Universidade Estadual de Campinas

INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO CIENTÍFICA

Departamento de Estatística

Dissertação de Mestrado

Estimação e Diagnóstico em Modelos Birnbaum-Saunders Skew-Normal

por

Lúcia Rolim Santana †

Mestrado em Estatística - Campinas - SP

Orientador: Prof. Dr. Filidor Edilfonso Vilca Labra

†Este trabalho contou com apoio financeiro da FAPEAM.

ESTIMAÇÃO e DIAGNÓSTICO EM MODELOS BIRNBAUM-SAUNDERS SKEW-NORMAL

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Lúcia Rolim Santana e aprovada pela comissão julgadora.

Campinas, 18 de Maio de 2009.

Prof. Dr. Filidor Edilfonso Vilca Labra
Orientador

Banca Examinadora:

1. Prof. Dr. Filidor Edilfonso Vilca Labra (IMECC-UNICAMP)
2. Prof. Dr. Gilberto Alvarenga Paula (IME-USP)
3. Prof. Dra. Hildete Prisco Pinheiro (IMECC-UNICAMP)

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

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IMECC DA UNICAMP
Bibliotecária: Crisllene Queiroz Custódio – CRB8 / 7966

Santana, Lucia Rolim
Sa59e Estimação e diagnóstico em modelos Birnbaum-Saunders skew-normal / Lucia Rolim Santana -- Campinas, [S.P. : s.n.], 2009.

Orientador : Filidor Edilfonso Vilca Labra
Dissertação (Mestrado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Análise de regressão - Modelos Birnbaum-Saunders. 2. Influencia local. 3. Algoritmo de Maximização de Expectativa. I. Vilca Labra, Filidor Edilfonso. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

(cqc/imecc)

Título em inglês: Estimation and diagnostic in Birnbaum-Saunders skew-normal models

Palavras-chave em inglês (Keywords): 1. Regression analysis - Birnbaum-Saunders models. 2. Local influence. 3. Expectation-maximization algorithms.

Área de concentração: Inferência

Titulação: Mestre em Estatística

Banca examinadora: Prof. Dr. Filidor Edilfonso Vilca Labra (IMECC-Unicamp)
Profa. Dra. Hildete Prisco Pinheiro (IMECC-Unicamp)
Prof. Dr. Gilberto Alvarenga Paula (IME-USP)

Data da defesa: 18/05/2009

Programa de Pós-Graduação: Mestrado em Estatística

Dissertação de Mestrado defendida em 18 de maio de 2009 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.

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

Prof(a). Dr(a). GILBERTO ALVARENGA PAULA

Hildete Pinheiro

Prof(a). Dr(a). HILDETE PRISCO PINHEIRO

Dedido este trabalho à minha mãe Maria de Lourdes Rolim Santana e em memória de meu pai Walter dos Santos Santana.

Agradecimentos

Agradeço primeiramente a Deus pois sem Ele nada teria acontecido.

Aos meus pais Walter dos Santos Santana (em memória) e Maria de Lourdes Rolim Santana, que sempre lutaram para dar uma boa educação e formação aos seus filhos.

Aos meus irmãos Adelaide, Wal, Dilce, Dilma, Carlos e Walter por respeitar minhas escolhas, pela amizade, amor, apoio e grandes incentivos em todos os momentos da minha vida.

A José Antônio Oliveira de Freitas, pela ajuda, companheirismo, carinho e total apoio que me fez alcançar este objetivo.

Aos meus amigos Alejandro Monzón e sua esposa Jenny, Cristiano Amâncio, Dianelis Torres, Fanny Roxana, Jonas Bodini, José Clelto Barros, Jussara Lobato, Lyvia Patrícia, Rosemeire Morokawa e Simoni Fernanda, pela convivência, apoio, companheirismo e principalmente pela grande amizade dedicada.

Ao meu orientador, Filidor Edilfonso Vilca Labra, pela dedicação, incentivo, paciência e amizade.

Ao professor Victor Hugo Lachos Dávila, professor Victor Leiva e Camila Borelli pela ajuda e principalmente pelas contribuições valorosas.

Aos professores do Departamento de Estatística da Universidade Estadual de Campinas, pelos ensinamentos concedidos.

Aos professores do Departamento de Estatística da Universidade Federal do Amazonas, pela recomendação e incentivo. Em especial ao professor Celso Rômulo Barbosa Cabral pela confiança, amizade e por ter colaborado com a minha formação acadêmica.

A Fundação de Amparo à Pesquisa do Estado do Amazonas pelo apoio financeiro, fundamental para o desenvolvimento da dissertação.

Aos participantes da banca examinadora, pelas sugestões.

Resumo

A classe de modelos Birnbaum-Saunders (BS) surgiu em problemas de fadiga dos materiais (que é um dano estrutural que ocorre quando um material é exposto a estresse e tensão). Nos últimos tempos, este modelo tem sido aplicado em áreas fora do contexto de fadiga dos materiais e engenharia, como por exemplo, em ciências da saúde, ambiental, florestal, demográficas, atuarial, financeira, entre outras. Tendo em vista que a distribuição BS tem a propriedade de descrever processos de degradação acumulativa.

Neste trabalho, apresentamos um estudo do modelo BS baseado na distribuição skew-normal. Como subproduto consideramos o modelo de regressão linear log-Birnbaum-Saunders (log-BS). Para obter as estimativas de máxima verossimilhança usamos o algoritmo EM. Além disso, apresentamos um estudo de análise de influência global e local, através da metodologia de Zhu e Lee (2001) para dados incompletos. Ilustramos a metodologia proposta com dados encontrados na literatura.

Abstract

The class of models Birnbaum-Saunders (BS) appeared in problems of fatigue of materials (which is a structural damage that occurs when a material is exposed to stress and tension). Recently, this model has been applied in areas outside the context of fatigue of materials and engineering, for example in health sciences, environmental, forestry, demographic, actuarial, financial, among others. As the BS distribution has the property to describe cumulative degradation processes.

In this work we present a study of the BS model based on Skew-normal distribution. As a byproduct consider the model of linear regression log-Birnbaum-Saunders (log-BS). To obtain estimates of maximum likelihood we use the EM algorithm. Furthermore, we present a study of the analysis of global and local influence, through the method of Zhu and Lee (2001) to incomplete data. Illustrate the proposed methodology with data found in literature.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

CAPÍTULO 1

Introdução

1.1 Introdução

Um dos grandes problemas na área industrial é a ruptura ou falha dos materiais. As falhas são ocasionadas por diversas causas e uma destas causas é dada por fadiga. A fadiga e a falha estrutural resultam quando o material é submetido a diferentes esforços dinâmicos repetidos. Um modelo sobre fadiga de materiais muito utilizado na área industrial é dado pela distribuicão de tempo de vida proposta por Birnbaum e Saunders (1969a), que motivados por problemas de vibração encontrados em aviões comerciais, usaram o conhecimento sobre um tipo particular de fadiga para derivar uma nova família de distribuições, a qual modela o tempo de vida de materiais e equipamentos sujeitos a cargas dinâmicas.

A seguir, descreveremos brevemente como o modelo Birnbaum-Saunders foi derivado. Considere um material que é sujeito a um padrão cíclico de tensão e força. Aqui, está implícito que, em qualquer momento, "carga" é uma função que representa o resultado da tensão lançada no material neste tempo. A falha do material ocorre devido ao desenvolvimento e ao crescimento de uma rachadura dominante dentro do material.

Um ciclo é definido como m oscilações e cada aplicação da i-ésima oscilação em um ciclo resulta em uma extensão aleatória da rachadura X_i. A distribuição desta variável aleatória depende somente da rachadura atual causada pela oscilação de carga neste ciclo. A extensão da rachadura devido ao j-ésimo ciclo é dada por

$$Y_j = \sum_{i=1}^{m} X_i,$$

em que Y_j é uma variável aleatória com média μ e variância σ^2, para todo $j = 1, 2, 3, \ldots$. A extensão total da rachadura após z ciclos é dada por

$$W_z = \sum_{j=1}^{z} Y_j,$$

com função de distribuição

$$H_z(\omega) = P(W_z \leq \omega), \quad \text{para} \quad z = 0, 1, 2, 3, \ldots.$$

Seja C o número de ciclos até que seja observada a falha, onde tal falha ocorre quando o comprimento da rachadura dominante excede um dado comprimento crítico ω. A função de distribuição da variável C é

$$P(C \leq z) = P\left(\sum_{j=1}^{z} Y_j > \omega\right) = 1 - H_z(\omega).$$

Assumindo que os Y_j's são variáveis aleatórias independentes e identicamente distribuídas, a função de distribuição de C pode ser expressa aproximada usando o Teorema Central do Limite, isto é,

$$\begin{aligned} P(C \leq z) &= 1 - P\left(\sum_{j=1}^{z} \frac{Y_j - \mu}{\sigma\sqrt{z}} \leq \frac{\omega - z\mu}{\sigma\sqrt{z}}\right) \\ &= 1 - P\left(\sum_{j=1}^{z} \frac{Y_j - \mu}{\sigma\sqrt{z}} \leq \frac{\omega}{\sigma\sqrt{z}} - \frac{\mu\sqrt{z}}{\sigma}\right) \\ &\cong 1 - \Phi\left(\frac{\omega}{\sigma\sqrt{z}} - \frac{\mu\sqrt{z}}{\sigma}\right) \\ &= \Phi\left(\frac{\mu\sqrt{z}}{\sigma} - \frac{\omega}{\sigma\sqrt{z}}\right), \end{aligned}$$

em que $\Phi(.)$ representa a função de distribuição acumulada normal padrão.

Segundo Birnbaum-Saunders (1969a), se z é substituído por uma variável real não negativa T, a variável aleatória T é a extensão contínua da variável aleatória discreta C. Então T representa o tempo total até que ocorra falha. Birnbaum e Saunders (1969a) mostraram que sob certas condições, a função de distribuição acumulada (fda) de T é dada por

$$F_T(t;\alpha,\beta) = P(T \leq t) = \Phi\left[\frac{1}{\alpha}\left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right)\right], \quad t > 0, \tag{1.1}$$

em que

$$\alpha = \frac{\sigma}{\sqrt{\omega\mu}} > 0 \quad \text{e} \quad \beta = \frac{\omega}{\mu} > 0.$$

Dizemos que T segue uma distribuição Birnbaum-Saunders, com parâmetros α e β, e denotamos esta distribuição por $T \sim \mathrm{BS}(\alpha,\beta)$, onde $\sim$ significa "tem distribuição".

Uma propriedade importante na construção desta distribuição é que a variável aleatória

$$Z = \alpha^{-1}\left(\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}}\right) \sim N(0,1). \tag{1.2}$$

Assim, a distribuição Birnbaum-Saunders, denotada por $T \sim \mathrm{BS}(\alpha,\beta)$, está estreitamente relacionada com a distribuição normal através da relação

$$T = \beta\left[\frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2}\right)^2 + 1}\right]^2, \tag{1.3}$$

em que $Z \sim \mathrm{N}(0,1)$.

Algumas das propriedades da distribuição acima podem ser encontradas em Birnbaum e Saunders (1969a). O parâmetro α é um parâmetro de forma, à medida que α tende a zero, a distribuição Birnbaum-Saunders tende para a distribuição normal de média β e variância τ, onde $\tau \to 0$ quando $\alpha \to 0$. O parâmetro β é um parâmetro de escala, isto é, $T/\beta \sim \mathrm{BS}(\alpha,1)$. Ainda, β é a mediana da distribuição, pois $F_T(\beta) = \Phi(0) = 1/2$.

Na Figura 1.1, apresentamos a função de distribuição acumulada, dada em (1.1), para alguns valores de α e considerando $\beta = 1$. Observamos que para diferentes escolhas dos parâmetros (α,β), as suas respectivas funções de distribuição são diferentes, isto é, a distribuição Birnbaum-Saunders é identificável, ou seja, se temos (α_1,β_1) e (α_2,β_2) com $\alpha_1 \neq \alpha_2$ ou $\beta_1 \neq \beta_2$, as probabilidades correspondentes a $F(.;\alpha_1,\beta_1)$ são diferentes das probabilidades correspondentes a $F(.;\alpha_2,\beta_2)$.

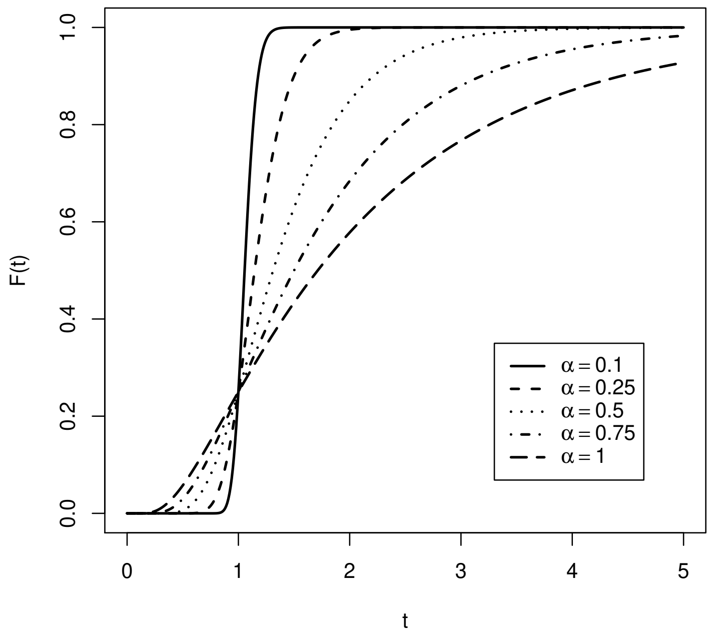


Figura 1.1: Função de distribuição Birnbaum-Saunders para diferentes valores de α e $\beta = 1.0$.

A distribuição Birnbaum-Saunders vem sendo amplamente usada na área de engenharia; consequentemente, há vários estudos referentes a esta distribuição. Birnbaum e Saunders (1969b) obtiveram originalmente os estimadores de máxima verossimilhança para α e β. Mann, Schafer e Singpurwalla (1974, p. 155) mostraram que a distribuição Birnbaum-Saunders é unimodal. Engelhardt, Bain e Wright (1981) propuseram intervalos de confiança para α considerando β como um parâmetro de perturbação desconhecido e, da mesma forma, para β considerando α como um parâmetro de perturbação desconhecido. Eles também apresentaram testes de hipóteses para estes parâmetros. Desmond (1985) derivou (1.1) tendo por base um modelo biológico. O autor também estendeu a justificação física para o uso desta distribuição, relaxando algumas suposições feitas em Birnbaum-Saunders (1969a). Desmond (1986) investigou a relação entre a distribuição Birnbaum-Saunders e a distribuição gaussiana inversa. Rieck e Nedelman (1991) formularam e desenvolveram um modelo log-linear para a distribuição BS(α, β) e mostraram que este modelo pode ser usado para testes de vida acelerada ou para comparar vidas medianas de algumas populações. Achcar (1993) desenvolveu procedimentos de estimação bayesiana para os parâmetros de (1.1) usando aproximações para as distribuições posterioris marginais de α e β. Lu e Chang (1997) utilizaram métodos bootstrap para construir intervalos de predição para as observações futuras de uma distribuição BS(α, β) e concluíram que os intervalos bootstrap são satisfatórios para tamanhos de amostras maiores que 30. Dupuis e Mills (1998) utilizaram métodos robustos de estimação para estimar os parâmetros da distribuição BS(α, β) na presença de outliers nos dados. Eles mostraram que o procedimento robusto é uma técnica alternativa poderosa quando os dados incluem observações distantes. Rieck (1999) derivou uma função geratriz de momentos para a distribuição Senh-Normal (seno hiperbólico normal) que pode ser usada para obter momentos de ordem inteira ou fracionária da distribuição BS(α, β).

Algumas generalizações e aplicações da distribuição Birnbaum-Saunders foram propostas recentemente. Os estudos foram considerados introduzindo novos parâmetros com o intuito de obter modelos mais robustos e mais flexíveis para serem utilizados em problemas reais. Por exemplo, Owen e Padgett (1999) desenvolveram a distribuição com três parâmetros. Barros, Paula e Leiva (2007) consideraram a distribuição t-Student $t_\nu(0, 1)$ em lugar da distribuição normal padrão, aqui o parâmetro ν permite controlar por exemplo, a curtose da distribuição e este novo modelo representa uma extensão robusta da distribuição proposta por Birnbaum e Saunders (1969a). Baseada nesta extensão, Leiva et al. (2007a), apresentaram um estudo de inferência e diagnóstico (baseado na metodologia de Cook, 1986) que representa uma extensão de alguns resultados desenvolvidos por Galea, Leiva e Paula (2004). Díaz–García e Leiva (2005) generalizaram a distribuição Birnbaum-Saunders baseado em uma classe de distribuições de contornos elípticos que inclui as distribuições como Cauchy, Laplace, logística, normal e t-Student com ν graus de liberdade. Recentemente, Barros et al. (2009) desenvolveram um pacote chamado *gbs* (incluído no *software* R) para a distribuição Binrbaum Saunders generalizada, o qual pode ser obtido em CRAN (http://CRAN.R-project.org/) na sua versão 1.0.

Os argumentos teóricos de que a BS é apropriada para descrever processos de degradação acumulativa têm transformado esta distribuição em um modelo que pode ser aplicado em outras áreas. Por exemplo, a distribuição BS sob a família de distribuições elíptica tem sido considerada em problemas ambientais, mais especificamente em concentrações de poluição atmosférica. Esta distribuição tem se mostrado mais adequada que a distribuição lognormal, como foi apontado por Leiva et al. (2008). Outra aplicação interessante nesta mesma área baseada no modelo BS sob a distribuição normal é em estudos da qualidade da água, veja por exemplo Leiva et al. (2009a).

Recentemente extensões da distribuição BS baseada na família de distribuições assimétricas tem sido considerado, especificamente sob a classe das distribuições skew-elíptica. Alguns resultados podem ser encontrados em Vilca-Labra e Leiva (2006), e Leiva et al. (2007b, 2008). Nesses trabalhos foram obtidos resultados teóricos, estendendo as propriedades das distribuições BS e log-BS.

A seguir, apresentaremos o algoritmo EM para a obtenção da estimativa de máxima verosimilhança (EMV) e em seguida mostraremos alguns critérios que serão utilizados para fazer comparações entre os modelos.

Algoritmo EM

O algoritmo EM (Dempster, Laird e Rubin, 1977) é um enfoque amplamente aplicável para o cálculo iterativo de estimativas de máxima verossimilhança, sendo bastante útil para problemas com dados incompletos. Muitos problemas em estatística podem ser encarados utilizando uma formulação de dados aumentados permitindo assim simplificar a obtenção das estimativas de máxima verossimilhança.

No algoritmo EM, cada iteração consiste de dois passos: Esperança (passo E) e Maximização (passo M). Sejam $\mathbf{Y}$ e $\mathbf{Z}$ o vetor de dados observados e o vetor de dados perdidos, respectivamente, desse modo o vetor de dados completos $\mathbf{Y}_c$ aumenta $\mathbf{Y}$ mediante $\mathbf{Z}$ como $\mathbf{Y_c} = (\mathbf{Y}^\top; \mathbf{Z}^\top)^\top$. Considere $\ell(\boldsymbol{\theta}, \mathbf{Y_c})$ o logaritmo da função de verossimilhança de dados completos para o vetor de parâmetros $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. O algoritmo EM aborda problemas com dados incompletos indiretamente mediante a substituição da parte não observável em $\ell(\boldsymbol{\theta}, \mathbf{Y}_c)$ por suas esperanças condicionais dado $\mathbf{Y}$, usando o ajuste atual para $\boldsymbol{\theta}$. Isto é, considera a função Q definida como

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \mathbb{E}[\ell(\boldsymbol{\theta}, \mathbf{Y}_c)|\mathbf{Y}, \widehat{\boldsymbol{\theta}})].$$

A (r+1)-ésima iteração do algoritmo EM é definida como:

Passo E: para $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(r)}$, calcular $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$;

Passo M: escolher $\boldsymbol{\theta}^{(r+1)}$, que maximize $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ tal que $Q(\boldsymbol{\theta}^{(r+1)}|\widehat{\boldsymbol{\theta}}) \geq Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$, para todo $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Deve-se alternar os passos E e M repetidamente até atingir a convergência. Cada iteração do algoritmo EM incrementa a função de verossimilhança de dados observados $\ell(\boldsymbol{\theta}, \mathbf{Y})$ e sob condições apropriadas o algoritmo EM apresenta convergência monótona ao máximo global ou local de $L(\boldsymbol{\theta}, \mathbf{Y})$, veja, por exemplo, Wu (1983) e McLachlan e Krishnan (1997).

Seleção de Modelos

Os critérios AIC, SIC e HQC serão utilizados para comparar as distribuições BSSN e BS. Tais critérios foram desenvolvidos sob bases teóricas distintas, mas têm o objetivo comum de comparar a adequação de modelos às observações. Nas aplicações têm sido observado que o AIC apresenta uma tendência de selecionar modelos com dimensão maior do que aqueles selecionados pelo SIC (veja, Kass e Raftery, 1995). Os critérios AIC, SIC e HQC são definidos como

$$\begin{aligned} \text{AIC} &= -\frac{\ell(\theta)}{n} + \frac{p}{n} \\ \text{SIC} &= -\frac{\ell(\theta)}{n} + \frac{p\log(n)}{2n} \\ \text{HQC} &= -\frac{\ell(\theta)}{n} + \frac{p\log(\log(n))}{n} \end{aligned}$$

onde $\ell(\theta)$ é a função *log*-verossimilhança associada ao modelo, n é o tamanho amostral e p é o número de parâmetros envolvidos.

Cada um destes critérios baseia-se numa penalização da verossimilhança na medida em que o modelo se torna mais complexo, isto é, modelos com um grande número de parâmetros. Desta forma, o modelo que apresente o menor valor do critério de informação (o modelo mais parcimonioso) será o modelo selecionado, para mais detalhes veja Akaike (1974), Schwarz (1978) e Spieglhaiter et al. (2002).

Fator de Bayes

Suponha que os dados D são provenientes dos modelos M_1 e M_2 representados pelas hipóteses H_1 (distribuição com o menor valor do Critério de Informação de Schwartz (SIC)) e H_2 (distribuicao a ser comparada com H_1), cujas densidades são correspondentes a $P(D|H_1)$ e $P(D|H_2)$. Dadas as probabilidades a priori $P(H_1)$ e $P(H_2) = 1 - P(H_1)$ temos que as probabilidades a posteriori são dadas por $P(H_1|D)$ e $P(H_2|D) = 1 - P(H_1|D)$. Um método que pode ser utilizado para fazer comparações entre os modelos é o Fator de Bayes (B_{12}) e pode ser definido, segundo Kass e Raftery (1995), como:

$$B_{12} = \frac{P(Y|H_1)}{P(Y|H_2)} = \frac{\int L(D|\theta_1, H_1)P(\theta_1|H_1)d\theta_1}{\int L(D|\theta_2, H_2)P(\theta_2|H_2)d\theta_2} \tag{1.4}$$

em que $L(D|\theta_1, H_1)$ e $L(D|\theta_2, H_2)$ são a função de verossimilhança dos modelos representados pelas hipóteses H_1 e H_2 respectivamente.

Uma interpretação do Fator de Bayes sugerido por Kass e Raftery (1995) é apresentada na Tabela 1.1.

Tabela 1.1: Interpretação do Fator de Bayes.

B_{12}	$2\log(B_{12})$	Evidência a favor de H_1
< 1	< 0	Negativa (apóia H_2)
$1 \vdash 3$	$0 \vdash 2$	Fraca
$3 \vdash 20$	$2 \vdash 6$	Positiva
$20 \vdash 150$	$6 \vdash 10$	Forte
> 150	> 10	Muito forte

De uma maneira geral, o Fator de Bayes é informativo, pois apresenta uma escala de valores que permite quantificar o grau de superioridade de um modelo em relação ao outro (veja Kass e Raftery, 1995).

Aproximação do Fator de Bayes mediante o SIC

É possível obter as distribuições a priori dadas na equação (1.4), usando

$$S = \log\left(P(D|\widehat{\theta_1}, H_1)\right) - \log\left(P(D|\widehat{\theta_2}, H_2)\right) - \frac{1}{2}(d_1 - d_2)\log(n), \tag{1.5}$$

em que $\widehat{\theta_k}$, $k = 1, 2$, são os estimadores de máxima verossimilhança de θ_k sob as hipóteses H_k, d_k é a dimensão de θ_k e n é o tamanho amostral. Assim, quando $n \to \infty$ a quantidade $n(\mathrm{SIC}_2 - \mathrm{SIC}_1)$ (SIC_1 e SIC_2 correspondem aos valores do critério de informação de Schwartz dos modelos dados em H_1 e H_2, respectivamente) satisfaz

$$\frac{S - \log(B_{12})}{\log(B_{12})} \to 0.$$

Segundo Raftery (1995) é mais conveniente reescrever a equação (1.5) como

$$\log(B_{12}) \approx 2\left(\log P(D|\widehat{\theta_1}, H_1) - \log P(D|\widehat{\theta_2}, H_2)\right) - (d_1 - d_2)\log(n). \tag{1.6}$$

Neste caso, se o modelo sob a hipótese H_1 for um caso particular do modelo sob a hipótese H_2 então

$$\log(B_{12}) \approx \chi^2_{12} - d_{f_{12}}\log(n),$$

onde a estatística do teste da razão de verossimilhança (χ^2_{12}) tem distribuição qui-quadrado com $d_{f_{12}} = d_1 - d_2$ graus de liberdade. Para mais detalhes, veja Raftery (1995).

1.2 Objetivos do trabalho

O objetivo deste trabalho é apresentar um estudo de inferência da distribuição Birnbaum-Saunders Skew-Normal e também apresentaremos um estudo de diagnóstico em modelos Senh-Skew-Sormal. O estudo será baseado em análises de influência local e global, através da metodologia de Zhu e Lee (2001) para dados incompletos. Será estudado o modelo de regressão log-Birnbaum-Saunders Skew-Normal. Será utilizado o pacote estatístico **R** em sua versão 2.8.1 para as programações dos processos de estimação e diagnóstico. Na parte de estimação, embora possam ser usadas técnicas de estimação direta, a obtenção das estimativas de máxima verossimilhança será via algoritmo EM proposto por Dempster, Laird e Rubin (1977). Para comparar os modelos Birnbaum-Saunders (BS) e Birnbaum-Saunders Skew-Normal (BSSN) será adotado um procedimento baseado na *teoria geral de seleção de modelos*, cuja idéia é estimar a dimensão de forma que uma dada função critério, envolvendo a qualidade do ajuste e a complexidade do modelo, seja otimizada. Três critérios serão utilizados: o *Critério de Informação de Akaike* (AIC), *Critério de Informação de Schwartz* (SIC) e o *Critério de Informação de Hannan-Queen* (HQC). Além disso, utilizaremos outro método, chamado Fator de Bayes, para fazer comparações entre os modelos.

1.3 Organização do trabalho

O Capítulo 2 descreve a distribuição Birnbaum-Saunders introduzida por Birnbaum e Saunders (1969a), destacando suas propriedades, momentos, função geradora de momentos e representação estocástica. Em seguida, é abordada a distribuição Senh-Normal e algumas generalizações da distribuição Birnbaum-Saunders.

O Capítulo 3 descreve a distribuição Birnbaum-Saunders Skew-Normal, destacando suas propriedades, momentos, representação estocástica e função de confiabilidade. Em seguida, é abordado um processo de estimação usando o algoritmo EM, o qual usa a função de verossimilhança completa da distribuição Birnbaum-Saunders Skew-Normal e logo após é mostrada a matriz de informação de Fisher observada. Um estudo de simulação é realizado para demonstrar a robustez do algoritmo EM no sentido de estimação dos parâmetros da distribuição Binrbaum-Saunders Skew-Normal. Por último, uma aplicação do modelo é realizada, considerando o conjunto de dados de poluição atmosférica, previamente analisado por Leiva et al. (2008).

O Capítulo 4 apresenta a distribuição Senh-Skew-Normal, que permite aplicações no modelo de regressão linear log-Birnbaum-Saunders Skew-Normal. Propriedades da distribuição Senh-Skew-Normal são estudadas, como representação estocástica e função geradora de momentos. Em seguida, é abordado o modelo de regressão linear log-Birnbaum-Saunders Skew-Normal. Distribuições condicionais para o algoritmo EM são desenvolvidas para a estimação dos parâmetros do modelo. Posteriormente, é calculada a matriz de informação de Fisher observada. Um estudo de convergência do algoritmo EM é desenvolvido, determinando a eficácia dos estimadores calculados neste trabalho. Em seguida, é feito um estudo de diagnóstico via influência global e local, através da metodologia desenvolvida por Zhu e Lee (2001), utilizando a função de verossimilhança completa. A matriz hessiana é calculada e analisadas as perturbações de ponderação de casos, perturbação nas variáveis resposta e explicativa. Um estudo de simulção utilizando o modelo log-Birnbaum-Saunders Skew-Normal é conduzido, sob perturbação de algum ponto para verificar a capacidade do modelo de detectar pontos influentes. Ainda, uma aplicação utilizando os dados de fadiga é desenvolvida, ajustando o modelo e aplicando a técnica de diagnóstico.

CAPÍTULO 2

A Distribuição Birnbaum-Saunders

A distribuição Binrbaum-Saunders foi inicialmente proposta por Birnbaum e Saunders (1969a), que motivados por problemas de vibração encontrados nos aviões comerciais, propuseram uma nova família de distribuições para modelar o tempo de vida de materiais expostos a provas de fadigas. A distribuição Birnbaum-Saunders assumiu um importante papel na modelagem do tempo de falha em processos de fadiga. Também vem sendo largamente utilizada como um meio eficaz de modelar falhas de desgaste estocástico em geral. Esta distribuição foi obtida a partir de um modelo que mostra que as falhas acontecem devido ao desenvolvimento e ao crescimento de uma rachadura dominante.

Função Densidade de Probabilidade

Considerando a função distribuição acumulada da variável aletória T dada em (1.1), a sua correspondente função densidade de probabilidade (fdp) é dada por

$$f_T(t;\alpha,\beta) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2\alpha^2}\left(\frac{t}{\beta}+\frac{\beta}{t}-2\right)\right\}\frac{t^{-3/2}(t+\beta)}{2\alpha\sqrt{\beta}}, \quad t>0, \tag{2.1}$$

em que $\alpha, \beta > 0$. Agora, fazendo

$$a_t(\alpha,\beta) = \frac{1}{\alpha}\left[\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right] \quad \text{e} \quad A_t(\alpha,\beta) \;=\; \frac{d}{dt}a_t(\alpha,\beta) = \frac{t^{-3/2}(t+\beta)}{2\alpha\beta^{1/2}}, \tag{2.2}$$

a função densidade de probabilidade $f_T(t)$ pode ser escrita como

$$f_T(t) = \phi(a_t(\alpha, \beta))A_t(\alpha, \beta),$$

em que $\phi(.)$ é a função densidade da distribuição $N(0, 1)$. Além disso, de (1.1) vamos ter que a função de distribuição acumulada $F_T(t)$ pode ser escrita como

$$F_T(t) = \Phi(a_t(\alpha, \beta)), \ \ t > 0.$$

Observação 2.1. *Pode-se mostrar facilmente que se $T \sim BS(\alpha, \beta)$, então*

i) $aT \sim BS(\alpha, a\beta)$, para toda constante positiva a ;

ii) $T^{-1} \sim BS(\alpha, \beta^{-1})$.

Para mais detalhes, veja Saunders (1974).

Agora suponhamos que $T \sim BS(\alpha, \beta)$, então de (1.2) temos que a variável aleatória Z definida por

$$Z = \frac{1}{\alpha}\left[\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right] \tag{2.3}$$

tem distribuição $N(0, 1)$ ou que $X = \dfrac{\alpha}{2}Z \sim N\left(0, \dfrac{\alpha^2}{4}\right)$. Assim, os números pseudo-aleatórios provenientes de uma distribuição Birnbaum-Saunders podem ser gerados a partir de uma distribuição normal através da relação

$$\begin{aligned} T &= \beta\left\{1 + 2X^2 + 2X(1 + X^2)^{1/2}\right\} = \beta\left[X + (1 + X^2)^{1/2}\right]^2 \\ &= \beta\left[\frac{\alpha}{2}Z + (1 + \frac{\alpha}{2}Z)^{1/2}\right]. \end{aligned}$$

A média, variância, coeficientes de assimetria e curtose da distribuição BS dada em (2.1) são importantes para determinar seu comportamento sob diferentes valores dos parâmetros. Para determinar estas quantidades é necessário obter os momentos de T, que podem ser obtidos usando a expressão (2.3), o Teorema Binomial e $Z^2 \sim \chi^2_1$. Assim, desde que βT^{-1} e $\beta^{-1}T$ tenham a mesma distribuição, o valor esperado de $(T/\beta)^r$, proposto por Johnson et al. (1995) é dado por

$$\mathbb{E}\left[\left(\frac{T}{\beta}\right)^r\right] = \sum_{j=0}^{r}\begin{pmatrix} 2r \\ 2j \end{pmatrix}\sum_{i=0}^{j}\begin{pmatrix} j \\ i \end{pmatrix}\frac{[2(r-j+i)]}{2^{r-j+i}(r-j+i)}\left(\frac{1}{2}\alpha\right)^{2(r-j+i)}.$$

Consequentemente, o valor esperado, a variância e os coeficientes de assimetria (μ_3) e curtose (μ_4) são dados, respectivamente, por

$$\mathbb{E}(T) = \beta\left(1 + \frac{\alpha^2}{2}\right), \quad \mathrm{Var}(T) = (\alpha\beta)^2\left(1 + \frac{5}{4}\alpha^2\right),$$

$$\mu_3 = \frac{16\alpha^2\,(11\alpha^2 + 6)}{(5\alpha^2 + 4)^3} \quad \text{e} \quad \mu_4 = 3 + \frac{6\alpha^2(93\alpha^2 + 41)}{(5 + \alpha^2 + 4)}.$$

As quantidades μ_3 e μ_4 medem a assimetria e o peso das caudas da distribuição, respectivamente. A densidade dada em (2.1) é caracteristicamente assimétrica à direita. Entretanto, à medida que o valor de α decresce, particularmente para valores menores que um, a densidade se torna mais simétrica. Na Figura 2.1 apresentamos o gráfico da densidade (2.1) para alguns valores de α considerando $\beta = 1$. Observe que à medida que o valor de α decresce, a curva tende a ficar mais simétrica em torno de β, que é a mediana da distribuição. Da mesma forma, note que a variância também decresce com α.

