mostrado em Castro *et al.* (2008), se λ , *a priori*, tem uma distribuição própria $\pi(\lambda)$, a distribuição *a posteriori* para λ , a densidade preditiva *a priori* de \mathbf{X} e a densidade preditiva *a posteriori* de \mathbf{Y} são dadas, respectivamente, por

$$\pi(\lambda|\mathbf{x}) = \frac{\pi(\lambda) \prod_{i=1}^n \Phi(\lambda x_i)}{E_\lambda \left[\prod_{i=1}^n \Phi(\lambda x_i) \right]}, \quad (3.11)$$

$$f_{SN}(\mathbf{x}) = 2^n \left[\prod_{i=1}^n \phi(x_i) \right] E_\lambda \left[\prod_{i=1}^n \Phi(\lambda x_i) \right] \quad \text{e} \quad (3.12)$$

$$f_{SN}(\mathbf{y}|\mathbf{x}) = 2^m \left[\prod_{i=1}^m \phi(y_i) \right] E_{\lambda|\mathbf{x}} \left[\prod_{i=1}^m \Phi(\lambda y_i) \right], \quad (3.13)$$

onde $E_{\lambda}[T]$ é o valor esperado de T em relação à distribuição $\pi(\lambda)$, ou seja,

$$E_{\lambda}[T] = \int_{-\infty}^{\infty} t\pi(\lambda)d\lambda.$$

Nesse caso, o valor esperado de λ *a posteriori* é dado por

$$E_{\lambda|\mathbf{x}}[\lambda] = \frac{E_{\lambda}[\lambda \prod_{i=1}^n \Phi(\lambda x_i)]}{E_{\lambda}[\prod_{i=1}^n \Phi(\lambda x_i)]}. \quad (3.14)$$

Liseo e Loperfido (2006) propõem a utilização da distribuição *a priori* de referência, baseados no método de Berger e Bernardo (1992), para o parâmetro de assimetria e mostram que, para a família de densidades em (2.1), essa distribuição *a priori* de referência coincide com a distribuição *a priori* de Jeffreys. Liseo e Loperfido (2006) derivam propriedades da distribuição *a priori* de Jeffreys e mostram que tal distribuição é própria. Bayes e Branco (2007) propõem a utilização de uma aproximação para a distribuição *a priori* de Jeffreys desenvolvida por Liseo e Loperfido (2006) e também uma análise bayesiana não informativa alternativa utilizando uma distribuição *a priori* uniforme para a reparametrização $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$ do parâmetro de assimetria λ .

Para estabelecer notação, assuma que um vetor aleatório de dimensão n cuja distribuição é normal n -variada com vetor de médias μ e matriz de variância e covariância Σ , denotada por $N_n(\mu, \Sigma)$, possui fdp e fda denotadas por, respectivamente, $\phi_n(\cdot; \mu, \Sigma)$ e $\Phi_n(\cdot; \mu, \Sigma)$. Quando $n = 1$ serão utilizadas as notações $\phi(\cdot; \mu, \sigma^2)$ e $\Phi(\cdot; \mu, \sigma^2)$ para denotar, respectivamente, a fdp e a fda de uma distribuição normal com média μ e variância σ^2 .

Neste trabalho assume-se que, *a priori*, $\lambda \sim N(m, v)$, $m \in \mathbb{R}$ e $v \in \mathbb{R}^+$. Para essa especificação *a priori*, Castro *et al.* (2008) mostram que (3.11), (3.12) e (3.14) se tornam, respectivamente:

$$\pi(\lambda|\mathbf{x}) = \phi(\lambda; m, v) \frac{\Phi_n(\lambda\mathbf{x})}{\Phi_n(m\mathbf{x}; \mathbf{0}, I_n + v\mathbf{x}\mathbf{x}^t)}, \quad (3.15)$$

$$f_{SN}(\mathbf{x}) = 2^n \phi_n(\mathbf{x}) \Phi_n(m\mathbf{x}; \mathbf{0}, I_n + v\mathbf{x}\mathbf{x}^t), \quad (3.16)$$

$$E(\lambda|\mathbf{x}) = m + v\mathbf{x}^t\xi, \quad (3.17)$$

onde $\phi_n(\mathbf{x}) = \prod_{i=1}^n \phi(x_i)$, $\Phi_n(\lambda\mathbf{x}) = \prod_{i=1}^n \Phi(\lambda x_i)$, o i -ésimo componente de $\xi = (\xi_1, \dots, \xi_n)^t$

é tal que

$$\xi_i = \phi(mx_i; 0, 1 + vx_i^2) \frac{\Phi_{n-1}\left(\frac{m\mathbf{x}_{(i)}}{1+vx_i^2}; \mathbf{0}, I_{n-1} + \frac{v\mathbf{x}_{(i)}\mathbf{x}_{(i)}^t}{1+vx_i^2}\right)}{\Phi_n(m\mathbf{x}; \mathbf{0}, I_n + v\mathbf{x}\mathbf{x}^t)},$$

e $\mathbf{x}_{(i)}$ é o subvetor de \mathbf{x} sem o i -ésimo componente.

3.2 Inferência na família de distribuições normais bimodais

Nesta seção serão apresentados alguns métodos inferenciais para a família de distribuições normais bimodais.

3.2.1 Inferência clássica na família normal bimodal padrão

Suponha que, fixado α , as variáveis aleatórias X_1, \dots, X_n são iid com distribuição normal bimodal padrão dada em (2.7). Dessa forma, para cada amostra observada $\mathbf{x} = (x_1, \dots, x_n)^t$ do vetor aleatório $\mathbf{X} = (X_1, \dots, X_n)^t$ tem-se que um possível estimador de α via método dos momentos é dado por

$$\hat{\alpha} = \frac{1 - \bar{x}^2}{\bar{x}^2 - 3}, \quad \text{se } 1 \leq \bar{x}^2 < 3, \quad (3.18)$$

onde $\bar{x}^2 = 1/n \sum_{i=1}^n x_i^2$.

Em uma amostra selecionada independentemente de tamanho n tem-se que a função de verossimilhança é dada por

$$f(\mathbf{x}|\alpha) = \left[\prod_{i=1}^n \frac{1 + \alpha x_i^2}{1 + \alpha} \phi(x_i) \right]. \quad (3.19)$$

A Figura 3.2 apresenta exemplos da função de verossimilhança dada em (3.19) quando todas as observações são maiores que um, menores que um e quando há observações tanto menores quanto maiores que um, em módulo, respectivamente. Para construção da Figura 3.2 foram feitas modificações nos 50 valores amostrais que formam os dados de fronteira, que serão analisados no Capítulo 5, para avaliar o comportamento da função de verossimilhança dada em (3.19). Nota-se que, quando todas as observações são, em módulo, maiores (menores) que um, a função de verossimilhança é monótona crescente (decrecente) em α . Logo, a estimativa de máxima verossimilhança para o parâmetro de forma α é infinito (zero). Tal fato ocorre com probabilidade positiva. Similar ao que

ocorre na distribuição normal assimétrica padrão essa probabilidade decresce rapidamente com o aumento do tamanho da amostra n .

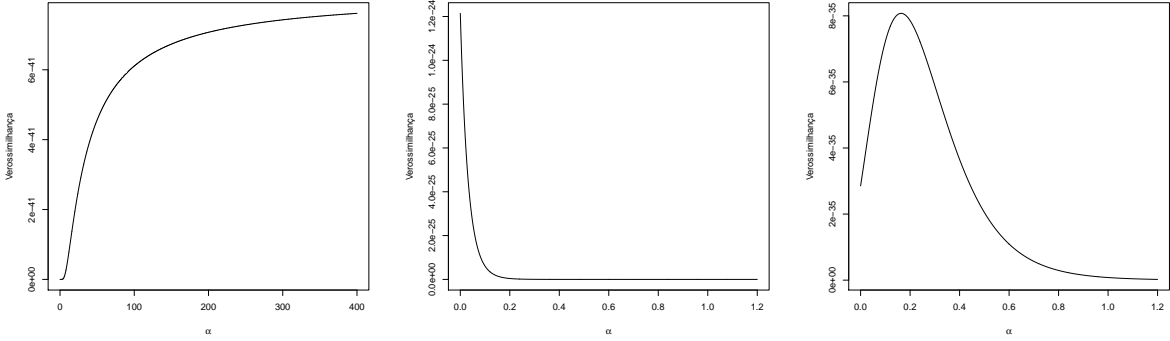


Figura 3.2: Funções de verossimilhança da distribuição normal bimodal padrão quando todas as observações são maiores que um, menores que um e quando há observações tanto menores quanto maiores que um, em módulo, respectivamente.

Quando existem observações tanto menores que um, em módulo, quanto maiores ou iguais a um, em módulo, é de interesse encontrar o valor do parâmetro de forma α que maximiza a função de verossimilhança em (3.19). Assim como acontece com a distribuição normal assimétrica padrão não há, geralmente, soluções analíticas para encontrar o EMV para α e nem provas de que será único. Porém, existem condições que garantem sua existência. A Proposição 2 a seguir mostra tais condições e os possíveis valores que o EMV para α pode assumir.

Proposição 2. *Seja (x_1, \dots, x_n) uma amostra iid da distribuição normal bimodal padrão. Denote por $I_{<1}$ o conjunto de índices das observações tais que $|x_i| < 1$ e por $I_{>1}$ o conjunto de índices das observações tais que $|x_i| > 1$. Seja C_α o conjunto dos diferentes valores de α , $\alpha > 0$, para os quais as seguintes condições são satisfeitas:*

$$\sum_{i \in I_{<1}} \frac{x_i^2 - 1}{(1 + \alpha x_i^2)(1 + \alpha)} = - \sum_{i \in I_{>1}} \frac{x_i^2 - 1}{(1 + \alpha x_i^2)(1 + \alpha)} \quad e$$

$$\left| \sum_{i \in I_{>1}} (x_i^2 - 1)K(x_i, \alpha_j) \right| > \left| \sum_{i \in I_{<1}} (x_i^2 - 1)K(x_i, \alpha_j) \right|,$$

onde $K(x, \alpha) = \frac{x^2 + 2\alpha x^2 + 1}{(1 + \alpha x^2)^2 (1 + \alpha)^2} > 0$ para todo $x \in \mathbb{R}$ e $\alpha \geq 0$. Então,

1. se $I_{<1} = \emptyset$ e $I_{>1} \neq \emptyset$ o EMV para α é infinito;

2. se $I_{<1} \neq \emptyset$ e $I_{>1} = \emptyset$ o EMV para α é zero;

3. se $I_{<1} \neq \emptyset$, $I_{>1} \neq \emptyset$ e $C_\alpha \neq \emptyset$ então uma das seguintes situações é observada:

(a) o EMV para α ocorrerá no extremo do intervalo, isto é, $\hat{\alpha} = 0$ se

$$f(\mathbf{x}|\alpha) < \prod_{i=1}^n \phi(x_i), \quad \forall \alpha \in C_\alpha;$$

(b) o EMV para α poderá não ser único. Nesse caso, os EMV para α serão os valores $\hat{\alpha}_1, \dots, \hat{\alpha}_k \in C_\alpha$, $k \geq 1$, tais que:

i.

$$f(\mathbf{x}|\hat{\alpha}_j) > \prod_{i=1}^n \phi(x_i), \quad j = 1, \dots, k;$$

ii.

$$f(\mathbf{x}|\hat{\alpha}_j) = \max_{\alpha \in C_\alpha} f(\mathbf{x}|\alpha), \quad j = 1, \dots, k;$$

(c) o EMV para α poderá não ser único e serão os valores 0 e $\hat{\alpha}_1, \dots, \hat{\alpha}_k \in C_\alpha$, $k \geq 1$, se

$$f(\mathbf{x}|\hat{\alpha}_j) = \prod_{i=1}^n \phi(x_i) = \max_{\alpha \in C_\alpha} f(\mathbf{x}|\alpha), \quad j = 1, \dots, k.$$

Prova: Seja (x_1, \dots, x_n) uma amostra iid da distribuição normal bimodal padrão. Denote por $I_{<1}$ o conjunto de índices das observações tais que $|x_i| < 1$ e por $I_{>1}$ o conjunto de índices das observações tais que $|x_i| > 1$.