A função de sobrevivência e a função risco da distribuição Birnbaum-Saunders são expressas, respectivamente, como

$$S_T(t) = 1 - F_T(t) = 1 - \Phi\left[\frac{1}{\alpha}\left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right)\right] \tag{2.4}$$

e

$$h_T(t) = \frac{f_T(t)}{S_T(t)}.$$

A função risco $h_T(t)$ da distribuição Birnbaum-Saunders é zero em $t = 0$, cresce até um máximo para algum t_0 e finalmente decresce até um valor finito. De fato, o comportamento assintótico de $h_T(t)$ é dado por

$$\lim_{t\to\infty} h_T(t) = \frac{1}{2\alpha^2\beta}.$$

Na Figura 2.2 apresentamos o gráfico de $h(t)$ para alguns valores de α fixando $\beta = 1.0$.

Na Figura 2.3 apresentamos o gráfico da densidade (2.1) para alguns valores de β, considerando os mesmos valores de α dados na Figura 2.1. Note que o comportamento das

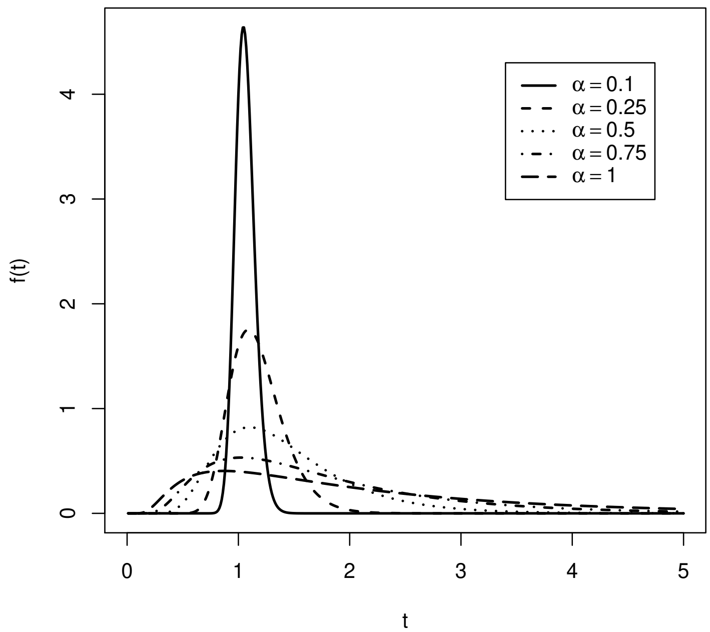


Figura 2.1: Densidade Birnbaum-Saunders para diferentes valores de α e $\beta = 1.0$.

curvas nos quatro gráficos são os mesmos da Figura 2.1, havendo apenas uma mudança na escala dos dados. Isto evidencia o fato de que β é um parâmetro de escala.

Considere as variáveis aleatórias T_1, $T_2 \sim \mathrm{BS}(\alpha, \beta)$ independentes e as variáveis aleatórias $T_{(1)}$ e $T_{(2)}$ definidas, respectivamente por

$$T_{(1)} = \min(T_1, T_2) \text{ e } T_{(2)} = \max(T_1, T_2).$$

Então suas respectivas fdp são dadas por

$$f_{T_{(1)}}(t) = 2\phi(a_t(\alpha,\beta))\Phi(-a_t(\alpha,\beta))A_t(\alpha,\beta), \quad t > 0 \tag{2.5}$$

e

$$f_{T_{(2)}}(t) = 2\phi(a_t(\alpha,\beta))\Phi(a_t(\alpha,\beta))A_t(\alpha,\beta), \quad t > 0, \tag{2.6}$$

que têm uma expressão similar à fdp de uma variável aleatória Skew-Normal. Na Figura 2.4 apresentamos os gráficos das densidades (2.5) e (2.6) comparado com a densidade da

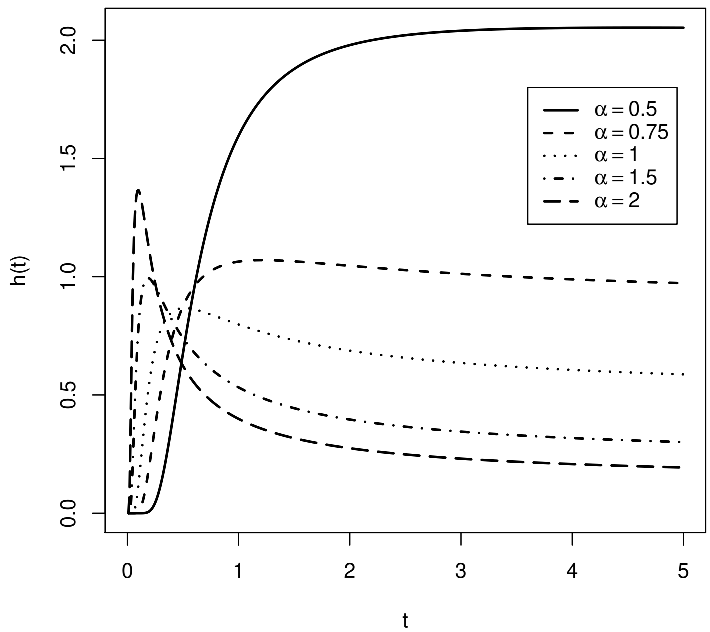


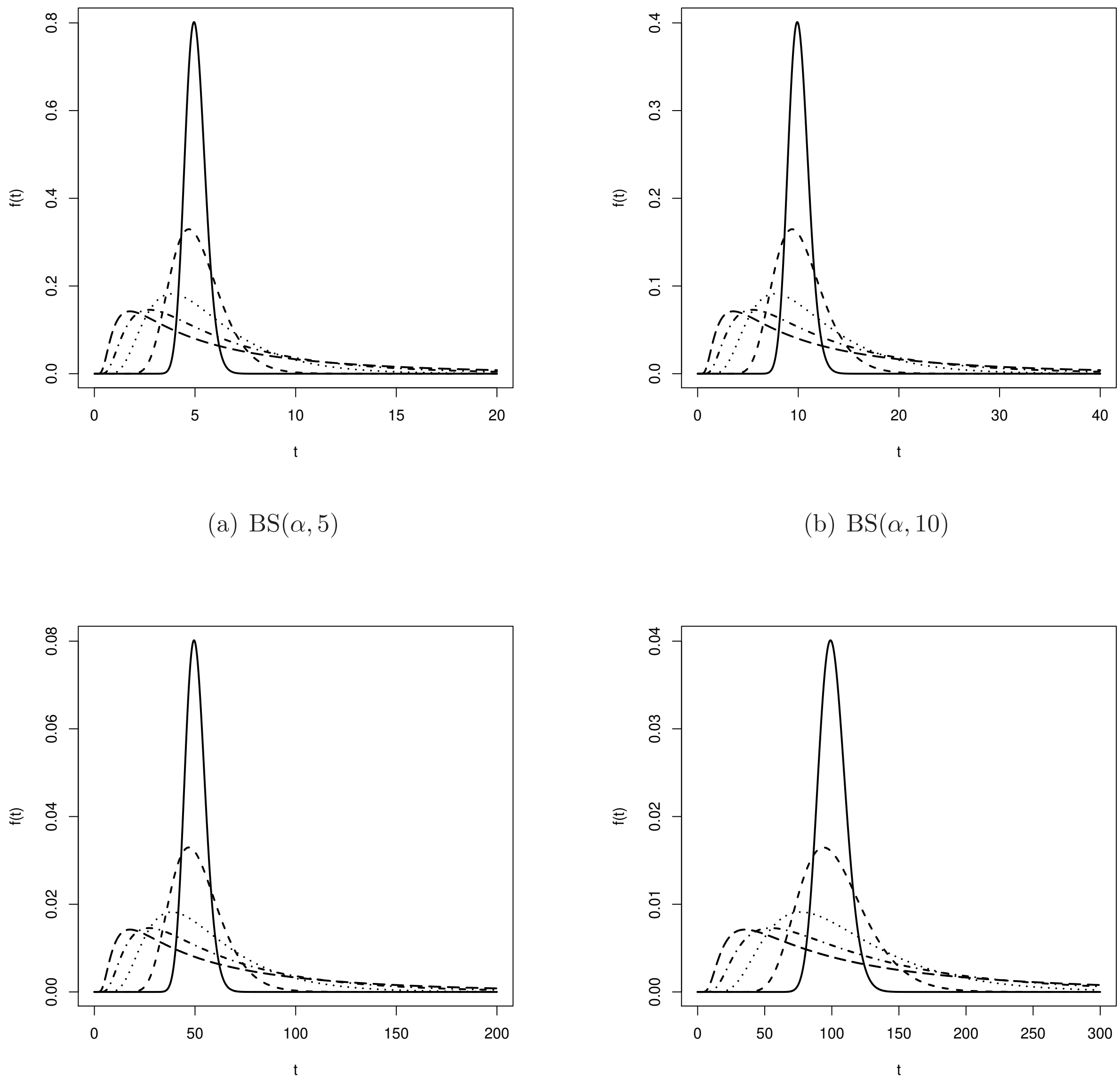
Figura 2.2: Função risco da Birnbaum-Saunders para diferentes valores de α e $\beta = 1.0$.

distribuição Birnbaum-Saunders, definida em (2.1), considerando $\alpha = 0.5$ e $\beta = 0.8$. Note que a estatística $T_{(1)} = \min(T_1, T_2)$ ajusta melhor os percentis baixos da distribuição do que a distribuição Birnbaum-Saunders, enquanto que a estatítica $T_{(2)} = \max(T_1, T_2)$ ajusta melhor os percentis altos da distribuição.

2.1 Distribuição Senh-Normal

Rieck e Nedelman (1991) desenvolveram a distribuição *seno hiperbólico normal* ou simplesmente Senh-Normal (SHN), que foi obtida através do log da distribuição BS(α, β), e mostraram que esta distribuição pode ser usada em modelos para testes de vida acelerada ou para comparar vidas medianas de algumas populações.

Alternativamente a distribuição SHN pode ser definida através de uma transfomação da



(a) BS($\alpha, 5$)

(b) BS($\alpha, 10$)

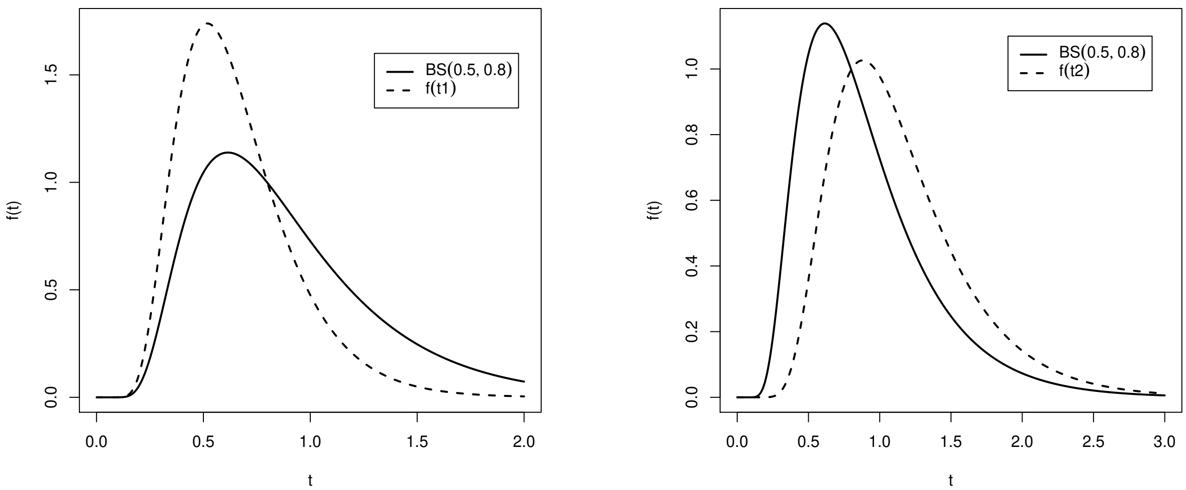
(c) BS($\alpha, 50$)

(d) BS($\alpha, 100$)

Figura 2.3: Densidades da Birnbaum-Saunders para diferentes valores de β considerando os mesmos valores de α dados na Figura 2.1.

distribuição normal baseda na relação

$$Y = \gamma + \sigma \text{arcsenh}\left(\frac{\alpha Z}{2}\right), \tag{2.7}$$



(a) BS$(0.5, 0.8)$ e $f_{T_{(1)}}(t)$ (b) BS$(0.5, 0.8)$ e $f_{T_{(2)}}(t)$

Figura 2.4: Densidades da Birnbaum-Saunders e das estatísticas $T_{(1)} = \min(T_1, T_2)$ e $T_{(2)} = \max(T_1, T_2)$ considerando $\alpha = 0.5$ e $\beta = 0.8$.

em que $Z \sim \mathrm{N}(0, 1)$. Assim a função densidade de probabilidade é dada por

$$\begin{aligned} f_Y(y) &= \left(\frac{2}{\alpha\sigma\sqrt{2\pi}}\right) \exp\left[-2\alpha^{-2}\mathrm{senh}^{\,2}\left(\frac{y-\gamma}{\sigma}\right)\right] \cosh\left(\frac{y-\gamma}{\sigma}\right), \\ &= \phi(b_y)b_y^{'}, \quad y \in \mathbb{R}. \end{aligned} \tag{2.8}$$

em que $\phi(.)$ denota a fdp da distribuição normal padrão. Além disso,

$$b_y = \frac{2}{\alpha}\mathrm{senh}\left(\frac{y-\gamma}{\sigma}\right) \quad \text{e} \quad b_y^{'} = \frac{db_y}{dy} = \frac{2}{\alpha\sigma} cosh\left(\frac{y-\gamma}{\sigma}\right). \tag{2.9}$$

A distribuição de Y é denotada por $Y \sim SHN(\alpha, \gamma, \sigma)$. Algumas propriedades desta distribuição que serão de utilidade no decorrer deste trabalho são, por exemplo:

(A1) $Z = \dfrac{2}{\alpha}\mathrm{senh}\left(\dfrac{Y-\gamma}{\sigma}\right) \sim \mathrm{N}(0, 1)$;

(A2) $V = a + b\,Y \sim \mathrm{SHN}(\alpha, a + b\,\gamma, b\,\sigma)$, em que $a \in \mathbb{R}$ e $b \in \mathbb{R} - \{0\}$. Em particular, se $Y \sim \mathrm{SHN}(\alpha, \gamma, \sigma)$, então $[Y - \gamma]/\sigma \sim \mathrm{SHN}(\alpha, 0, 1)$. Isto é, com esta transformação obtemos a distribuição SHN padrão.

Analogamente à distribuição BS, a função de distribuição acumulada de Y é dada por

$$F_Y(y) = \Phi\left[\frac{2}{\alpha}\text{senh}\left(\frac{y-\gamma}{\sigma}\right)\right], \quad y \in \mathbb{R} \tag{2.10}$$

em que $\Phi(.)$ denota a fda da distribuição normal padrão. Agora, se $T \sim \text{BS}(\alpha, \beta)$ então a distribuição de $Y = \log(T)$ permite o estudo de modelos de regressão linear log-Birnbaum-Saunders, e neste caso $Y \sim SHN(\alpha, \gamma, \sigma = 2)$. Rieck (1989) mostrou que a distribuição Senh-Normal é simétrica em torno de λ, unimodal para $\alpha \leq 2$ e bimodal para $\alpha > 2$. Além disso, se $Y \sim SN(\alpha, \gamma, \sigma = 2)$, então $Z = Y - \gamma$ converge em distribuição para a distribuição normal padrão quando $\alpha \to 0$.

A média e a variância da distribuição $SHN(\alpha, \gamma, \sigma = 2)$ podem ser obtidas usando a função geradora de momentos dada por

$$m(s) = \exp(\mu s)\left[\frac{K_a(\delta^{-2}) + K_b(\delta^{-2})}{2K_{1/2}(\delta^{-2})}\right], \tag{2.11}$$

onde $a = (\sigma s + 1)/2$, $b = (\sigma s - 1)/2$ e $K(.)$ é a função de Bessel modificada do terceiro tipo, dada por $K_\lambda(\omega) = (1/2)(\omega/2)^\lambda \int_0^\infty y^{-\lambda-1} e^{-y-(\omega^2/4y)} dy$, (veja, Gradshteyn e Randzhik (2000), p. 907).

Rieck e Nedelman (1991) provaram que se $T \sim \text{BS}(\alpha, \beta)$, então $Y = \log(T) \sim SN(\alpha, \gamma, \sigma = 2)$, onde $\gamma = \log(\beta)$. Por isso, a distribuição Senh-Normal é denominada distribuição log-Birnbaum-Saunders. Na equação acima $\gamma = \log \beta$ e $\sigma = 2$. Neste caso, dizemos que Y tem distribuição Senh-Normal. Esta distribuição será denotada por $SHN(\alpha, \gamma, \sigma = 2)$.

2.2 Algumas generalizações da distribuição BS

A suposição de normalidade no estudo da distribuição BS tem permitido aplicações com sucesso em muitas situações, no entanto esta suposição não é adequada especialmente quando desejamos modelar os percentis mais baixos ou mais altos da distribuição. Para modelar os percentis baixos, podemos usar no lugar da distribuição Normal uma distribuição assimétrica com parâmetro de assimetria negativa. Recentemente, Leiva et al. (2008) apresentaram um estudo teórico do modelo log-Birnbaum-Saunders, eles assumiram no lugar da distribuição normal uma distribuição Skew-elíptica. Embora estas generalizações representem uma contribuição positiva no aspecto teórico, as aplicações não tem sido estudadas com muito detalhes, como por exemplo, a metodologia de estimação e de análise de dados.

As extensões da distribuição BS podem ser baseada na distribuição de Z dada na equação (1.2), por exemplo, Díaz–García e Leiva (2005), apresentaram um estudo teórico da distribuição BS sob a classe das distribuições elípticas proporcionando distribuições de vida que cresçam mais rapidamente e que possuam caudas com menor ou maior curtose do que a distribuição BS. Baseado em um procedimento similar, Vilca-Labra e Leiva (2006) obtiveram uma maior generalização ao desenvolver a distribuição BS sob a classe das distribuições Skew-elíptica que tem a propriedade de serem mais flexíveis na curtose e assimetria. A generalização do modelo BS sob a classe skew-elíptica, proposto por Leiva et al. (2008), envolve um número maior de parâmetros. Sob normalidade, esta nova distribuição é definida por

$$f_T(t) = \phi(\nu + a_t)A_t; \quad t > 0,\ \nu \in \mathbb{R}, \tag{2.12}$$

em que a_t e A_t são como em (2.2). Esta distribuição é denotada por $T \sim \text{BSE}(\alpha, \beta, \sigma = 2, \nu)$. Quando $\nu = 0$, vamos ter que $\text{BSE}(\alpha, \beta, \sigma = 2, \nu) = \text{BS}(\alpha, \beta, \sigma = 2) = \text{BSE}(\alpha, \beta)$.

Distribuição Simétrica

A classe que inclui distribuições simétricas com caudas mais ou menos pesadas que a distribuição normal é denominada de distribuições de contornos elípticos ou simplesmente elípticas. As distribuições elípticas correspondem a todas as distribuições simétricas em $\mathbb{R}$. Dizemos que uma variável aleatória X segue uma distribuição elíptica se sua função de densidade de probabilidade é dada por

$$f_Z(z) = \frac{1}{\sigma} f\left(\frac{(z-\mu)^2}{\sigma^2}\right), \quad z \in \mathbb{R}, \tag{2.13}$$

em que $g(u)$, $u > 0$ é o núcleo da função densidade de probabilidade de Z. A distribuição de Z é denotada por Z $\sim EC(\mu, \sigma^2; f)$.

A generalização da distribuição Birnbaum-Saunders clássica, proposta por Díaz-García e Leiva (2005), foi obtida a partir de distribuições de contornos elípticos, a qual denominaram distribuição Birnbaum-Saunders generalizada. Este modelo foi obtido a partir de (1.2), assumindo que Z $\sim EC(0,1;f)$. Neste caso, T segue uma distribuição Birnbaum-Saunders generalizada, que é denotada por $T \sim BSG(\alpha,\beta;f)$, e sua função de densidade tem a seguinte forma:

$$f_T(t) = f_Z(a_t(\alpha,\beta))A_t(\alpha,\beta), \tag{2.14}$$

em que f_Z é dada em (2.13), $a_t(\alpha,\beta)$ e $A_t(\alpha,\beta)$ são dadas em (2.2).

CAPÍTULO 3

A Distribuição Birnbaum-Saunders Skew-Normal

Neste capítulo apresentamos um novo modelo para tempos de falhas, definido em termos da distribuição skew-normal padrão (SN) como uma proposta alternativa ao modelo clássico proposto por Birnbaum-Saunders (1969a), destacando suas propriedades e procedimento de estimação de máxima verossimilhança baseado no algorimo EM. Além disso, apresentamos a relação entre a BS baseado na distribuição normal e a BS baseado na Skew-Normal. Inicialmente faremos uma breve descrição sobre a distribuição assimétrica skew-normal univariada.

Distribuição Assimétrica

Embora a suposição de assumir que Z tenha distribuição elíptica represente uma boa alternativa ao modelo normal, existem situações práticas em que a variável resposta é assimétrica (veja, por exemplo, Hill e Dixon, 1982). Nestes casos, uma nova classe de distribuições denominada normal assimétrica proposta por Azzalini (1985), pode ser postulada para estes tipos de situações. A função de densidade de probabilidade da normal assimétrica possui um parâmetro que permite controlar a assimetria e, consequentemente, se ajusta melhor a dados de natureza assimétrica.

Quando a distribuição dos dados tem um comportamento assimétrico esta pode ser modelada através da classe das distribuições chamadas de *distribuições assimétricas*. Dentre esta classe, a distribuição skew-normal univariada é talvez a pioneira nesta idéia, inicialmente introduzida por O'Hagan e Leonard (1976) como uma distribuição a priori em análise bayesiana e posteriormente formalmente estudada por Azzalini (1985) (do ponto de vista clássico) como uma extensão natural da distribuição normal para modelar a estrutura assimética presente nos dados. Segundo Azallini (1985), uma variável aleatória Z tem distribuição *skew-normal* com parâmetro de locação μ, parâmetro de escala σ, com $\sigma > 0$, e parâmetro de assimetria λ se sua função densidade de probabilidade (fdp) é dada por

$$f(z) = 2\phi(z|\mu, \sigma^2)\Phi\left(\frac{\lambda}{\sigma}(z-\mu)\right), \quad z \in \mathbb{R}, \tag{3.1}$$

em que $\phi(.|\mu, \sigma^2)$ é a fdp de uma distribuição $N(\mu, \sigma^2)$ e $\Phi(.)$ é função de distribuição acumulada de uma distribuição normal padrão. O parâmetro $\lambda \in \mathbb{R}$ regula a assimetria da função densidade de Z. Em particular, se $\lambda = 0$, $f(z)$ é a função densidade da distribuição $N(\mu, \sigma^2)$. A distribuição em (3.1) é denotada por $Z \sim SN(\mu, \sigma^2, \lambda)$.

A representação estocástica de Z é dada por

$$Z \stackrel{d}{=} \mu + \sigma[\delta|X_0| + (1-\delta^2)^{1/2}X_1], \quad \text{com} \quad \delta = \frac{\lambda}{\sqrt{1+\lambda^2}}, \tag{3.2}$$

em que $X_0 \sim N_1(0,1)$ e $X_1 \sim \mathrm{N}(0,1)$ são independentes. A representação estocástica acima é importante, para a obtenção da estimativa de máxima verossimilhança (EMV) baseada no algoritmo EM, proposto por Dempster, Laird e Rubin (1977). Agora fazendo, $H = |X_0|$, temos que $H \sim HN(0,1)$ e consequentemente vamos ter que

$$Z|H = h \sim N(\mu + \sigma\delta h, \sigma^2(1-\delta^2)). \tag{3.3}$$

Além disso, de (3.2) segue que a média e a variância de Z são dada, respectivamente, por

$$E[Y] = \mu + \sqrt{\frac{2}{\pi}}\sigma\delta \quad \text{e} \quad Var[Y] = \sigma^2 - \frac{2}{\pi}\sigma^2\delta^2. \tag{3.4}$$

3.1 Definição e propriedades

Seja T o tempo total até que ocorra falha, então em lugar da suposição de normalidade definida em (1.2), aqui vamos assumir que

$$Z = \frac{1}{\alpha}\left(\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}}\right) \sim \mathrm{SN}(0, 1, \lambda). \tag{3.5}$$

Isto é, Z tem distribuição skew normal padrão com parâmetro de assimtetria λ. Assim a distribuição da variável aleatória T é definida em termos de Z

$$T = \beta\left[\frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2}\right)^2 + 1}\right]^2, \tag{3.6}$$

e consequentemente dependerá da distribuição da $SN(0, 1, \lambda)$. Dizemos que a variável aleatória T baseada na distribuição skew-normal segue uma distribuição Birnbaum-Saunders Skew-Normal (BSSN), esta distribuição será denotada por $T \sim \mathrm{BSSN}(\alpha, \beta, \lambda)$. Esta distribuição representa uma generalização da distribuição Birnbaum-Saunders clássica, pois a distribuição de T para $\lambda = 0$ coincide com a distribuição Birnbaum-Saunders, dada em (2.1). Assim, formalmente vamos ter a seguinte definição

Definição 3.1. *Uma variável aleatória T segue uma distribuição na classe $BSSN$ com parâmetro de forma $\alpha > 0$, parâmetro de escala $\beta > 0$ e parâmetro de assimetria $\lambda \in \mathbb{R}$, se sua fdp for dada por*

$$f_T(t) = 2\phi(a_t(\alpha, \beta))A_t(\alpha, \beta)\Phi(\lambda a_t(\alpha, \beta)) \tag{3.7}$$

em que $\phi(.)$ e $\Phi(.)$ são respectivamente a densidade e a função distribuição acumulada da normal padrão, $a_t(\alpha, \beta)$ e $A_t(\alpha, \beta)$ são dadas em (2.2).

Analogamente à distribuição Birnbaum-Saunders, os parâmetros α e β são, respectivamente, os parâmetros de forma e escala e λ é o parâmetro de assimetria. À medida que α e λ tendem a zero, a distribuição skew-normal Birnbaum-Sauders tende para a distribuição normal de média β e variância τ, onde $\tau \to 0$ quando $\alpha \to 0$.

Na Figura 3.1 apresentamos a densidade da distribuição Birnbaum-Saunders, denotada por $\mathrm{BS}(0.75, 1)$, junto com as densidades da Birnaum-Saunders skew-normal $\mathrm{BSSN}(0.75, 1, \lambda)$, considerando $\lambda = 1, 2$ e 3. Os valores das densidades são transformados de forma que tenham os mesmos valores na origem, para efeito de comparação das curvas. Note que as densidades são caracteristicamente assimétricas à direita e que as densidades da BSSN$(0.75, 1, \lambda)$, considerando $\lambda = 1$, 2 e 3 tem as caudas muito mais pesadas do que a distribuição Birnbaum-Saunders.

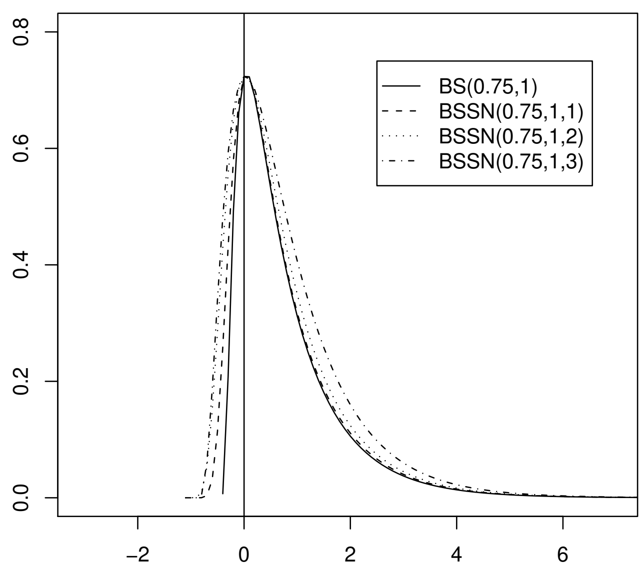


Figura 3.1: Funções de densidade das distribuições BS$(0.75, 1)$ e BSSN$(0.75, 1, \lambda)$, considerando $\lambda = 1$, 2 e 3.

Na Figura 3.2 apresentamos os gráficos da densidade (3.7) para alguns valores de λ positivos e negativos, considerando $\alpha = 0.2$ (Figuras (a) e (b)), $\alpha = 0.8$ (Figuras (c) e (d)) e $\beta = 1.0$. Observe que, quando $\lambda > 0$ as caudas da distribuição Birnbaum-Saunders Skew-Normal são mais pesadas em relação a distribuição Birnbaum-Saunders, obtida quando $\lambda = 0$. Quando $\lambda < 0$, a distribuição Birnbaum-Saunders Skew-Normal (BSSN) modela melhor os percentis baixos, sendo mais flexível na curtose e assimetria do que a distribuição

Birnbaum-Saunders.

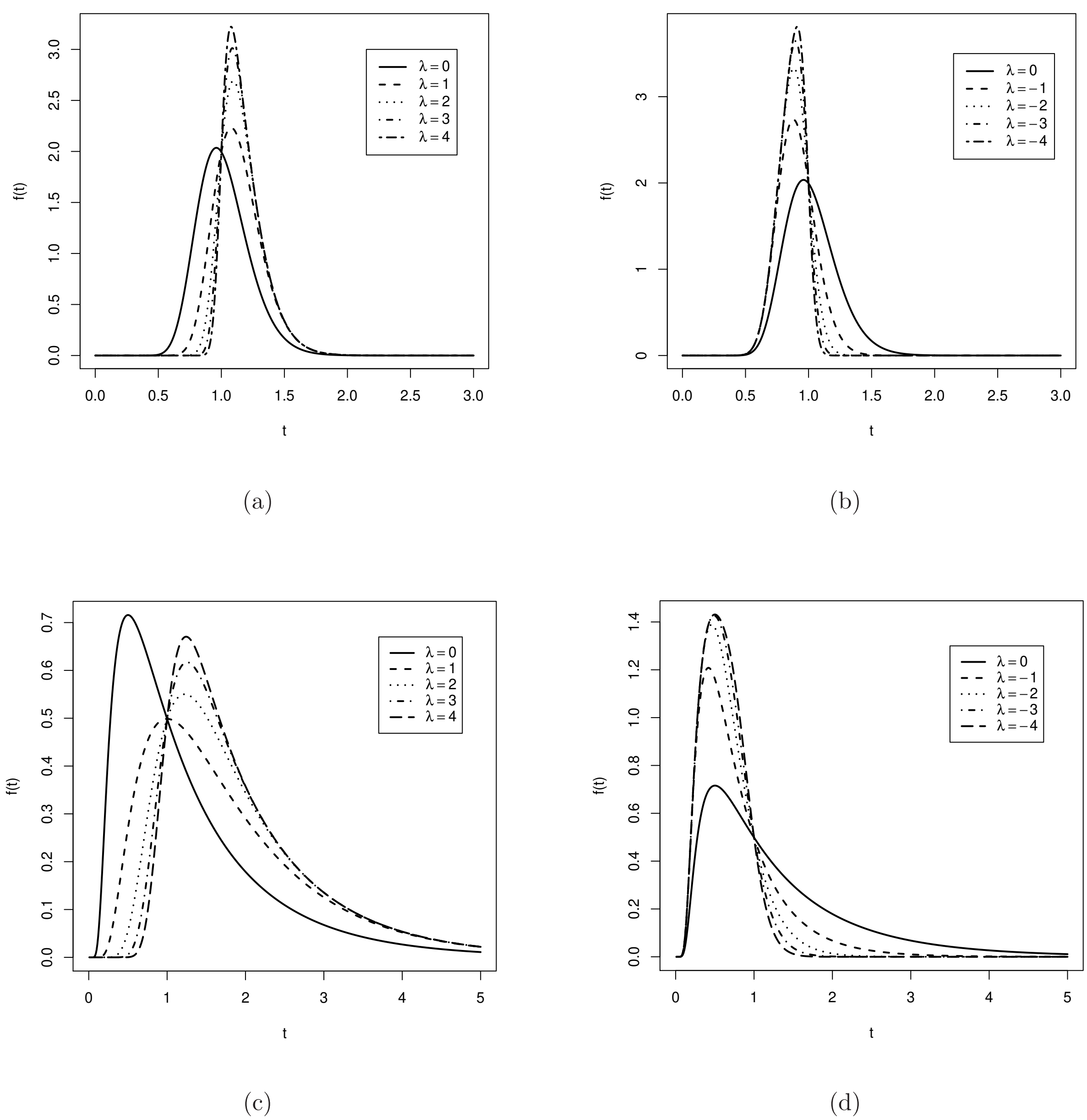
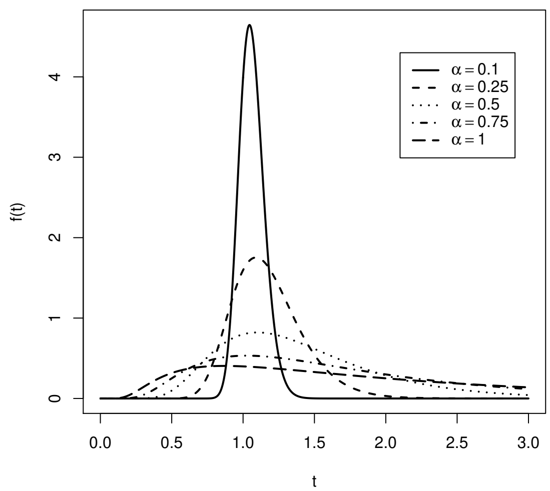
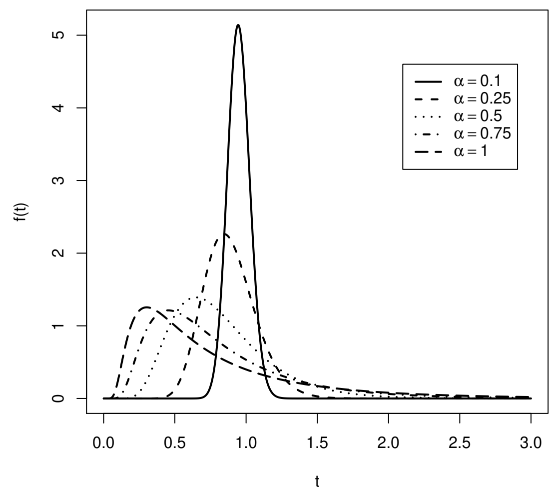


Figura 3.2: Densidades da BSSN para diferentes valores de λ, considerando $\beta = 1.0$, (a)-(b) $\alpha = 0.2$ e (c)-(d) $\alpha = 0.8$.