Para encontrar os possíveis pontos de máximo e, conseqüentemente, os estimadores de máxima verossimilhança para α , deve-se realizar um estudo em $\frac{d}{d\alpha} f(\mathbf{x}|\alpha)$ ou, equivalentemente, em $\frac{d}{d\alpha} l(\alpha)$, onde $l(\alpha) = \ln[f(\mathbf{x}|\alpha)]$ é a log-verossimilhança da função de verossimilhança em (3.19). Seja

$$\begin{aligned} \frac{d}{d\alpha} l(\alpha) &= \sum_{i=1}^n \frac{x_i^2 - 1}{(1 + \alpha x_i^2)(1 + \alpha)} \\ &= \sum_{i \in I_{>1}} \frac{x_i^2 - 1}{(1 + \alpha x_i^2)(1 + \alpha)} + \sum_{i \in I_{<1}} \frac{x_i^2 - 1}{(1 + \alpha x_i^2)(1 + \alpha)} \end{aligned} \quad (3.20)$$

a primeira derivada de $l(\alpha)$ em relação a α . Portanto:

1. se $I_{<1} = \emptyset$ e $I_{>1} \neq \emptyset$ então $\frac{d}{d\alpha} l(\alpha) > 0$ para todo $\alpha \geq 0$. Logo, $f(\mathbf{x}|\alpha)$ é estritamente crescente e, portanto, o EMV para α é infinito;

2. se $I_{<1} \neq \emptyset$ e $I_{>1} = \emptyset$ então $\frac{d}{d\alpha}l(\alpha) < 0$ para todo $\alpha \geq 0$. Logo, $f(\mathbf{x}|\alpha)$ é estritamente decrescente e, portanto, o EMV para α é zero;

3. se $I_{<1} \neq \emptyset$ e $I_{>1} \neq \emptyset$ então deve-se encontrar, primeiramente, os valores de α tais que $\frac{d}{d\alpha}l(\alpha) = 0$. Logo, os valores de α tais que $\frac{d}{d\alpha}l(\alpha) = 0$ são os valores de $\alpha > 0$ tais que

$$\sum_{i \in I_{<1}} \frac{x_i^2 - 1}{(1 + \alpha x_i^2)(1 + \alpha)} = - \sum_{i \in I_{>1}} \frac{x_i^2 - 1}{(1 + \alpha x_i^2)(1 + \alpha)}. \quad (3.21)$$

Para que esses valores de $\alpha > 0$ sejam pontos de máximo deve-se ter que $\frac{d^2}{d\alpha^2}l(\alpha)$ avaliada nesses valores seja negativa. Seja

$$\begin{aligned} \frac{d^2}{d\alpha^2}l(\alpha) &= - \sum_{i=1}^n (x_i^2 - 1)K(x_i, \alpha) \\ &= - \sum_{i \in I_{>1}} (x_i^2 - 1)K(x_i, \alpha) - \sum_{i \in I_{<1}} (x_i^2 - 1)K(x_i, \alpha), \end{aligned}$$

onde $K(x, \alpha) = \frac{x^2 + 2\alpha x^2 + 1}{(1 + \alpha x^2)^2(1 + \alpha)^2} > 0$ para todo $x \in \mathbb{R}$ e $\alpha \geq 0$. Como $|x_i| > 1$ para todo $i \in I_{>1}$, tem-se $\sum_{i \in I_{>1}} (x_i^2 - 1)K(x_i, \alpha) > 0$. Analogamente, como $|x_i| < 1$ para todo $i \in I_{<1}$, tem-se $\sum_{i \in I_{<1}} (x_i^2 - 1)K(x_i, \alpha) < 0$. Logo, para que $\frac{d^2}{d\alpha^2}l(\alpha) < 0$ deve-se ter

$$\left| \sum_{i \in I_{>1}} (x_i^2 - 1)K(x_i, \alpha_j) \right| > \left| \sum_{i \in I_{<1}} (x_i^2 - 1)K(x_i, \alpha_j) \right|. \quad (3.22)$$

Portanto, os valores de $\alpha > 0$ que satisfazem (3.21) e (3.22) são pontos de máximo de $f(\mathbf{x}|\alpha)$ e formam o conjunto C_α . Sejam $\hat{\alpha}_1, \dots, \hat{\alpha}_k \in C_\alpha$, $k \geq 1$, os pontos de máximo global em $(0, \infty)$, ou seja, são os valores de C_α tais que

$$f(\mathbf{x}|\hat{\alpha}_j) = \max_{\alpha \in C_\alpha} f(\mathbf{x}|\alpha) \quad j = 1, \dots, k.$$

Como o domínio de α é $[0, \infty)$, deve-se verificar a relação entre $f(\mathbf{x}|0) = \prod_{i=1}^n \phi(x_i)$ e $f(\mathbf{x}|\hat{\alpha}_j)$, $j = 1, \dots, k$. Logo, nos casos onde $C_\alpha \neq \emptyset$,

- (a) se $f(\mathbf{x}|0) = \prod_{i=1}^n \phi(x_i) > f(\mathbf{x}|\alpha_j)$ para todo $\alpha_j \in C_\alpha$, então 0 é o ponto de máximo global para $f(\mathbf{x}|\alpha)$ e, portanto, é o EMV para α ;
- (b) se $f(\mathbf{x}|0) = \prod_{i=1}^n \phi(x_i) < f(\mathbf{x}|\hat{\alpha}_j)$, $j = 1, \dots, k$, então os valores $\hat{\alpha}_1, \dots, \hat{\alpha}_k \in C_\alpha$ são pontos de máximo global em todo o domínio de α e, portanto, estimadores de máxima verossimilhança para α ;

(c) se $f(\mathbf{x}|0) = \prod_{i=1}^n \phi(x_i) = f(\mathbf{x}|\hat{\alpha}_j)$, $j = 1, \dots, k$, então o valor 0, juntamente com $\hat{\alpha}_1, \dots, \hat{\alpha}_k \in C_\alpha$, são pontos de máximo global em todo o domínio de α e, portanto, estimadores de máxima verossimilhança para α .

□

Embora a Proposição 2 forneça condições para a existência do estimador de máxima verossimilhança para α , estes não podem ser obtidos analiticamente. Também nesse caso não se consegue garantir a unicidade do EMV para α . Assim como foi feito com o parâmetro de assimetria λ da distribuição normal assimétrica padrão, será considerado o algoritmo EM para obter um valor aproximado para o EMV para α , o qual será apresentado na próxima seção.

3.2.2 Construção do algoritmo EM para a família normal bimodal padrão

Para a construção do algoritmo EM é considerada a representação estocástica de Arellano-Valle *et al.* (2008) dada na expressão (2.9), ou seja, considera-se que, se $X \sim BN(\alpha)$, então

$$X \stackrel{d}{=} \sqrt{\frac{\alpha}{1+\alpha}} Z(2M-1) + \sqrt{\frac{1}{1+\alpha}} Y,$$

onde Z , M e Y são variáveis aleatórias independentes, $Z \sim \chi_3^2$, $M \sim \text{Bernoulli}(1/2)$, $Y \sim N(0, 1)$ e $\stackrel{d}{=}$ significa igualdade em distribuição. Seja $G = \sqrt{\frac{\alpha}{1+\alpha}} Z(2M-1)$. Considerando tal definição, o modelo normal bimodal padrão pode ser representado hierarquicamente da seguinte forma:

$$\begin{aligned} X|G = g, \alpha &\sim N(g, (1+\alpha)^{-1}) \\ G|\alpha &\sim f_{G|\alpha}(g) = \left(\frac{1+\alpha}{\alpha}\right)^{3/2} g^2 \phi\left(\sqrt{\frac{1+\alpha}{\alpha}} g\right). \end{aligned} \quad (3.23)$$

Logo, utilizando o Teorema de Bayes pode-se mostrar que

$$G|X = x, \alpha \sim f_{G|x,\alpha}(g) = \frac{(1+\alpha)^2}{\alpha(1+\alpha x^2)} g^2 \phi\left(g; \frac{\alpha}{1+\alpha} x, \frac{\alpha}{(1+\alpha)^2}\right). \quad (3.24)$$

Seja $\mathbf{G} = (G_1, \dots, G_n)^t$ a representação para as variáveis não observadas, latentes. Tem-se, de (3.23), que a função de log-verossimilhança aumentada associada a (\mathbf{X}, \mathbf{G}) é

dada por

$$\begin{aligned} \ln f(\mathbf{X}, \mathbf{G}|\alpha) &= -n \ln 2\pi + 2n \ln(1 + \alpha) - \frac{3}{2}n \ln \alpha + 2 \sum_{i=1}^n \ln G_i - \\ &\frac{1 + \alpha}{2} \left[\sum_{i=1}^n X_i^2 - 2 \sum_{i=1}^n X_i G_i + \frac{1 + \alpha}{\alpha} \sum_{i=1}^n G_i^2 \right]. \end{aligned} \quad (3.25)$$

Para o passo E do algoritmo EM deve-se calcular a esperança condicional de (3.25) dados a amostra observada \mathbf{x} e o parâmetro estimado $\hat{\alpha}$. Esta esperança é dada por

$$\begin{aligned} Q(\alpha, \hat{\alpha}) &= E [\ln f(\mathbf{X}, \mathbf{G}|\alpha) | \mathbf{X} = \mathbf{x}, \alpha = \hat{\alpha}] \\ &= -n \ln 2\pi + 2n \ln(1 + \alpha) - \frac{3}{2}n \ln \alpha + 2 \sum_{i=1}^n E [\ln G_i | x_i, \hat{\alpha}] - \\ &\frac{1 + \alpha}{2} \left[\sum_{i=1}^n x_i^2 - 2 \sum_{i=1}^n x_i E [G_i | x_i, \hat{\alpha}] + \frac{1 + \alpha}{\alpha} \sum_{i=1}^n E [G_i^2 | x_i, \hat{\alpha}] \right]. \end{aligned} \quad (3.26)$$

No passo M do algoritmo EM deve-se maximizar (3.26) em relação a α para que se possa atualizar o valor de $\hat{\alpha}$. Logo, deve-se encontrar os valores de α que são raízes da seguinte expressão:

$$\begin{aligned} \frac{d}{d\alpha} Q(\alpha, \hat{\alpha}) &= \frac{2n}{1 + \alpha} - \frac{3n}{2\alpha} - \frac{\sum_{i=1}^n x_i^2}{2} + \sum_{i=1}^n x_i E [G_i | x_i, \hat{\alpha}] - \\ &\frac{1 + \alpha}{2\alpha} \sum_{i=1}^n E [G_i^2 | x_i, \hat{\alpha}] - \frac{1 + \alpha}{2\alpha^2} \sum_{i=1}^n E [G_i^2 | x_i, \hat{\alpha}]. \end{aligned} \quad (3.27)$$

De (3.27) nota-se que se faz necessário conhecer $E [G_i | x_i, \hat{\alpha}]$ e $E [G_i^2 | x_i, \hat{\alpha}]$. A partir de (3.24) tem-se a seguinte expressão geral para os momentos de uma variável aleatória com tal distribuição:

$$E [G^k | X = x, \alpha] = \frac{(1 + \alpha)^2}{\alpha(1 + \alpha x^2)} E [M^{k+2}], \quad (3.28)$$

onde $M \sim N(\alpha x(1 + \alpha)^{-1}, \alpha(1 + \alpha)^{-2})$, para $k = 1, 2, \dots$

Logo, de (3.28) tem-se que

$$E [G | X = x, \alpha] = \frac{\alpha(\alpha x^3 + 3x)}{(1 + \alpha)(\alpha x^2 + 1)} \quad \text{e} \quad (3.29)$$

$$E [G^2 | X = x, \alpha] = \frac{\alpha(\alpha^2 x^4 + 6\alpha x^2 + 3)}{(1 + \alpha)^2(\alpha x^2 + 1)}. \quad (3.30)$$

Portanto, o algoritmo EM se resume a:

- **Passo E:** Dado $\alpha = \hat{\alpha}$, calcular $E[G_i|x_i, \hat{\alpha}]$ e $E[G_i^2|x_i, \hat{\alpha}]$, de acordo com as expressões (3.29) e (3.30), respectivamente;
- **Passo M:** Atualizar $\hat{\alpha}$ maximizando $Q(\alpha, \hat{\alpha})$ com respeito a α .

3.2.3 Inferência bayesiana na família normal bimodal padrão

Sejam as variáveis aleatórias X_1, \dots, X_n e Y_1, \dots, Y_m que, condicional em α , são iid com distribuição normal bimodal padrão dada em (2.7). É fácil observar que, se α , *a priori*, tem uma distribuição própria $\pi(\alpha)$, a distribuição *a posteriori* para α , a densidade preditiva *a priori* de \mathbf{X} e a densidade preditiva *a posteriori* de \mathbf{Y} são dadas, respectivamente, por

$$\pi(\alpha|\mathbf{x}) = \frac{\pi(\alpha) \prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha}}{E_\alpha \left[\prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha} \right]}, \quad (3.31)$$

$$f_{BN}(\mathbf{x}) = \left[\prod_{i=1}^n \phi(x_i) \right] E_\alpha \left[\prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha} \right] \quad \text{e} \quad (3.32)$$

$$f_{BN}(\mathbf{y}|\mathbf{x}) = \left[\prod_{i=1}^m \phi(y_i) \right] E_{\alpha|\mathbf{x}} \left[\prod_{i=1}^m \frac{1+\alpha y_i^2}{1+\alpha} \right], \quad (3.33)$$

onde $E_\alpha[T]$ denota o valor esperado de T com respeito à distribuição $\pi(\alpha)$. Nesse caso, a esperança *a posteriori* de α é dada por

$$E_{\alpha|\mathbf{x}}[\alpha] = \frac{E_\alpha \left[\alpha \prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha} \right]}{E_\alpha \left[\prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha} \right]}. \quad (3.34)$$

Neste trabalho assume-se que, *a priori*, $\alpha \sim NT(c, d^2; 0, \infty)$. Considerando essa distribuição *a priori* para α segue de (3.31) e (3.34) que

$$\begin{aligned} \pi(\alpha|\mathbf{x}) &= \frac{\phi(\alpha; c, d^2)}{1 - \Phi(0; c, d^2)} \frac{\prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha}}{E_\alpha \left[\prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha} \right]} \\ &= \frac{\phi(\alpha; c, d^2) \prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha}}{\int_0^\infty \phi(\alpha; c, d^2) \prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha} d\alpha}, \end{aligned} \quad (3.35)$$

$$\begin{aligned} E_{\alpha|\mathbf{x}}[\alpha] &= \frac{\int_0^\infty \alpha \phi(\alpha; c, d^2) \prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha} d\alpha}{[1 - \Phi(0; c, d^2)] E_\alpha \left[\prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha} \right]} \\ &= \frac{\int_0^\infty \alpha \phi(\alpha; c, d^2) \prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha} d\alpha}{\int_0^\infty \phi(\alpha; c, d^2) \prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha} d\alpha}. \end{aligned} \quad (3.36)$$

3.3 Inferência na família de distribuições normais bimodais assimétricas

Esta seção se destina ao problema de inferência sobre os parâmetros (α, λ) da distribuição normal bimodal assimétrica padrão. Seguindo as seções anteriores, serão estabelecidas as condições de existência de estimadores de máxima verossimilhança para (α, λ) assim como serão obtidas as distribuições *a posteriori* para (α, λ) , as distribuições preditivas *a priori* e *a posteriori*, bem como a esperança *a posteriori* de (α, λ) .

3.3.1 Inferência clássica na família normal bimodal assimétrica padrão

Suponha que, fixado (α, λ) , as variáveis aleatórias X_1, \dots, X_n são iid com distribuição normal bimodal assimétrica padrão dada em (2.10). Logo, para cada amostra observada $\mathbf{x} = (x_1, \dots, x_n)^t$ do vetor aleatório $\mathbf{X} = (X_1, \dots, X_n)^t$, possíveis estimadores via método dos momentos são obtidos ao resolver o seguinte sistema em relação a α e λ :

$$\begin{cases} \bar{x} &= E(X|\alpha, \lambda) \\ \bar{x}^2 &= E(X^2|\alpha, \lambda) \end{cases} \quad (3.37)$$

De acordo com as expressões (2.11) e (2.12) o sistema em (3.37) se torna

$$\begin{cases} \bar{x} &= \frac{1}{1+\alpha} \sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{1+\lambda^2}} + \frac{\alpha}{1+\alpha} \sqrt{\frac{2}{\pi}} \frac{\lambda}{(1+\lambda^2)^{3/2}} (3 + 2\lambda^2) \\ \bar{x}^2 &= \frac{1+3\alpha}{1+\alpha} \end{cases} \quad (3.38)$$

Nota-se, de (3.38), que, a partir da segunda expressão em tal sistema consegue-se encontrar uma solução para α e, nesse caso, o estimador via método dos momentos para α é o mesmo que em (3.18). Porém, não há solução analítica real para λ .

Em uma amostra de tamanho n selecionada independentemente de uma distribuição normal bimodal assimétrica padrão com parâmetros α e λ tem-se que a função de verossimilhança é dada por

$$\begin{aligned} f(\mathbf{x}|\alpha, \lambda) &= 2^n \left[\prod_{i=1}^n \frac{1 + \alpha x_i^2}{1 + \alpha} \phi(x_i) \Phi(\lambda x_i) \right] \\ &= \frac{f(\mathbf{x}|\alpha) f(\mathbf{x}|\lambda)}{\prod_{i=1}^n \phi(x_i)}, \end{aligned} \quad (3.39)$$

onde $f(\mathbf{x}|\alpha)$ e $f(\mathbf{x}|\lambda)$ são as mesmas definidas em (3.2) e (3.19), respectivamente. Como pode-se observar de (3.39) a função de verossimilhança da distribuição normal bimodal as-

simétrica padrão é função das funções de verossimilhança da distribuição normal bimodal padrão e normal assimétrica padrão.

Para uma dada amostra \mathbf{x} nota-se que (3.39) se trata de uma função de duas variáveis, α e λ . Logo, deve-se encontrar valores para o par (α, λ) que maximizam a função de verossimilhança (3.39). O resultado a seguir fornece condições de existência para tais pontos que maximizam a função de verossimilhança e quais seus possíveis valores.

Proposição 3. *As condições de existência para os estimadores de máxima verossimilhança para os parâmetros de forma α e de assimetria λ na família de distribuições normais bimodais assimétricas padrão são as mesmas verificadas para a existência dos estimadores de máxima verossimilhança para os parâmetros de forma e assimetria, respectivamente, nas famílias de distribuições normal bimodal padrão e normal assimétrica padrão e, se existirem, serão os mesmos encontrados em tais famílias.*

Prova: Seja (x_1, \dots, x_n) uma amostra iid da distribuição normal bimodal assimétrica padrão definida em (2.10). Assim como ocorre com funções de uma variável, para encontrar possíveis pontos de máximo e, conseqüentemente, os EMV para α e λ , deve-se, primeiramente, realizar um estudo nas derivadas de primeira ordem $\frac{\partial}{\partial \alpha} l(\alpha, \lambda)$ e $\frac{\partial}{\partial \lambda} l(\alpha, \lambda)$, onde

$$\begin{aligned} l(\alpha, \lambda) &= \ln[f(\mathbf{x}|\alpha, \lambda)] \\ &= l(\alpha) + l(\lambda) - \sum_{i=1}^n \phi(x_i), \end{aligned}$$

$l(\alpha)$ e $l(\lambda)$ são as log-verossimilhanças das funções de verossimilhança das distribuições normal bimodal padrão e normal assimétrica padrão, respectivamente.

Tem-se que

$$\frac{\partial}{\partial \alpha} l(\alpha, \lambda) = \sum_{i=1}^n \frac{x_i^2 - 1}{(1 + \alpha x_i^2)(1 + \alpha)} \quad \text{e} \quad (3.40)$$

$$\frac{\partial}{\partial \lambda} l(\alpha, \lambda) = \sum_{i=1}^n \frac{\phi(\lambda x_i)}{\Phi(\lambda x_i)} x_i. \quad (3.41)$$

Observa-se de (3.40) e (3.41) que $\frac{\partial}{\partial \alpha} l(\alpha, \lambda)$ e $\frac{\partial}{\partial \lambda} l(\alpha, \lambda)$ não dependem de, respectivamente, λ e α e que tais expressões são iguais a, respectivamente, (3.20) e (3.3). Logo, os valores

que podem resolver $\frac{\partial}{\partial \alpha} l(\alpha, \lambda) = 0$ e $\frac{\partial}{\partial \lambda} l(\alpha, \lambda) = 0$ são os mesmos que resolveriam, respectivamente, $\frac{d}{d\alpha} l(\alpha) = 0$ e $\frac{d}{d\lambda} l(\lambda) = 0$, ou seja, os pares formados por um ponto crítico de $l(\alpha)$ e por um ponto crítico de $l(\lambda)$ é um ponto crítico de $l(\alpha, \lambda)$. Além disso, as condições que levam a função de verossimilhança da distribuição normal bimodal assimétrica padrão a ser crescente ou decrescente em relação a cada um dos parâmetros α e λ também são as mesmas considerando as famílias normal bimodal padrão e normal assimétrica padrão, respectivamente.

As derivadas de segunda ordem de $l(\alpha, \lambda)$ são

$$\begin{aligned}\frac{\partial^2}{\partial \alpha^2} l(\alpha, \lambda) &= -\sum_{i=1}^n (x_i^2 - 1)K(x_i, \alpha), \\ \frac{\partial^2}{\partial \lambda^2} l(\alpha, \lambda) &= -\sum_{i=1}^n H(x_i, \lambda)[\lambda x_i \Phi(\lambda x_i) + \phi(\lambda x_i)] \quad e \\ \frac{\partial^2}{\partial \alpha \partial \lambda} l(\alpha, \lambda) &= 0 = \frac{\partial^2}{\partial \lambda \partial \alpha} l(\alpha, \lambda),\end{aligned}$$

onde $H(x, \lambda) = \frac{\phi(\lambda x)}{\Phi^2(\lambda x)} x^2 > 0$ para todo $x \in \mathbb{R}$ e $\lambda \in \mathbb{R}$ e $K(x, \alpha) = \frac{x^2 + 2\alpha x^2 + 1}{(1 + \alpha x^2)^2 (1 + \alpha)^2} > 0$ para todo $x \in \mathbb{R}$ e $\alpha \geq 0$. Assim como ocorre com as derivadas de primeira ordem nota-se que as derivadas de segunda ordem de $l(\alpha, \lambda)$ em relação a α e λ são as mesmas derivadas de segunda ordem de, respectivamente, $l(\alpha)$ em relação a α e $l(\lambda)$ em relação a λ .