Na Figura 3.3 apresentamos os gráficos da densidade (3.7) para alguns valores de α considerando $\beta = 1.0$, $\lambda = 1.0$ e $\lambda = -1.0$. Observe que, à medida que o valor de α decresce, a curva da densidade (3.7) tende a ficar mais simétrica em torno de β, que é a mediana da distribuição e a variância também decresce com α da mesma forma que a distribuição Birnbaum-Saunders (BS).



(a) $\lambda = 1.0$



(b) $\lambda = -1.0$

Figura 3.3: Densidades BSSN para diferentes valores de α, considerando $\beta = 1.0$, (a) $\lambda = 1.0$ e (b) $\lambda = -1.0$.

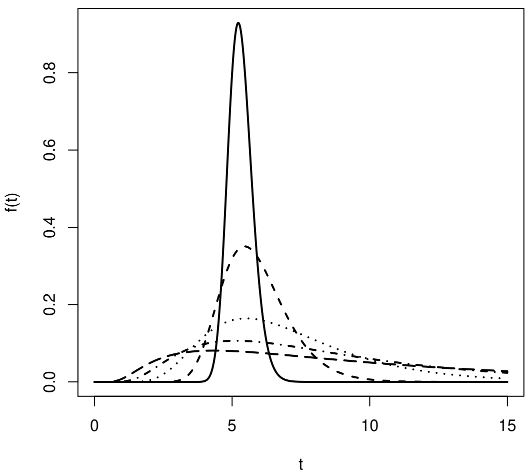
Na Figura 3.4 apresentamos o gráfico da densidade (3.7) para alguns valores de β, considerando os mesmos valores de α dados na Figura 3.3. Note que o comportamento das curvas nos quatro gráficos são as mesmas que na Figura 3.3, havendo apenas uma mudança na escala dos dados. Isto evidencia o fato de que β também é um parâmetro de escala.

A seguir apresentamos algumas propriedades da distribuição Birnbaum-Saunders Skew-Normal ($BSSN$), algumas das quais foram apresentadas no Capítulo 2 para a distribuição Birnbaum-Saunders, quando $Z \sim N(0, 1)$.

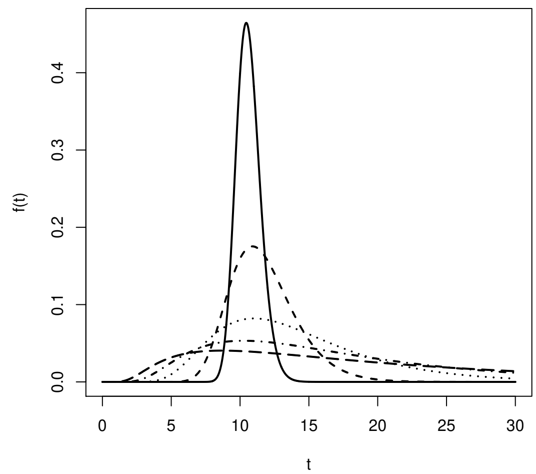
3.1.1 Propriedades

Teorema 3.1. *Seja* $T \sim BSSN(\alpha, \beta, \lambda)$*. Então,*

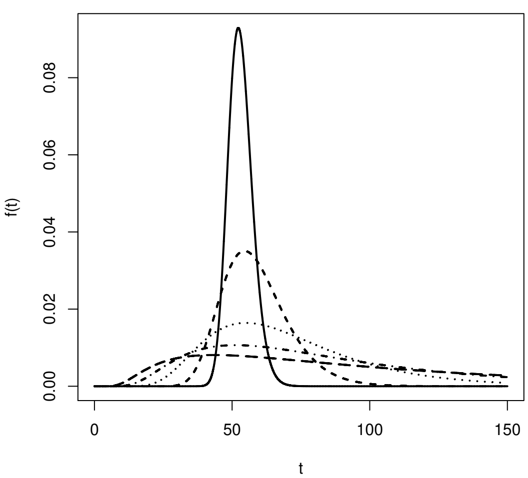
i) $aT \sim BSSN(\alpha, a\beta, \lambda), \;\; a > 0,$



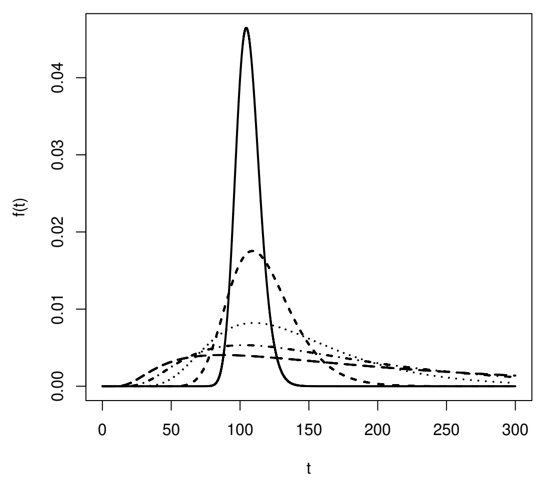
(a) BSSN$(\alpha, 5, 1)$



(b) BSSN$(\alpha, 10, 1)$



(c) BSSN$(\alpha, 50, 1)$



(d) BSSN$(\alpha, 100, 1)$

Figura 3.4: Densidades da BSSN para diferentes valores de β considerando os mesmos valores de α dados na Figura 3.3 e $\lambda = 1$.

ii) $T^{-1} \sim BSSN(\alpha, \beta^{-1}, -\lambda)$.

Prova: As demonstrações de *i)* e *ii)* são obtidas diretamente do teorema de mudança de variável. ■

Observação 3.1. *O item ii) do Teorema 3.1 indica que a distribuição BSSN*(α, β, λ) *também pertence a família de variáveis aleatórias fechada sob reciprocidade (veja, Saunders (1974)). Além disso, a distribuição de* T^{-1} *tem o mesmo coeficiente de assimetria da distribuição* SN$(-\lambda)$.

Sejam $T \sim \text{BSSN}(\alpha, \beta, \lambda)$ e Z dado em (3.5). Então

$$D = \frac{1}{\alpha^2}\left(\frac{T}{\beta} + \frac{\beta}{T} - 2\right) \sim \chi_1^2, \tag{3.8}$$

onde χ_1^2 denota a distribuição Qui-Quadrado com 1 grau de liberdade. Esse resultado permite checar o modelo na prática através do gráfico de envelope simulado.

Teorema 3.2. *Sejam* $T \sim BSSN(\alpha, \beta, \lambda)$, $F_T(t)$ *a função de distribuição acumulada de* T *e* F_Z *é a função de distribuição acumulada de* $Z \sim \text{SN}(0, 1, \lambda)$. *Então,*

i) $F_T(t; \lambda) = F_Z(a_t(\alpha, \beta); \lambda)$. *Em particular, para* $\lambda = 0$, $F_T(t; \lambda = 0) = \Phi(a_t(\alpha, \beta))$;

ii) $F_T(t; -\lambda) = 2F_T(t; \lambda = 0) - F_T(t; \lambda)$, *em que* $F_T(t; \lambda = 0) = \Phi(t)$;

iii) Para $\lambda = 1$, $F_T(t; \lambda) = [\Phi(a_t(\alpha, \beta))]^2$;

iv) O p-ésimo percentil da distribuição, $t_p = F_T^{-1}(p; \lambda)$ *é dado por*

$$t_p = \frac{\beta}{4}\left(\alpha z_p + \sqrt{\alpha^2 z_p^2 + 4}\right)^2,$$

em que z_p *é o p-ésimo percentil da distribuição Skew-Normal. Para* $\lambda = 0$, *isto é, quando* $Z \sim N(0, 1)$ t_p *coincide com o resultado proposto em Jeng (2003) relacionado a este tópico.*

3.1.2 Momentos

A distribuição Birnbaum-Saunders Skew-Normal possui momentos e para determinar os momentos de $T \sim \text{BSSN}(\alpha, \beta, \lambda)$ usaremos o Teorema Binomial. Então mostraremos que os momentos de k-ésima ordem de T dependem da existência dos momentos de uma função da variável aleatória Z, ou seja, a distribuição $BSSN$ possui momentos se, e somente se, a função da distribuição skew-normal possui momentos.

Teorema 3.3. *Seja $T \sim BSSN(\alpha, \beta, \lambda)$ e Z dado em (3.5). Então, $\mathbb{E}(T^n)$ existe e depende dos momentos*

$$\mathbb{E}\left[\left(\frac{\alpha Z}{2}\right)^{k+l}\left(\left(\frac{\alpha Z}{2}\right)^2+1\right)^{\frac{k-l}{2}}\right],$$

com $k = 0, \ldots, n$ e $l = 0, \ldots, k$.

Prova: De (3.6), temos que

$$\begin{aligned}
\mathbb{E}\left(\left[\frac{T}{B}\right]^n\right) &= \mathbb{E}\left(\left[\left\{\frac{\alpha}{2}Z+\sqrt{\left(\frac{\alpha}{2}Z\right)^2+1}\right\}^2\right]^n\right)\\
&= \mathbb{E}\left(\left[1+\left\{\frac{\alpha^2}{2}Z^2+\alpha Z\sqrt{\left(\frac{\alpha}{2}Z\right)^2+1}\right\}\right]^n\right).
\end{aligned}$$

Usando o Teorema Binomial $(a+b)^m = \sum_{k=0}^{m}\begin{pmatrix} m \\ k \end{pmatrix} a^{m-k}b^k$, temos que

$$\mathbb{E}\left(\left[\frac{T}{B}\right]^n\right) = \sum_{k=0}^{n}\begin{pmatrix} n \\ k \end{pmatrix}\mathbb{E}\left(\left\{\frac{\alpha^2}{2}Z^2+\alpha Z\sqrt{\left(\frac{\alpha}{2}Z\right)^2+1}\right\}^k\right).$$

Expandindo o binômio de Newton, obtemos

$$\mathbb{E}\left(\left[\frac{T}{B}\right]^n\right) = \sum_{k=0}^{n}\begin{pmatrix} n \\ k \end{pmatrix}\sum_{l=0}^{k}\begin{pmatrix} k \\ l \end{pmatrix}2^k\mathbb{E}\left[\left(\frac{\alpha Z}{2}\right)^{k+l}\left(\left(\frac{\alpha Z}{2}\right)^2+1\right)^{\frac{k-l}{2}}\right].$$

Portanto, $\mathbb{E}\left(\left[\frac{T}{B}\right]^n\right)$ existe e depende dos momentos $\mathbb{E}\left[\left(\frac{\alpha Z}{2}\right)^{k+l}\left(\left(\frac{\alpha Z}{2}\right)^2+1\right)^{\frac{k-l}{2}}\right]$, com $k = 0, \ldots, n$ e $l = 0, \ldots, k$, que também existe e são finitos.

De fato, seja $W_{kl} = \left(\frac{\alpha Z}{2}\right)^{k+l}\left(\left(\frac{\alpha Z}{2}\right)^2+1\right)^{\frac{k-l}{2}}$. Então utilizando a desigualdade de Markov, temos que

$$\begin{aligned}
\mathbb{E}(|W_{kl}|) &\leq \mathbb{E}^{1/2}\left[\left(\frac{\alpha Z}{2}\right)^{2(k+l)}\right]\mathbb{E}^{1/2}\left[\left((\frac{\alpha Z}{2}\right)^2+1)^{k-l}\right]\\
&< \infty,
\end{aligned}$$

pois $Z^2 \sim \chi_1^2$ e consequentemente $\mathbb{E}\left[\left(\frac{\alpha Z}{2}\right)^{2(k+l)}\right]$ e $\mathbb{E}\left[\left((\frac{\alpha Z}{2}\right)^2+1)^{k-l}\right]$ são finitos. ■

Corolário 3.1. *Seja $T \sim BSSN(\alpha, \beta, \lambda)$ e Z dado em (3.5). Então*

$$\mathbb{E}(T^n) = \beta^n \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} \sum_{l=0}^{k} \begin{pmatrix} k \\ l \end{pmatrix} 2^k \mathbb{E}\left[\left(\frac{\alpha Z}{2} \right)^{k+l} \left(\left(\frac{\alpha Z}{2} \right)^2 + 1 \right)^{\frac{k-l}{2}} \right].$$

Prova: A prova segue facilmente do Teorema 3.3. ∎

Corolário 3.2. *Seja $T \sim BSSN(\alpha, \beta, \lambda)$ e $W_k = \mathbb{E}(Z^K \sqrt{\alpha^2 Z^2 + 4})$. Então*

i) $\mathbb{E}(T) = \frac{\beta}{2} \left(2 + \alpha^2 + \alpha W_1\right);$

ii) $\mathbb{E}(T^2) = \frac{\beta^2}{2} \left(2 + 4\alpha^2 + 3\alpha^4 + 2\alpha W_1 + \alpha^3 W_3\right);$

iii) $\mathbb{E}(T^3) = \frac{\beta^3}{2} \left(2 + 9\alpha^2 + 18\alpha^4 + 15\alpha^6 + 3\alpha W_1 + 4\alpha^3 W_3 + \alpha^5 W_5\right);$

iv) $\mathbb{E}(T^4) = \frac{\beta^4}{2} \left(2 + 16\alpha^2 + 60\alpha^4 + 120\alpha^6 + 105\alpha^8 + 4\alpha W_1 + 10\alpha^3 W_3 + 6\alpha^5 W_5 + \alpha^7 W_7\right).$

Prova: A demonstração é imediata aplicando o Corolário 3.1, para $n = 1, 2, 3, 4$. ∎

Para obter os valores de W_1, W_3, W_5 e W_7, as integrais envolvidas nos cálculos desses momentos devem ser resolvidas utilizando métodos de integração numérica.

Agora, seja $T \sim \text{BSSN}(\alpha, \beta, \lambda)$. Então, do Corolário 3.1, temos que a variância de T é dada por

$$\text{Var}(T) = \left(\frac{\beta}{2} \right)^2 \left[4\alpha^2 - \alpha^2 W_1^2 + 2\alpha^3 W_3 - 2\alpha^3 W_1 + 5\alpha^4 \right].$$

Por outro lado, do Teorema 3.1, sabemos que $T^{-1} \sim \text{BSSN}(\alpha, \beta^{-1}, -\lambda)$. Consequentemente, dos resultados acima, temos que

$$\mathbb{E}(T^{-1}) = \frac{1}{2\beta} \left(2 + \alpha^2 + \alpha W_1 \right)$$

e

$$\text{Var}(T^{-1}) = \frac{1}{(2\beta)^2} \left[4\alpha^2 - \alpha^2 W_1^2 + 2\alpha^3 W_3 - 2\alpha^3 W_1 + 5\alpha^4 \right].$$

Note que do Teorema 3.1 e Corolário 3.2, temos que os coeficientes de assimetria e curtose são dados, respectivamente, por

$$\alpha_3(T) = \sqrt{\beta_1(T)} = \frac{\mu_3}{(\mu_2)^{3/2}} \quad \text{e} \quad \alpha_4(T) = \sqrt{\beta_2(T)} = \frac{\mu_4}{(\mu_2)^2}, \tag{3.9}$$

onde $\mu_k = \mathbb{E}[(T-\mathbb{E}[T])^k]$, com $k = 2, 3, 4$, de acordo com a notação de Johnson et al. (1993, p. 42). Generalizando, temos que $\alpha_k(T) = \mu_k(\mu_2)^{-k/2}$.

Seja $T \sim \text{BSSN}(\alpha, \beta, \lambda)$. Então, os coeficientes de assimetria e curtose de T são dados, respectivamente, por

$$\begin{aligned}
\alpha_3(T) &= -2\alpha^3 \left(\frac{6W_1 - W_1^3 - 2W_3 - 12\alpha - 3W_1^2\alpha + 3W_1W_3\alpha - 3W_1\alpha^2 + 9W_1\alpha^2}{A} \right. \\
&\quad \left. + \frac{3W_3\alpha^2 - 2W_5\alpha^2 - 22\alpha^3}{A} \right)
\end{aligned}$$

$$\begin{aligned}
\alpha_4(T) &= -\left(\frac{-48 - 24W_1^2 + 3W_1^4 + 168W_1\alpha^3 + 16W_1W_3 + 96W_1\alpha + 12W_1^3\alpha + 16W_3\alpha}{B} \right. \\
&\quad - \frac{12W_1^2W_3\alpha - 16W_5\alpha - 18W_1^2\alpha^2 - 24W_1W_3\alpha^2 + 16W_1W_5\alpha^2 + 16W_5\alpha^3}{B} \\
&\quad \left. + \frac{360\alpha^2 - 8W_7\alpha^3 - 360\alpha^2 - 633\alpha^4}{B} \right),
\end{aligned}$$

em que $A = (-\alpha^2(W_1^2 + \alpha^2 - 2\alpha(W_3 + 3\alpha) + (-4 + 2W_1\alpha)))^{3/2}$ e $B = (-4 + W_1^2 + 2W_1\alpha - 2W_3\alpha - 5\alpha^2)^2$. As quantidades $\alpha_3(T)$ e $\alpha_4(T)$ medem a assimetria e a largura das caudas da distribuição, respectivamente.

Observação 3.2. *Segundo Vilca-Labra e Leiva (2006), se $Z \sim SN(0, 1, \lambda)$ e $\lambda = 0$, isto é, Z segue uma distribuição normal padrão, então a média e a variância coincidem com os resultados propostos em Ng et al. (2003) e Johnson et al. (1995, p. 653) relacionados a este tópico. Entretanto, os coeficientes de assimetria e curtose são dados respectivamente por*

$$\alpha_3(T) = \frac{4\alpha(11\alpha^2 + 6)}{(5\alpha^2 + 4)^{3/2}}$$

e

$$\alpha_4(T) = 3 + \frac{6\alpha^2(93\alpha^2 + 40)}{(5\alpha^2 + 4)^2}.$$

Nas Tabelas 3.1-3.2 encontram-se a média, o desvio padrão e os coeficientes de assimetria e curtose da distribuição $\text{BSSN}(\alpha, 1, \lambda)$. Os valores de W_1, W_3, W_5 e W_7 foram obtidos através de integração numérica. Consideramos diferentes valores para α e λ. Podemos observar que, para valores de α pequeno e λ negativo, a distribuição torna-se assimétrica negativa.

A Proposição a seguir e os seguintes Lemas serão de grande utilidade para obter a distribuição condicional de H dado T, que será utilizada na implementação do algoritmo EM.

Tabela 3.1: Os quatro primeiros momentos da distribuição BSSN($\alpha, 1, \lambda$) para $\beta = 1$ e diferentes valores de λ e α.

λ	α	μ_{BSSN}	σ_{BSSN}	**Assimetria**	**Curtose**
0.00	0.0500	1.0012	0.0500	0.1499	3.0374
	0.1	1.005	0.1006	0.2998	3.1497
	0.5	1.125	0.5728	1.4547	6.4421
	1.0	1.5000	1.5000	2.5185	12.8518
	1.5	2.125	2.9288	3.0980	14.4686
0.20	0.05	1.0090	0.0498	0.1495	3.0375
	0.1	1.0207	0.1009	0.2967	3.1464
	0.5	1.2100	0.6047	1.4021	6.1452
	1.0	1.7037	1.6394	2.3573	11.4597
	1.5	2.4976	3.2504	2.8517	14.0235
0.50	0.05	1.0191	0.0476	0.1639	3.0519
	0.1	1.0408	0.0974	0.3028	3.1609
	0.5	1.3180	0.6198	1.3207	5.7744
	1.0	1.9595	1.7412	2.1513	9.9723
	1.5	2.9609	3.5032	2.5637	12.6179
1.00	0.05	1.0294	0.0426	0.2631	3.1551
	0.1	1.0615	0.0881	0.3886	3.2995
	0.5	1.4278	0.5990	1.3034	5.7910
	1.0	2.2110	1.7431	2.0266	9.3206
	1.5	3.4042	3.5568	2.3765	10.4223
1.50	0.05	1.0344	0.0388	0.4173	3.3290
	0.1	1.0715	0.0810	0.5342	3.5261
	0.5	1.4795	0.5727	1.3918	6.2145
	1.0	2.3244	1.7029	2.0619	9.6363
	1.5	3.5959	3.5053	2.3812	10.5817
2.00	0.05	1.0369	0.0366	0.5644	3.5059
	0.1	1.0765	0.0765	0.6752	3.7464
	0.5	1.5050	0.5548	1.4924	6.6316
	1.0	2.3786	1.6724	2.1254	10.0331
	1.5	3.6846	3.4621	2.4232	12.8976

Tabela 3.2: Os quatro primeiros momentos da distribuição BSSN$(\alpha, 1, -\lambda)$ para $\beta = 1$ e diferentes valores de $-\lambda$ e α.

$-\lambda$	α	μ_{BSSN}	σ_{BSSN}	**Assimetria**	**Curtose**
−0.05	0.05	0.9992	0.0499	0.1499	3.0374
	0.1	1.0010	0.1001	0.2998	3.1499
	0.5	1.1033	0.5621	1.4637	6.4508
	1.0	1.4479	1.4569	2.5559	13.2156
	1.5	2.0297	2.8312	3.1609	18.1625
−0.20	0.05	0.9934	0.0490	0.1467	3.0359
	0.1	0.9892	0.0978	0.2956	3.1460
	0.5	1.0399	0.5254	1.4745	6.6204
	1.0	1.2962	1.3149	2.6502	14.2732
	1.5	1.7523	2.5125	3.3401	20.3905
−0.50	0.05	0.9833	0.0459	0.1170	3.0260
	0.1	0.9691	0.0905	0.2588	3.1143
	0.5	0.9319	0.4443	1.4143	6.5316
	1.0	1.0404	1.0225	2.7156	16.1460
	1.5	1.2890	1.8670	3.5967	25.7571
−1.00	0.05	0.9730	0.0400	−0.0102	3.0222
	0.1	0.9484	0.0779	0.1168	3.0398
	0.5	0.8221	0.3375	1.1500	6.0644
	1.0	0.7889	0.6709	2.4624	16.0722
	1.5	0.8457	1.1101	3.5892	20.7379
−1.50	0.05	0.9680	0.0359	−0.1834	3.0409
	0.1	0.9384	0.0694	−0.0671	3.1143
	0.5	0.7704	0.2770	0.8502	4.0012
	1.0	0.6755	0.4904	2.0271	9.8445
	1.5	0.6540	0.7357	3.1942	21.7345
−2.00	0.05	0.9655	0.0335	−0.3439	3.0649
	0.1	0.9334	0.0643	−0.2351	3.0791
	0.5	0.7449	0.2439	0.5973	3.5065
	1.0	0.6213	0.3985	1.6263	8.4835
	1.5	0.5653	0.5522	2.7021	14.6602

Proposição 3.1. *Seja Z como em (3.5) e $T \sim BSSN(\alpha, \beta, \lambda)$. Assuma que Z tem representação estocástica como em (3.2) dada por $Z = \delta H + (1-\delta^2)^{1/2} X_1$. Então*

$$T|H = h \sim BSE(\alpha_\delta, \beta, \sigma = 2, \nu_h),$$

em que $\alpha_\delta = \alpha\sqrt{1-\delta^2}$ e $\nu_h = -\frac{\delta}{\sqrt{1-\delta^2}} h$. A função de densidade condicional de T dada $H = h$ é dada por

$$f_{T/H}(t/h) = \phi\left(\nu_h + a_t(\alpha_\delta, \beta)\right) A_t(\alpha_\alpha, \beta), \tag{3.10}$$

em que $a_t(\alpha_\delta, \beta)$ e $A_t(\alpha_\delta, \beta)$ são como em (2.2).

Prova: Como $Z \sim SN(0, 1, \lambda)$, então usando a representação estocástica dada em (3.2) podemos expressar Z como

$$Z \;=\; \delta H + (1-\delta^2)^{1/2} X_1 = \frac{1}{\alpha}\left(\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}}\right).$$

Assim a distribuição condicional de Z dado $H = h$ é dada por

$$Z/(H = h) \;=\; \frac{1}{\alpha}\left(\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}}\right) \sim N(\mu_h, \sigma^2),$$

onde $\mu_h = \delta h$ e $\sigma^2 = 1 - \delta^2$. Logo, dado $H = h$ a distribuição de $W = (Z - \mu_h)/\sigma$ é dada por

$$W \;=\; -\frac{\mu_h}{\sigma} + \frac{1}{\sigma\alpha}\left(\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}}\right) \sim N(0, 1). \tag{3.11}$$

Assim, do resultado (3.11), temos a prova. ■

Lema 3.1. *Sejam $Y \sim N(\mu_y, \sigma_y^2)$ e $X \sim N(\mu_x, \sigma_x^2)$. Então*

$$\phi(y|\mu_y + ax, \sigma_y^2)\phi(x|\mu_x, \sigma_x^2) \;=\; \phi(y|\mu_y + a\mu_x, \sigma_y^2 + a^2\sigma_x^2)\phi\left(x|\mu_x + \Lambda\frac{a}{\sigma_y^2}(y - \mu_y - a\mu_x), \Lambda\right),$$

em que $\Lambda = \dfrac{\sigma_x^2\sigma_y^2}{\sigma_y^2 + a\sigma_x^2}$.

Lema 3.2. *Sejam $X \sim N(\eta, \tau^2)$. Então para qualquer constante real a*

$$E(X|X>a) \;=\; \eta + \frac{\phi\left(\frac{a-\eta}{\tau}\right)}{1-\Phi\left(\frac{a-\eta}{\tau}\right)}\tau$$

e

$$E(X^2|X>a) \;=\; \eta^2+\tau^2+\frac{\phi\left(\frac{a-\eta}{\tau}\right)}{1-\Phi\left(\frac{a-\eta}{\tau}\right)}(\eta+a)\tau.$$

Proposição 3.2. *Seja Z como em (3.5) e $T \sim \mathrm{BSSN}(\alpha, \beta, \lambda)$. Então*

i) $f_{H/T}(h/t) = \dfrac{1}{\Phi(\lambda a_t)}\phi(h|\delta a_t, (1-\delta^2))I_{[0,\infty)}(h),$

ii) $\mathbb{E}[H|T=t] = \delta a_t + W_\Phi\left(\dfrac{\delta a_t}{\sqrt{1-\delta^2}}\right)\sqrt{1-\delta^2},$

iii) $\mathbb{E}[H^2|T=t] = (\delta a_t)^2 + (1-\delta^2) + W_\Phi\left(\dfrac{\delta a_t}{\sqrt{1-\delta^2}}\right)\delta a_t\sqrt{1-\delta^2},$

em que $W_\Phi\left(\dfrac{\delta a_t}{\sqrt{1-\delta^2}}\right) = \phi\left(\dfrac{\delta a_t}{\sqrt{1-\delta^2}}\right)/\Phi\left(\dfrac{\delta a_t}{\sqrt{1-\delta^2}}\right).$

Prova:

i) Como $h_H(h) = 2\phi(h|0,1)I_{[0,\infty)}(h)$ e

$$\phi\left(-\frac{\delta}{\sqrt{1-\delta^2}}h + \frac{1}{\sqrt{1-\delta^2}}a_t|0,1\right) = \sqrt{1-\delta^2}\,\phi\left(a_t|\delta h, 1-\delta^2\right).$$

Então usando o Lema 3.1, temos que $\phi(a_t|\delta h, 1-\delta^2)\phi(t|0,1) = \phi(a_t|0,1)\phi(h|\delta a_t, (1-\delta^2))$. Assim a prova segue da relação $f_{H|T}(h|t) = f_{T|H}(t|h)f_H(h)/f_T(t)$.

ii)- iii) Para $k = 1, 2$, temos que

$$\begin{aligned}
\mathbb{E}(H^k/T) &= \int_0^\infty \frac{h^k}{\Phi(\lambda a_t)}\phi(h|\delta a_t, 1-\delta^2)dh \\
&= \frac{1}{\Phi(\lambda a_t)}\int_0^\infty h^k\phi(h|\delta a_t, 1-\delta^2)dh \\
&= E[X^k|X>0].
\end{aligned}$$

Agora, usando as propriedades da distribuição Half-Normal dada no Lema 3.2, segue que

$$\mathbb{E}(H/T) = \delta a_t + W_\Phi\left(\frac{\delta a_t}{\sqrt{1-\delta^2}}\right)\sqrt{1-\delta^2}$$

e

$$\mathbb{E}[H^2|T=t] = (\delta a_t)^2 + (1-\delta^2) + W_\Phi\left(\frac{\delta a_t}{\sqrt{1-\delta^2}}\right)\delta a_t\sqrt{1-\delta^2}.$$

■

3.2 Função de Confiabilidade da Distribuição BSSN

Nesta seção apresentaremos uma breve descrição da função de confiabilidade da distribuição Birnbaum-Saunders Skew-Normal e em seguida faremos uma comparação com a Birnbaum-Saunders com respeito a certa ordem estocástica.

A confiabilidade é um aspecto importante que aparece nos requisitos da qualidade de produtos, estando associada à capacidade dos mesmos de "funcionarem de maneira satisfatória durante um longo período de tempo"(Freitas e Colosimo, 1997). Uma das ferramentas utilizadas para medir a confiabilidade de produtos é o Teste de Vida Acelerada, no qual estes são submetidos a testes em condições diferentes de uso, e permitem que a falha ocorra mais rapidamente.

Considerando a variável de interesse T, tempo até a ocorrência da falha ou tempo de falha, a função de confiabilidade ou função de sobrevivência $R_T(t)$ é dada por

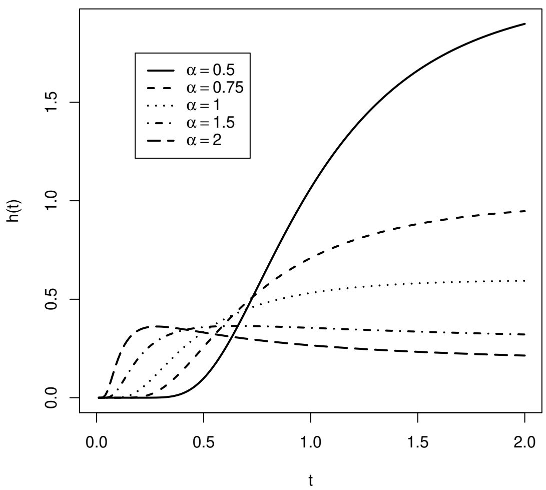
$$R_T(t) = 1 - F_Z(a_t),$$

em que $F_Z(.)$ é a função de distribuição acumulada da variável aleatória skew-normal padrão e a_t é como em (2.2).

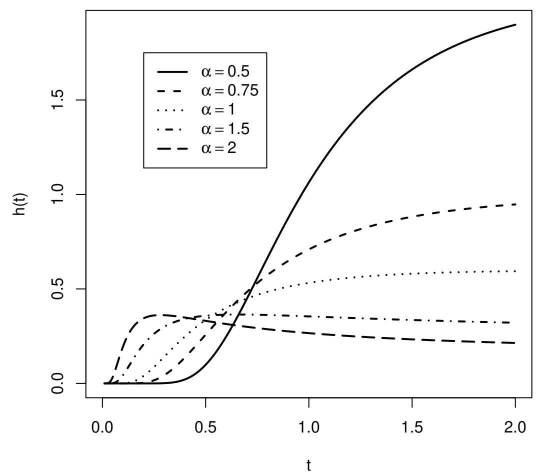
Segue de Leiva et al. (2008), que se $t \to \infty$, temos que a função risco da distribuição BSSN é dada por

$$\lim_{t\to\infty} h_T(t) = \frac{1+\lambda^2}{2\alpha^2\beta}.$$

A função risco $h_T(t)$ da distribuição Birnbaum-Saunders skew-normal é crescente em λ. Na Figura 3.5 apresentamos os gráficos de $h_T(t)$ para alguns valores de α fixando $\beta = 1.0$ e $\lambda = 1$ e $\lambda = -1$. Observe que o risco aproxima-se de uma constante quando $t \to \infty$ e para valores de α e t grandes a função $h_T(t)$ é decrescente. Na Figura 3.6 apresentamos os gráficos de $h_T(t)$ para valores de λ positivos e negativos. Observe que o risco é uma função crescente em λ, tanto para $\lambda > 0$ quanto para $\lambda < 0$.



(a) $\lambda = 1.0$



(b) $\lambda = -1.0$

Figura 3.5: Função risco da Birnbaum-Saunders Skew-Normal para diferentes valores de α considerando $\beta = 1.0$, $\lambda = 1$ e $\lambda = -1$.

Outra função útil em análise de dados de sobrevivência e confiabilidade é a função de taxa de falha acumulada. Esta função, como o próprio nome sugere, fornece a taxa de falha acumulada do indivíduo e é definida por:

$$H_T(t) = \int_0^t h_T(t)dt.$$

A função taxa de falha acumulada, $H_T(t)$, não tem uma interpretação direta, mas pode ser útil na avaliação da função de maior interesse que é a taxa de falha, $h_T(t)$. Isto acontece essencialmente na estimação não-paramétrica em que $H_T(t)$ apresenta um estimador com propriedades ótimas e $h_T(t)$ é difícil de ser estimada.