Para verificar se os pontos críticos, que são os pontos que anulam as derivadas de primeira ordem, de funções de duas variáveis são pontos de máximo deve-se fazer um estudo na matriz formada pelas derivadas de segunda ordem. A matriz $M(\alpha, \lambda)$ das derivadas de segunda ordem de $l(\alpha, \lambda)$ é

$$\begin{aligned}M(\alpha, \lambda) &= \begin{bmatrix} \frac{\partial^2}{\partial \alpha^2} l(\alpha, \lambda) & \frac{\partial^2}{\partial \alpha \partial \lambda} l(\alpha, \lambda) \\ \frac{\partial^2}{\partial \lambda \partial \alpha} l(\alpha, \lambda) & \frac{\partial^2}{\partial \lambda^2} l(\alpha, \lambda) \end{bmatrix} \\ &= \begin{bmatrix} -\sum_{i=1}^n (x_i^2 - 1)K(x_i, \alpha) & 0 \\ 0 & -\sum_{i=1}^n H(x_i, \lambda)[\lambda x_i \Phi(\lambda x_i) + \phi(\lambda x_i)] \end{bmatrix}.\end{aligned}$$

Suponha que (α_c, λ_c) é um ponto crítico de $l(\alpha, \lambda)$. Para que esse ponto seja ponto de máximo, de acordo com Stewart (2002), deve-se ter $\det[M(\alpha_c, \lambda_c)] > 0$ e $\frac{\partial^2}{\partial \alpha^2} l(\alpha_c, \lambda_c) < 0$, onde $\det[M(\alpha, \lambda)]$ denota o determinante de $M(\alpha, \lambda)$ e dado por

$$\det[M(\alpha, \lambda)] = \left[\frac{\partial^2}{\partial \alpha^2} l(\alpha, \lambda) \right] \left[\frac{\partial^2}{\partial \lambda^2} l(\alpha, \lambda) \right].$$

Como para existir pontos de máximo deve-se ter $\frac{\partial^2}{\partial \alpha^2} l(\alpha, \lambda) < 0$ então, para que $\det[M(\alpha, \lambda)] > 0$, também deve-se ter $\frac{\partial^2}{\partial \lambda^2} l(\alpha, \lambda) < 0$. Logo, para que (α_c, λ_c) seja ponto de máximo de

$l(\alpha, \lambda)$, α_c deve satisfazer as condições para ser ponto de máximo em $l(\alpha)$ e λ_c deve satisfazer as condições para ser ponto de máximo em $l(\lambda)$. Portanto, o EMV para λ na família normal assimétrica padrão também será EMV para λ na família normal bimodal assimétrica padrão e o EMV para α na família normal bimodal padrão também será EMV para α na família normal bimodal assimétrica padrão. \square

De acordo com a Proposição 3 estimar, via máxima verossimilhança, os parâmetros de forma α e de assimetria λ da distribuição normal bimodal assimétrica padrão dada em (2.10) é equivalente a estimar cada um dos parâmetros individualmente na suposição de que os dados vem das famílias mais simples, distribuição normal bimodal padrão e normal assimétrica padrão, respectivamente. Dessa forma, pode-se utilizar os algoritmos EM vistos nas Seções 3.1.1 e 3.2.1 para se obter estimativas de máxima verossimilhança para λ e α , respectivamente.

3.3.2 Inferência bayesiana na família normal bimodal assimétrica

Sejam as variáveis aleatórias X_1, \dots, X_n e Y_1, \dots, Y_m que, condicional em α e λ , são iid com distribuição normal bimodal assimétrica padrão dada em (2.10). Assim como em Elal-Olivero *et al.* (2009), considere α e λ independentes, com distribuições *a priori* $\pi(\alpha)$ e $\pi(\lambda)$, respectivamente. A distribuição conjunta *a posteriori* para (α, λ) é dada por

$$\pi(\alpha, \lambda | \mathbf{x}) = \frac{\pi(\alpha) \prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha}}{E_\alpha \left[\prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha} \right]} \frac{\pi(\lambda) \prod_{i=1}^n \Phi(\lambda x_i)}{E_\lambda \left[\prod_{i=1}^n \Phi(\lambda x_i) \right]} = \pi(\alpha | \mathbf{x}) \pi(\lambda | \mathbf{x}), \quad (3.42)$$

onde $E_\alpha[T]$ e $E_\lambda[T]$ denotam o valor esperado de T com respeito às distribuições $\pi(\alpha)$ e $\pi(\lambda)$, respectivamente.

Como pode-se ver em (3.42) a condição de independência *a priori* de α e λ induz à condição de independência *a posteriori* para α e λ . E mais, $\pi(\alpha | \mathbf{x})$ e $\pi(\lambda | \mathbf{x})$ são as mesmas de (3.31) e (3.11), respectivamente. Isso quer dizer que, assim como ocorre com o EMV para α e λ , sob a família normal bimodal assimétrica padrão as distribuições *a posteriori* para cada um dos parâmetros α e λ também são as obtidas considerando-se famílias mais simples - a distribuição normal bimodal padrão para α e a distribuição normal assimétrica padrão para λ . Logo, os valores esperados *a posteriori* de α , $E_{\lambda, \alpha | \mathbf{x}}[\alpha]$,

e λ , $E_{\lambda, \alpha | \mathbf{x}}[\lambda]$, na família normal bimodal assimétrica padrão são os mesmos obtidos nas famílias mais simples, ou seja,

$$E_{\lambda, \alpha | \mathbf{x}}[\lambda] = \frac{E_{\lambda} [\lambda \prod_{i=1}^n \Phi(\lambda x_i)]}{E_{\lambda} [\prod_{i=1}^n \Phi(\lambda x_i)]} \quad \text{e}$$

$$E_{\lambda, \alpha | \mathbf{x}}[\alpha] = \frac{E_{\alpha} \left[\alpha \prod_{i=1}^n \frac{1 + \alpha x_i^2}{1 + \alpha} \right]}{E_{\alpha} \left[\prod_{i=1}^n \frac{1 + \alpha x_i^2}{1 + \alpha} \right]}.$$

A função de densidade preditiva *a priori* para \mathbf{X} e a função de densidade preditiva *a posteriori* para \mathbf{Y} são dadas por

$$f_{SBN}(\mathbf{x}) = 2^n \left[\prod_{i=1}^n \phi(x_i) \right] E_{\lambda} \left[\prod_{i=1}^n \Phi(\lambda x_i) \right] E_{\alpha} \left[\prod_{i=1}^n \frac{1 + \alpha x_i^2}{1 + \alpha} \right]$$

$$= \frac{f_{SN}(\mathbf{x}) f_{BN}(\mathbf{x})}{\prod_{i=1}^n \phi(x_i)} \quad \text{e} \quad (3.43)$$

$$f_{SBN}(\mathbf{y} | \mathbf{x}) = 2^m \left[\prod_{i=1}^m \phi(y_i) \right] E_{\lambda | \mathbf{x}} \left[\prod_{i=1}^m \Phi(\lambda y_i) \right] E_{\alpha | \mathbf{x}} \left[\prod_{i=1}^m \frac{1 + \alpha y_i^2}{1 + \alpha} \right]$$

$$= \frac{f_{SN}(\mathbf{y} | \mathbf{x}) f_{BN}(\mathbf{y} | \mathbf{x})}{\prod_{i=1}^m \phi(y_i)}, \quad (3.44)$$

onde $f_{SN}(\mathbf{x})$, $f_{BN}(\mathbf{x})$, $f_{SN}(\mathbf{y} | \mathbf{x})$ e $f_{BN}(\mathbf{y} | \mathbf{x})$ são as mesmas definidas em (3.12), (3.32), (3.13) e (3.33), respectivamente. Como pode-se ver a partir de (3.43) e (3.44) as funções preditivas *a priori* e *a posteriori* para os dados na família normal bimodal assimétrica padrão são funções das densidades preditivas *a priori* e *a posteriori* obtidas para as famílias normal assimétrica padrão e normal bimodal padrão. No entanto, neste caso, não é apenas o produto dessas preditivas.

Elal-Olivero *et al.* (2009) consideram testes e estimações na família de distribuições normais bimodais assimétricas. São assumidas distribuições *a priori* que são misturas de uma distribuição contínua com uma distribuição pontual. Para testes de hipóteses consideram para λ uma distribuição *a priori* que é uma mistura entre uma distribuição pontual em 0 e uma distribuição normal com pequena variância. Para α , *a priori*, consideram uma mistura entre uma distribuição pontual em 0 e ou uma distribuição normal truncada em valores maiores que 0 ou uma distribuição uniforme com limite superior bem alto. Em relação à estimação consideram distribuições *a priori* que são misturas entre distribuições pontuais em 0 e distribuições normais padrão (truncada em valores positivos quando necessário) e/ou distribuições uniformes para ambos os parâmetros.

O trabalho desenvolvido aqui estende os resultados de Elal-Olivero *et al.* (2009) por apresentar uma contribuição teórica para a família normal bimodal assimétrica padrão tanto em relação à inferência clássica quanto em relação à inferência bayesiana.

De acordo com as expressões em (3.15), (3.35), (3.16) e (3.32) quando, *a priori*, α e λ são independentes, $\alpha \sim NT(c, d^2; 0, \infty)$ e $\lambda \sim N(m, v)$, a distribuição *a posteriori* para (α, λ) em (3.42) e a distribuição preditiva *a priori* para os dados em (3.43) se tornam

$$\begin{aligned}\pi(\alpha, \lambda|\mathbf{x}) &= \frac{\phi(\alpha; c, d^2)}{1 - \Phi(0; c, d^2)} \frac{\prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha}}{E_\alpha \left[\prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha} \right]} \phi(\lambda; m, v) \frac{\Phi_n(\lambda\mathbf{x})}{\Phi_n(m\mathbf{x}; \mathbf{0}, I_n + v\mathbf{x}\mathbf{x}^t)} \\ &= \frac{\phi(\alpha; c, d^2) \prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha}}{\int_0^\infty \phi(\alpha; c, d^2) \prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha} d\alpha} \phi(\lambda; m, v) \frac{\Phi_n(\lambda\mathbf{x})}{\Phi_n(m\mathbf{x}; \mathbf{0}, I_n + v\mathbf{x}\mathbf{x}^t)} \quad e \\ f_{SBN}(\mathbf{x}) &= 2^n \phi_n(\mathbf{x}) \Phi_n(m\mathbf{x}; \mathbf{0}, I_n + v\mathbf{x}\mathbf{x}^t) E_\alpha \left[\prod_{i=1}^n \frac{1 + \alpha x_i^2}{1 + \alpha} \right].\end{aligned}$$

Sob essas especificações *a priori* para α e λ e a independência *a priori* entre os parâmetros tem-se, de acordo com as expressões (3.36) e (3.17), que as esperanças *a posteriori* para α e λ sob a família normal bimodal assimétrica padrão são dadas, respectivamente, por

$$\begin{aligned}E_{\lambda, \alpha|\mathbf{x}}[\alpha] &= \frac{\int_0^\infty \alpha \phi(\alpha; c, d^2) \prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha} d\alpha}{[1 - \Phi(0; c, d^2)] E_\alpha \left[\prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha} \right]} \\ &= \frac{\int_0^\infty \alpha \phi(\alpha; c, d^2) \prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha} d\alpha}{\int_0^\infty \phi(\alpha; c, d^2) \prod_{i=1}^n \frac{1+\alpha x_i^2}{1+\alpha} d\alpha} \quad e\end{aligned}\tag{3.45}$$

$$E_{\lambda, \alpha|\mathbf{x}}[\lambda] = m + v\mathbf{x}^t\xi,\tag{3.46}$$

onde ξ é como definido na Seção 3.1.3.

Capítulo 4

Análise de dados simulados

Neste Capítulo serão realizadas inferências sobre os parâmetros de forma e assimetria da distribuição normal bimodal assimétrica padrão assumindo resultados vistos no Capítulo 3, Seção 3.3 e considerando dados simulados.

Primeiramente, será realizado um estudo Monte Carlo com o objetivo de avaliar as estimativas obtidas pelo estimador de máxima verossimilhança via algoritmo EM e pelos estimadores bayesianos, média e moda *a posteriori*.

Em seguida é feito um novo estudo Monte Carlo para realizar uma análise de sensibilidade nas estimativas resultantes dos estimadores bayesianos quando, *a priori*, considera-se uma distribuição pouco informativa (entendida aqui como uma distribuição com variância grande) para uma distribuição informativa.

São utilizadas as linguagens de programação Ox para as programações dos processos de geração de amostras e estimação dos parâmetros e R para geração de gráficos.

4.1 Estudo Monte Carlo

Nesta Seção são realizados estudos Monte Carlo para avaliar a qualidade dos estimadores propostos. Foram consideradas 10.000 réplicas Monte Carlo da distribuição normal bimodal assimétrica padrão dada em (2.10) com parâmetros α iguais a 0, 1; 0, 25; 5 e 10 e λ iguais a 0, 5; 1; 5 e 10 e assumidos dois tamanhos de amostras: $n = 30$ e 100.