Outra quantidade de interesse é a vida média residual (MRLF), que é definida como

$$\mu_F(t) = E(X - t|X > t) = \int_t^\infty \frac{(x-t)f(x)}{R_T(t)}dx,$$

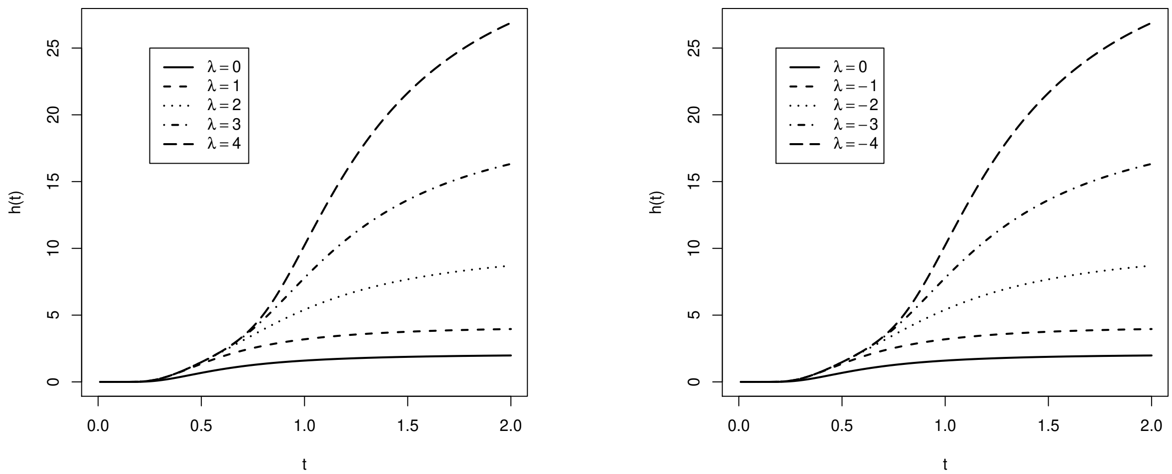


Figura 3.6: Função risco da Birnbaum-Saunders Skew-Normal para diferentes valores de λ considerando $\alpha = 0.5$ e $\beta = 1.0$.

ou seja, a vida média residual é definida condicional a um certo tempo de vida t.

A seguir apresentamos uma comparação entre a distribuição Birnbaum-Saunders Skew-Normal (BSSN) e Birnbaum-Saunders (BS) com respeito a certa ordem estocástica.

3.2.1 Comparação entre a distribuição Birnbaum-Saunders Skew-Normal e Birnbaum-Saunders

Na modelagem estatística em áreas como engenharia, economia, medicina, bem como em outros campos, vê-se frequentemente a necessidade de comparar variáveis aleatórias. A forma mais simples de confrontar duas variáveis aleatórias é comparando os respectivos valores esperados. No entanto, esta comparação é pouco informativa já que se baseia apenas em duas quantidades. Além disso, na maior parte das situações dispõe-se de informação muito mais detalhada acerca do comportamento das variáveis aleatórias, como é o caso das suas funções de distribuição, transformadas de Laplace, funções geradoras de momentos, funções taxa de falha e outros funcionais. A comparação das características de variáveis aleatórias resulta no estabelecimento de diferentes relações de ordem estocástica entre essas mesmas variáveis aleatórias, que são mais informativas que a mera comparação dos seus valores esperados.

Neste seção, apresentamos alguns resultados de ordem estocásticas para comparar as variáveis BSSN e BS, seguindo um procedimento similar ao trabalho considerado por Gupta e Brown (2001).

Para esta comparação, consideramos algumas definições e resultados mostrados em Shaked e Shanthikumar (1994). Sejam T e S duas variáveis aleatórias independentes com funções densidades de probabilidades $f_T(t)$ e $f_S(t)$ e funções de confiabilidade dadas por $R_T(t)$ e $R_S(t)$, respectivamente. Então

1. T é dito ser maior que S em relação a ordem de razão de verossimilhanças, escrita como $T \underset{LR}{\geq} S$, se $\dfrac{f_T(t)}{f_S(t)}$ é não descrescente em t;
2. T é dito ser maior que S em relação a ordem de taxa de falha, escrita como $T \underset{FR}{\geq} S$, se $h_T(t) \leq h_S(t)\ \forall t$, onde $h_T(t)$ e $h_S(t)$ são as taxas de falha de T e S, respectivamente;
3. T é dito ser maior que S em relação a ordem estocástica, escrita como $T \underset{ST}{\geq} S$, se $R_T(t) \geq R_S(t)\ \forall t$, onde $R_T(t)$ e $R_S(t)$ são as funções de confiabilidade de T e S, respectivamente;
4. T é dito ser maior que S em relação a ordem de vida média residual, escrita como $T \underset{MRL}{\geq} S$, se $\mu_{FT}(t) \geq \mu_{FS}(t)\ \forall t$, onde $\mu_{FT}(t)$ e $\mu_{FS}(t)$ são as funções de vida média residual (MRLF) de T e S, respectivamente.

Usando os resultados de Gupta and Kirmani (1998), sabemos que [1] $\Rightarrow$ [2] $\Rightarrow$ [3] e [2] $\Rightarrow$ [4].

Sejam T_{BSSN} e T_{BS} variáveis aleatórias com distribuicões Birnbaum-Saunders Skew-Normal e Birnbaum-Saunders, respectivamente. Considere a razão das funções densidades de probabilidade de T_{BSSN} e T_{BS} dadas por

$$\begin{aligned}\frac{f_{BSSN}(t)}{f_{BS}(t)} &= \frac{2\phi(a_t(\alpha,\beta))\Phi(\lambda a_t(\alpha,\beta))A_t(\alpha,\beta)}{\phi(a_t(\alpha,\beta))A_t(\alpha,\beta)} \\ &= 2\Phi(\lambda a_t(\alpha,\beta)).\end{aligned}$$

Além disso, $\forall t_1, t_2 > 0$, tal que $t_1 < t_2$, $a_{t_1}(\alpha,\beta) \leq a_{t_2}(\alpha,\beta)$, então $\dfrac{f_{BSSN}(t)}{f_{BS}(t)}$ é não decrescente se $\lambda < 0$ e consequentemente temos que

1. $T_{BSSN} \underset{LR}{\geq} T_{BS}$;

2. $T_{BSSN} \underset{FR}{\geq} T_{BS}$;

3. $T_{BSSN} \underset{ST}{\geq} T_{BS}$;

4. $T_{BSSN} \underset{MRL}{\geq} T_{BS}$.

Similarmente, se $\lambda < 0$ temos que

1. $T_{BSSN} \underset{LR}{\leq} T_{BS}$;

2. $T_{BSSN} \underset{FR}{\leq} T_{BS}$;

3. $T_{BSSN} \underset{ST}{\leq} T_{BS}$;

4. $T_{BSSN} \underset{MRL}{\leq} T_{BS}$.

3.3 Estimação de Máxima Verossimilhança

Suponha que temos n observações independentes, $T_1, \ldots, T_n$, com $T_i \sim \text{BSSN}(\alpha, \beta, \lambda)$, $i = 1, \ldots, n$, onde α, β e λ são parâmetros desconhecidos. Isto é, o vetor de parâmetros é $\boldsymbol{\theta} = (\alpha, \beta, \lambda)^\top$ e a função log-verossimilhança para $\boldsymbol{\theta}$ é dada por

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}),$$

em que $\ell_i(\boldsymbol{\theta}) = \log 2 + \log A_{t_i}(\alpha, \beta) + \ell_{1_i}(\boldsymbol{\theta}) + \log\left[\Phi(\lambda a_{t_i}(\alpha, \beta))\right]$, com $\ell_{1_i}(\boldsymbol{\theta}) = \log \phi(a_{t_i}(\alpha, \beta))$, $a_{t_i}(\alpha, \beta)$ e $A_{t_i}(\alpha, \beta)$ são como em (2.2).

Para encontrar a EMV do vetor de parâmetros $\boldsymbol{\theta}$ será utilizado um procedimento de estimação usando o algoritmo EM, uma ferramenta popular para estimação de máxima verossimilhança para modelos com dados incompletos. A seguir, discutimos a implementação do algoritmo EM para encontrar o EMV de $\boldsymbol{\theta}$. A metodologia proposta para a classe de modelos BSSN tratada aqui não existe na literatura.

3.3.1 Estimação via algoritmo EM

Para efeito de estimação vamos considerar uma reparametrização em lugar de estimar $\boldsymbol{\theta} = (\alpha, \beta, \lambda)^\top$, vamos estimar $\boldsymbol{\theta}$ dado por $\boldsymbol{\theta} = (\alpha, \beta, \delta)^\top$, onde $\delta = \dfrac{\lambda}{\sqrt{1+\lambda^2}}$. Isto é, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \equiv \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}$.

Utilizando a Proposição 3.1, a distribuição Birnbaum-Saunders Skew-Normal tem a seguinte representação hierárquica

$$\begin{aligned} T_i|H_i = h_i &\overset{ind}{\sim} \text{BSE}\left(\alpha_\delta, \beta, \sigma = 2, \nu_{h_i}\right) \quad \text{e} \\ H_i &\overset{ind}{\sim} HN(0,1); \quad i = 1, \ldots, n, \end{aligned} \tag{3.12}$$

em que $\alpha_\delta = \alpha\sqrt{1-\delta^2}$, $\nu_{h_i} = -\frac{\delta}{\sqrt{1-\delta^2}} h_i$ e $HN(0,1)$ denota a distribuição half-normal padrão univariada (Johnson et al., 1995).

Sejam $\mathbf{t} = (t_1, \ldots, t_n)^\top$ e $\mathbf{h} = (h_1, \ldots, h_n)^\top$ e tratando $\mathbf{h}$ como dado faltante. Então, de acordo com a representação hierárquica dada em (3.12), segue que a função log-verossimilhança completa associada com $\mathbf{t}_c = (\mathbf{t}^\top, \mathbf{h}^\top)^\top$ é dada por

$$\begin{aligned} \ell_c(\boldsymbol{\theta}|\mathbf{t}_c) &= \sum_{i=1}^n \log f_{T/H}(t_i/h_i) + \sum_{i=1}^n f_H(h_i) \\ &\propto -n\log\alpha - \frac{n}{2}\log\beta + \sum_{i=1}^n \log(t_i+\beta) - \frac{n}{2}\log(1-\delta^2) - \frac{1}{2(1-\delta^2)}\sum_{i=1}^n a_{t_i}^2(\alpha,\beta) \\ &+ \frac{\delta}{(1-\delta^2)}\sum_{i=1}^n a_{t_i}(\alpha,\beta)h_i - \frac{\delta^2}{2(1-\delta^2)}\sum_{i=1}^n h_i^2. \end{aligned}$$

Fazendo $\widehat{h}_i = E[H_i|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, t_i]$ e $\widehat{h_i^2} = E[H_i^2|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, t_i]$ e aplicando as propriedades de esperança condicional de H dado $T = t$, que foram obtidas na Proposição 3.2 e usando os momentos da distribuição half-normal (Lachos, 2004), tem-se que

$$\widehat{h}_i = \delta a_{t_i} + W_\Phi\left(\frac{\delta a_{t_i}}{\sqrt{1-\delta^2}}\right)\sqrt{1-\delta^2} \tag{3.13}$$

e

$$\widehat{h_i^2} = (\delta a_{t_i})^2 + (1-\delta^2) + W_\Phi\left(\frac{\delta a_{t_i}}{\sqrt{1-\delta^2}}\right)\delta a_{t_i}\sqrt{1-\delta^2}, \tag{3.14}$$

em que $W_\Phi\left(\frac{\delta a_{t_i}}{\sqrt{1-\delta^2}}\right) = \phi\left(\frac{\delta a_{t_i}}{\sqrt{1-\delta^2}}\right)/\Phi\left(\frac{\delta a_{t_i}}{\sqrt{1-\delta^2}}\right)$.

Denote por $\boldsymbol{\theta}^{(r)} = (\alpha^{(r)}, \beta^{(r)}, \delta^{(r)})^\top$ a estimativa de $\boldsymbol{\theta}$ para a k-ésima iteração. Segue que a esperaça com respeito a t, condicionada em h, da função log-verossimilhança completa (Passo E), tem a forma

$$\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) &= \mathbb{E}[\ell_c(\boldsymbol{\theta}|\mathbf{t}_c)|\mathbf{t}, \widehat{\boldsymbol{\theta}}^{(r)}] \\
&= c - n\log\alpha^{(r)} - \frac{n}{2}\log\beta^{(r)} + \sum_{i=1}^{n}\log(t_i + \beta^{(r)}) - \frac{n}{2}\log(1 - {\delta^{(r)}}^2) \\
&- \frac{1}{2{\alpha^{(r)}}^2(1 - {\delta^{(r)}}^2)}\sum_{i=1}^{n} b_{t_i}^2 + \frac{\delta^{(r)}}{\alpha^{(r)}(1 - {\delta^{(r)}}^2)}\sum_{i=1}^{n}\widehat{h}_i^{(r)} b_{t_i} \\
&- \frac{{\delta^{(r)}}^2}{2(1 - {\delta^{(r)}}^2)}\sum_{i=1}^{n}\widehat{h_i^2}^{(r)},
\end{aligned} \tag{3.15}$$

em que $\delta^{(r)} = \dfrac{\lambda^{(r)}}{\sqrt{1 - \lambda^{(r)}}}$ e $b_{t_i} = b_{t_i}(\alpha^{(r)}, \beta^{(r)}) = \sqrt{\dfrac{t_i}{\beta^{(r)}}} - \sqrt{\dfrac{\beta^{(r)}}{t_i}}$.

A etapa M do algoritmo, é implementada maximizando a função de log-verossimilhança completa dada em (3.15). Derivando $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})$ com relação a $\boldsymbol{\theta}$ e após algumas manipulações algébricas, obtemos as seguintes equações de estimação,

$$\widehat{\alpha}^2 + \alpha\frac{\delta}{(1 - \delta^2)}\bar{b}_{ht}(\beta) - \frac{1}{(1 - \delta^2)}\bar{b}_t^2(\beta) = 0, \tag{3.16}$$

$$n + (1 - \delta^2)\alpha^2\beta\sum_{i=1}^{n}\frac{1}{t_i + \beta} - \beta\sum_{i=1}^{n}\frac{1}{t_i} - \delta\alpha\beta^{1/2}\sum_{i=1}^{n}\widehat{h}_i t_i^{-1/2} = 0, \tag{3.17}$$

$$\widehat{\delta} = \frac{\sum_{i=1}^{n}\widehat{h}_i b_{t_i}}{\widehat{\alpha}\sum_{i=1}^{n}\widehat{h_i^2}}, \tag{3.18}$$

em que

$$\bar{b}_{ht}(\beta) = \frac{1}{n}\sum_{i=1}^{n}\widehat{h}_i b_{t_i} \;\text{ e }\; \bar{b_t^2}(\beta) = \frac{1}{n}\sum_{i=1}^{n} b_{t_i}^2.$$

Neste caso, β não tem forma fechada então na etapa M do algoritmo EM obtemos a EMV de β através

$$\widehat{\beta} = \arg\max\{Q(\beta, \widehat{\alpha}, \widehat{\lambda}|\widehat{\boldsymbol{\theta}})\}. \tag{3.19}$$

Portanto, se $\boldsymbol{\theta}^{(r)}$ maximiza (3.15) na r-ésima iretação, então na iteração $r+1$ o algoritmo EM procede como segue:

Etapa E: Calcule $\widehat{h}_i(\boldsymbol{\theta}^{(r)})$ e $\widehat{h_i^2}(\boldsymbol{\theta}^{(r)})$, $i = 1, \ldots, n$, dados por (3.13) e (3.14) respectivamente, com $\boldsymbol{\theta} = \boldsymbol{\theta}^{(r)}$.

Etapa M: Calcule $(\boldsymbol{\theta} = \boldsymbol{\theta}^{(r+1)})$ a partir de (3.16), (3.18) e (3.19).

Observação 3.3. *Note que quando $\lambda = 0$ as equações (3.16) e (3.17) podem ser escritas, respectivamente, como*

$$\widehat{\alpha}^2 = \left(\frac{1}{n}\sum_{i=1}^{n} b_{t_i}^2\right) = \left[\frac{S}{\widehat{\beta}} + \frac{\widehat{\beta}}{R} - 2\right] \quad \text{e} \quad \widehat{\beta}^2 + \widehat{\beta}[\kappa(\widehat{\beta}) + 2\mathrm{R}] + \mathrm{R}[\kappa(\widehat{\beta}) + \mathrm{S}] = 0,$$

em que $S = \frac{1}{n}\sum_{i=1}^{n} t_i$, $R = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} t_i}$ e $\kappa(x) = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x+t_i}}$. Note também que este algoritmo generaliza os resultados encontrados em Birnbaum e Saunders (1969b).

Como estimativas iniciais $\boldsymbol{\theta} = \boldsymbol{\theta}^{(0)}$ para α e β podemos utilizar as estimativas de máxima verossimilhança da distribuição BS obtidas usando a função **est1bs** do pacote **bs** proposto por Leiva et al. (2006) e como critério de convergência podemos usar

$$||\boldsymbol{\theta}^{(r)} - \boldsymbol{\theta}^{(r-1)}|| < \epsilon,$$

onde $||\mathbf{a}||$ indica norma do vetor $\mathbf{a}$ e $\epsilon > 0$.

3.4 Matriz de Informação de Fisher Observada

Considere $T_1, \ldots, T_n$, onde $T_i \sim \text{BSSN}(\alpha, \beta, \lambda)$, com $i = 1, \ldots, n$. Então a função log-verossimilhança $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta})$ é da forma

$$\ell_i(\boldsymbol{\theta}) = \log[2A_t(\alpha, \beta)] + \ell_{1_i}(\boldsymbol{\theta}) + \log[\Phi(\ell_{2_i}(\boldsymbol{\theta}))], \;\; \boldsymbol{\theta} = (\alpha, \beta, \lambda)^\top,$$

em que $\ell_{1_i}(\boldsymbol{\theta}) = \phi(a_t(\alpha, \beta))$, $\ell_{2_i}(\boldsymbol{\theta}) = \lambda a_t(\alpha, \beta)$, $a_t(\alpha, \beta)$ e $A_t(\alpha, \beta)$ são dadas em (2.2). Então a primeira derivada de $\ell_i(\boldsymbol{\theta})$ é dada por

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \psi} = \frac{\partial 2A_t(\alpha, \beta)}{\partial \psi} + \frac{\partial \ell_{1_i}(\boldsymbol{\theta})}{\partial \psi} + W_\Phi(\ell_{2_i}(\boldsymbol{\theta}))\frac{\partial \ell_{2_i}(\boldsymbol{\theta})}{\partial \psi}, \;\; \psi = \alpha, \beta, \lambda.$$

A segunda derivada é dada por

$$\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial\gamma\partial\psi} = \frac{2\partial^2 A_t(\alpha,\beta)}{\partial\gamma\partial\psi} + \frac{\partial^2 \ell_{1_i}(\boldsymbol{\theta})}{\partial\gamma\partial\psi} + W_\Phi(\ell_{2_i}(\boldsymbol{\theta}))\frac{\partial^2 \ell_{2_i}(\boldsymbol{\theta})}{\partial\gamma\partial\psi} + W_\Phi^{(1)}(\ell_{2_i}(\boldsymbol{\theta}))\frac{\partial \ell_{2_i}(\boldsymbol{\theta})}{\partial\gamma}\frac{\partial \ell_{2_i}(\boldsymbol{\theta})}{\partial\psi},$$

em que $W_\Phi^{(1)}(x) = -W_\Phi(x)(x + W_\Phi(x))$ é a derivada de $W_\Phi(x)$.

Portanto, a matriz de informação observada para $\boldsymbol{\theta}$ pode ser escrita como

$$I(\boldsymbol{\theta}) = \begin{pmatrix} I_{\alpha\alpha} & & \\ I_{\alpha\beta} & I_{\beta\beta} & \\ I_{\alpha\lambda} & I_{\beta\lambda} & I_{\lambda\lambda} \end{pmatrix},$$

em que

$$\begin{aligned}
I_{\alpha\alpha} &= \frac{n}{\alpha^2} + \frac{6n}{\alpha^4} - \frac{3}{\alpha^4\beta}\sum_{i=1}^n t_i - \frac{3\beta}{\alpha^4}\sum_{i=1}^n \frac{1}{t_i} + \frac{2\lambda}{\alpha^3}\sum_{i=1}^n \left(\sqrt{\frac{t_i}{\beta}} - \sqrt{\frac{\beta}{t_i}}\right) W_\Phi(\lambda a_{t_i}(\alpha,\beta)) \\
&+ \frac{\lambda^2}{\alpha^4}\sum_{i=1}^n \left(\frac{t_i}{\beta} + \frac{\beta}{t_i} - 2\right) W_\Phi^{(1)}(\lambda a_{t_i}(\alpha,\beta)), \\
I_{\alpha\beta} &= -\frac{1}{\alpha^3\beta^2}\sum_{i=1}^n t_i + \frac{1}{\alpha^3}\sum_{i=1}^n \frac{1}{t_i} + \frac{\lambda}{2\alpha^2}\sum_{i=1}^n \left(\frac{t_i^{1/2}}{\beta^{3/2}} + \frac{\beta^{-1/2}}{t_i^{1/2}}\right) W_\Phi(\lambda a_{t_i}) \\
&+ \frac{\lambda^2}{2\alpha^3}\sum_{i=1}^n \left(\frac{t_i}{\beta^2} - \frac{1}{t_i}\right) W_\Phi^{(1)}(\lambda a_{t_i}), \\
I_{\beta\beta} &= \frac{n}{2\beta^2} - \sum_{i=1}^n \frac{1}{(t_i+\beta)^2} - \frac{1}{\alpha^2\beta^3}\sum_{i=1}^n t_i + \frac{\lambda}{4\alpha}\sum_{i=1}^n \left(\frac{3t_i^{1/2}}{\beta^{5/2}} + \frac{t_i^{-1/2}}{\beta^{3/2}}\right) W_\Phi(\lambda a_{t_i}(\alpha,\beta)) \\
&+ \frac{\lambda^2}{4\alpha^2}\sum_{i=1}^n \left(\frac{t_i}{\beta^3} + \frac{2}{\beta^2} + \frac{1}{\beta t_i}\right) W_\Phi^{(1)}(\lambda a_{t_i}), \\
I_{\alpha\lambda} &= -\frac{1}{\alpha^2}\sum_{i=1}^n \left(\sqrt{\frac{t_i}{\beta}} - \sqrt{\frac{\beta}{t_i}}\right) W_\Phi(\lambda a_{t_i}) - \frac{\lambda}{\alpha^3}\sum_{i=1}^n \left(\frac{t_i}{\beta} + \frac{\beta}{t_i} - 2\right) W_\Phi^{(1)}(\lambda a_{t_i}), \\
I_{\beta\lambda} &= -\frac{1}{2\alpha}\sum_{i=1}^n \left(\frac{t_i^{1/2}}{\beta^{3/2}} + \frac{\beta^{-1/2}}{t_i^{1/2}}\right) W_\Phi(\lambda a_{t_i}) + \frac{\lambda}{2\alpha^2}\sum_{i=1}^n \left(\frac{1}{t_i} - \frac{t_i}{\beta^2}\right) W_\Phi^{(1)}(\lambda a_{t_i}), \\
I_{\lambda\lambda} &= \frac{1}{\alpha^2}\sum_{i=1}^n \left(\frac{t_i}{\beta} + \frac{\beta}{t_i} - 2\right) W_\Phi^{(1)}(\lambda a_{t_i}).
\end{aligned}$$

3.5 Estudo de Simulação

Para demonstrar a robustez do algoritmo EM no sentido de estimação para a distribuição Binrbaum-Saunders skew-normal, um estudo de simulação foi conduzido. Para tanto, foi considerado o tamanho de amostra $n = 300$ e número de replicações $M = 5000$. Os parâmetros considerados foram $\boldsymbol{\theta}^\top = (\alpha, \beta, \lambda) = (0.5, 1, 1.5)$. Na tabela (3.3) apresentamos a média das M estimativas ($\bar{\widehat{\boldsymbol{\theta}}} = \sum_{k=1}^{M} \widehat{\boldsymbol{\theta}_k}/M$) nas M replicações, juntamente com a média dos erros padrões (EP) assintóticos obtido através da matriz de informação observada (MI) (Seção 3.4), dado por

$$EP(\widehat{\boldsymbol{\theta}}) = \left[Diag(I^{-1}(\widehat{\boldsymbol{\theta}}))\right]^{1/2}$$

e também o erro padrão empírico das M estimativas, dado por

$$EP_{emp} = \sum_{k=1}^{M} (\widehat{\boldsymbol{\theta}_k} - \bar{\widehat{\boldsymbol{\theta}}})^2/M.$$

De acordo com a Tabela 3.3, percebe-se que o algoritmo EM se comporta bem pois as estimativas dos parâmetros se aproximam dos valores simulados e os erros padrões, tanto o empírico como o teórico, são muito similares, o que demonstra a eficiência da matriz de informação observada. O critério de parada adotado neste trabalho é $||\boldsymbol{\theta}^{r+1} - \boldsymbol{\theta}^{r}|| < 10^{-6}$.

Tabela 3.3: Distribuição Birnbaum-Saunders Skew-Normal: média das M estimativas de $\widehat{\boldsymbol{\theta}}$, erro padrão assintótico e erro padrão empírico das M estimativas.

	$\bar{\widehat{\boldsymbol{\theta}}}$	$EP(\widehat{\boldsymbol{\theta}})$	EP_{emp}
α	0.499	0.054	0.057
β	1.005	0.088	0.097
λ	1.534	0.539	0.579

A seguir apresentamos um outro estudo de simulação para analisar o desempenho dos estimadores de máxima verossimilhança dos parâmetros da distribuição Birnbaum-Saunders Skew-Normal. Tal análise foi feita através de simulações de Monte Carlo considerando diferentes tamanhos amostrais. Todo o processo de simulação de Monte Carlo foi realizado utilizando o pacote estatístico **R** em sua versão 2.8.1.

Os números pseudo-aleatórios da distribuição Birnbaum-Saunders skew-normal podem ser obtidos usando 3.6. Os tamanhos amostrais considerados foram $n = 10, 20, 40$ e 60, os valores considerados para o parâmetro α sendo $\alpha = 0.1, 0.25, 0.5, 0.75$ e 1.0. O parâmetro de escala β foi fixado em 1.0, isto é $\beta = 1$ em todos os experimentos de Monte Carlo. Já os valores considerados para o parâmetro λ sendo $\lambda = 0.2, 0.5, 1.0$ e 1.5. Consideramos $R = 5000$ (número de réplicas de Monte Carlo).

Para cada réplica de Monte Carlo geramos uma amostra aleatória da distribuição Birnbaum-Saunders Skew-Normal $\mathbf{t} = (t_1, t_2, \ldots, t_n)^\top$, onde $t_i \sim \text{BSSN}(\alpha, \beta, \lambda)$, $i = 1, \ldots, n$. As estimativas de máxima verossimilhança dos parámetros da distribuição são obtidas usando o algoritmo EM. Os valores iniciais para α e β foram obtidos considerando os estimadores do método dos momentos modificados, para mais detalhes veja Ng et al. (2003), dados por

$$\widehat{\alpha} = \left[2\left(\frac{S}{R}\right)^{1/2} - 1\right]^{1/2} \quad \text{e} \quad \widehat{\beta} = (SR)^{1/2},$$

em que $S = \frac{1}{n}\sum_{i=1}^{n} Y_i$ e $R = \left(\frac{1}{n}\sum_{i=1}^{n} Y_i^{-1}\right)^{-1}$.

Para análise de resultados da estimação pontual foram calculados para cada tamanho da amostra: a média, o viés relativo (o viés relativo de um estimador $\widehat{\theta}$ de um parâmetros escalar é definido como $\{\mathbb{E}(\widehat{\theta}) - \theta\}/\theta$ e o viés relativo estimado é obtido estimando $\mathbb{E}(\widehat{\theta})$ via Monte Carlo) e a raiz quadrada do erro quadrático médio ($\sqrt{\text{EQM}}$) das 5000 estimativas.

Nas Tabelas **A.1-A.4**, apresentadas no Apêndice **A**, encontram-se as estatísticas obtidas das simulações para os estimadores dos parâmetros α, β e λ. Podemos observar que as estimativas do viés relativo e da raiz do erro quadrático médio de todas as estimativas dos parâmetros, diminuem à medida que aumentamos o tamanho da amostra, como era esperado. Nota-se também que, em módulo, as estimativas dos vieses relativos de λ são consideravelmente maiores que os viéses relativos de α e β.

3.6 Aplicação

Estudo de contaminação atmosférica de Nova York

Para esta aplicação consideramos o conjunto de dados considerado em Nadarajah (2007) e Leiva et al. (2008), sobre medidas diárias do nível de ozônio da cidade de Nova York, no período de Maio à Setembro de 1973. Os dados correspondentes às medidas são apresentados na Tabela 3.4.

Tabela 3.4: Medidas diárias do nível de ozônio em Nova York, Maio–Setembro 1973.

41	36	12	18	28	23	19	8	7	16	11	14	18	14	34	6	30
1	11	4	32	23	45	115	37	29	71	39	23	21	37	20	12	13
49	32	64	40	77	97	97	85	10	27	7	48	35	61	79	63	16
108	20	52	82	50	64	59	39	9	16	78	35	66	122	89	110	44
65	22	59	23	31	44	21	9	45	168	73	76	118	84	85	96	78
91	47	32	20	23	21	24	44	21	28	9	13	46	18	13	24	16
23	36	7	14	30	14	18	20	11	135	80	28	73	13			

A média da concentração de poluentes atmosféricos tem sido utilizada como um indicador do nível de poluição atmosférica e seus efeitos sobre os seres humanos são doenças como a bronquite crônica. A distribuição desta concentração é positiva e assimétrica. Para uma completa descrição de várias distribuições aplicadas a dados de concentração de poluentes atmosféricos veja Gokhale e Khare (2004).

Para dados de concentração de poluição do ar, é usualmente assumido que são não correlacionados e independentes. Utilizamos a distribuição Birnbaum-Saunders e a Birnbaum-Saunders Skew-Normal para modelar o nível de ozônio da cidade de Nova York. O histograma da variárel de interesse é apresentado na Figura 3.7. As estatísticas descritivas dos dados são apresentadas na Tabela 3.5.

Tabela 3.5: Estatísticas descritivas para os dados de poluição atmosférica.

Média	Mediana	Mínimo	Máximo	DP	C. Variação	Assimetria	Curtose
42.13	31.50	1.00	168.00	32.99	78.30	1.21	1.11

Para estimação dos parâmetros α, β e λ, utilizamos o algoritmo EM. Como valores iniciais para α e β usamos as estimativas de máxima verossimilhança da distribuição Birnbaum-

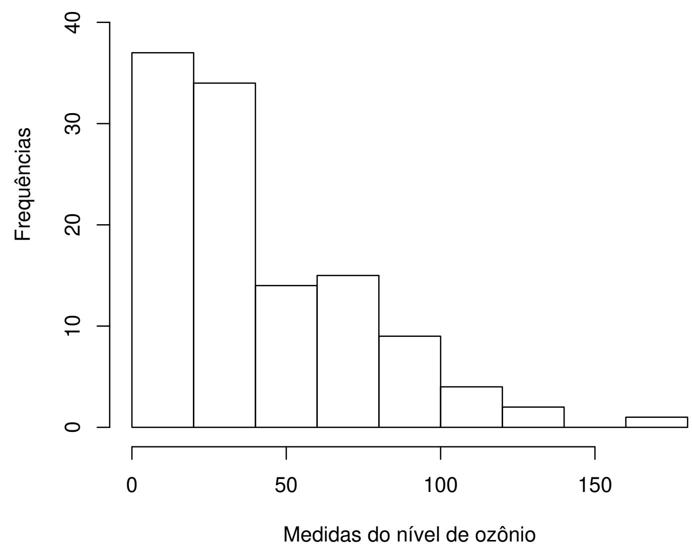


Figura 3.7: Histograma da variável nível de ozônio.

Saunders clássica obtidas através da função *est1bs* do pacote incluso no **R** chamado **bs**, veja Leiva et al. (2006). O valor inicial para λ foi pré-fixado em $\lambda = 0.0$. As estimativas dos parâmetros, os valores das funções de log-verossimilhanças e os critérios de seleção de Akaike (AIC), Schwartz (SIC) e Hannan-Queen (HQC) são apresentados na Tabela 3.6. De acordo com os critérios AIC, SIC e HQC, a distribuição BSSN apresenta um melhor ajuste em relação à distribuição BS.

Realizando um teste assintótico (teste da razão de verossimilhança) para testar $H_0 : \lambda = 0$ *vs* $H_1 : \lambda \neq 0$ (comparando os modelos BSSN e BS), tem-se que $RV = 2(\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_0)) = 6.982$, com um p-valor igual a 0.008, levando a rejeição da hipótese nula. Ou seja, a distribuição BSSN é mais apropriada para ajustar os dados. Além disso, aplicamos a aproximação do fator de Bayes mediante o SIC (veja Tabela 3.7), confirmando o resultado do teste da razão de verossimilhanças.

Como temos n observações provenientes de uma distribuição $F(t)$, podemos construir

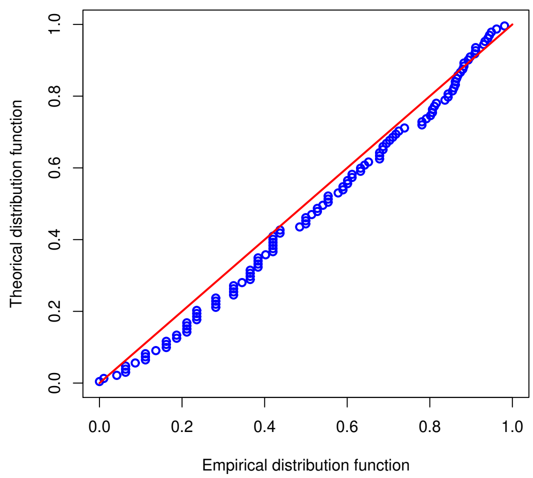
Tabela 3.6: EMV e critérios AIC, SIC e HQC das distribuições BSSN e BS para os dados de contaminação atmosférica.

	BSSN	**BS**
Parâmetros	Estimativa	Estimativa
α	1.270	0.982
β	14.835	28.023
λ	1.066	–
$-\ell(\widehat{\boldsymbol{\theta}})$	545.605	549.097
AIC	4.729	4.750
SIC	4.764	4.774
HQC	4.743	4.760

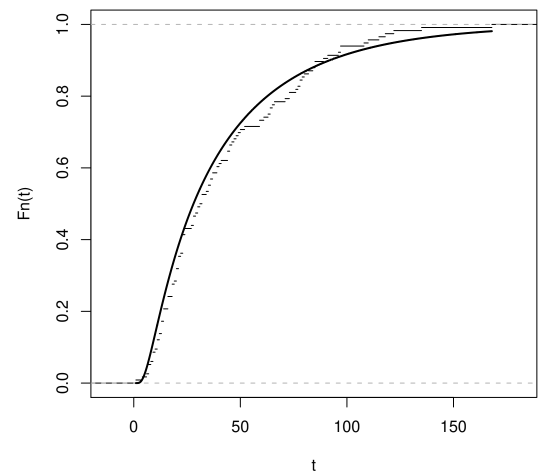
Tabela 3.7: Fator de Bayes (aproximação mediante o SIC) para os dados de contaminação atmosférica.

H_1	H_2	$2\log(B_{12})$	Evidência a favor de H_1
BSSN	**BS**	2.229	Positiva

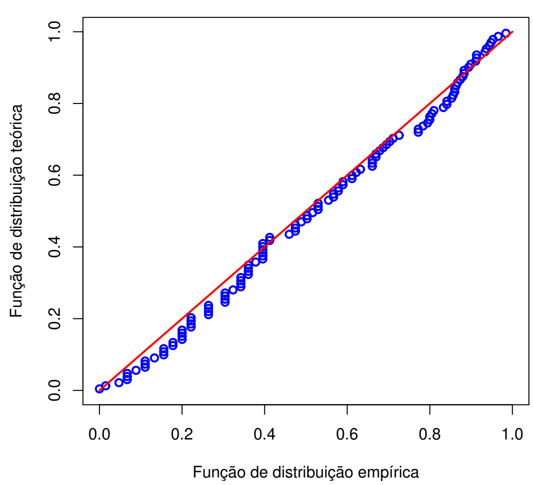
o gráfico das probabilidades empíricas versus as probabilidades teóricas dessas observações. Comparando os gráficos Q-Q (veja Figura 3.8 (a) e (c)) e os gráficos da função de distribuição acumulada empírica das distribuições BSSN e BS (veja Figura 3.8 (b) e (d)), observamos que a distribuição BSSN é mais adequada para ajustar os dados de poluição atmosférica.



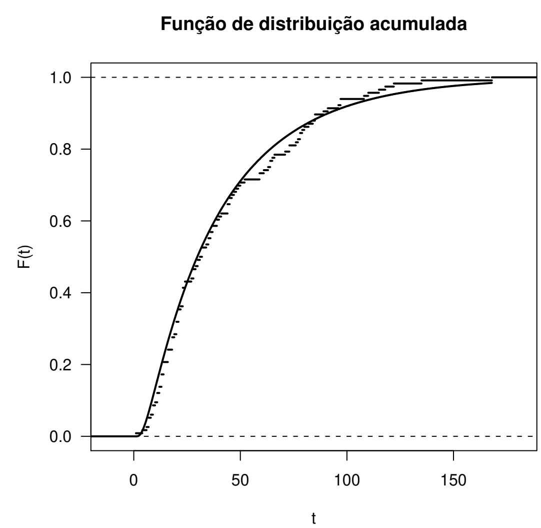
(a)



(b)



(c)



(d)

Figura 3.8: (a) Gráfico Q-Q usando a distribuição BS, (b)Gráfico da função distribuição estimada usando a distribuição BS, (c) Gráfico Q-Q usando a distribuição BSSN e (d) Gráfico da função distribuição estimada usando a distribuição BSSN.

Substituindo as EMV de $\boldsymbol{\theta}$ (de acordo com cada distribuição – BSSN e BS) na representação dada na equação (3.8), construímos os gráficos de envelope simulado apresentados na Figura 3.9 (linhas representam o percentil-5, a média e o percentil-95 de 200 pontos simulados para cada observação). Os gráficos fornecem evidência, avaliado junto aos critérios mostrados anteriormente, de que o modelo BSSN proporciona melhor ajuste aos dados.

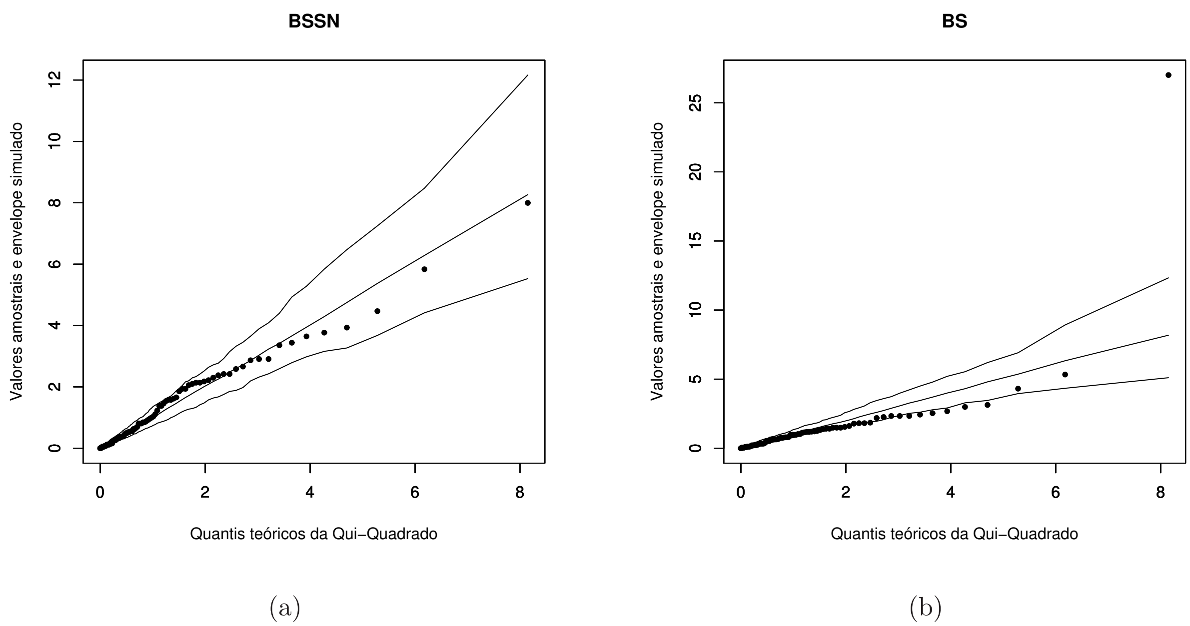


Figura 3.9: Envelope simulado para os dados de fadiga. (a) Modelo BSSN e (b) Modelo BS.

CAPÍTULO 4

Modelo de Regressão Log-Birnbaum-Saunders Skew-Normal

A distribuição Birnbaum-Saunders surge no contexto de engenharia de materiais, mas a sua aplicabilidade foi extendida a vários campos da engenharia para modelar tempos de falha. Nestas áreas, geralmente encontramos covariáveis que normalmente explicam adequadamente o tempo de falha. Diferentemente de outras áreas de aplicação estatística, na engenharia a relação funcional entre o tempo de falha e as covariáveis (tais como temperatura, umidade, velocidade, funcionamento, tensão, esforço, etc.) é não-linear. Este problema tem sido resolvido por uma família de modelos de regressão, conhecida como "modelos de vida acelerada", onde as covariáveis são conhecidas como fatores de stress, veja por exemplo Bagdonavicius e Nikulin (2002) e Nelson (2004) para mais detalhes. Estes modelos de vida acelerada (MVA) usam argumentos teóricos que vêm da Física e Química para justificar a sua utilização. Uma proposta foi recentemente formulada por Leiva et al. (2009a), em que os MVA podem ser visualizados como uma classe geral de modelos log-lineares, dado na equação (4.1) (para mais informações sobre modelos log-lineares, veja por exemplo, Christensen 1997). Esta classe de modelos contém os casos mais importantes do MVA conhecido na literatura. Assim, o parâmetro de interesse θ pode ser modelado através de um conjunto de covariáveis que chamamos $\mathbf{X}$ da seguinte forma:

$$\theta(\mathbf{X}) = \exp\left(\beta_0 + \beta_1 h(\mathbf{X})\right); \quad \beta_0 > 0, \quad \beta_1 > 0, \quad \mathbf{X} \in A \subseteq \mathbb{R}, \tag{4.1}$$

em que

$$g(\mathbf{X}) = \begin{cases} 1/\mathbf{X}, & \text{modelo Arrenhius,} \\ \mathbf{X}, & \text{modelo exponencial ou log-linear,} \\ 1/\mathbf{X} + \log(\mathbf{X}), & \text{modelo Eyring,} \\ \log(\mathbf{X}), & \text{modelo } \textit{power law}. \end{cases}$$

Alguns resultados considerando as expressões dadas na equação (4.1) foram considerados para a distribuição BS, veja por exemplo, Owen e Padgett (1999) e Leiva et al. (2009b). Nos últimos tempos, a distribuição BS tem sido aplicada em áreas fora do contexto de fadiga dos materiais e engenharia. Em vista que a distribuição BS tem a propriedade de descrever processos de degradação acumulativa, o que tem permitido a sua aplicação em outras áreas, como por exemplo, em ciências da saúde, ambiental, florestal, demográficas, atuarial, financeira, entre outras, para a justificativa teórica dos resultados e para mais detalhes sobre estas propostas, veja Leiva et al. (2009b). Todos estes aspectos mencionados acima têm permitido considerar a distribuição BS como um modelo de probabilidade, em vez de um modelo considerado apenas para descrever tempos de vida, conhecido na literatura como "distribuição de vida", veja Marshall e Olkin (2007). Assim, a família de modelos fornecidos na Equação (4.1) pode ser encarada como uma classe geral de modelos estatísticos não lineares fora do contexto da engenharia e de modelos de vida acelerada.

Como em Leiva et al. (2009a, b), o modelo BS(α, β) na presença de covariáveis pode ser definido assumindo que o parâmetro β depende de covariáveis. Por exemplo, podemos assumir que β pode ser escrito como

$$\beta = w(\boldsymbol{x}^\top \boldsymbol{\gamma}),$$

em que $\boldsymbol{x}^\top = (x_1, \ldots, x_p)$ é um vetor $p \times 1$ de variáveis explicativas e $\boldsymbol{\gamma} : p \times 1$ é um vetor de coeficientes de regressão. Em geral, na literatura é considerado que $w(\boldsymbol{x}^\top \boldsymbol{\gamma}) = \exp(\boldsymbol{x}^\top \boldsymbol{\gamma})$, veja por exemplo Rieck e Nedelman (1991), que consideraram $Y = \log(T)$, em que $T = \exp(\boldsymbol{x}^\top \boldsymbol{\beta}) T_0$, com $T_0 \sim BS(\alpha, 1)$.

A distribuição Senh-Normal proposta por Rieck e Nedelman (1991) permitiu o desenvolvimento dos modelos de regressão linear log-Birnbaum-Saunders, definido pela relação

$$y_j = \boldsymbol{x}_j^\top \boldsymbol{\beta} + \epsilon_j, \tag{4.2}$$

onde y_j é o logaritmo do tempo de sobrevivência ou do tempo de censura para a i-ésima unidade experimental, os vetores $\boldsymbol{x}_j^\top$, $j = 1, ..., n$ são conhecidos e são as linhas da matriz $\boldsymbol{X} : n \times p$ e $\boldsymbol{\beta}$ é um vetor de parâmetros de dimensão p. Rieck e Nedelman (1991) assumem que os erros ϵ_j's são independentes com distribuição comum $\epsilon_j \sim SHN(\alpha, 0, 2)$. Neste caso $\epsilon_j = log(T_j)$, onde T_j tem uma distribuição Birnbaum-Saunders $BS(\alpha, 1)$. Assim, Y_j segue uma distribuição Senh-Normal $SHN(\alpha, \boldsymbol{x}_j^\top \boldsymbol{\beta}, \sigma = 2)$. Recentemente Galea, Leiva e Paula (2004) apresentaram um estudo de influência local e diagnóstico considerando que as observações são provenientes de tempos de vida.

Neste capítulo vamos considerar o modelo BSSN quando o parâmetro de escala depende de covariáveis. Especificamente vamos estudar propriedades da distribuição de $Y = \log(T)$, quando $T \sim \text{BSSN}(\alpha, \beta, \lambda)$, esta distribuição será chamada de Senh-Skew-Normal (SHSN), ou log-BSSN. Além disso, apresentaremos um estudo de influência local no contexto de um modelo de regressão linear log-BSSN definido pela relação

$$y_j = \mathbf{x}_j^\top \beta + \epsilon_j, \tag{4.3}$$

quando $\epsilon_j \sim \text{SHSN}(\alpha, 0, \sigma = 2, \lambda)$, $j = 1, \ldots, n$.

4.1 A distribuição Senh-Skew-Normal

Nesta seção vamos apresentar uma generalização da distribuição Senh-Normal (SHN), desenvolvida por Rieck e Nedelman (1991), que é baseada na distribuição de $Y = log(T)$, com $T \sim \text{BSSN}(\alpha, \beta, \lambda)$. Esta distribuição permite aplicações no modelo de regressão linear log-Birnbaum-Saunders skew-normal. Alternativamente esta distribuição pode ser definida através de uma transformação da distribuição skew-normal baseada na relação

$$Y = \gamma + 2arcsenh\left(\frac{\alpha Z}{2}\right), \tag{4.4}$$

em que $Z \sim SN(0,1,\lambda)$, $\alpha > 0$ é o parâmetro de forma e $\gamma \in R$ é o parâmetro de locação. Ou seja, em lugar de assumir normalidade (caso senh-normal), vamos assumir que $Z \sim SN(0,1,\lambda)$. A distribuição de Y será denotada por $Y \sim SHSN(\alpha,\gamma,\sigma=2;\lambda)$.

Definição 4.1. *Dizemos que uma variável aleatória Y tem distribuição senh-skew-normal (SHSN) com parâmetro de forma $\alpha > 0$, parâmetro de locação $\gamma \in \mathbb{R}$, parâmetro de escala $\sigma = 2$ e parâmetro de assimetria $\lambda \in \mathbb{R}$, se sua função densidade de probabilidade for dada por*

$$
\begin{aligned}
f_Y(y) &= \left(\frac{2}{\alpha\sqrt{2\pi}}\right)\exp\left[-2\alpha^{-2}\text{senh}^{\,2}\left(\frac{y-\gamma}{2}\right)\right]\cosh\left(\frac{y-\gamma}{2}\right)\Phi(\lambda b_y), \\
&= 2\phi(b_y)b_y^{'}\Phi(\lambda b_y),\ y \in \mathbb{R},
\end{aligned}
\tag{4.5}
$$

em que b_y e $b_y^{'}$ são como em (2.9).

Na Figura 4.1 apresentamos os gráficos da densidade (4.5) para alguns valores de λ positivos e negativos, considerando $\gamma = 0$, (a)-(b) $\alpha = 0.5$ e (c)-(d) $\alpha = 1$. Observe que a distribuição Senh-Skew-Normal é assimétrica e o parâmetro λ caracteriza a forma da distribuição, pois valores negativos de λ indicam assimetria negativa e valores positivos de λ indicam assimetria positiva, como no caso da distribuição skew-normal. Se $\lambda = 0$, a densidade (4.5) coincide com a densidade da distribuição senh-normal e portanto é simétrica. Na Figura 4.2 apresentamos os gráficos da densidade (4.5) para alguns valores de α considerando $\beta = 1.0$, $\lambda = 1.0$ e $\lambda = -1.0$. Observe que, para o valor de $\lambda = 1$ a curva da densidade (4.5) é assimétrica positiva e para $\lambda = -1$ assimétrica negativa.

De (4.4) temos que se $Y \sim SHSN(\alpha,\gamma,\sigma=2,\lambda)$, então

$$
Z = \frac{2}{\alpha}\text{senh}\left[\frac{y-\gamma}{2}\right] \sim SN(0,1,\lambda). \tag{4.6}
$$

Observe que a distribuição de $Y = log(T) \sim SHSN(\alpha,\gamma,\sigma=2,\lambda)$ com $\gamma = log(\beta)$.

Proposição 4.1. *Seja $Y \sim$ SHSN$(\alpha,\gamma,\sigma=2,\lambda)$. Então*

a) A função de distribuição acumulada da Senh-Skew-Normal é dada por

$$
F_Y(y) = \Phi_\lambda\left[\frac{2}{\alpha}\text{senh}\left(\frac{y-\gamma}{2}\right)\right] = \Phi_\lambda(b_y),
$$

com b_y como em (2.9) e $\Phi_\lambda(.)$ é uma fda de uma variável aleatória SN$(0,1,\lambda)$. *Em particular, se $\lambda = 1$, então $F_Y(y) = \left[\Phi\left(\frac{2}{\alpha}\text{senh}\left(\frac{y-\gamma}{2}\right)\right)\right]^2$.*

b) Temos que

$$D = \frac{4}{\alpha^2}\text{senh}^{\,2}\left(\frac{Y-\gamma}{2}\right) \sim \chi_1^2, \tag{4.7}$$

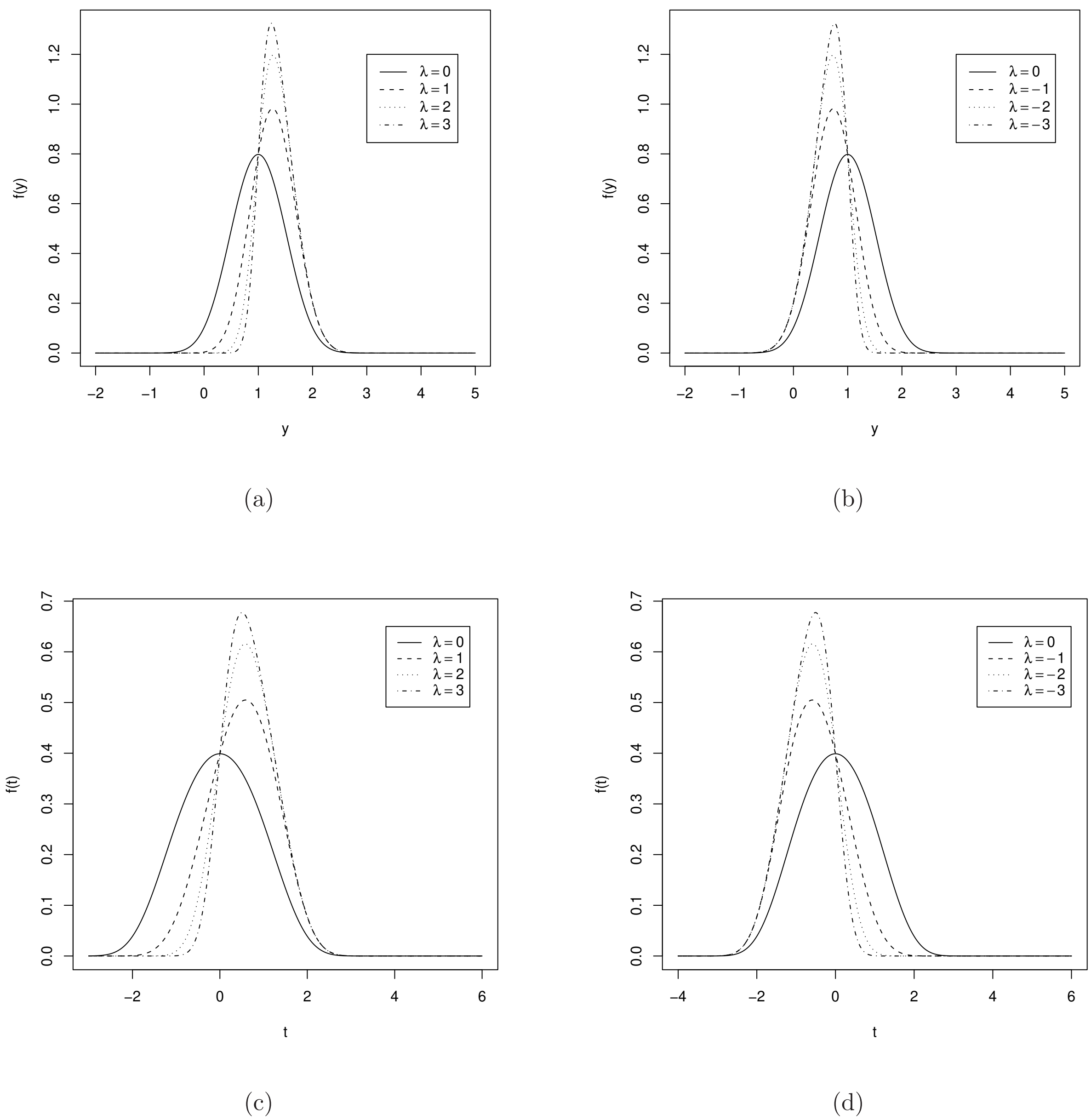
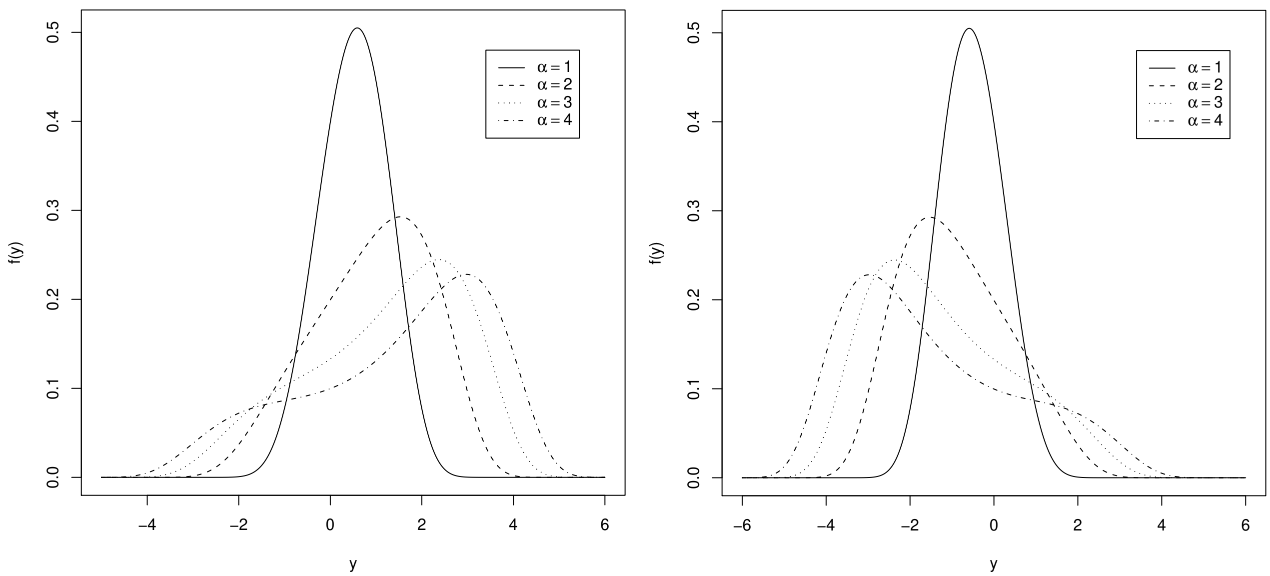


Figura 4.1: Densidades da senh-skew-normal para diferentes valores de λ, considerando $\gamma = 0$, (a)-(b) $\alpha = 0.5$ e (c)-(d) $\alpha = 1$.



(a) $\lambda = 1.0$ (b) $\lambda = -1.0$

Figura 4.2: Densidades da senh-skew-normal para diferentes valores de α, considerando $\gamma = 0.0$, (a) $\lambda = 1.0$ e (b) $\lambda = -1.0$.

onde χ^2_1 denota a distribuição Qui-Quadrado com 1 grau de liberdade. Esse resultado permite checar o modelo na prática através do gráfico de envelope simulado.

c) A função geradora de momentos da Senh-Skew-Normal é dada por

$$\begin{aligned} M_Y(s) &= \mathbb{E}[e^{sY}] \\ &= \exp(\gamma s)\left[\frac{K_a(\alpha^{-2};\lambda) + K_b(\alpha^{-2};\lambda)}{K_{\frac{1}{2}}(\alpha^{-2})}\right], \quad s \in \mathbb{R}, \end{aligned} \tag{4.8}$$

em que $K_\nu(\omega;\lambda) = (1/2)\int_0^\infty \exp(-\omega\cosh(u) - \nu u)\Phi\left(\frac{2\lambda}{\alpha}\text{senh}\left(\frac{u}{2}\right)\right)du$, $a = (2s+1)/2$, $b = (2s-1)/2$ e $K_\nu(\omega, \lambda = 0) = (1/2)K_\nu(\omega)$, com $K_\nu(.)$ é a função de Bessel modificada do terceiro tipo. Para mais detalhes veja Gradshteyn e Randzhik (2000), p. 907.

A demonstração da função geradora de momentos encontra-se no Apêndice **B**. A média e a variância da distribuição Senh-Skew-Normal (SHSN) podem ser obtidas usando esta função geradora de momentos. Quando $\lambda = 0$ obtêm-se a função geratriz de momentos da distribuição Senh-Normal, dada em (2.11).

Proposição 4.2. *Seja Z como em (4.6) e $Y \sim SHSN(\alpha, \gamma, \sigma = 2, \lambda)$. Assuma que Z tem representação estocástica como em (3.2) dada por $Z = \delta H + (1-\delta^2)^{1/2} X_1$. Então a função de densidade condicional de Y dado $H = h$ é dada por*

$$f_{Y/H}(y/h) = \phi\left(\nu_h + b_y(\alpha_\delta, \gamma)\right) b_y^{'}(\alpha_\delta, \gamma), \tag{4.9}$$

onde $b_y(\alpha_\delta, \gamma)$ e $b_y^{'}(\alpha_\delta, \gamma)$ são como em (2.9), com α_δ no lugar de α.

Prova: Como $Z \sim SN(0, 1, \lambda)$, então usando a representação estocástica dada em (3.2) podemos expressar Z como

$$Z \;=\; \delta H + (1-\delta^2)^{1/2} X_1 = \frac{2}{\alpha}\text{senh}\left(\frac{Y-\gamma}{2}\right).$$

Assim da distribuição condicional de Z dado $H = h$ é dada por

$$Z/(H = h) \;=\; \frac{2}{\alpha}\text{senh}\left(\frac{Y-\gamma}{2}\right) \sim N(\delta h, 1-\delta^2).$$

Logo, dado $H = h$ a distribuição de $W = (Z - \delta h)/\sqrt{1-\delta^2}$ é dada por

$$W \;=\; -\frac{\delta}{\sqrt{1-\delta^2}}h + \frac{2}{\alpha\sqrt{1-\delta^2}}\text{senh}\left(\frac{Y-\gamma}{2}\right) \sim N(0,1) \tag{4.10}$$

Assim, do resultado (4.10), temos a prova. ■

Proposição 4.3. *Seja Z como em (4.6) e $Y \sim$ SHSN$(\alpha, \gamma, \sigma = 2, \lambda)$. Então*

i) $f_{H/Y}(h/y) = \dfrac{1}{\Phi\left(\lambda b_y\right)}\phi\left(h|\delta b_y, (1-\delta^2)\right) I_{[0,\infty)}(h)$,

ii) $\mathbb{E}[H|T = t] = \delta b_t + W_\Phi\left(\dfrac{\delta b_t}{\sqrt{1-\delta^2}}\right)\sqrt{1-\delta^2}$;

iii) $\mathbb{E}[H^2|T = t] = (\delta b_t)^2 + (1-\delta^2) + W_\Phi\left(\dfrac{\delta b_t}{\sqrt{1-\delta^2}}\right)\delta b_t\sqrt{1-\delta^2}$,

em que $W_\Phi\left(\dfrac{\delta b_t}{\sqrt{1-\delta^2}}\right) = \phi\left(\dfrac{\delta b_t}{\sqrt{1-\delta^2}}\right)/\Phi\left(\dfrac{\delta b_t}{\sqrt{1-\delta^2}}\right)$.

Prova:

i) Como $h_H(h) = 2\phi(h/0,1)I_{[0,\infty)}(h)$ e

$$\phi\left(-\frac{\delta}{\sqrt{1-\delta^2}}h + \frac{1}{\sqrt{1-\delta^2}}b_t|0,1\right) = \sqrt{1-\delta^2}\phi\left(b_t|\delta h, 1-\delta^2\right),$$

e usando o Lema 3.1, temos que $\phi(b_t/\delta h, 1-\delta^2)\phi(t|0,1) = \phi(b_t/0,1)\phi(h/\delta b_t,(1-\delta^2))$. Assim a prova segue da relação $f_{H|T}(h|t) = f_{T|H}(t|h)f_H(h)/f_T(t)$.

ii) e iii) Usando a definição de esperança condicional, temos que

$$\begin{aligned}
\mathbb{E}(H/T) &= \int_0^\infty \frac{h}{\Phi(\lambda b_t)}\phi(h|\delta b_t,(1-\delta^2))dh \\
&= \frac{1}{\Phi(\lambda b_t)}\int_0^\infty h\phi(h|\delta b_t,(1-\delta^2))dh.
\end{aligned}$$

Agora usando as propriedades da distribuição half-normal dada no Lema 3.2, segue

$$\mathbb{E}(H/T) = \delta b_t + W_\Phi\left(\frac{\delta b_t}{\sqrt{1-\delta^2}}\right)\sqrt{1-\delta^2}$$

e

$$\mathbb{E}[H^2|Y=y] = (\delta b_t)^2 + \tau^2 + W_\Phi\left(\frac{\delta b_t}{\sqrt{1-\delta^2}}\right)\delta b_t\sqrt{1-\delta^2}.$$

■

Proposição 4.4. *Seja* $Y \sim SHSN(\alpha,\gamma,\sigma=2,\lambda)$ *e* $W = \frac{\theta}{2}Y$, $\theta > 0$. *Então*

$$W \sim SHSN\left(\alpha, \frac{\theta}{2}\gamma, \sigma = 2, \lambda\right).$$

Ou seja, a distribuição de $W = \frac{\theta}{2}Y$ *pode ser associada a uma distribuição Senh-Skew-Normal.*

Prova: Usando o método do Jacobiano e a densidade de Y dada em (4.5), temos que a função densidade de probabilidade de W é dada por

$$\begin{aligned}
f_W(w) &= \frac{2}{\theta}f_Y\left(\frac{2}{\theta}y\right) \\
&= \frac{2}{\theta}\left(\frac{2}{\alpha\sqrt{2\pi}}\right)\exp\left[-2\alpha^{-2}\text{senh}^2\left(\frac{w-\frac{\theta}{2}\gamma}{2}\right)\right]\cosh\left(\frac{w-\frac{\theta}{2}\gamma}{2}\right)\Phi\left(\lambda\frac{2}{\alpha}\text{senh}\left(\frac{w-\frac{\theta}{2}\gamma}{2}\right)\right) \\
&= \left(\frac{4}{\theta\alpha\sqrt{2\pi}}\right)\exp\left[-2\alpha^{-2}\text{senh}^2\left(\frac{w-\frac{\theta}{2}\gamma}{2}\right)\right]\cosh\left(\frac{w-\frac{\theta}{2}\gamma}{2}\right)\Phi\left(\lambda\frac{2}{\alpha}\text{senh}\left(\frac{w-\frac{\theta}{2}\gamma}{2}\right)\right) \\
&= \left(\frac{4}{\theta\alpha\sqrt{2\pi}}\right)\exp\left[-2\alpha^{-2}\text{senh}^2\left(\frac{w-\delta}{2}\right)\right]\cosh\left(\frac{w-\delta}{2}\right)\Phi\left(\lambda\frac{2}{\alpha}\text{senh}\left(\frac{w-\delta}{2}\right)\right).
\end{aligned}$$

Portanto, $W \sim \text{SHSN}(\alpha, \delta, \sigma = 2, \lambda)$, onde $\delta = \dfrac{\theta}{2}\gamma$, $\theta > 0$. ■

Proposição 4.5. *Seja* $Y \sim SHSN(\alpha, \gamma, \sigma, \lambda)$ *e* $W = \mu + Y$, $\mu \in \mathbb{R}$. *Então*

$$W \sim SHSN(\alpha, \mu + \gamma, \sigma = 2, \lambda).$$

Prova: Usando o método do Jacobiano e a densidade de Y dada em (4.5), temos que a função densidade de probabilidade de W é dada por

$$\begin{aligned} f_W(w) &= \left(\frac{2}{\alpha\sqrt{2\pi}}\right) \exp\left[-2\alpha^{-2}\text{senh}^2\left(\frac{w-(\mu+\gamma)}{2}\right)\right] \cosh\left(\frac{w-(\mu+\gamma)}{2}\right) \\ &\quad \Phi\left(\lambda\frac{2}{\alpha}\text{senh}\left(\frac{w-(\mu+\gamma)}{2}\right)\right). \end{aligned}$$

Portanto, $W \sim \text{SHSN}(\alpha, \mu + \gamma, \sigma = 2, \lambda)$, $\mu \in \mathbb{R}$. ■

O resultado na proposição 4.5 acima permite introduzir o modelo de regressão linear que discutiremos na seção seguinte.

4.2 O Modelo de Regressão Log-Birnbaum-Saunders Skew-Normal

Suponha que se tenha um conjunto de n observações independentes, denotadas por $Y_1, \ldots, Y_n$, onde $Y_i \sim \text{SHSN}(\alpha, \mu_i, \sigma = 2, \lambda)$, $i = 1, \ldots, n$. Para definir o modelo log-Birnbaum-Saunders Skew-Normal (log-BSSN) procedemos como segue: associado com a observação i, considere um vetor $p \times 1$ de covariáveis $\boldsymbol{x}_i$, através do qual especifica-se o preditor linear $\mu_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$, onde $\boldsymbol{\beta}$ é um vetor p-dimensional de coeficientes de regressão desconhecidos. Assim, usando a Proposição 4.5 e relacionando os dois conjuntos de variáveis, tem-se o modelo

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad \epsilon_i \sim \text{SHSN}(\alpha, 0, \sigma = 2, \lambda), \quad i = 1, \ldots, n. \tag{4.11}$$

Neste caso $\epsilon_i = log(T_i)$, onde T_i tem uma distribuição Birnbaum-Saunders skew-normal $\text{BSSN}(\alpha, 1, \lambda)$. Assim, a função log-verossimilhança para $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^\top, \lambda)$ pode ser escrita como

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log\left[2\phi(b_{y_i}) b'_{y_i} \Phi(\lambda b_{y_i})\right], \tag{4.12}$$

em que b_{y_i} e b'_{y_i} são dados em (2.9).