Para cada réplica Monte Carlo foram calculados os estimadores de máxima verossimilhança via algoritmo EM tendo como valor inicial para α e λ os próprios valores utilizados na geração dos dados. Assim como em Lin (2009), o algoritmo EM era encerrado quando havia diferenças absolutas na log-verossimilhança menores que 10^{-4} . Também foram cal-

culadas as esperanças e modas *a posteriori*, denotadas, respectivamente, por E e Mo . Para o cálculo das esperanças *a priori* envolvidas na distribuição *a posteriori* para (α, λ) , dada em (3.42), foi considerado um método Newton-Cotes, chamado regra de Simpson (Migon e Gamerman, 1999, página 144, Seção 5.4), para aproximá-las. Como distribuição *a priori* para λ foi considerada uma distribuição normal com média zero e variância 10.000 e, para α , uma distribuição normal truncada em valores não negativos com parâmetros de posição μ igual a zero e σ^2 igual a 10.000, o que garante uma alta variabilidade *a priori* para ambos os parâmetros. Após gerar as 10.000 amostras Monte Carlo tomou-se as estimativas medianas obtidas por cada estimador e os erros quadráticos médio empíricos, que estão entre parênteses. A adoção do erro quadrático médio para quantificar a diferença entre os valores reais e os valores estimados é devida à sua popularidade e fácil compreensão. As Tabelas 4.1 e 4.2 mostram os resultados obtidos para amostras de tamanho 30 e 100, respectivamente, onde os valores entre parênteses são os erros quadráticos médio.

Das Tabelas 4.1 e 4.2 nota-se que, em geral, os erros quadráticos médio referentes a cada estimador aumentam à medida que os valores dos parâmetros aumentam. Exceções ocorrem para estimadores de λ quando $n = 30$ e para a moda *a posteriori* quando amostras são geradas assumindo (α, λ) iguais a $(5, 10)$ e $(10, 10)$, onde os valores de EQM são menores do que quando amostras são geradas de valores $(\alpha, \lambda) = (5, 5)$ e $(10, 5)$, respectivamente. Para a esperança *a posteriori* de λ tal fato também ocorre para amostras geradas com par de parâmetros iguais a $(10, 10)$. Para amostras de tamanho 100 o erro quadrático médio obtido pelo estimador de máxima verossimilhança (EMV) para λ quando amostras são geradas assumindo $(5, 10)$ é menor que o gerado assumindo $(5, 5)$.

Na Tabela 4.1 observa-se, em relação ao EQM, que, para os casos onde amostras foram geradas com $\alpha = 0, 1$ e $0, 25$, ou seja, amostras unimodais, os melhores resultados para α foram obtidos para o EMV , com pouca diferença em relação à moda *a posteriori*. Para os outros valores de α considerados, os melhores resultados são obtidos pela moda *a posteriori* e os piores para o EMV , que se mostraram muito inferiores até mesmo em relação à esperança *a posteriori*. Em relação aos estimadores para λ , em todos os casos onde amostras são geradas com $\lambda = 0, 5$ os melhores resultados são obtidos pelo estimador de máxima verossimilhança, exceto quando $\alpha = 10$, onde o melhor resultado é observado

Tabela 4.1: Medianas e erros quadráticos médio para estimador de máxima verossimilhança (EMV), esperança (E) e moda (Mo) *a posteriori*, $n = 30$.

α	EMV	E	Mo	λ	EMV	E	Mo
0,1	0,0982	0,3147	0,0904	0,5	0,5196	0,5427	0,5201
	(0,0397)	(2,1993)	(0,0434)		(0,0911)	(0,1141)	(0,0920)
	0,0990	0,3147	0,0904	1	1,0387	1,1103	1,0389
	(0,0396)	(2,1993)	(0,0434)		(0,6416)	(2,1346)	(0,5338)
	0,0974	0,3147	0,0904	5	5,9369	7,9854	5,9521
	(0,0406)	(2,1993)	(0,0434)		(6,116,1001)	(1,305,3510)	(224,0917)
	0,0956	0,3147	0,0904	10	15,7649	38,1537	16,5754
	(0,0413)	(2,1993)	(0,0434)		(18,021,1747)	(2,877,5171)	(440,7827)
0,25	0,2456	0,5106	0,2429	0,5	0,5179	0,5403	0,5182
	(0,1071)	(11,4581)	(0,1142)		(0,0775)	(0,0977)	(0,0788)
	0,2469	0,5106	0,2429	1	1,0441	1,1167	1,0444
	(0,1065)	(11,4581)	(0,1142)		(2,7270)	(7,4599)	(1,4126)
	0,2447	0,5106	0,2429	5	6,0145	8,4037	6,1027
	(0,1104)	(11,4581)	(0,1142)		(5,575,1081)	(1,583,8590)	(257,2652)
	0,2439	0,5106	0,2429	10	15,8461	59,0920	17,3268
	(0,1104)	(11,4581)	(0,1142)		(23,545,4595)	(3,114,7523)	(443,9013)
5	5,2679	55,4109	5,2880	0,5	0,5109	0,5292	0,5119
	(4,844,6879)	(3,059,0135)	(205,6763)		(0,0536)	(1,5543)	(0,0621)
	5,2527	55,4109	5,2880	1	1,0433	1,1209	1,0438
	(4,902,9075)	(3,049,0135)	(205,6763)		(27,5277)	(141,6459)	(8,4953)
	5,2932	55,4109	5,2880	5	7,6950	81,4919	9,9330
	(5,048,7668)	(3,049,0135)	(205,6763)		(2,189,7800)	(4,313,2217)	(215,3769)
	5,2540	55,4109	5,2880	10	10,5739	81,8802	12,7613
	(4,841,4567)	(3,049,0135)	(205,6763)		(2,634,4116)	(4,602,5946)	(194,7672)
10	11,0461	75,6759	11,2502	0,5	0,5097	0,5266	0,5099
	(13,157,1947)	(3,888,8555)	(347,6735)		(0,9204)	(1,4863)	(0,3021)
	10,9612	75,6759	11,2502	1	1,0453	1,1226	1,0458
	(13,373,8386)	(3,888,8555)	(347,6735)		(25,5398)	(218,0350)	(9,2122)
	11,4688	75,6759	11,2502	5	7,0402	81,3415	8,8652
	(13,656,9993)	(3,888,8555)	(347,6735)		(1,167,1398)	(4,818,7902)	(153,5763)
	11,0978	75,6759	11,2502	10	10,1330	81,5228	10,2598
	(13,153,2270)	(3,888,8555)	(347,6735)		(1,261,7165)	(4,785,1361)	(124,2592)

para a moda *a posteriori*. Para todos os outros casos de λ considerados, os melhores resultados foram obtidos pela moda *a posteriori*. Em todos os casos onde λ é igual a 0, 5 e 1 os piores resultados são obtidos pela esperança *a posteriori*, assim como para λ e α iguais a 5 e 10.

Ainda considerando $n = 30$ nota-se da Tabela 4.1 que em relação às estimativas medianas, para λ , os melhores resultados são observados para as estimativas do estimador de máxima verossimilhança, similar ao que ocorre com os erros quadráticos médio apenas em três situações. Isso ocorreu quando (α, λ) são iguais a $(0, 1; 0, 5)$, $(0, 25; 0, 5)$ e $(5; 0, 5)$.

Tabela 4.2: Medianas e erros quadráticos médio para estimador de máxima verossimilhança (EMV), esperança (E) e moda (Mo) *a posteriori*, $n = 100$.

α	EMV	E	Mo	λ	EMV	E	Mo
0,1	0,1799	0,1588	0,0990	0,5	0,5005	0,5119	0,5051
	(0,0217)	(0,0148)	(0,0092)		(0,0009)	(0,0207)	(0,0197)
	0,2039	0,1588	0,0990	1	1,0034	1,0314	1,0113
	(0,0231)	(0,0148)	(0,0092)	(0,0085)	(0,0470)	(0,0421)	
	0,1781	0,1588	0,0990	5	5,0213	5,6453	5,2164
	(0,0207)	(0,0148)	(0,0092)	(0,0485)	(46,9331)	(18,2565)	
	0,1747	0,1588	0,0990	10	10,0434	12,9277	10,9699
	(0,0195)	(0,0148)	(0,0092)	(0,0914)	(638,6752)	(196,5911)	
0,25	0,3496	0,3090	0,2482	0,5	0,5006	0,5113	0,5045
	(0,0429)	(0,0320)	(0,0206)		(0,0009)	(0,0183)	(0,0173)
	0,3852	0,3090	0,2482	1	1,0034	1,0327	1,0121
	(0,0453)	(0,0320)	(0,0206)	(0,0069)	(0,0456)	(0,0403)	
	0,3461	0,3090	0,2482	5	5,0227	5,7574	5,2615
	(0,0404)	(0,0320)	(0,0206)	(0,0422)	(88,1646)	(29,1948)	
	0,3433	0,3090	0,2482	10	10,0454	13,5687	11,2023
	(0,0383)	(0,0320)	(0,0206)	(0,0778)	(857,2160)	(240,6875)	
5	6,2705	14,6602	5,1555	0,5	0,5010	0,5086	0,5036
	(348,7661)	(849,8832)	(45,2071)		(0,0079)	(0,0102)	(0,0097)
	6,1798	14,6602	5,1555	1	1,0054	1,0375	1,0161
	(348,2741)	(849,8832)	(45,2071)	(0,0286)	(0,0408)	(0,0339)	
	5,9301	14,6602	5,1555	5	5,0420	10,5057	6,7488
	(422,9867)	(849,8832)	(45,2071)	(956,9158)	(2.180,7778)	(264,0720)	
	5,5575	14,6602	5,1555	10	10,0497	83,1303	18,5445
	(357,1527)	(849,8832)	(45,2071)	(871,5067)	(3.668,0878)	(393,4214)	
10	12,0432	49,0520	10,3679	0,5	0,5017	0,5087	0,5039
	(4.632,6535)	(2.380,8700)	(243,3872)		(0,0086)	(0,0096)	(0,0092)
	11,8221	49,0520	10,3679	1	1,0109	1,0360	1,0149
	(4.632,2105)	(2.380,8700)	(243,3872)	(0,0317)	(0,0405)	(0,0333)	
	13,2471	49,0520	10,3679	5	7,4227	58,2723	9,6670
	(5.407,9951)	(2.380,8700)	(243,3872)	(1.128,6963)	(3.191,4854)	(254,5815)	
	11,2951	49,0520	10,3679	10	10,2240	82,8625	16,3691
	(4.703,7454)	(2.380,8700)	(243,3872)	(1.407,17917)	(4.249,5327)	(291,9732)	

Para α nota-se que, apenas para α iguais a 5 e 10 as estimativas medianas, em dois casos, são melhores para a moda *a posteriori*. Isso ocorre quando $(\alpha, \lambda) = (5, 5)$ e $(10, 5)$. Por sinal, esses são os dois casos para α iguais a 5 e 10 que estão semelhantes ao que ocorre com os erros quadráticos médio. Para os outros valores de α tanto as estimativas medianas de cada estimador quanto os valores de EQM levam à mesma conclusão sobre os melhores estimadores.

Para amostras de tamanho 100 observa-se da Tabela 4.2 que, em relação a α e observando os valores do erro quadrático médio, melhores resultados são obtidos para a moda

a posteriori e os piores para o *EMV* quando $\alpha = 0, 1; 0, 25$ e 10 . Para os estimadores de λ os melhores resultados foram observados quando são utilizados o *EMV*, exceto quando amostras são geradas assumindo $(\alpha, \lambda) = (5, 5), (5, 10), (10, 5)$ e $(10, 10)$, onde os melhores resultados são notados quando é considerada a moda *a posteriori* como estimador de λ . Para todos os valores de λ os piores resultados são obtidos quando é utilizada a esperança *a posteriori*.

Levando em conta $n = 100$ verifica-se na Tabela 4.2 que, para α , os casos onde ocorrem as melhores estimativas medianas são os mesmos onde ocorrem os menores erros quadráticos médio, ou seja, todos ocorrem para a moda *a posteriori*. Para λ todas as melhores estimativas medianas são para o *EMV* e há quatro casos que não são similares com o EQM. Tal fato ocorre nos casos onde os menores valores de EQM são para a moda *a posteriori*, $(\alpha, \lambda) = (5, 5), (5, 10), (10, 5)$ e $(10, 10)$.