Maximizar diretamente a função log-verossimilhança, dada em (4.12), para encontrar as estimativas de máxima verossimilhança dos parâmetros pode ser complicado, pois as equações resultantes não têm forma fechada. Embora existam muitas metodologias para encontrar os EMV, iremos obter os EMV de $\boldsymbol{\theta}$ via algoritmo EM, o qual é fácil de implementar e computacionalmente conveniente.

Em seguida é proposto um algoritmo EM para obter a EMV do vetor de parâmetros $\boldsymbol{\theta}$. Mais ainda, estudos relacionados com influência local para dados incompletos (Zhu e Lee, 2001) podem ser facilmente estendidos. A metodologia de estimação e análise de diagnóstico para a classe de modelos SHSN não existe na literatura.

4.3 Estimação de Máxima Verossimilhança via Algoritmo EM

Para efeito de estimação vamos considerar uma reparametrização em lugar de estimar $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^\top, \lambda)^\top$, vamos estimar $\boldsymbol{\theta}$ dado por $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^\top, \delta)^\top$, onde $\delta = \dfrac{\lambda}{\sqrt{1+\lambda^2}}$.

Para implementar o algoritmo EM, primeiro note que, utilizando a representação estocástica da distribuição Skew-Normal, dada em (3.2) e a Proposição (4.2), o modelo de regressão (4.11) acima pode ser escrito como

$$\begin{aligned} Y_i|H = h_i &\overset{ind}{\sim} \mathrm{SHN}\left(\alpha_\delta, \mu_i, \sigma = 2, \nu_{h_i}\right), \\ H_i &\overset{ind}{\sim} HN(0,1), \;\; i = 1, \ldots, n \end{aligned}$$

onde $\alpha_\delta = (1-\delta^2)^{1/2}\alpha$, $\nu_{h_i} = -\dfrac{\delta h_i}{(1-\delta^2)^{1/2}}$ e $HN(0,1)$ denota a distribuição half-normal padrão univariada (Johnson et al., 1995).

A distribuição conjunta de y_i e h_i é dada por

$$f(y_i, h_i) = 2\phi(b^*_{y_i}) b^{*\prime}_{y_i} \phi(h_i) \mathbf{I}(h_i > 0),$$

em que

$$b^*_{y_i} = \nu_{h_i} + \frac{2}{\alpha_\delta} \mathrm{senh}\left(\frac{y_i - \mu_i}{2}\right) \quad \text{e} \quad b^{*\prime}_{y_i} = \frac{1}{\alpha_\delta} \cosh\left(\frac{y_i - \mu_i}{2}\right).$$

Sejam $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ e $\boldsymbol{h} = (h_1, \ldots, h_n)^\top$. Então tratando $\mathbf{h}$ como dados faltantes, segue que a função log-verossimilhança completa associada com $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{h}^\top)^\top$ é dada por

$$\begin{aligned} \ell_c(\boldsymbol{\theta}|\mathbf{y_c}) &= \sum_{i=1}^{n} \log f(y_i, h_i) \\ &\propto -n\log(\alpha) - \frac{n}{2}\log(1-\delta^2) + \sum_{i=1}^{n}\log\left[\cosh\left(\frac{y_i-\mu_i}{2}\right)\right] \\ &\quad -\frac{1}{2(1-\delta^2)}\sum_{i=1}^{n}(b_{y_i}-\delta h_i)^2. \end{aligned} \tag{4.13}$$

Fazendo $\widehat{h}_i = E[H_i|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, y_i]$ e $\widehat{h_i^2} = E[H_i^2|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, y_i]$ e aplicando as propriedades de esperança condicional de H dado $Y = y$, que foram obtidas na Proposição 4.3 e usando os momentos da distribuição half-normal (Lachos, 2004) tem-se que

$$\widehat{h}_i = \delta b_{t_i} + W_\Phi\left(\frac{\delta b_{t_i}}{\sqrt{1-\delta^2}}\right)\sqrt{1-\delta^2} \tag{4.14}$$

e

$$\widehat{h_i^2} = (\delta b_{t_i})^2 + \tau^2 + W_\Phi\left(\frac{\delta b_{t_i}}{\sqrt{1-\delta^2}}\right)\delta b_{t_i}\sqrt{1-\delta^2}. \tag{4.15}$$

Denote por $\boldsymbol{\theta}^{(r)} = (\alpha^{(r)}, {\boldsymbol{\beta}^{(r)}}^\top, \delta^{(r)})^\top$ a estimativa de $\boldsymbol{\theta}$ para a r-ésima iteração. Seque que a esperança da função de log-verossimilhança completa, que chamamos de função Q será dada por

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) = \mathbb{E}[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(r)}] = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(r)}), \tag{4.16}$$

em que

$$\begin{aligned} Q_i(\boldsymbol{\theta}, |\widehat{\boldsymbol{\theta}}^{(r)}) &= c - \frac{1}{2}\log(1-\delta^{2(r)}) - \log(\alpha^{(r)}) + \log\left[\cosh\left(\frac{y_i - \boldsymbol{x}_i^\top \beta^{(r)}}{2}\right)\right] \\ &- \frac{1}{2(1-\delta^{2(r)})}[b_{y_i}^2 - 2\delta^{(r)}\widehat{h}_i^{(r)} b_{y_i} + \delta^{2(r)}\widehat{h_i^2}^{(r)}], \end{aligned}$$

com $\delta^{(r)} = \dfrac{\lambda^{(r)}}{\sqrt{1-\lambda^{(r)}}}$ e $b_{y_i} = \dfrac{2}{\alpha^{(r)}}\text{senh}\left(\dfrac{y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(r)}}{2}\right)$.

Assim como na distribuição BSSN, estas equações não apresentam soluções analíticas, mas podemos obter as soluções por métodos iterativos. A etapa M do algoritmo EM é implementada maximizando a função de log-verossimilhança completa dada em (4.16). Derivando $Q(\boldsymbol{\theta}, |\widehat{\boldsymbol{\theta}})$ com relação a $\boldsymbol{\theta}$ e após algumas manipulações algébricas, obtemos as seguintes equações de estimação,

$$\widehat{\alpha}^2 + \widehat{\alpha}\frac{\delta}{(1-\delta^2)}\bar{d}_{hy}(\boldsymbol{\beta}) - \frac{1}{(1-\delta^2)}\bar{d}_y^2(\boldsymbol{\beta}) = 0, \tag{4.17}$$

$$\sum_{i=1}^{n}\left[\frac{1}{\alpha^2(1-\delta^2)}\text{senh}\left(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}\right) - \frac{1}{2}tagh\left(\frac{y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}}{2}\right) - \frac{\delta}{\alpha(1-\delta^2)}\widehat{h}_i\cosh\left(\frac{y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}}{2}\right)\right]\boldsymbol{x}_i^\top = 0,$$

$$\widehat{\delta} = \frac{\sum_{i=1}^n \widehat{h}_i d_{y_i}}{\widehat{\alpha}\sum_{i=1}^n \widehat{h_i^2}}, \tag{4.18}$$

em que

$$\bar{d}_{hy}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \widehat{h}_i d_{y_i}, \;\; \bar{d}_y^2(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n d_{y_i}^2 \;\text{ e }\; d_{y_i} = 2\text{senh}\left(\frac{y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}}{2}\right).$$

Assim como na distribuição BSSN, o parâmetro $\boldsymbol{\beta}$ não tem forma fechada então na etapa M do algoritmo EM obtemos a estimativa de máxima verossimilhança (EMV) de $\boldsymbol{\beta}$ através

$$\widehat{\boldsymbol{\beta}} = \arg\max\{Q(\boldsymbol{\beta}, \widehat{\alpha}, \widehat{\lambda}|\widehat{\boldsymbol{\theta}})\}. \tag{4.19}$$

Portanto, se $\boldsymbol{\theta}^{(r)}$ maximiza (4.16) na r-ésima iretação, então na iteração $r+1$ o algoritmo EM procede como segue:

Etapa E: Calcule $\widehat{h}_i(\boldsymbol{\theta}^{(r)})$ e $\widehat{h_i^2}(\boldsymbol{\theta}^{(r)})$, $i = 1, \ldots, n$, dados por (4.14) e (4.15) respectivamente, com $\boldsymbol{\theta} = \boldsymbol{\theta}^{(r)}$.

Etapa M: Calcule $(\boldsymbol{\theta} = \boldsymbol{\theta}^{(r+1)})$ a partir de (4.17), (4.18) e (4.19).

4.4 Matriz de Informação de Fisher Observada

Considere $Y_1, \ldots, Y_n$ observações independentes, onde $Y_i \sim \text{SHSN}(\alpha, \mu_i, \sigma = 2, \lambda)$, $i = 1, \ldots, n$. Então a função log-verossimilhança $\ell(\boldsymbol{\theta})$ é da forma $\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta})$ em que

$$\ell_i(\boldsymbol{\theta}) = \log(2b_{y_i}^{'}) + \ell_{1_i}(\boldsymbol{\theta}) + \log[\Phi(\ell_{2_i}(\boldsymbol{\theta}))], \;\; \boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \alpha, \lambda)^\top,$$

com $\ell_{1_i}(\boldsymbol{\theta}) = \phi(b_{y_i}|0,1)$, $\ell_{2_i}(\boldsymbol{\theta}) = \lambda b_{y_i}$, b_{y_i} e $b^{'}_{y_i}$ são dados em (2.9). Então a primeira derivada de $\ell_i(\boldsymbol{\theta})$ é dada por

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \psi} = \frac{\partial 2A_{t_i}(\alpha,\boldsymbol{\beta})}{\partial \psi} + \frac{\partial \ell_{1_i}(\boldsymbol{\theta})}{\partial \psi} + W_\Phi(\ell_{2_i}(\boldsymbol{\theta}))\frac{\partial \ell_{2_i}(\boldsymbol{\theta})}{\partial \psi}, \;\; \psi = \boldsymbol{\beta}, \alpha, \lambda.$$

A segunda derivada é dada por

$$\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \gamma \partial \psi^\top} = \frac{\partial^2 2A_{t_i}(\alpha,\boldsymbol{\beta})}{\partial \gamma \partial \psi^\top} + \frac{\partial^2 \ell_{1_i}(\boldsymbol{\theta})}{\partial \gamma \partial \psi^\top} + W_\Phi(\ell_{2_i}(\boldsymbol{\theta}))\frac{\partial^2 \ell_{2_i}(\boldsymbol{\theta})}{\partial \gamma \partial \psi^\top} + W_\Phi^{(1)}(\ell_{2_i}(\boldsymbol{\theta}))\frac{\partial \ell_{2_i}(\boldsymbol{\theta})}{\partial \gamma}\frac{\partial \ell_{2_i}(\boldsymbol{\theta})}{\partial \psi^\top},$$

em que $W_\Phi^{(1)}(x) = -W_\Phi(x)(x + W_\Phi(x))$ é a derivada de $W_\Phi(x)$, $x \in \mathbb{R}$.

Portanto, a matriz de informação observada para $\boldsymbol{\theta}$ pode ser escrita como

$$I(\boldsymbol{\theta}) = \begin{pmatrix} I_{\boldsymbol{\beta}\boldsymbol{\beta}} & I_{\alpha\boldsymbol{\beta}} & I_{\lambda\boldsymbol{\beta}} \\ & I_{\alpha\alpha} & I_{\lambda\alpha} \\ & & I_{\lambda\lambda} \end{pmatrix},$$

em que

$$\begin{aligned}
I_{\boldsymbol{\beta}\boldsymbol{\beta}} &= \mathbf{X}^\top\mathbf{V}\mathbf{X} + \frac{\lambda}{2\alpha}\mathbf{X}^\top\mathbf{D}(\mathbf{w1})\mathbf{D}(\mathbf{b})\mathbf{X} - \left(\frac{\lambda}{\alpha}\right)^2\mathbf{X}^\top\mathbf{f}\mathbf{f}^\top\mathbf{D}(\mathbf{w2})\mathbf{X}, \\
I_{\alpha\boldsymbol{\beta}} &= -\frac{2}{\alpha^3}\mathbf{X}^\top\mathbf{a} + \frac{\lambda}{\alpha^2}\mathbf{X}^\top\mathbf{D}(\mathbf{w1})\mathbf{f} + \frac{2\lambda^2}{\alpha^3}\mathbf{X}^\top\mathbf{D}(\mathbf{w2})\mathbf{D}(\mathbf{f})\mathbf{a}, \\
I_{\lambda\boldsymbol{\beta}} &= -\frac{1}{\alpha}\mathbf{X}^\top\mathbf{D}(\mathbf{w1})\mathbf{f} - \frac{2\lambda}{\alpha^2}\mathbf{X}^\top\mathbf{D}(\mathbf{w2})\mathbf{D}(\mathbf{f})\mathbf{b}, \\
I_{\alpha\alpha} &= \frac{n}{\alpha^2} - \frac{12}{\alpha^4}\sum_{i=1}^n \mathrm{senh}^2\left(\frac{y_i-\mu_i}{2}\right) + \frac{4\lambda}{\alpha^3}\sum_{i=1}^n W_\Phi(\lambda b_{y_i})\mathrm{senh}\left(\frac{y_i-\mu_i}{2}\right) \\
&+ \frac{4\lambda^2}{\alpha^4}\sum_{i=1}^n W_\Phi^{(1)}(\lambda b_{y_i})\mathrm{senh}^2\left(\frac{y_i-\mu_i}{2}\right), \\
I_{\lambda\alpha} &= -\frac{2}{\alpha^2}\sum_{i=1}^n \mathrm{senh}\left(\frac{y_i-\mu_i}{2}\right)W_\Phi(\lambda b_{y_i}) - \frac{2\lambda}{\alpha^2}\sum_{i=1}^n W_\Phi^{(1)}(\lambda b_{y_i})b_{y_i}\mathrm{senh}\left(\frac{y_i-\mu_i}{2}\right), \\
I_{\lambda\lambda} &= \sum_{i=1}^n W_\Phi^{(1)}(\lambda b_{y_i})b_{y_i}^2,
\end{aligned}$$

em que

$$
\begin{aligned}
\mathbf{a} &= (\operatorname{senh}(y_1 - x_1^\top\beta), \ldots, \operatorname{senh}(y_n - x_n^\top\beta))^\top, \\
\mathbf{b} &= \left(\operatorname{senh}\left(\frac{y_1 - x_1^\top\beta}{2}\right), \ldots, \operatorname{senh}\left(\frac{y_n - x_n^\top\beta}{2}\right)\right)^\top, \\
\mathbf{c} &= \left(sech^2\left(\frac{y_1 - x_1^\top\beta}{2}\right), \ldots, sech^2\left(\frac{y_n - x_n^\top\beta}{2}\right)\right)^\top, \\
\mathbf{d} &= (\cosh(y_1 - x_1^\top\beta), \ldots, \cosh(y_n - x_n^\top\beta))^\top, \\
\mathbf{f} &= \left(\cosh\left(\frac{y_1 - x_1^\top\beta}{2}\right), \ldots, \cosh\left(\frac{y_n - x_n^\top\beta}{2}\right)\right)^\top, \\
\mathbf{V} &= \frac{1}{4}\mathbf{D}(\mathbf{c}) - \frac{1}{\alpha^2}\mathbf{D}(\mathbf{d}),
\end{aligned}
$$

e os vetores $\mathbf{w1} = (w1_1, \ldots, w1_n)^\top$ e $\mathbf{w2} = (w2_1, \ldots, w2_n)^\top$ têm componentes $W_\Phi(\lambda b_{y_i})$ e $W_\Phi^{(1)}(\lambda b_{y_i})$, respectivamente e b_{y_i} é dado em (2.9).

4.5 Influência Local no Modelo Log-Birnbaum-Saunders Skew-Normal

Nesta seção vamos apresentar o método de diagnóstico proposto por Zhu e Lee (2001) para análise de influência local em modelos estatísticos com dados faltantes. Em seguida, apresentaremos o método de diagnóstico para o modelo Log-Birnbaum-Saunders Skew-Normal. No Apêndice **C** apresentamos de forma resumida o método de influência local proposto por Cook(1986).

Para dados incompletos, Zhu e Lee (2001) propuseram um enfoque para realizar os estudos de influência considerando o vetor de perturbações $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ variando numa região aberta $\boldsymbol{\Omega} \subset \mathbb{R}^n$.

Seja $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{Y_c})$, $\boldsymbol{\theta} \in \mathbb{R}^p$, a função de log-verossimilhança dos dados completos do modelo perturbado. Assumimos que existe um vetor de não perturbação $\boldsymbol{\omega}_0$ tal que $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}_0|\mathbf{Y_c}) = \ell_c(\boldsymbol{\theta}|Y_c)$ para todo $\boldsymbol{\theta}$. Zhu e Lee (2001) propõem a função Q-afastamento dada por

$$
f_Q(\boldsymbol{\omega}) = 2\left[Q(\widehat{\boldsymbol{\theta}}/\widehat{\boldsymbol{\theta}}) - Q(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})|\widehat{\boldsymbol{\theta}})\right], \tag{4.20}
$$

onde $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ maximiza a função $Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \mathbb{E}[\ell_c(\boldsymbol{\theta},\boldsymbol{\omega}_0|\mathbf{Y_c})|\mathbf{Y_0},\widehat{\boldsymbol{\theta}}]$. O gráfico de influência é definido como $\boldsymbol{\alpha}(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, f_Q(\boldsymbol{\omega}))^\top$.

Analogamente a Cook (1986), Zhu e Lee (2001) estudaram o comportamento da superfície $\boldsymbol{\alpha}(\boldsymbol{\omega})$ em $\boldsymbol{\omega}_0$ mediante a curvatura normal $C_{f_Q,\mathbf{h}}$ na direção de um vetor unitário $\mathbf{h}$ que pode ser usado para representar o comportamento da função $f_Q(\boldsymbol{\omega})$

$$\mathbf{C}_{f_Q,\mathbf{h}} = -2\mathbf{h}^\top \ddot{Q}_{\boldsymbol{\omega}_0}\mathbf{h}, \quad -\ddot{Q}_{\boldsymbol{\omega}_0} = \boldsymbol{\Delta}^\top_{\boldsymbol{\omega}_0}\{-\ddot{Q}(\widehat{\boldsymbol{\theta}})\}^{-1}\boldsymbol{\Delta}_{\boldsymbol{\omega}_0} \tag{4.21}$$

onde

$$\ddot{Q}(\widehat{\boldsymbol{\theta}}) = \frac{\partial^2 Q(\boldsymbol{\theta}/\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \quad \text{e} \quad \boldsymbol{\Delta}_{\boldsymbol{\omega}} = \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}/\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}.$$

Como em Cook (1986), a expressão $-Q\ddot{}_{\boldsymbol{\omega}_0}$ é importante para detectar observações localmente influentes.

Baseado na metodologia de Poon e Poon (1999), a curvatura normal conformalizada B_{f_Q} em $\boldsymbol{\omega}_0$ do gráfico $\boldsymbol{\alpha}(\boldsymbol{\omega})$ na direção $\mathbf{h}$ é definida por

$$B_{f_Q,\mathbf{h}} = -\frac{2\mathbf{h}^\top\ddot{Q}_{\boldsymbol{\omega}_0}\mathbf{h}}{tr(-2\ddot{Q}_{\boldsymbol{\omega}_0})}. \tag{4.22}$$

A norma de $-2\ddot{Q}_{\boldsymbol{\omega}_0}$ é obtida através do traço da matriz, diferentemente de Poon e Poon (1999). Sob condições de regularidade, $\ddot{Q}_{\boldsymbol{\omega}_0}$ é semidefinida positiva. Ainda, Zhu e Lee (2001) provam que $0 \leq B_{f_Q,\mathbf{h}} \leq 1$. Como em Cook (1986), a expressão $-\ddot{Q}_{\boldsymbol{\omega}_0}$ é a matriz fundamental para detectar observações influentes. Para tanto, considera-se a decomposição espectral de $-\ddot{Q}_{\boldsymbol{\omega}_0}$ dada por

$$-2\ddot{Q}_{\boldsymbol{\omega}_0} = \sum_{k=1}^{q}\lambda_k\mathbf{e_k}\mathbf{e}^{'},$$

onde $(\lambda_1,\mathbf{e_1}),\ldots,(\lambda_q,\mathbf{e_q})$ são os pares (autovalores-autovetores) da matriz $-\ddot{Q}_{\boldsymbol{\omega}_0}$ com $\lambda_1 \geq \ldots \geq \lambda_r, \lambda_{r+1} = \ldots = \lambda_q = 0$ e $\mathbf{e_1},\ldots,\mathbf{e_q}$ são os vetores da base ortonormal associada. Lesaffre e Verbeke (1998), Poon e Poon (1999) propuseram inspecionar todas $C_{f_Q,\mathbf{u}_j}$, onde $\mathbf{u}_j$ é um vetor de perturbação básica, com j-ésima entrada igual a 1 e restantes iguais a 0. Seja $\widehat{\lambda_k} = \widehat{\lambda_k}/(\lambda_1+,\ldots,+\lambda_q)$. Desde que $tr(\ddot{Q}_{\boldsymbol{\omega}_0}) = \sum_{i=1}^{r}\lambda_i$, pode ser visto que

$$C_{f_Q,\mathbf{u}_j} = \sum_{i=1}^{r}\lambda_i\mathbf{e}^2_{ij}, \quad B_{f_Q,\mathbf{u}_j} = \sum_{i=1}^{r}\lambda_i\mathbf{e}^2_{ij}.$$

Assim, tem-se que $B_{f_Q,\mathbf{e}_k} = \lambda_k$. Segundo Zhu e Lee (2001), um autovetor $\mathbf{e}_i$ de $-2\ddot{Q}_{\boldsymbol{\omega}_0}$ é chamado m_0-influente se $B_{f_Q,\mathbf{e}_i} \geq m_0/r$. Considerando $\mathbf{e}_k^2 = (e_{k1}^2, \ldots, e_{kq}^2)$, define-se como *vetor de contribuição agregada de todos os autovetores m_0-influentes* a soma ponderada

$$M(m_0) = \sum_{i:\lambda_i \geq m_0/r} \widehat{\lambda}_i \mathbf{e}_i^2.$$

Em particular, quando $m_0 = 0$, $M(0) = \sum_{i=1}^r \lambda_i \mathbf{e}_i^2$ e $M(0)_j = B_{f_Q,\mathbf{u}_j}$, para todo j e sua média é $\overline{M(0)} = 1/q$. Ver Zhu e Lee (2001) para outras propriedade teóricas de $B_{f_Q,\mathbf{u}_l}$, tal como a invariância sob reparametrização de $\boldsymbol{\theta}$.

Assim, a avaliação de casos influentes é baseada em $\{M(0)_l,\ \ l = 1, \ldots, q\}$. Zhu e Lee (2001) propõem como ponto de corte o valor $\overline{M(m_0)} + c^* DP(M(m_0))$, onde $DP(m_0)$ é o desvio padrão de $\{M(0)_l,\ \ l = 1, \ldots, q\}$ e c^* é uma constante arbitrária. Neste trabalho consideramos $c^* = 2$.

Diagnósticos via exclusão de casos serão também considerados neste trabalho. Na literatura, as medidas clássicas são a distância de Cook e o afastamento pela verossimilhança. Baseado nessas idéias, Lee e Xu (2004) propõem as medidas análogas D_i e LD_i para a função $\mathbf{Q}$. Tais medidas são, respectivamente, dadas por

$$\begin{aligned} D_i &= (\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_{(i)})^\top [-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\boldsymbol{\theta})](\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_{(i)}) \quad \text{e} \\ LD_i &= 2\left[Q(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) - Q(\widehat{\boldsymbol{\theta}}_{(i)}|\widehat{\boldsymbol{\theta}})\right], \end{aligned}$$

onde $\widehat{\boldsymbol{\theta}}_{(i)}$ é o maximizador da função $Q_{(i)}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$, $i = i, \ldots, n$. Neste trabalho, utiliza-se $M(0)$ como diagóstico para influência local e D_i e LD_i como medidas de influência global.

A Matriz Hessiana

A seguir, vamos desenvolver o método de influência local para o modelo Log-Birnbaum-Saunders Skew-Normal. Neste caso, serão calculadas a matriz hessiana $\ddot{Q}(\widehat{\boldsymbol{\theta}})$ e a matriz $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$ (avaliadas em $\widehat{\boldsymbol{\theta}}$) para cada esquema de perturbação proposto. Estas matrizes são os elementos básicos para a aplicação da metodologia de análise de diagnóstico proposta por Zhu e Lee (2001). As perturbações consideradas são: perturbação de ponderação de casos, perturbação na variável resposta e perturbação em uma variável explicativa. Para obter as medidas de diagnóstico baseadas na metodologia de Zhu e Lee (2001), primeiramente vamos obter a matriz hessiana

$$\ddot{Q}(\widehat{\boldsymbol{\theta}}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}},$$

em que $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^\top, \lambda)^\top$. Sejam $\widehat{\mathbf{h}} = (\widehat{h}_1, \ldots, \widehat{h}_n)^\top$ e $\widehat{\mathbf{h_2}} = (\widehat{h_1^2}, \ldots, \widehat{h_n^2})^\top$, assim os elementos de $\ddot{Q}(\widehat{\boldsymbol{\theta}})$ são dados por

$$\begin{aligned}
\frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\alpha\partial\alpha} &= \frac{n}{\alpha^2} - \frac{12}{\alpha^4}(1+\lambda^2)\mathbf{b}^\top\mathbf{b} + \frac{4}{\alpha^3}(1+\lambda^2)^{1/2}\lambda\widehat{\mathbf{h}}^\top\mathbf{b},\\
\frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\alpha\partial\boldsymbol{\beta}^\top} &= -\frac{2}{\alpha^3}(1+\lambda^2)\mathbf{a}^\top\mathbf{X} + \frac{\lambda}{\alpha^2}(1+\lambda^2)^{1/2}\mathbf{h}^\top\mathbf{D}(\mathbf{f})\mathbf{X},\\
\frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\alpha\partial\lambda} &= \frac{8}{\alpha^3}\lambda\mathbf{b}^\top\mathbf{b} - \frac{2(1+2\lambda^2)}{\alpha^2(1+\lambda^2)^{1/2}}\widehat{\mathbf{h}}^\top\mathbf{b},\\
\frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} &= \frac{1}{4}\mathbf{X}^\top\mathbf{D}(\mathbf{c})\mathbf{X} - \frac{1}{\alpha^2}(1+\lambda^2)\mathbf{X}^\top\mathbf{D}(\mathbf{d})\mathbf{X} + \frac{1}{2\alpha}(1+\lambda^2)^{1/2}\lambda\mathbf{X}^\top\mathbf{D}(\mathbf{b})\mathbf{D}(\widehat{\mathbf{h}})\mathbf{X},\\
\frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\lambda} &= \frac{2\lambda}{\alpha^2}\mathbf{X}^\top\mathbf{a} - \frac{(1+2\lambda^2)}{\alpha(1+\lambda^2)^{1/2}}\mathbf{X}^\top\mathbf{D}(\mathbf{f})\mathbf{h},\\
\frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\lambda\partial\lambda} &= -\frac{n(1-\lambda^2)}{(1+\lambda^2)^2} - \frac{4}{\alpha^2}\mathbf{b}^\top\mathbf{b} + \frac{2\lambda(3+2\lambda^2)}{\alpha(1+\lambda^2)^{3/2}}\widehat{\mathbf{h}}^\top\mathbf{b} - \widehat{\mathbf{h_2}}^\top\mathbf{1}_n,
\end{aligned}$$

em que $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$ e $\mathbf{f}$ são dados na Seção 4.4 e

$$\mathbf{e} = \left(tagh\left(\frac{y_1 - x_1^\top\beta}{2}\right), \ldots, tagh\left(\frac{y_n - x_n^\top\beta}{2}\right)\right)^\top.$$

4.5.1 Perturbação de ponderação de casos

Esse esquema de perturbação permite identificar aqueles indivíduos que exercem um grande impacto no processo de estimação do parâmetro $\boldsymbol{\theta}$. Seja $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ um vetor de dimensão $n \times 1$ com $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$. Então a esperança da função log-verossimilhança de dados completos perturbada (função Q perturbada), pode ser escrita como

$$Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \mathbb{E}[l_c(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{y}_c)] = \sum_{i=1}^n \omega_i \mathbb{E}[l_i(\boldsymbol{\theta}|\mathbf{y_c})] = \sum_{i=1}^n \omega_i Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}).$$

Nesse caso, a matriz $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0} = \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0}$ é dada por

$$\boldsymbol{\Delta}_{\boldsymbol{\omega}_0} = \left(\boldsymbol{\Delta}_\beta^\top, \boldsymbol{\Delta}_\alpha^\top, \boldsymbol{\Delta}_\lambda^\top\right)^\top,$$

em que

$$\begin{aligned}
\mathbf{\Delta}_\beta &= \mathbf{X}^\top \left[-\frac{1}{2}\mathbf{D}^{-1}(\mathbf{f})\mathbf{D}(\mathbf{b}) + \frac{1}{\alpha^2}(1+\lambda^2)\mathbf{D}(\mathbf{a}) - \frac{1}{\alpha}(1+\lambda^2)^{1/2}\lambda\mathbf{D}(\mathbf{f})\mathbf{D}(\mathbf{h})\right],\\
\mathbf{\Delta}_\alpha &= -\frac{1}{\alpha}\mathbf{1}_n^\top + \frac{4}{\alpha^3}(1+\lambda^2)\mathbf{b}^\top\mathbf{D}(\mathbf{b}) - \frac{2}{\alpha^2}(1+\lambda^2)^{1/2}\lambda\mathbf{h}^\top\mathbf{D}(\mathbf{b}),\\
\mathbf{\Delta}_\lambda &= -\frac{\lambda}{(1+\lambda^2)}\mathbf{1}_n^\top - \frac{4}{\alpha^2}\lambda\mathbf{b}^\top\mathbf{D}(\mathbf{b}) + \frac{2(1+2\lambda^2)}{\alpha(1+\lambda^2)^{1/2}}\mathbf{h}^\top\mathbf{D}(\mathbf{b}) - \lambda\mathbf{1}_n^\top\mathbf{D}(\widehat{\mathbf{h_2}}).
\end{aligned}$$

4.5.2 Perturbação na variável resposta

Este esquema de perturbação permite verificar a existência de observações atípicas na variável resposta. Uma característica de interesse deste esquema de perturbação é a sua conexão com a matriz de alavancas generalizadas (Wei et al., 1998). A perturbação em $\mathbf{y} = (y_1, \dots, y_n)^\top$ é introduzida substituindo y_i por $y_{i\omega} = y_i + \omega_i S_y$, $i = 1, \dots, n$, onde S_y é o desvio padrão das componentes de $\mathbf{y}$. Assim segue que o vetor de perturbação é dado por $\boldsymbol{\omega} = (\omega_1, ..., \omega_n)^\top$ e $\boldsymbol{\omega}_0 = \mathbf{0}$. Assim

$$Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}, \omega_i|\widehat{\boldsymbol{\theta}}),$$

em que

$$Q_i(\boldsymbol{\theta}, \omega_i|\widehat{\boldsymbol{\theta}}) = c - \frac{1}{2}\log(1-\delta^2) - \log(\alpha) + \log\cosh\left(\frac{y_{i\omega} - \boldsymbol{x}_i^\top\beta}{2}\right) - \frac{1}{2(1-\delta^2)}[b_{y_{i\omega}}^2 - 2\delta\widehat{h}_i b_{y_{i\omega}} + \delta^2\widehat{h}_i^2],$$

com $b_{y_{i\omega}} = \frac{2}{\alpha}\text{senh}\left(\frac{y_{i\omega} - \boldsymbol{x}_i^\top\beta}{2}\right)$. Aqui a matriz $\mathbf{\Delta}_{\boldsymbol{\omega}_0}$ é dada por

$$\mathbf{\Delta}_{\boldsymbol{\omega}_0} = \left(\mathbf{\Delta}_\beta^\top, \mathbf{\Delta}_\alpha^\top, \mathbf{\Delta}_\lambda^\top\right)^\top,$$

em que seus elementos são dado por

$$\begin{aligned}
\mathbf{\Delta}_\beta &= S_y\left[\frac{1}{\alpha^2}(1+\lambda^2)\mathbf{X}^\top\mathbf{D}(\mathbf{d}) - \frac{1}{4}\mathbf{X}^\top\mathbf{D}(\mathbf{c}) - \frac{\lambda}{2\alpha}(1+\lambda^2)^{1/2}\mathbf{X}^\top\mathbf{D}(\widehat{\mathbf{h}})\mathbf{D}(\mathbf{b})\right],\\
\mathbf{\Delta}_\alpha &= S_y\left[\frac{2}{\alpha^3}(1+\lambda^2)\mathbf{1}_n^\top\mathbf{D}(\mathbf{a}) - \frac{1}{\alpha^2}(1+\lambda^2)^{1/2}\lambda\widehat{\mathbf{h}}^\top\mathbf{D}(\mathbf{f})\right],\\
\mathbf{\Delta}_\lambda &= S_y\left[\frac{(1+2\lambda^2)}{\alpha(1+\lambda^2)^{1/2}}\widehat{\mathbf{h}}^\top\mathbf{D}(\mathbf{f}) - \frac{2\lambda}{\alpha^2}\mathbf{1}_n^\top\mathbf{D}(\mathbf{a})\right].
\end{aligned}$$

4.5.3 Perturbação em uma variável explicativa

Nesse caso o interesse está em perturbar uma variável explicativa específica permitindo, por exemplo, detectar possíveis maus condicionamentos de alguma coluna da matriz de desenho $\mathbf{X}$ (Belsley, 1991). Sem perda de generalidade, vamos considerar a perturbação aditiva da variável explicativa X_p. A saber

$$x_{ip\omega} = x_{ip} + \omega_i S_x,\ \omega_i \in \mathbb{R}, i = 1, \ldots, n,$$

em que S_x é o desvio padrão da variável explicativa X_p. O vetor de pertubação é dado por $\boldsymbol{\omega} = (\omega_1, ..., \omega_n)^\top$. Nesse caso, $\boldsymbol{\omega}_0 = \mathbf{0}$ e consequentemente a função Q perturbada é dada por

$$Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) \;=\; \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}, \omega_i|\widehat{\boldsymbol{\theta}}),$$

em que

$$Q_i(\boldsymbol{\theta}, \omega_i|\widehat{\boldsymbol{\theta}}) = c - \frac{1}{2}\log(1-\delta^2) - \log(\alpha) + \log\cosh\left(\frac{y_i - \boldsymbol{x}_{i\omega}^\top\beta}{2}\right) - \frac{1}{2(1-\delta^2)}[b_{y_i}^2(\omega) - 2\delta\widehat{h}_i b_{y_i}(\omega) + \delta^2\widehat{h}_i^2],$$

com $b_{y_i}(\omega) = \dfrac{2}{\alpha}\text{senh}\left(\dfrac{y_i - \boldsymbol{x}_{i\omega}^\top\beta}{2}\right)$. A matriz $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$ pode ser escrita por $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0} = \left(\boldsymbol{\Delta}_\beta^\top, \boldsymbol{\Delta}_\alpha^\top, \boldsymbol{\Delta}_\lambda^\top\right)^\top$, cujos elementos são dados por

$$\begin{aligned}
\boldsymbol{\Delta}_\beta &= S_x\left[\beta_p \mathbf{X}^\top D(\mathbf{r}) + \mathbf{1}_p(p)\mathbf{r}^\top\right], \\
\boldsymbol{\Delta}_\alpha &= S_x\beta_p\left[-\frac{2}{\alpha^3}(1+\lambda^2)\mathbf{a}^\top + \frac{\lambda}{\alpha^2}(1+\lambda^2)^{1/2}\widehat{\mathbf{h}}^\top\mathbf{D}(\mathbf{f})\right], \\
\boldsymbol{\Delta}_\lambda &= S_x\beta_p\left[\frac{2\lambda}{\alpha^2}\mathbf{a}^\top - \frac{1+2\lambda^2}{\alpha(1+\lambda^2)^{1/2}}\widehat{\mathbf{h}}^\top\mathbf{D}(\mathbf{f})\right],
\end{aligned}$$

em que

$$\mathbf{r} = \frac{1}{4}\mathbf{c} - \frac{1}{\alpha^2}(1+\lambda^2)\mathbf{d} + \frac{1}{2\alpha}(1+\lambda^2)^{1/2}\lambda\mathbf{D}(\mathbf{b})\mathbf{h}$$

e

$$D(\mathbf{r}) = \frac{1}{4}D(\mathbf{c}) - \frac{1}{\alpha^2}(1+\lambda^2)D(\mathbf{d}) + \frac{1}{2\alpha}(1+\lambda^2)^{1/2}\lambda\mathbf{D}(\mathbf{b})\mathbf{D}(\mathbf{h}).$$

4.6 Estudo de Simulação

Para verificar a validade do algoritmo EM proposto e análise de diagnóstico, foram simulados dados com a presença de uma variável regressora, da forma $y = \beta_0 + \beta_1 x_i + \epsilon_i$, onde $\epsilon_i \sim \text{SHSN}(\alpha, 0, \sigma = 2, \lambda)$, $i = 1, \ldots, 200$ e a variável x foi gerada da distribuição $U(0, 1)$. Os valores escolhidos dos parâmetros foram $\beta_0 = 1$, $\beta_1 = 2$, $\alpha = 0.5$ e $\lambda = 3$. A Tabela 4.1 apresenta os valores escolhidos e estimados dos parâmetros do modelo. Os erros padrões são calculados utilizando a matriz de informação de Fisher observada, na forma $EP(\widehat{\boldsymbol{\theta}}) = \left[Diag(I^{-1}(\widehat{\boldsymbol{\theta}}))\right]^{1/2}$.

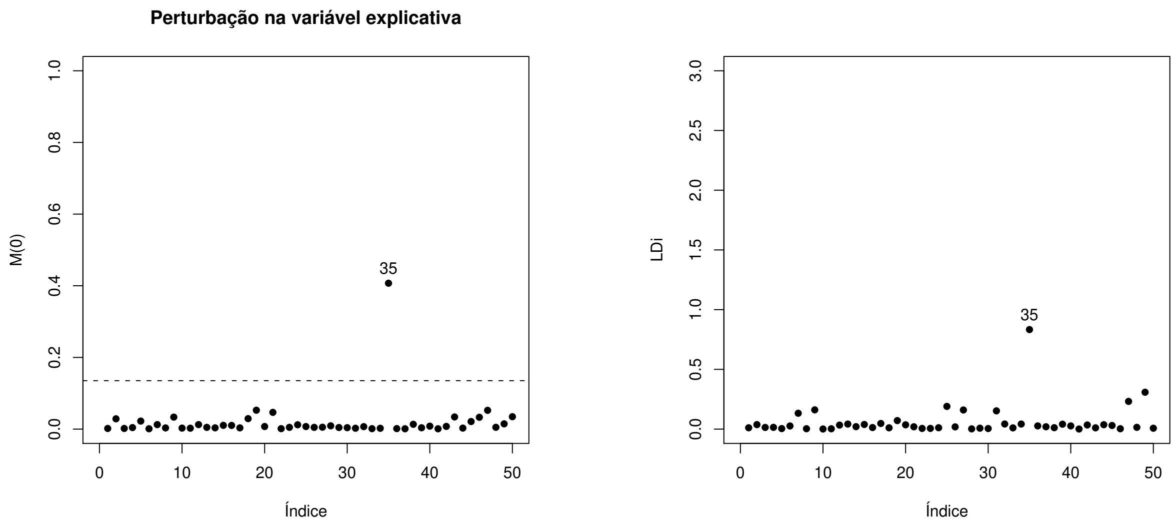
Tabela 4.1: Estimativas dos parâmetros do modelo com seus respectivos erros padrões assintóticos entre parênteses.

	β_0	β_1	α	λ
Valores verdadeiros	1	2	0.5	3
Estimativas	1.04 (0.05)	1.96 (0.12)	0.49 (0.02)	2.95 (0.44)

A seguir, será apresentado um estudo de simulação de pequena escala para estudar o comportamento do esquema de perturbação em uma particular variável explicativa. Seguindo o mesmo procedimento sugerido por Ortega et al. (2003), foram simulados dados (considerando (4.4)) de um modelo com uma variável regressora, da forma $y_i = \alpha + \beta x_i + \epsilon_i$, onde $\epsilon_i \sim \text{SHSN}(\alpha, 0, \sigma = 2, \lambda)$, $i = 1, \ldots, 50$ e a variável x foi gerada da distribuição $U(0, 1)$. Os valores escolhidos dos parâmetros foram $\beta_0 = 1$, $\beta_1 = -2$, $\alpha = 0.5$ e $\lambda = 0.5$. Depois de gerar valores da variável explicativa x_i, $i = 1, \ldots, 50$, substituímos o valor máximo de x_i por

$$x_{max} \leftarrow x_{max} + 0.2\sqrt{||\mathbf{x}||}, \tag{4.23}$$

em que $||\mathbf{x}|| = \mathbf{x}^\top \mathbf{x}$. Então ajustamos o modelo de regressão SHSN e aplicamos o esquema de perturbação na variável explicativa. O gráfico de influência local e global é mostrado na Figura 4.3 (a)-(b), respectivamente. Observamos que a observação 35 (perturbada através de (4.23)) exerce grande influência, confirmando a sensibilidade da metodologia proposta.



(a) (b)

Figura 4.3: Diagnóstico para o modelo de regressão com dados simulados (Perturbação na variável explicativa) (b) Influência local e (b) Influência global LD_i.

4.7 Aplicação

Para ilustrar a aplicação do modelo de regressão log-Birnbaum-Saunders Skew-Normal e a análise de diagnóstico, foi ajustado um modelo a um conjunto de dados previamente analisado em Rieck e Nedelman (1991), Galea, Leiva e Paula (2004) e Xie e Wei (2007), que correspondem a 46 peças de metal que foram submetidas à fadiga em ciclos. O objetivo do estudo foi verificar o número de ciclos até que ocorresse a falha. A variável resposta N é o número de ciclos até que ocorra a falha e a variável explicativa X é o trabalho por ciclo (mJ/m^3). Os dados são apresentados no Apêndice **C**, Tabela 5. As estatísticas descritivas para o número de ciclos até que ocorra a falha e para o trabalho por ciclo são apresentadas na Tabela 4.2.

Sob normalidade, este conjunto de dados foi analisado por Rieck e Nedelman (1991) considerando o seguinte modelo de regressão

$$y_i = \beta_0 + \beta_1 \log(x_i) + \epsilon_i,$$

em que $y_i = \log N_i$ e $\epsilon_i \sim \text{SHN}(\alpha, 0, \sigma = 2)$, $i = 1, \ldots, 46$. Na Figura 4.4 apresentamos a

Tabela 4.2: Conjunto de dados de fadiga: Estatísticas descritivas para o número de ciclos até que ocorra a falha N e para o trabalho por ciclo X.

	Mínimo	Média	Máximo	Percentil 25	Mediana	Percentil 75
N	125	943.6522	5046	242.75	566	1086.25
X	11.5	40.293	100.5	24	33.2	60.125

reta de ajuste para este caso. Em lugar de $\log(x_i)$ consideramos outras transformações, por exemplo as citados na equação (4.1).

Neste trabalho estamos propondo o seguinte modelo de regressão:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 \exp(-\gamma_0/x_i) + \epsilon_i \\ &= \beta_0 + \beta_1 g(x_i) + \epsilon_i, \quad i = 1, \ldots, 46, \end{aligned}$$

em que γ_0 é uma quantidade positiva conhecida. A quantidade γ_0 é estimada previamente: geramos m valores para γ, isto é, $\gamma_1, \ldots, \gamma_m$. Então o valor de γ_0 é obtido minimizando a equação dos mínimos quadrados dada por

$$MQ_k(\boldsymbol{\theta}) = \sum_{i=i}^{n} (y_{ik} - \widehat{\beta}_{0k} - \widehat{\beta}_{1k} \exp(-\gamma_k/x_i))^2, \quad k = 1, \ldots, m.$$

Ou seja, o valor de γ_0 escolhido será o $\min\limits_{1 \leq k \leq m} MQ_k(\boldsymbol{\theta})$. Os valores de $\widehat{\beta}_0$ e $\widehat{\beta}_1$ foram obtidos usando as estimativas de mínimos quadrados. Neste caso, o γ_0 obtido foi $\gamma_0 = 20.076$. Portanto o modelo utilizado nesta aplicação será

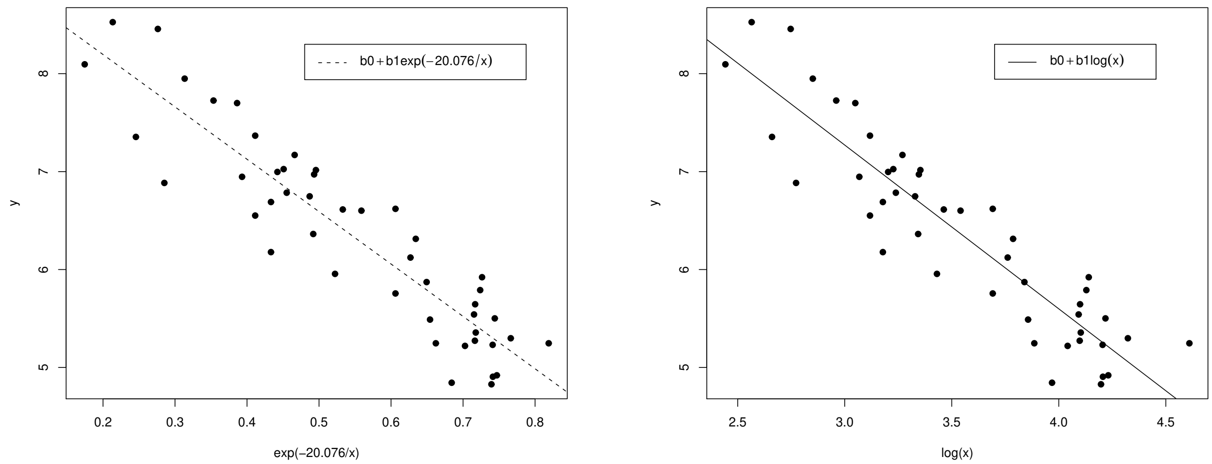
$$y_i = \beta_0 + \beta_1 \exp(-20.076/x_i) + \epsilon_i, \quad i = 1, \ldots, 46. \tag{4.24}$$

em que $\epsilon_i \sim SHN(\alpha, 0, \sigma = 2)$.

Nas Figuras 4.4 (a)-(b) apresentamos a reta ajustada do modelo (4.24) considerando $\epsilon_i \sim SHN(\alpha, 0, \sigma = 2)$ e o modelo proposto por Rieck e Nedelman (1991), dado por

$$y_i = \beta_0 + \beta_1 \log(x_i) + \epsilon_i, \tag{4.25}$$

em que $y_i = \log N_i$ e $\epsilon_i \sim SHN(\alpha, 0, \sigma = 2)$, $\quad i = 1, \ldots, 46$. Note que o modelo (4.24) considerando $\epsilon_i \sim SHN(\alpha, 0, \sigma = 2)$ também é apropriado para ajustar esse conjunto de dados.



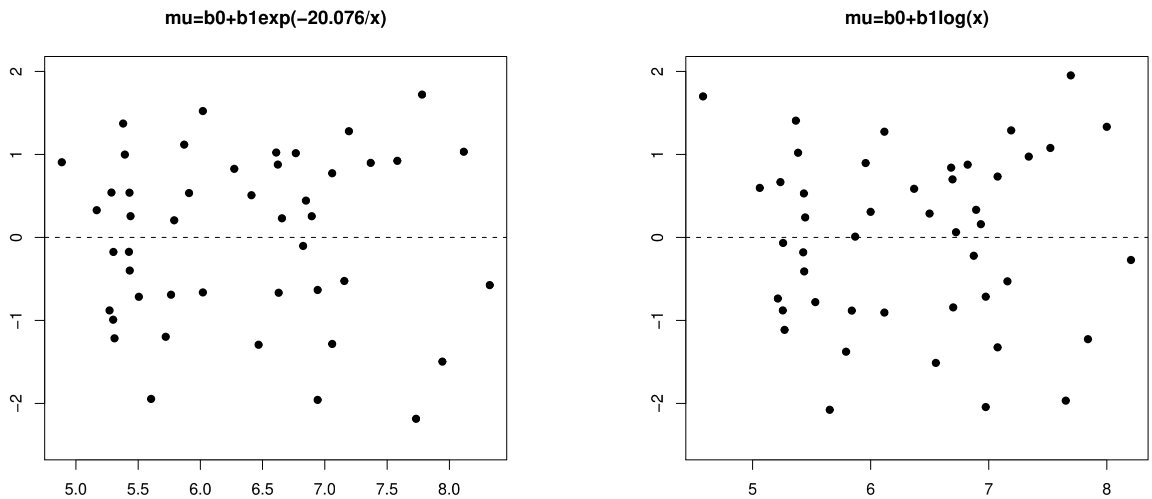
(a) (b)

Figura 4.4: Scatter-plot dos modelos: (a) – modelo dado em (4.24), considerando $\epsilon_i \sim SHN(\alpha, 0, \sigma = 2)$ e (b) – modelo dado em (4.25).

Seguindo o mesmo procedimento sugerido por Xie e Wei (2007), nós obtemos o resíduo $\widehat{\epsilon}_i = y_i - \widehat{\mu}_i$ e $\widehat{R}_i = 2\widehat{\alpha}^{-1}\text{senh}\,(\widehat{\epsilon}_i/2)$. Na Figura 4.5 apresentamos o gráfico de $\widehat{R}_i$ versus os valores preditos $\widehat{\mu}_i$ para o modelo dado em (4.24), considerando $\epsilon_i \sim SHN(\alpha, 0, \sigma = 2)$ e o modelo dado em (4.25). Nota-se que os pontos estão aleatoriamente dispersos para os dois modelos.

Para estimação dos parâmetros $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^\top, \lambda)^\top$, utilizamos o algoritmo EM. Como valores iniciais para os parâmetros α e $\boldsymbol{\beta}$ consideramos as estimativas sob o modelo de regressão dado em (4.24), considerando $\epsilon_i \sim SHN(\alpha, 0, \sigma = 2)$ e o valor inicial de λ foi estimado através do método dos momentos utilizando a variável $T_i^* = N_i \exp(-\widehat{\beta}_0 - \widehat{\beta}_1 \log(x_i))$, $i = 1, \ldots, 46$ e o fato de que Z dado em (4.6) tem distribuição SN$(0, 1, \lambda)$. Na Figura 4.6 apresentamos o histograma da variável T_i^*, nota-se que existe uma assimetria negativa.

Na Figura 4.7 (a)-(b) apresentamos alguns dos gráficos de diagnóstico para o modelo (4.25) obtidos por Galea, Leiva e Paula (2004) e na Figura 4.8 apresentamos os gráficos de diagnóstico para o modelo (4.24), considerando $\epsilon_i \sim SHN(\alpha, 0, \sigma = 2)$. Nota-se que o



(a) (b)

Figura 4.5: Gráfico de $\widehat{R}_i$ versus $\widehat{\mu}_i$.

modelo (4.25) é mais sensível à presença de observações discrepantes do que o modelo que estamos propondo (veja os gráficos 4.7 e 4.8).

Agora vamos supor que ϵ_i em (4.24) tem distribuição $\epsilon_i \sim \text{SHSN}(\alpha, 0, \sigma = 2, \lambda)$.

As estimativas dos parâmetros e os valores da funções log-verossimilhanças avaliados nas estimativas são apresentados na Tabela 4.3. Além disso, são apresentados os respectivos erros padrões das estimativas que foram obtidos utilizando a matriz de informação de Fisher observada, na forma $EP(\widehat{\boldsymbol{\theta}}) = \left[Diag(I^{-1}(\widehat{\boldsymbol{\theta}}))\right]^{1/2}$. De acordo com os critérios AIC, SIC e HQC, o modelo (4.24), considerando $\epsilon_i \sim SHSN(\alpha, 0, \sigma = 2, \lambda)$, apresenta um melhor ajuste em relação ao modelo que considera $\epsilon_i \sim SHN(\alpha, 0, \sigma = 2)$. Além disso, aplicamos a aproximação do fator de Bayes mediante o SIC (veja Tabela 4.4). De acordo com esta tabela nota-se que há evidências de que modelo (4.24), considerando $\epsilon_i \sim SHSN(\alpha, 0, \sigma = 2, \lambda)$, é mais apropriado para ajustar os dados.

Realizando um teste assintótico (teste da razão de verossimilhança) para testar $H_0 : \lambda = 0$ *vs* $H_1 : \lambda \neq 0$, tem-se que a estatística $RV = 2(\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_0)) = 5.830$, com um p-valor igual a 0.015, levando à rejeição da hipótese nula, considerando um nível de significância de 5%.

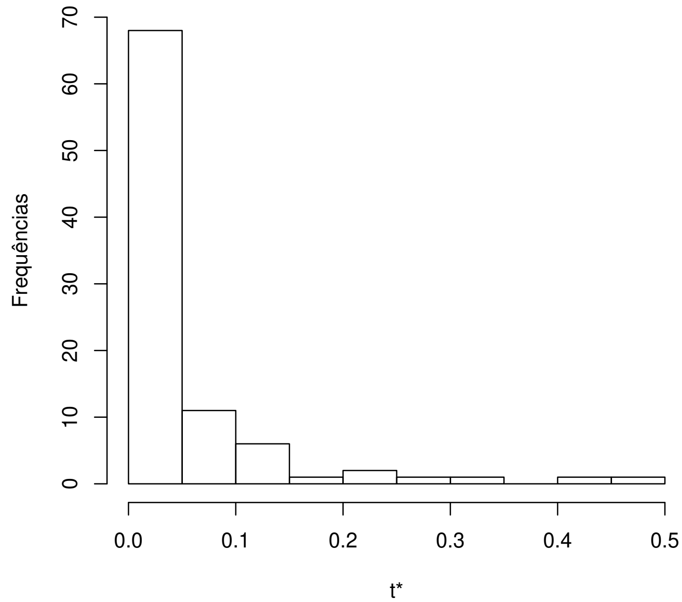


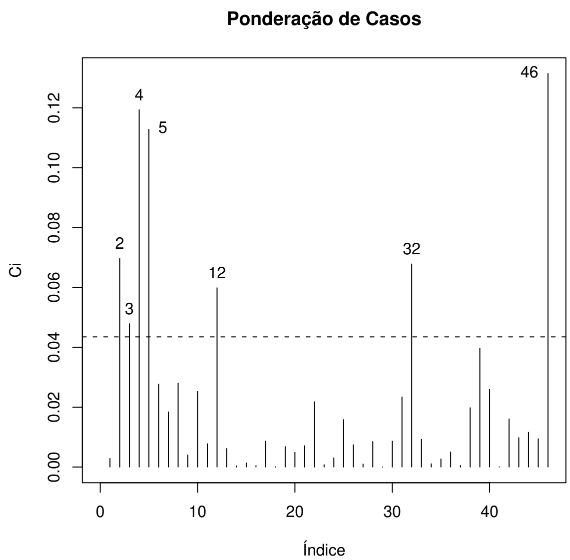
Figura 4.6: Histograma da variável T_i^*.

Portanto, o modelo (4.24) considerando $\epsilon_i \sim SHSN(\alpha, 0, \sigma = 2, \lambda)$ é mais apropriado para ajustar os dados.

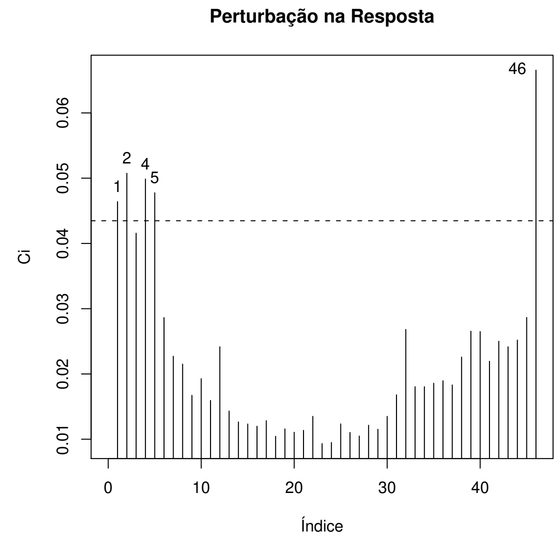
Substituindo as EMV de $\boldsymbol{\theta}$ (de acordo com cada modelo – SHSN e SHN) na representação dada na equação (4.7), construímos os gráficos de envelope apresentados na Figura 4.9 (linhas representam o percentil-5, a média e o percentil-95 de 200 pontos simulados para cada observação). Os gráficos fornecem evidência, avaliado junto aos critérios mostrados anteriormente, de que modelo (4.24) considerando $\epsilon_i \sim SHSN(\alpha, 0, \sigma = 2, \lambda)$ proporciona melhor ajuste aos dados.

A análise de diagnóstico através de influência local detectou os pontos 3, 5, 12 e 32 como potencialmente influentes (veja Figura 4.10 (a)-(c)). O gráfico de influência global (veja Figura 4.11 - (b)) para D_i identificou apenas um ponto 4. Os demais não exercem grande influência.

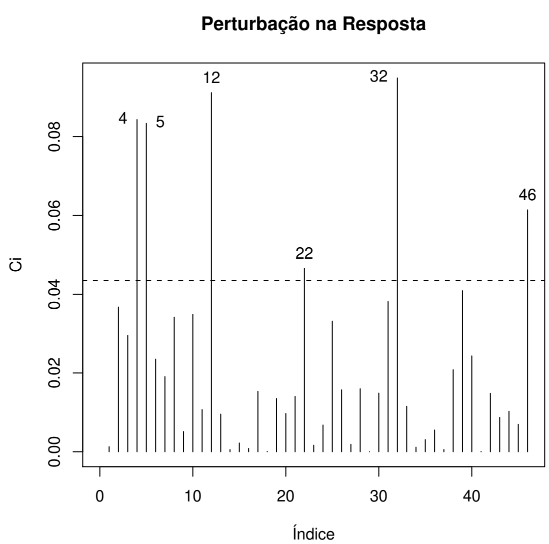
Eliminamos as observações 3, 4, 5, 12 e 32 e reajustamos o modelo. Na Tabela 4.5



(a)



(b)



(c)

Figura 4.7: (a) Perturbação de ponderação de casos para a estimativa $\widehat{\boldsymbol{\beta}}$, (b) Perturbação na variável resposta para a estimativa $\widehat{\boldsymbol{\beta}}$ e (c) Perturbação na variável resposta para a estimativa $\widehat{\alpha}$, considerando o modelo (4.25).

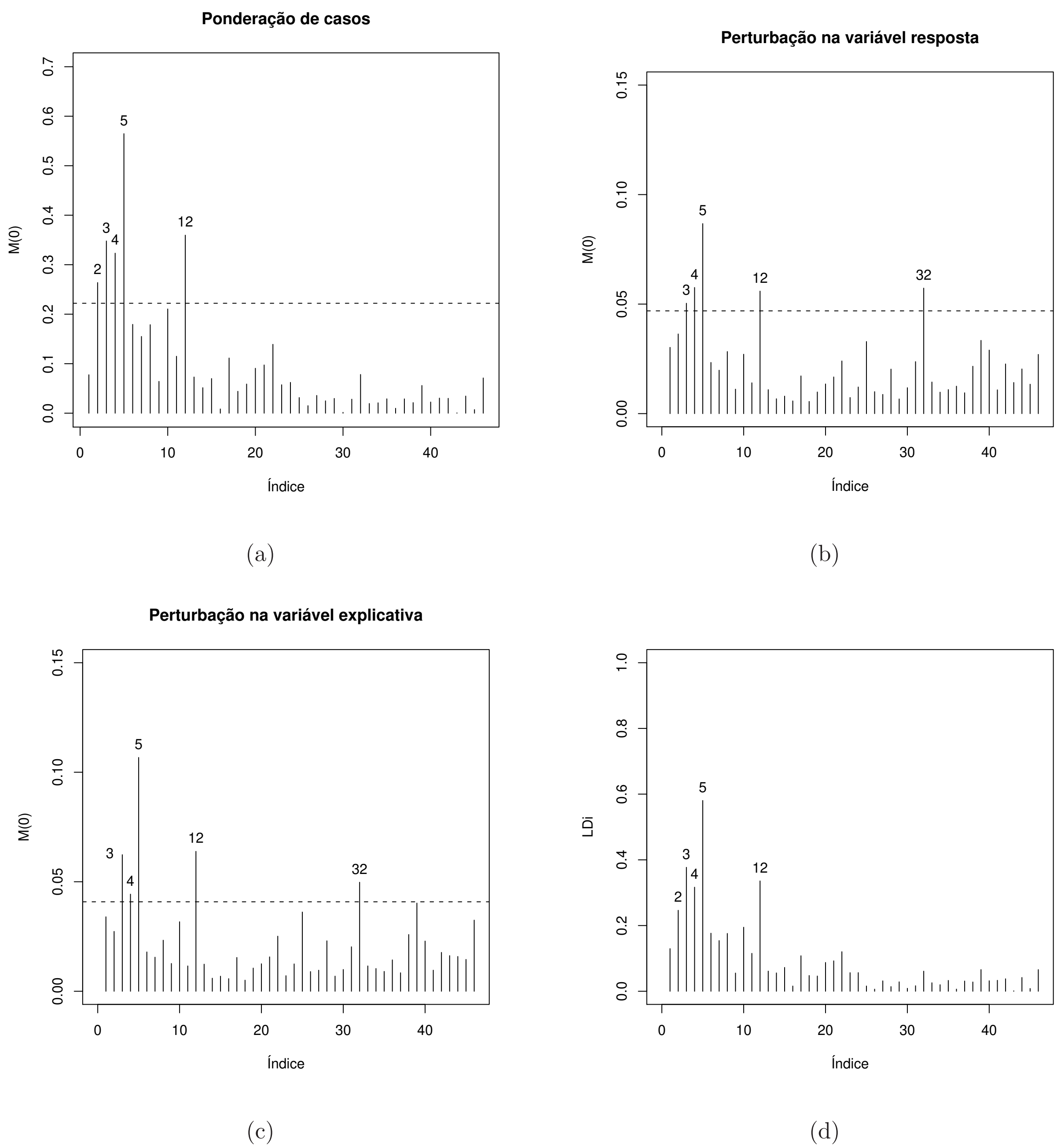


Figura 4.8: (a) Perturbação de ponderação de casos, (b) Perturbação na variável resposta e (c) Perturbação na variável explicativa e (d) Influência global para conjunto de dados de fadiga, considerando o modelo (4.24).

Tabela 4.3: Estimativas dos parâmetros com seus respectivos erros padrões assintóticos entre parênteses e critérios AIC, SIC e HQC dos modelos SHSN e SHN para os dados de fadiga.

	SHSN	**SHN**
Parâmetros	Estimativa	Estimativa
α	0.700 (0.073)	0.400 (0.0417)
β_0	9.920(0.314)	9.263 (0.201)
β_1	-5.544 (0.579)	-5.340 (0.351)
λ	-6.807 (2.947)	–
$-\ell(\widehat{\boldsymbol{\theta}})$	19.375	22.290
AIC	0.508	0.549
SIC	0.587	0.609
HQC	0.537	0.572

Tabela 4.4: Fator de Bayes (aproximação mediante o SIC) para os dados de fadiga.

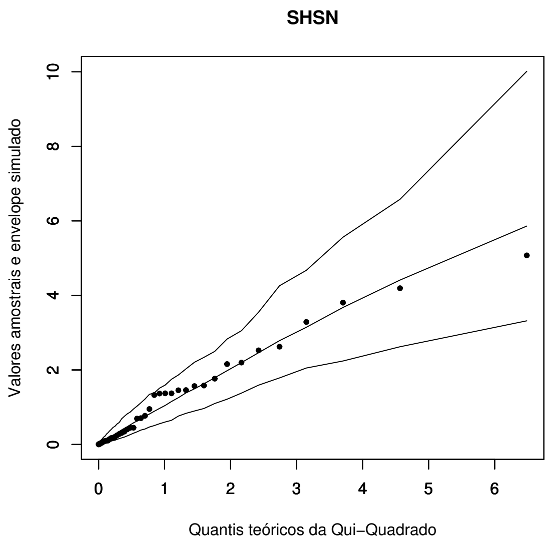
H_1	H_2	$2\log(B_{12})$	Evidência a favor de H_1
SHSN	**SHN**	2.002	Positiva

apresentamos a variação das estimativas de β_0 e β_1, α e λ obtidas através de

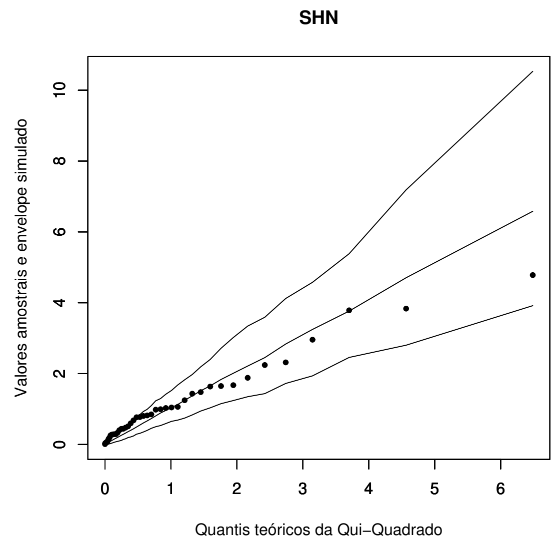
$$RC = \frac{|\widehat{\theta}_{j(i)} - \widehat{\theta}_j|}{\widehat{\theta}_j} \times 100, \quad j = 0, \ldots, 4 \ \text{e} \ i = 1, \ldots, n,$$

em que $\widehat{\theta}_{j(i)}$ denota a EMV de θ_j após a retirada da i-ésima observação. Por exemplo, para $j = 1$ temos que $\theta_1 = \beta_0$ e $\widehat{\beta}_{0(i)}$ é a EMV para β_0 após a retirada da i-ésima observação. De acordo com a Tabela 4.5, observamos que apenas o ponto 4 exerce maior mudança na estimativa do parâmetro λ.

Nas Figuras 4.12-4.14 apresentamos, respectivamente, os gráficos de diagnóstico para os esquemas: perturbação na variável resposta, perturbação na variável explicativa e pon-



(a)



(b)

Figura 4.9: Envelope simulado para os dados de fadiga. (a) Modelo SHSN e (b) Modelo SHN.

deração de casos, considerando as partições $\boldsymbol{\beta}$, α e λ.