A partir das Tabelas 4.1 e 4.2 também é possível notar que para amostras de tamanho 100 os erros quadráticos médio são menores que os casos onde amostras tem tamanho 30, exceto para a moda *a posteriori* de λ com amostras geradas sob $(5, 5), (5, 10), (10, 5)$ e $(10, 10)$ e para o valor esperado *a posteriori* de λ quando é considerado o caso onde $(\alpha, \lambda) = (10, 10)$. Devido ao alto número de casos onde o EQM é grande pode-se dizer que não é recomendável utilizar apenas estimação pontual para inferências sobre os parâmetros de assimetria e de forma da distribuição normal bimodal assimétrica padrão.

4.2 Análise de sensibilidade

Nesta Seção são realizados estudos Monte Carlo com o intuito de analisar o comportamento das estimativas bayesianas quando, *a priori*, considera-se distribuições com diferentes variâncias. Foram consideradas 10.000 réplicas Monte Carlo da distribuição normal bimodal assimétrica padrão dada em (2.10) com parâmetros (α, λ) iguais a $(0, 0), (0, 25; 1)$ e $(10, 10)$ e assumido tamanho amostral $n = 100$.

Para cada réplica Monte Carlo foram calculadas as esperanças e modas *a posteriori*, denotadas, respectivamente, por E e Mo , onde $i = \alpha$ ou λ . Assim como na Seção anterior, para o cálculo das esperanças *a priori* envolvidas na distribuição *a posteriori* para (α, λ) , dada em (3.42), foi considerada a regra de Simpson para aproximá-las. Como distribuição

a priori para λ foram consideradas distribuições normais com média zero e variâncias 10.000, 10 e 1 e, para α , distribuições normais truncadas em valores não negativos com parâmetros de posição μ igual a zero e $\sigma^2 = 10.000, 10$ e 1, o que garante diferentes variabilidades *a priori* para ambos os parâmetros. A Tabela 4.3 mostra as especificações *a priori* consideradas e suas respectivas média e variância.

Tabela 4.3: Médias e variâncias para as distribuições *a priori* utilizadas.

α			λ		
Distribuição	Média	Variância	Distribuição	Média	Variância
$NT(0, 10.000; 0, \infty)$	79,7885	3.633,8023	$N(0, 10.000)$	0	10.000
$NT(0, 10; 0, \infty)$	2,5231	3,6338	$N(0, 10)$	0	10
$NT(0, 1; 0, \infty)$	0,7979	0,3634	$N(0, 1)$	0	1

Tomou-se as estimativas medianas de cada estimador e os erros quadráticos médio empíricos. A Tabela 4.4 apresenta as estimativas medianas apresentadas pelo valor esperado (E) e modal (Mo) *a posteriori* e, entre parênteses, os erros quadráticos médio (EQM) de tais estimadores considerando as diferentes distribuições *a priori* (D.P.). Assim como ocorreu na Seção 4.1, a escolha pela mediana é devido ao alto número de estimativas atípicas, como se pode observar nas Figuras 4.1, 4.2 e 4.3. Tais valores atípicos influenciam fortemente o EQM.

Tabela 4.4: Medianas e erros quadráticos médio para esperança e moda *a posteriori* considerando diferentes especificações *a priori*.

(α, λ)		(0, 0)		(0, 25; 1)		(10, 10)	
Parâmetro	D.P.	E	Mo	E	Mo	E	Mo
α	$NT(0, 10.000; 0, \infty)$	0,1094 (0,0168)	0,0000 (0,0036)	0,3170 (0,0315)	0,2482 (0,0206)	49,0520 (2.380,8707)	10,3679 (243,3873)
	$NT(0, 10; 0, \infty)$	0,0803 (0,0114)	0,0000 (0,0035)	0,3075 (0,0313)	0,2477 (0,0204)	5,4596 (21,7401)	4,8731 (27,4576)
	$NT(0, 1; 0, \infty)$	0,0787 (0,0110)	0,0000 (0,0035)	0,2999 (0,0257)	0,2438 (0,0184)	2,6043 (54,9894)	2,4942 (56,6205)
λ	$N(0, 10.000)$	0,0021 (0,0173)	0,0005 (0,0169)	1,0327 (0,0456)	1,0121 (0,0403)	82,8625 (4.249,5327)	16,3691 (291,9732)
	$N(0, 10)$	0,0002 (0,0172)	0,0005 (0,0168)	1,0293 (0,0441)	1,0087 (0,0391)	4,7738 (27,4358)	3,9491 (36,5715)
	$N(0, 1)$	0,0002 (0,0167)	0,0005 (0,0163)	0,9988 (0,0340)	0,9798 (0,0314)	2,3976 (57,7613)	2,2470 (60,0755)

Quando são geradas amostras da distribuição normal bimodal assimétrica padrão com

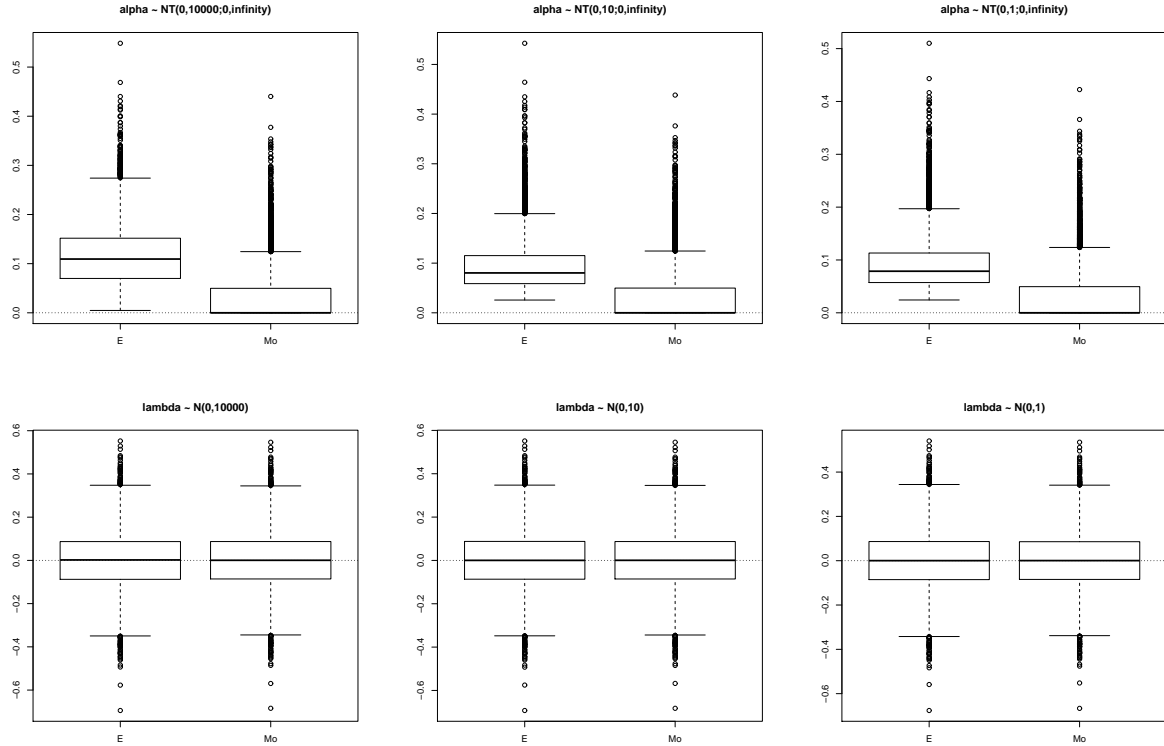


Figura 4.1: Esperanças (E) e modas (Mo) *a posteriori* para α e $\lambda - X_1, \dots, X_n \stackrel{iid}{\sim} SBN(0,0)$.

par de parâmetros (α, λ) igual a $(0,0)$ nota-se, na Figura 4.1, que o valor de α sempre é superestimado pela esperança *a posteriori*. Os valores modais *a posteriori* de α também estão superestimados, porém, aproximadamente 75% de todos os valores estimados estão concentrados entre 0 e 0,1. Esses resultados ocorrem independente da distribuição *a priori* para α considerada. Da Tabela 4.4 observa-se, nos valores do EQM, que, para α , a moda *a posteriori* fornece melhores resultados que a esperança *a posteriori*. Para a moda *a posteriori* pequenas diferenças aparecem ao considerar diferentes distribuições *a priori*, enquanto para a esperança *a posteriori* essa diferença é um pouco mais significativa. Em relação a λ conclui-se, a partir da Figura 4.1, que tanto o comportamento dos valores esperados quanto dos valores modais *a posteriori* são semelhantes e parecem estimar bem o valor do parâmetro de assimetria, para todas as distribuições *a priori* consideradas. Tal fato pode ser observado na Tabela 4.4, que mostra que os erros quadráticos médio para a esperança e moda *a posteriori* não se diferenciam muito. Também nota-se que, assim como ocorreu com α , melhores resultados são obtidos quando é considerada a distribuição

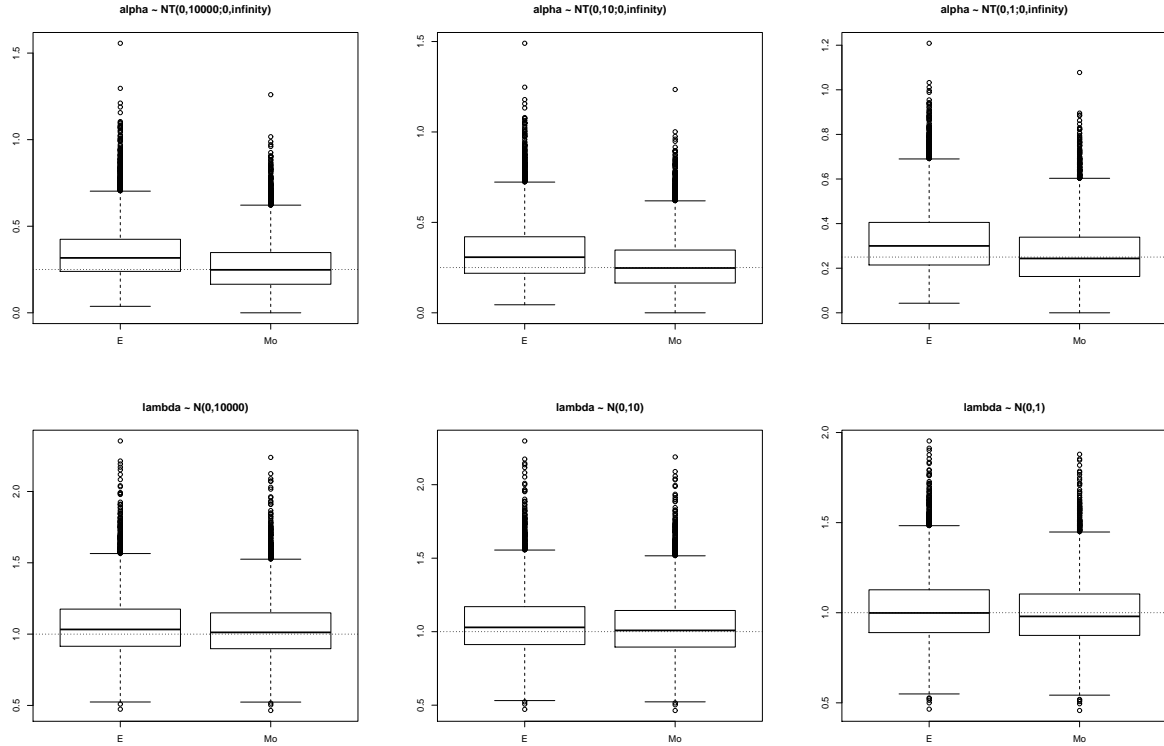


Figura 4.2: Esperanças (E) e modas (Mo) *a posteriori* para α e $\lambda - X_1, \dots, X_n \stackrel{iid}{\sim} SBN(0, 25; 1)$.

a priori com menor parâmetro de escala.

Nas amostras da distribuição em (2.10) com α igual a 0,25 e λ igual a 1 nota-se, na Figura 4.2, que, em relação a α , as esperanças *a posteriori* tendem a superestimar o valor do parâmetro. Já as modas *a posteriori* estão mais concentradas ao redor do valor real, 0,25, e estão menos dispersas que os valores obtidos pela esperança *a posteriori*. As distribuições empíricas dos estimadores apresentam, aproximadamente, um comportamento semelhante quando, *a priori*, o valor esperado para α é, de acordo com a Tabela 4.3, aproximadamente 79,7885 e 2,5231. Quando é considerada a distribuição *a priori* com menor valor do parâmetro de escala observa-se um pequeno decréscimo na variabilidade das estimativas. Em relação a λ observa-se que os valores medianos obtidos por ambos os estimadores bayesianos são aproximadamente iguais a 1, que é o valor de λ utilizado para gerar as amostras. Assim como ocorre para α , nota-se que a moda *a posteriori* tende a ser um melhor estimador e que quando é considerada uma distribuição *a priori* com menor variabilidade as estimativas ficam mais próximas entre si. Na Tabela 4.4 nota-se, em

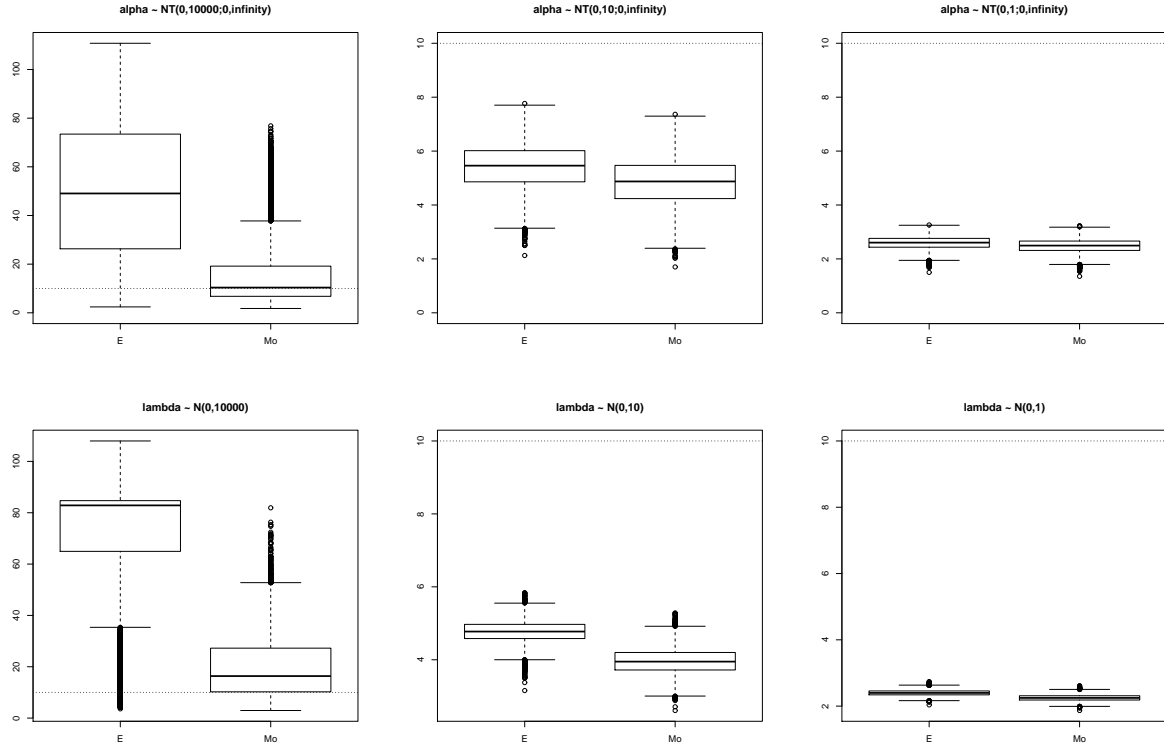


Figura 4.3: Esperanças (E) e modas (Mo) *a posteriori* para α e λ - $X_1, \dots, X_n \stackrel{iid}{\sim} SBN(10, 10)$.

relação ao EQM, que tanto para α quanto para λ melhores resultados foram obtidos pela moda *a posteriori*. Para as distribuições *a priori* consideradas observa-se que, à medida que a variabilidade *a priori* diminui, tanto para a moda e esperança *a posteriori* o EQM também diminui, indicando que as estimativas vão ficando mais próximas do valor real.

Para as amostras que foram geradas assumindo os parâmetros (10, 10) observa-se, na Figura 4.3, que, tanto para α quanto para λ , a escolha da distribuição *a priori* influencia muito as estimativas que podem ser obtidas. Nos casos considerados onde a distribuição *a priori* apresenta maior variabilidade nota-se que os valores esperados *a posteriori* estão, em sua maioria, superestimados. As modas *a posteriori* estão menos dispersas e mais próximas do valor real, em especial para α . Para as outras distribuições *a priori* consideradas todas as estimativas estão subestimadas. Quanto menor o parâmetro de escala das distribuições *a priori* menor é a dispersão das estimativas e mais subestimados estão os valores obtidos. Nesses casos, a esperança *a posteriori* apresenta melhores resultados por estarem mais próximos do valor real, como também pode ser observado pela Tabela

4.4. Na Tabela 4.4 nota-se que, quando são consideradas as distribuições *a priori* com maior variabilidade, $NT(0, 10.000; 0, \infty)$ para α e $N(0, 10.000)$ para λ , o EQM resultante é o maior. Os casos onde os estimadores bayesianos apresentaram os menores valores para o erro quadrático médio foram observados quando são consideradas as distribuições *a priori* com variâncias aproximadamente igual a 3,6338 para α e igual a 10 para λ , de acordo com a Tabela 4.3.

Capítulo 5

Análise dos dados de fronteira

Neste capítulo serão considerados os dados de fronteira, que são apresentados por Azzalini em seu site, <http://azzalini.stat.unipd.it/SN/>, e que foi discutido em Azzalini e Capitanio (1999). Esses dados consistem de uma amostra de 50 observações de uma distribuição normal assimétrica com parâmetro de posição igual a 0, parâmetro de escala igual a 1 e parâmetro de assimetria λ igual a 5, ou seja, são 50 observações de uma distribuição normal bimodal assimétrica padrão com parâmetros $\alpha = 0$ e $\lambda = 5$. Considerando a família normal assimétrica com parâmetros de posição e escala esses dados são interessantes pois o estimador de máxima verossimilhança para λ é infinito. Sartori (2006) obteve o valor 6,243 como estimativa para λ utilizando o estimador alternativo ao *EMV*, baseado na correção de viés apresentada em Firth (1993), com um intervalo de 95% de confiança igual a (1,799; 61,340). Bayes e Branco (2007), utilizando a aproximação para o fator de correção de viés proposto por Sartori (2006), obtiveram o valor 8,67 como estimativa para o parâmetro de assimetria. Utilizando inferência bayesiana Liseo e Loperfido (2006) estimaram pontualmente λ em 15,9 através do uso da mediana *a posteriori* quando, *a priori*, é considerada a distribuição não informativa de Jeffreys. O intervalo de 95% de credibilidade de mais alta probabilidade *a posteriori* foi dado por (4,2; 52,5). No trabalho de Bayes e Branco (2007) também foi realizada inferência para λ utilizando inferência bayesiana e, considerando a aproximação proposta para a distribuição *a priori* de Jeffreys, obtiveram as estimativas 6,85, ∞ e 31,27 para a moda, média e mediana *a posteriori*, respectivamente. Quando consideram a distribuição *a priori* uniforme para a reparametrização $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$ do parâmetro de assimetria λ as estimativas obtidas pelos estimadores moda, média e mediana *a posteriori* são, respectivamente, 3,97; 7,61 e 5,26.

5.1 Inferência nos dados de fronteira

Todos os trabalhos citados no início deste capítulo assumem a função de verossimilhança da distribuição normal assimétrica com parâmetros de posição e escala para os dados de fronteira. Neste trabalho os dados serão tratados como se fossem provindos da distribuição normal bimodal assimétrica padrão dada em (2.10). Serão realizadas inferências para os parâmetros de assimetria λ e de forma α . Os métodos inferenciais realizados serão os vistos no Capítulo 3. Para o algoritmo EM é utilizado como valor inicial para α o valor obtido pelo estimador via método dos momentos, dado em (3.18), cuja estimativa é aproximadamente 0,2076. Para λ alguns valores foram gerados aleatoriamente como valores iniciais para verificar a convergência. Assim como na Seção 4.1, o algoritmo EM era encerrado quando havia diferenças na log-verossimilhança menores que 10^{-4} . Como distribuição *a priori* para α foram consideradas distribuições normais truncadas em valores positivos e para λ foram utilizadas distribuições *a priori* normais. Para ambas as distribuições *a priori* serão utilizados $\mu = 0$ e $\sigma^2 = 10.000$. A Figura 5.1 mostra o comportamento da função de verossimilhança da distribuição normal bimodal assimétrica padrão em função dos parâmetros α e λ . A Tabela 5.1 apresenta as estimativas obtidas pelo estimador de máxima verossimilhança (*EMV*), esperança (*E*) e moda (*Mo*) *a posteriori*.

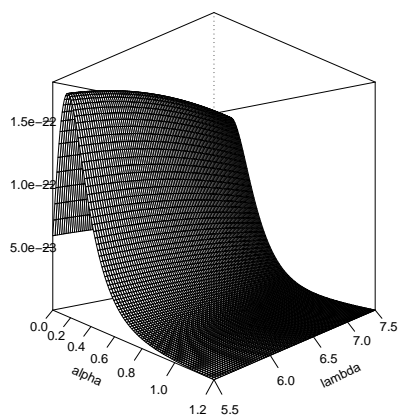


Figura 5.1: Função de verossimilhança da distribuição normal bimodal assimétrica padrão para os dados de fronteira.

Tabela 5.1: Inferência pontual para os dados de fronteira.

Parâmetro	EMV	E	Mo
α	0,1711	0,2696	0,1638
λ	6,4691	8,2660	6,5444

Na Tabela 5.1 observa-se que as estimativas encontradas pelos estimadores de máxima verossimilhança e pelas modas *a posteriori* são bem próximas. O ponto formado pelos EMV para α e λ parece realmente ser o ponto de máximo da função de verossimilhança quando observa-se a Figura 5.1. A Figura 5.2 mostra o histograma dos dados de fronteira. Nela também estão desenhadas a densidade original considerada para gerar a amostra, normal bimodal assimétrica padrão com parâmetros $(0, 5)$, a densidade preditiva *a posteriori* quando são consideradas as especificações *a priori* citadas acima e as densidades da normal bimodal assimétrica padrão utilizando, como parâmetros, as estimativas apresentadas na Tabela 5.1.

Como era de se esperar nota-se, a partir da Figura 5.2, que as densidades obtidas quando são consideradas as estimativas encontradas pelo EMV e pela moda *a posteriori* são bem próximas uma da outra. Tais densidades são as que mais se aproximam da densidade real e parecem se adequar bem aos dados. As densidades obtidas quando são utilizadas as estimativas pela esperança *a posteriori* e pela densidade preditiva *a posteriori* também estão próximas uma da outra e, apesar de mais distantes da densidade da normal bimodal assimétrica padrão com parâmetros $(0, 5)$, também parecem se adequar bem aos dados de fronteira.

De acordo com as densidades apresentadas na Figura 5.2 pode-se concluir que a distribuição normal bimodal assimétrica padrão pode ser considerada uma boa opção para esse conjunto de dados, uma vez que se trata de um modelo mais parcimonioso, com dois parâmetros a serem estimados, e apresentou resultados tão satisfatórios como os obtidos quando é considerado o modelo normal assimétrico com parâmetros de posição e escala, que possui três parâmetros.

Dados de fronteira

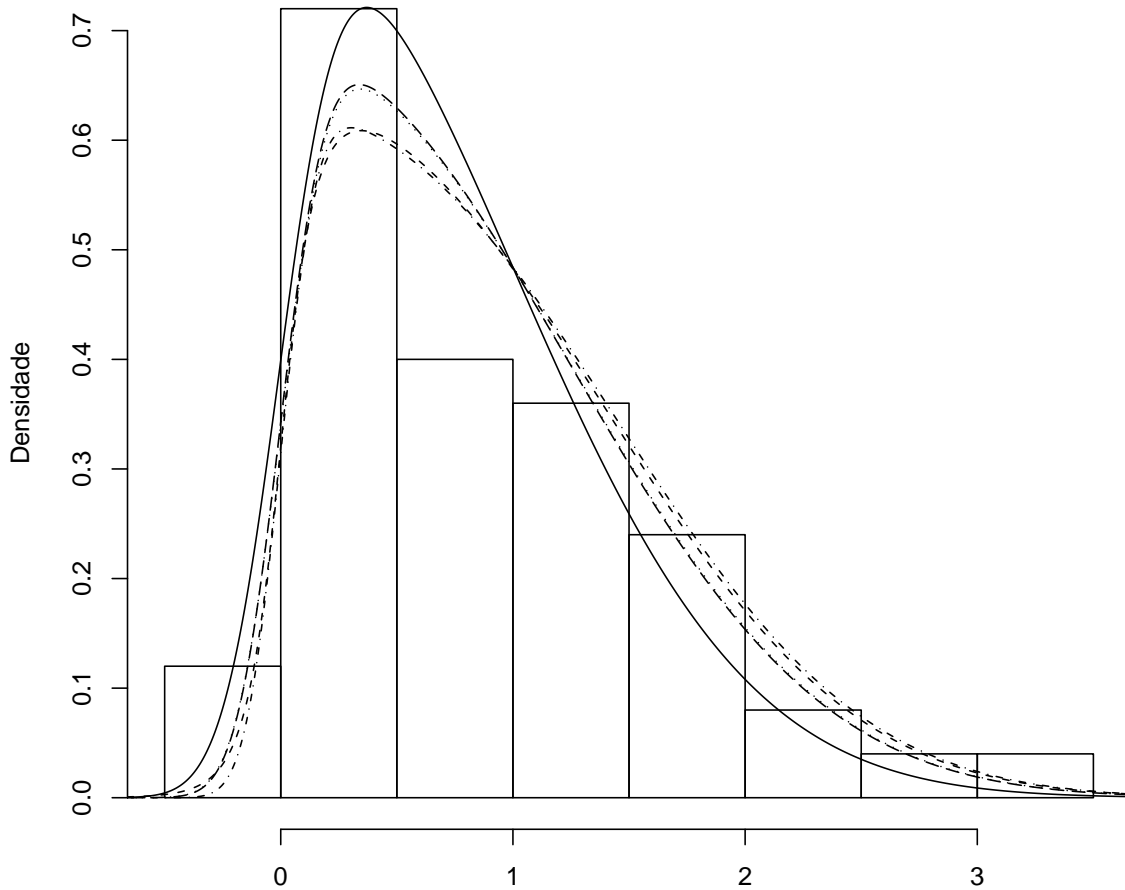


Figura 5.2: Histograma dos dados de fronteira com a densidade considerada para gerar a amostra (linha sólida), densidade preditiva *a posteriori* (linha com traços curtos) e densidades normais bimodais assimétricas padrão com os parâmetros estimados pelos estimadores de máxima verossimilhança (linha pontilhada), esperanças (linha tracejada com pontos) e modas (linha com traços longos) *a posteriori*.

Capítulo 6

Conclusão

Neste trabalho foram realizados estudos nas famílias de distribuições normais assimétricas, normais bimodais e normais bimodais assimétricas sem considerar os parâmetros de posição e escala. Para cada uma das famílias foram vistas algumas de suas propriedades e suas representações estocásticas. Para cada família foram realizadas inferências sobre os parâmetros de assimetria das distribuições normais assimétricas e normais bimodais assimétricas e sobre os parâmetros de forma das distribuições normais bimodais e normais bimodais assimétricas. Foram utilizados os estimadores via método dos momentos e de máxima verossimilhança na abordagem clássica e, na abordagem bayesiana, foram apresentadas as distribuições *a posteriori* para ambos os parâmetros e as densidades preditivas *a priori* e *a posteriori*. Para cada uma das famílias consideradas foram estabelecidas condições para a existência de estimadores de máxima verossimilhança para os parâmetros. Como, geralmente, não há soluções analíticas para obter os estimadores de máxima verossimilhança para cada um dos parâmetros foram construídos os algoritmos EM para cada um deles. Foram realizados estudos Monte Carlo em dados simulados para verificar a qualidade dos estimadores de máxima verossimilhança e dos estimadores bayesianos esperança e moda *a posteriori*. Outro estudo Monte Carlo é realizado para averiguar as mudanças no comportamento dos estimadores bayesianos quando se altera a variabilidade *a priori* de cada um dos parâmetros. Por fim, foram realizadas inferências nos dados de fronteira de Azzalini.

Em relação à inferência notou-se que tanto as distribuições *a posteriori* quanto os estimadores de máxima verossimilhança para os parâmetros de assimetria e forma da família normal bimodal assimétrica podem ser obtidos considerando famílias mais simples.

No estudo Monte Carlo para avaliar a qualidade dos estimadores considerados verificou-se que, para amostras pequenas, o melhor estimador para α é a moda *a posteriori* quando há uma bimodalidade mais evidente e o estimador de máxima verossimilhança caso os dados sejam unimodais. Para estimar λ o melhor estimador é a moda *a posteriori*. No caso onde foram geradas amostras grandes o melhor estimador é a moda *a posteriori* para α e o estimador de máxima verossimilhança para λ . Também notou-se que, em geral, para amostras grandes os erros quadráticos médio obtidos são menores do que os obtidos para amostras pequenas. Pelo fato de haver diversas estimativas atípicas para cada um dos estimadores considerados não é recomendável apenas realizar estimações pontuais.

Para avaliar a sensibilidade dos estimadores bayesianos considerando diferentes variabilidades *a priori* notou-se que a utilização de distribuições *a priori* com alta variabilidade sempre resulta em boas estimativas quando é considerada a moda *a posteriori* como estimador dos parâmetros. Quando as distribuições *a priori* possuem parâmetros de posição mais próximos dos valores utilizados para os parâmetros α e λ na geração das amostras os resultados se mostraram bons.

Analisando os dados de fronteira viu-se que os dados podem ser considerados como vindos de uma distribuição normal bimodal assimétrica padrão, uma vez que, considerando os resultados obtidos, as distribuições obtidas parecem se adequar bem aos dados e fornece resultados similares ao encontrado na literatura usando modelos menos parcimoniosos.

Como continuação do trabalho aqui realizado sugere-se utilizar os estimadores via método dos momentos assim como possíveis modificações e verificar suas propriedades e estimativas geradas. Em relação à inferência bayesiana pode-se considerar as distribuições *a priori* de referência, como por exemplo a distribuição obtida baseada no método de Berger e Bernardo (1992) e a distribuição de Jeffreys, para o parâmetro de forma da distribuição normal bimodal e da distribuição normal bimodal assimétrica. Em um projeto futuro sugere-se o estudo das famílias estendidas de distribuições normais com parâmetros de posição e escala. Também podem ser realizados testes de hipóteses sobre os parâmetros dessas famílias. Pretende-se avaliar a aplicação de misturas de distribuições normais bimodais e de distribuições normais bimodais assimétricas na estimação de densidades. Pode-se investigar a existência de famílias conjugadas nas famílias estendidas de distribuições normais e de propriedades, como, por exemplo, a distribuição

de transformações lineares, de potências, etc.

Referências Bibliográficas

Arellano-Valle, R. B. e Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, **33**, 561-574.

Arellano-Valle, R. B. e Genton, M. G. (2005). On fundamental skew distributions. *Journal of Multivariate Analysis*, **96** (1), 93-116.

Arellano-Valle, R. B., Branco, M. D. e Genton, M. G. (2006). A unified view on skewed distributions arising from selections. *The Canadian Journal of Statistics*, **34** (4), 581-601.

Arellano-Valle, R. B., Del Pino, G. e San Martin, E. (2002). Definition and probabilistic properties of skew-distributions. *Statistics and Probability Letters*, **58** (2), 111-121.

Arellano-Valle, R. B., Elal-Oliviero, D. e Gómez, H. W. (2008). A new class of symmetric distributions. *não publicado*.

Arellano-Valle, R. B., Gómez, H. W. e Quintana, F. A. (2005). Statistical inference for a general class of asymmetric distributions. *J. Statist. Plann. Inference*, **128** (2), 427-443.

Arnold, B. C., Castillo, E. e Sarabia, J. M. (2002). Conditionally specified multivariate skewed distributions. *Sankhya. Indian J. Statist. Ser. A*, **64** (2), 206-226. Selected articles from San Antonio Conference in honour of C. R. Rao (San Antonio, TX, 2000).

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandi-*

navian Journal of Statistics, **12**, 171-178.

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica*, **46**, 199-208.

Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, **32** (2) 159-200 (with discussion by Marc G. Genton and a rejoinder by the author).

Azzalini, A. e Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *J. Roy. Statist. Soc. B*, **61**, 579-602.

Azzalini, A. e Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. *J. Roy. Statist. Soc. B*, **65**, 367-389.

Azzalini, A. e Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**, 715-726.

Bayes, C. L. e Branco, M. D. (2007). Bayesian inference for the skewness parameter of the scalar skew-normal distribution. *Brazilian Journal of Probability and Statistics*, **21**, 141-163.

Berger, J. O. e Wolpert, R. L. (1984). *The Likelihood Principle*. Series Editor: Shanti S. Gupta, Lecture Notes–Monograph Series, Volume 6, Hayward, CA.

Berger, J. O. e Bernardo, J. M. (1992). On the development of reference priors. *Bayesian Statistics 4*. Oxford University Press, 35-60.

Castro, L. M., Arellano-Valle, R. B. e Loschi, R. H. (2008). Bayesian inference for the skew-normal shape parameter: An application to change point problems. *não publicado*.

Dempster, A. P., Laird, N. M. e Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society B*, **39** (1), 1-38.

Elal-Olivero, D., Gomez, H. W. e Quintana, F. A. (2009). Bayesian modeling using a class of bimodal skew-elliptical distributions. *Journal of Statistical Planning and Inference*, **139** (4), 1484-1492.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, **80**, 27-38.

Genton, M. e Loperfido, N. (2005). Generalized skew-elliptical distributions and their quadratic forms. *Annals of the Institute of Statistical Mathematics*, **57** (2), 389-401.

Genton, M. G. (2004). *Skew-elliptical distributions and their applications: A Journey Beyond Normality*. Edited Volume, Chapman & Hall, CRC Press, Boca Raton, FL.

Gómez, H. W., Venegas, O. e Bolfarine, H. (2007). Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics*, **18** (4), 395-407.

Gupta, A. K., Chang, F. C. e Huang, W. J. (2002). Some skew-symmetric models. *Random Operators and Stochastic Equations*, **10** (2), 133-140.

Henze, N. (1986). A probabilistic representation of the skew-normal distribution. *Scandinavian Journal of Statistics*, **13**, 271-275.

Jones, C. I. (1997). On the evolution of the world income distribution. *Journal of Economic Perspectives*, **11** (3), 19-36.

Lin, T. I. (2009). Maximum likelihood estimation for multivariate skew normal mixture models. *Journal of Multivariate Analysis*, **100**, 257-265.

Liseo, B. (1990). The skew-normal class of densities: inferential aspects from a Bayesian viewpoint. *Statistica*, **50**, 59-70. (Italian).

Liseo, B. e Loperfido, N. (2006). A note on reference priors for the scalar skew-normal distribution. *Journal of Statistical Planning and Inference*, **136**, 373-389.

Ma, Y. e Genton, M. G. (2004). A flexible class of skew-symmetric distributions. *Scandinavian Journal of Statistics*, **31**, 459-468.

Migon, H. S. e Gamerman, D. (1999). *Statistical Inference: An Integrated Approach*. Arnold, New York.

Nadarajah, S. e Kotz, S. (2003). Skewed distributions generated by the normal kernel. *Statist. Probab. Lett.*, **65**, 269-277.

Sartori, N. (2003). *Bias reduction of maximum likelihood estimates: skew normal and skew t distributions*. Technical Report, Università di Padova, Italy.

Sartori, N. (2006). Bias prevention of maximum likelihood estimates: skew normal and skew-t distributions. *Journal of Statistical Planning and Inference*, **136**, 4259-4275.

Stewart, J. (2002). *Cálculo, vol. 2*, 4a edição, São Paulo: Pioneira Thomson Learning.

Wang, J., Boyer, J. e Genton, M. G. (2004). A skew-symmetric representation of multivariate distributions. *Statistica Sinica*, **14** (4), 1259-1270.