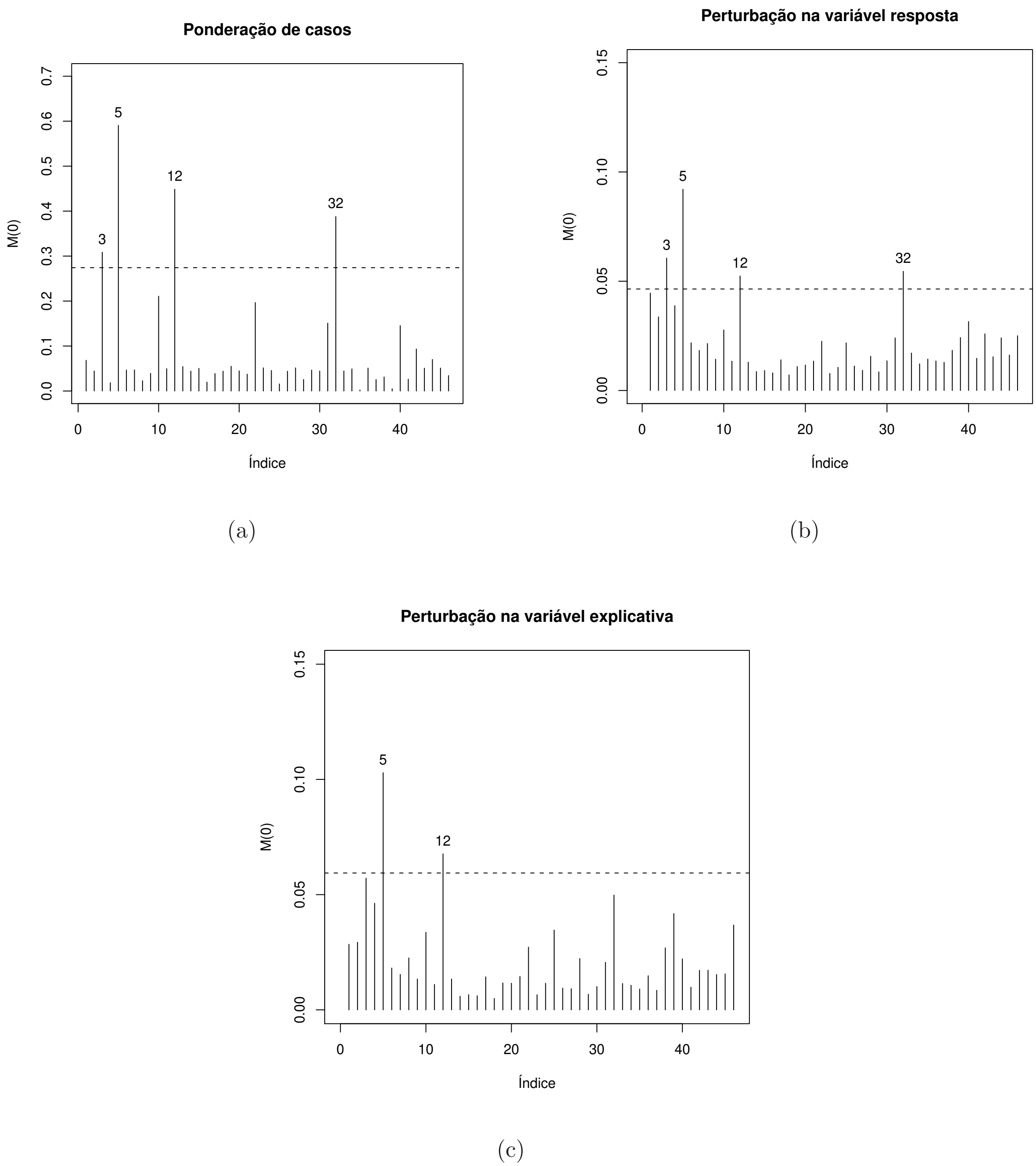
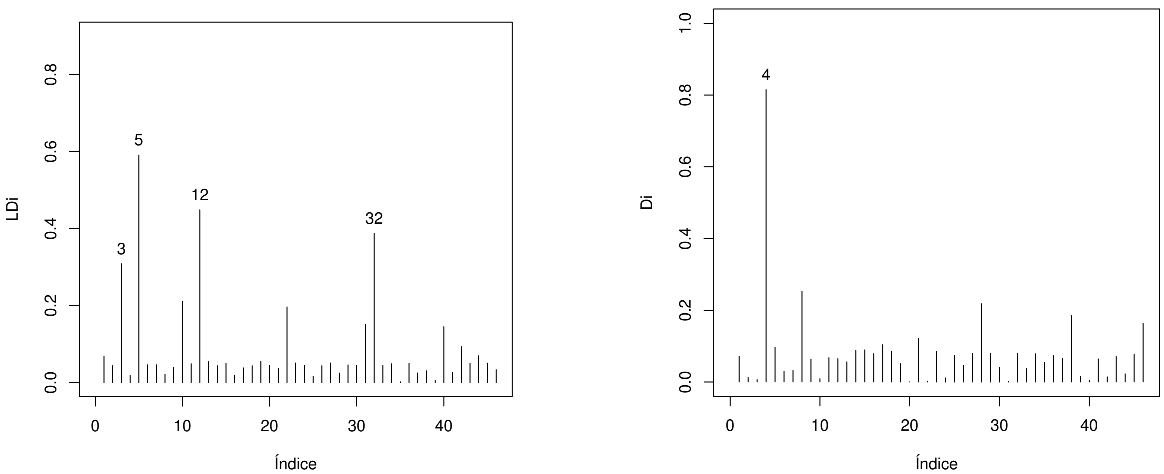


Figura 4.10: Diagnóstico para o conjunto de dados de fadiga. (a) Perturbação de ponderação de casos, (b) Perturbação na variável resposta e (c) Perturbação na variável explicativa.



(a) (b)

Figura 4.11: Influência global para conjunto de dados de fadiga. (a) Influência global LD_i e (b) Influência global D_i.

Tabela 4.5: Mudança nas estimativas dos parâmetros (em %), excluindo as observações potencialmente influentes.

Obs. Eliminada	$\widehat{\beta}_0$	RC	$\widehat{\beta}_1$	RC	$\widehat{\alpha}$	RC	$\widehat{\lambda}$	RC
3	9.946	0.267	-5.592	0.8703	0.682	2.554	-6.675	1.938
4	9.728	1.928	-5.216	5.919	0.696	0.665	-8.200	20.466
5	9.944	0.2497	-5.600	1.0155	0.663	5.391	-6.269	7.896
12	9.926	0.068	-5.562	0.331	0.673	3.975	-6.447	5.279
32	9.901	0.1908	-5.514	0.538	0.675	3.569	-6.437	5.430

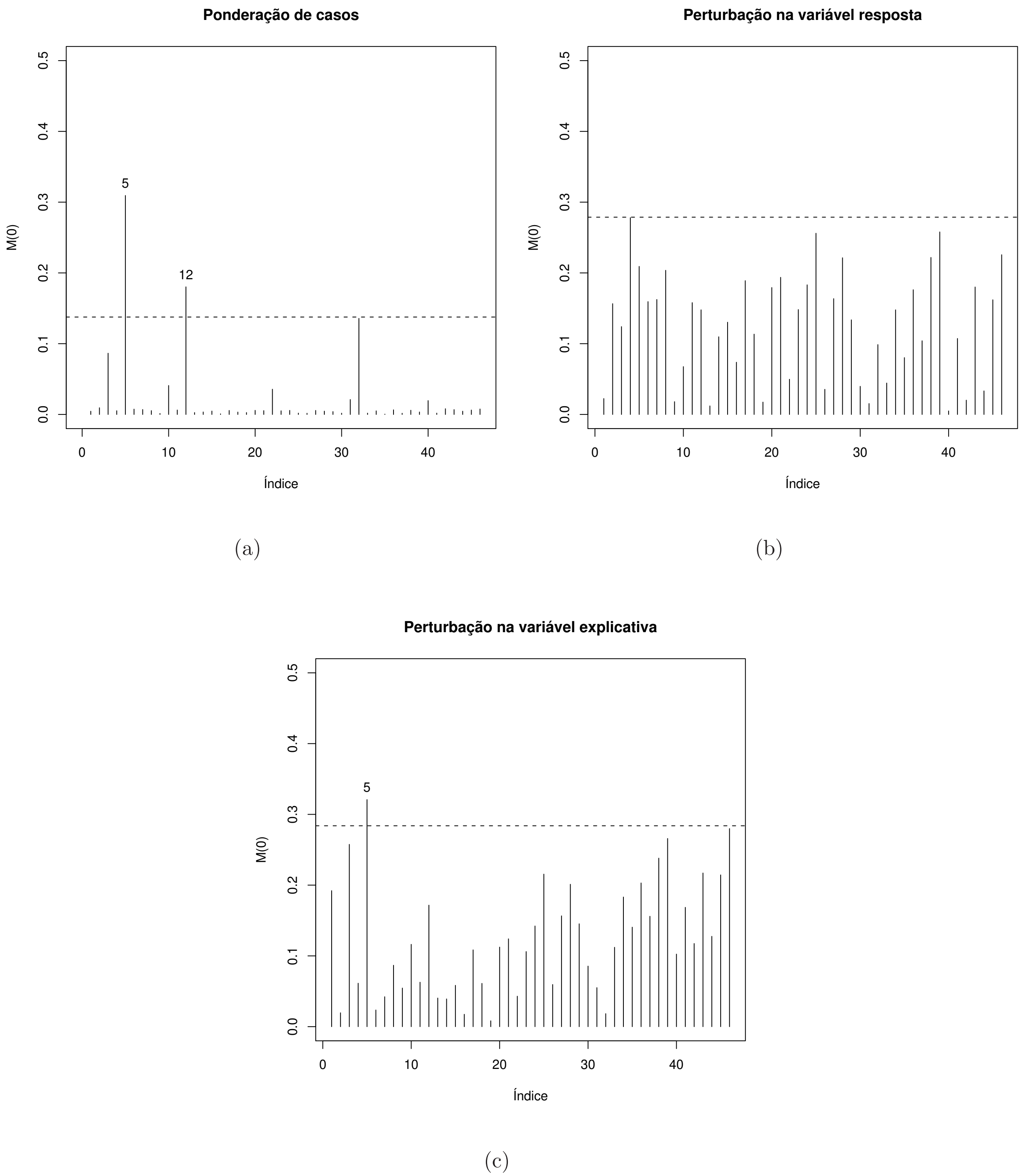


Figura 4.12: Diagnóstico para o conjunto de dados de fadiga considerando a partição $\boldsymbol{\beta}$. (a) Ponderação de casos, (b) Perturbação na variável resposta e (c) Perturbação na explicativa.

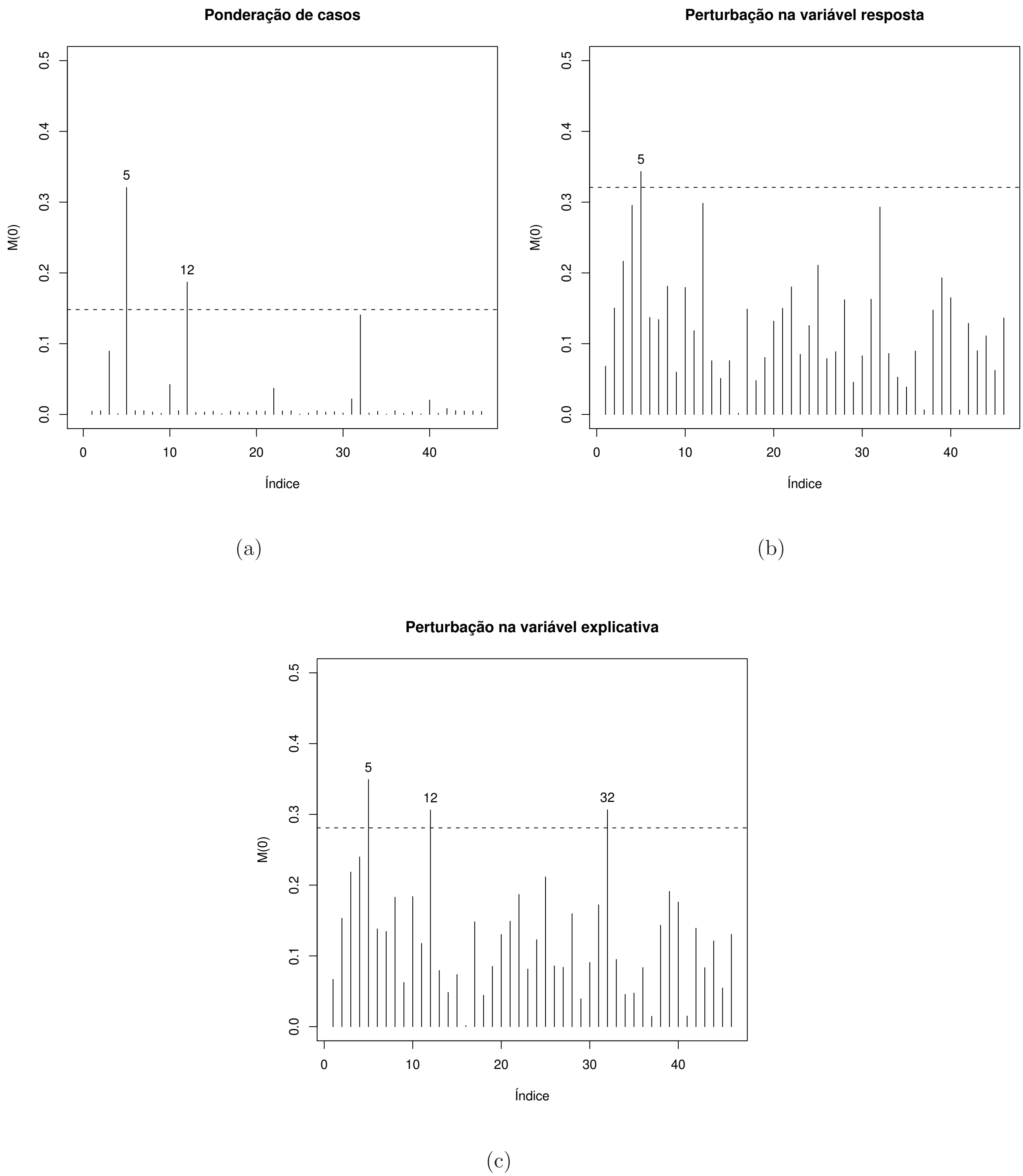


Figura 4.13: Diagnóstico para o conjunto de dados de fadiga considerando a partição α. (a) Ponderação de casos, (b) Perturbação na variável resposta e (c) Perturbação na explicativa.

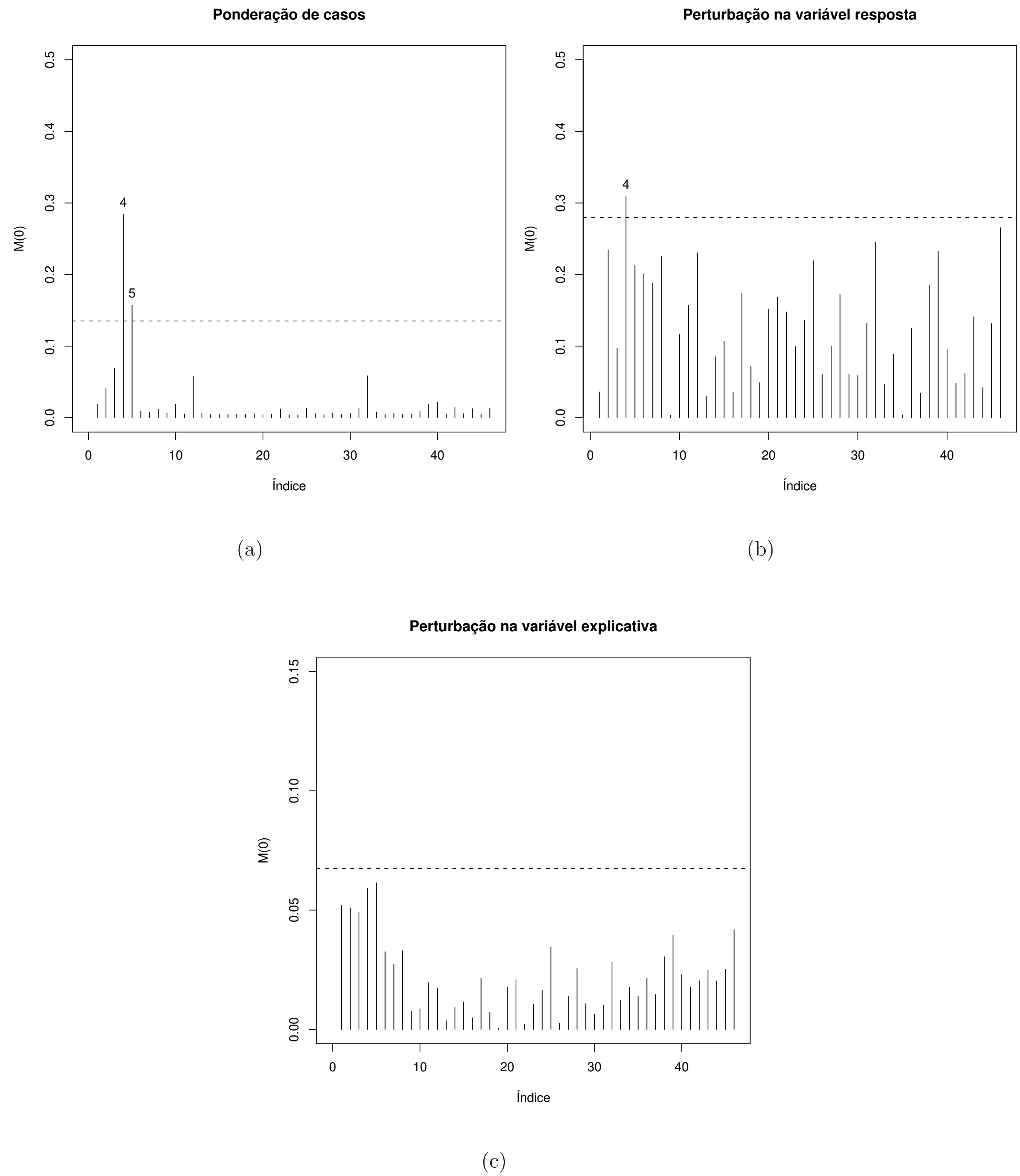


Figura 4.14: Diagnóstico para o conjunto de dados de fadiga considerando a partição λ. (a) Ponderação de casos, (b) Perturbação na variável resposta e (c) Perturbação na explicativa.

CAPÍTULO 5

Considerações Finais

Neste trabalho foi desenvolvido um estudo de estimação e diagnóstico em modelos Birnbaum-Saunders Skew-Normal e em modelos de regressão Log-Birnbaum-Saunders Skew-Normal(log-BSSN). A análise de diagnóstico foi baseada na metodologia de influência local, que foi proposta por Zhu e Lee (2001) para dados incompletos. Nosso resultado generalizou alguns resultados propostos por Birnbaum e Saunders (1969a) e Rieck and Nedelman (1991) para o contexto skew-normal.

Testamos alguns modelos e escolhemos o que apresentou melhor ajuste em relação a distribuição SHSN. Quanto à estimação dos parâmetros dos modelos, foram obtidas soluções numéricas dos estimadores de máxima verossimilhança via algoritmo EM. Foi utilizado o pacote estatístico **R** para as programações dos processos de estimação e diagnósticos do modelo estudado. No caso do modelo de regressão log-BSSN, os valores iniciais para os parâmetros α e $\boldsymbol{\beta}$ foram obtidos usando as estimativas sob o modelo de regressão considerando o caso normal e o valor inicial de λ foi estimado através do método dos momentos. Observamos que o sinal do paâmetro λ é importante, pois determina a assimetria. A inspiração de considerar modelos alternativos ao proposto por Rieck e Nedelman (1991) foi baseada em modelos de vida acelerado (MVA), que podem ser visualizados como uma classe geral de modelos log-lineares.

Em termos de pesquisa futura, no contexto de diagnóstico, estamos considerando outros esquemas de perturbação, por exemplo, perturbar o parâmetro α ou o parâmetro assimetria λ. Além disso, considerar o *leverage* generalizado baseado na verossimilhança completa.

Apêndice A

Neste apêndice apresentamos as tabelas referentes às simulações do Capítulo 3.

Tabela 1: Estimativas dos parâmetros considerando $\lambda = 0.2$.

			Estimativas de α			Estimativas de β			Estimativas de λ		
n	λ	α	Média	Viés rel.	$\sqrt{\text{EQM}}$	Média	Viés rel.	$\sqrt{\text{EQM}}$	Média	Viés rel.	$\sqrt{\text{EQM}}$
10	0.20	0.1	0.0941	-0.0583	0.0220	1.0013	0.0013	0.0135	0.2165	0.0825	0.1065
		0.25	0.2350	-0.0597	0.0551	1.0047	0.0047	0.0343	0.2197	0.0988	0.1077
		0.5	0.4737	-0.0525	0.1119	1.0120	0.0120	0.0684	0.2232	0.1164	0.1111
		0.75	0.7025	-0.0633	0.1659	1.0159	0.0159	0.0957	0.2240	0.1204	0.1173
		1.00	0.9298	-0.0701	0.2238	1.0196	0.0196	0.1245	0.2258	0.1292	0.1254
20	0.20	0.1	0.0972	-0.0279	0.0155	1.0008	0.0008	0.0089	0.2083	0.0418	0.0733
		0.25	0.2433	-0.0265	0.0390	1.0027	0.0027	0.0228	0.2096	0.0483	0.0742
		0.5	0.4858	-0.0282	0.0764	1.0050	0.0050	0.0444	0.2092	0.0462	0.0765
		0.75	0.7274	-0.0300	0.1168	1.0065	0.0065	0.0657	0.2082	0.0413	0.0799
		1.00	0.9692	-0.0307	0.1580	1.0116	0.0116	0.0858	0.2113	0.0568	0.0866
40	0.20	0.1	0.0988	-0.0117	0.0110	1.0003	0.0003	0.0062	0.2034	0.0172	0.0516
		0.25	0.2464	-0.0140	0.0274	1.0010	0.0010	0.0156	0.2038	0.0191	0.0528
		0.5	0.4932	-0.0135	0.0550	1.0023	0.0023	0.0307	0.2039	0.0199	0.0531
		0.75	0.7377	-0.0163	0.0836	1.0040	0.0040	0.0450	0.2047	0.0239	0.0570
		1.00	0.9814	-0.0185	0.1096	1.0060	0.0060	0.0590	0.2063	0.0316	0.0607
60	0.20	0.1	0.0990	-0.0094	0.0090	1.0003	0.0003	0.0050	0.2032	0.0161	0.0422
		0.25	0.2474	-0.0101	0.0223	1.0008	0.0008	0.0127	0.2028	0.0144	0.0432
		0.5	0.4953	-0.0093	0.0450	1.0018	0.0018	0.0246	0.2034	0.0173	0.0434
		0.75	0.7431	-0.0091	0.0684	1.0017	0.0017	0.0359	0.2017	0.0086	0.0457
		1.00	0.9888	-0.0111	0.0905	1.0031	0.0031	0.0485	0.2022	0.0114	0.0510

Tabela 2: Estimativas dos parâmetros considerando $\lambda = 0.5$.

			Estimativas de α			Estimativas de β			Estimativas de λ		
n	λ	α	Média	Viés rel.	$\sqrt{\text{EQM}}$	Média	Viés rel.	$\sqrt{\text{EQM}}$	Média	Viés rel.	$\sqrt{\text{EQM}}$
10	0.50	0.1	0.0946	-0.0539	0.0206	1.0032	0.0032	0.0151	0.5494	0.0989	0.1815
		0.25	0.2383	-0.0467	0.0518	1.0077	0.0077	0.0378	0.5451	0.0902	0.1851
		0.5	0.4727	-0.0544	0.1014	1.0158	0.0158	0.0730	0.5472	0.0944	0.1857
		0.75	0.7097	-0.0536	0.1534	1.0259	0.0259	0.1095	0.5548	0.1096	0.2009
		1.00	0.9415	-0.0584	0.1957	1.0341	0.0341	0.1388	0.5627	0.1254	0.2166
20	0.50	0.1	0.0974	-0.0252	0.0144	1.0017	0.0017	0.0105	0.5211	0.0423	0.1177
		0.25	0.2430	-0.0276	0.0368	1.0040	0.0040	0.0261	0.5210	0.0420	0.1208
		0.5	0.4860	-0.0278	0.0727	1.0092	0.0092	0.0508	0.5244	0.0488	0.1227
		0.75	0.7283	-0.0288	0.1071	1.0107	0.0107	0.0721	0.5214	0.0428	0.1244
		1.00	0.9693	-0.0306	0.1418	1.0190	0.0190	0.0953	0.5291	0.0583	0.1317
40	0.50	0.1	0.0986	-0.0132	0.0103	1.0008	0.0008	0.0071	0.5102	0.0204	0.0803
		0.25	0.2468	-0.0125	0.0259	1.0021	0.0021	0.0182	0.5098	0.0197	0.0813
		0.5	0.4924	-0.0150	0.0516	1.0043	0.0043	0.0352	0.5114	0.0228	0.0832
		0.75	0.7400	-0.0133	0.0757	1.0072	0.0072	0.0508	0.5123	0.0246	0.0849
		1.00	0.9871	-0.0128	0.1000	1.0088	0.0088	0.0662	0.5114	0.0228	0.0907
60	0.50	0.1	0.0991	-0.0087	0.0083	1.0005	0.0005	0.0058	0.5062	0.0125	0.0647
		0.25	0.2478	-0.0085	0.0214	1.0015	0.0015	0.0145	0.5072	0.0145	0.0650
		0.5	0.4947	-0.0105	0.0427	1.0027	0.0027	0.0289	0.5068	0.0137	0.0675
		0.75	0.7426	-0.0097	0.0637	1.0040	0.0040	0.0417	0.5074	0.0148	0.0696
		1.00	0.9902	-0.0097	0.0839	1.0053	0.0053	0.0524	0.5079	0.0158	0.0724

Tabela 3: Estimativas dos parâmetros considerando $\lambda = 1.0$.

			Estimativas de α			Estimativas de β			Estimativas de λ		
n	λ	α	Média	Viés rel.	$\sqrt{\text{EQM}}$	Média	Viés rel.	$\sqrt{\text{EQM}}$	Média	Viés rel.	$\sqrt{\text{EQM}}$
10	1.00	0.1	0.0948	-0.0510	0.0180	1.0030	0.0030	0.0158	1.0819	0.0819	0.3219
		0.25	0.2388	-0.0446	0.0458	1.0082	0.0082	0.0396	1.0870	0.0870	0.3301
		0.5	0.4758	-0.0482	0.0896	1.0155	0.0155	0.0776	1.0861	0.0861	0.3375
		0.75	0.7133	-0.0488	0.1329	1.0243	0.0243	0.1129	1.0939	0.0939	0.3447
		1.00	0.9583	-0.0416	0.1724	1.0306	0.0306	0.1426	1.0928	0.09280	0.3510
20	1.00	0.1	0.0974	-0.0258	0.0129	1.0015	0.0015	0.0111	1.0418	0.0418	0.2200
		0.25	0.2440	-0.0236	0.0322	1.0045	0.0045	0.0285	1.0452	0.0452	0.2251
		0.5	0.4874	-0.0250	0.0645	1.0090	0.0090	0.0545	1.0477	0.0477	0.2262
		0.75	0.7338	-0.0215	0.0933	1.0131	0.0131	0.0795	1.0455	0.0455	0.2264
		1.00	0.9759	-0.0240	0.1232	1.0182	0.0182	0.0992	1.0535	0.0535	0.2341
40	1.00	0.1	0.0987	-0.0127	0.0091	1.0009	0.0009	0.0079	1.0206	0.0206	0.1504
		0.25	0.2468	-0.0127	0.0231	1.0019	0.0019	0.0199	1.0208	0.0208	0.1506
		0.5	0.4939	-0.0120	0.0462	1.0038	0.0038	0.0384	1.0194	0.0194	0.1521
		0.75	0.7407	-0.0123	0.0672	1.0075	0.0075	0.0549	1.0271	0.0271	0.1552
		1.00	0.9882	-0.0117	0.0876	1.0076	0.0076	0.0687	1.0222	0.0222	0.1555
60	1.00	0.1	0.0991	-0.0087	0.0075	1.0005	0.0005	0.0065	1.0119	0.0119	0.1199
		0.25	0.2477	-0.0089	0.0188	1.0012	0.0012	0.0162	1.0119	0.0119	0.1203
		0.5	0.4955	-0.0088	0.0375	1.0032	0.0032	0.0314	1.0165	0.0165	0.1220
		0.75	0.7429	-0.0094	0.0552	1.0042	0.0042	0.0457	1.0148	0.0148	0.1253
		1.00	0.9899	-0.0100	0.0714	1.0049	0.0049	0.0569	1.0170	0.0170	0.1296

Tabela 4: Estimativas dos parâmetros considerando $\lambda = 1.5$.

			Estimativas de α			Estimativas de β			Estimativas de λ		
n	λ	α	Média	Viés rel.	$\sqrt{\text{EQM}}$	Média	Viés rel.	$\sqrt{\text{EQM}}$	Média	Viés rel.	$\sqrt{\text{EQM}}$
10	1.50	0.1	0.0956	-0.0435	0.0155	1.0021	0.0021	0.0144	1.6077	0.0718	0.4425
		0.25	0.2390	-0.0438	0.0394	1.0054	0.0054	0.0366	1.6057	0.0705	0.4450
		0.5	0.4789	-0.0420	0.0795	1.0111	0.0111	0.0699	1.6077	0.0718	0.4465
		0.75	0.7181	-0.0424	0.1167	1.0167	0.0167	0.1019	1.6165	0.0776	0.4551
		1.00	0.9598	-0.0401	0.1528	1.0189	0.0189	0.1256	1.6139	0.0759	0.4567
20	1.50	0.1	0.0976	-0.0231	0.0113	1.0010	0.0010	0.0102	1.5559	0.0373	0.3128
		0.25	0.2438	-0.0246	0.0280	1.0025	0.0025	0.0254	1.5588	0.0392	0.3134
		0.5	0.4889	-0.0221	0.0550	1.0059	0.0059	0.0497	1.5633	0.0422	0.3175
		0.75	0.7344	-0.0206	0.0816	1.0096	0.0096	0.0722	1.5676	0.0451	0.3262
		1.00	0.9789	-0.0210	0.1072	1.01172	0.0117	0.0901	1.5686	0.0457	0.3267
40	1.50	0.1	0.0987	-0.0128	0.0078	1.0005	0.0005	0.0074	1.5314	0.0209	0.2203
		0.25	0.2472	-0.0108	0.0193	1.0012	0.0012	0.0184	1.5276	0.0184	0.2211
		0.5	0.4943	-0.0112	0.0390	1.0035	0.0035	0.0355	1.5324	0.0216	0.2215
		0.75	0.7418	-0.0109	0.0579	1.0038	0.0038	0.0517	1.5287	0.0191	0.2231
		1.00	0.9865	-0.0134	0.0760	1.0042	0.0042	0.0644	1.5324	0.0216	0.2306
60	1.50	0.1	0.0993	-0.0069	0.0062	1.0004	0.0004	0.0059	1.5190	0.0127	0.1760
		0.25	0.2480	-0.0076	0.0123	1.0008	0.0008	0.0150	1.5185	0.0123	0.1792
		0.5	0.4957	-0.0084	0.0317	1.0017	0.0017	0.0288	1.5197	0.0131	0.1800
		0.75	0.7451	-0.0064	0.0461	1.0023	0.0023	0.0416	1.5173	0.0115	0.1809
		1.00	0.9914	-0.0085	0.0613	1.0026	0.0026	0.0522	1.5240	0.0160	0.1873

Apêndice B

Neste apêndice apresentamos a demonstração da função geradora de momentos da distribuição Senh-Skew-Normal (SHSN) dada na seção 4.1.

$$
\begin{aligned}
M_Y(s) &= \mathbb{E}[e^{sY}] \\
&= \int_{-\infty}^{\infty} e^{sy}\left[\left(\frac{2}{\alpha\sqrt{2\pi}}\right)\exp\left\{-2\alpha^{-2}senh^2\left(\frac{y-\gamma}{2}\right)\right\}cosh\left(\frac{y-\gamma}{2}\right)\Phi(\lambda b_y)\right]dy \\
&= \int_{-\infty}^{\infty} e^{sy}\left[\left(\frac{2}{\alpha\sqrt{2\pi}}\right)\exp\left\{-2\alpha^{-2}senh^2\left(\frac{y-\gamma}{2}\right)\right\}\left(\frac{e^{-\frac{y+\gamma}{2}}+e^{\frac{y-\gamma}{2}}}{2}\right)\Phi(\lambda b_y)\right]dy \\
&= \int_{-\infty}^{\infty} e^{sy}\left[\left(\frac{2}{\alpha\sqrt{2\pi}}\right)\exp\left\{-\frac{1}{\alpha^2}\left(cosh2\left(\frac{y-\gamma}{2}\right)-1\right)\right\}\left(\frac{e^{-\frac{y+\gamma}{2}}+e^{\frac{y-\gamma}{2}}}{2}\right)\Phi(\lambda b_y)\right]dy \\
&= \frac{1}{\alpha\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{sy}\exp\left\{\frac{1}{\alpha^2}\right\}\exp\left\{-\frac{1}{\alpha^2}cosh(y-\gamma)\right\}\left(e^{-\frac{y+\gamma}{2}}+e^{\frac{y-\gamma}{2}}\right)\Phi(\lambda b_y)dy
\end{aligned}
$$

fazendo $y = z + \gamma$, temos que

$$\begin{aligned}
M_Y(s) &= \frac{1}{\alpha\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{s(z+\gamma)} \exp\left\{\frac{1}{\alpha^2}\right\} \exp\left\{-\frac{1}{\alpha^2} cosh(z)\right\} \left(e^{-\frac{z}{2}} + e^{\frac{z}{2}}\right) \Phi\left(\lambda \frac{2}{\alpha} senh\left(\frac{z}{2}\right)\right) dz \\
&= \frac{e^{\alpha^{-2}+s\gamma}}{\alpha\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} e^{sz} \exp\left\{-\frac{1}{\alpha^2} cosh(z)\right\} \Phi\left(\lambda \frac{2}{\alpha} senh\left(\frac{z}{2}\right)\right) e^{-\frac{z}{2}} dz \right. \\
&\quad \left. + \int_{-\infty}^{\infty} e^{sz} \exp\left\{-\frac{1}{\alpha^2} cosh(z)\right\} \Phi\left(\lambda \frac{2}{\alpha} senh\left(\frac{z}{2}\right)\right) e^{\frac{z}{2}} dz \right] \\
&= \frac{e^{\alpha^{-2}+s\gamma}}{\alpha\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{\alpha^2} cosh(z) - z\left(\frac{1}{2} - s\right)\right\} \Phi\left(\lambda \frac{2}{\alpha} senh\left(\frac{z}{2}\right)\right) dz \right. \\
&\quad \left. + \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{\alpha^2} cosh(z) + z\left(\frac{1}{2} + s\right)\right\} \Phi\left(\lambda \frac{2}{\alpha} senh\left(\frac{z}{2}\right)\right) dz \right] \\
&= \frac{e^{\alpha^{-2}+s\gamma}}{\alpha\sqrt{2\pi}} \left[\frac{1}{2} \int_{-\infty}^{\infty} 2\exp\left\{\frac{(2s-1)}{2} z - \frac{1}{\alpha^2} cosh(z)\right\} \Phi\left(\lambda \frac{2}{\alpha} senh\left(\frac{z}{2}\right)\right) dz \right. \\
&\quad \left. + \frac{1}{2} \int_{-\infty}^{\infty} 2\exp\left\{\frac{(2s+1)}{2} z - \frac{1}{\alpha^2} cosh(z)\right\} \Phi\left(\lambda \frac{2}{\alpha} senh\left(\frac{z}{2}\right)\right) dz \right] \\
&= 2\frac{\exp\left(\alpha^{-2} + s\gamma\right)}{\alpha\sqrt{2\pi}} \left[K_a(\alpha^{-2};\lambda) + K_b(\alpha^{-2};\lambda)\right] \\
&= \exp\left(\gamma s\right) \left[\frac{K_a(\alpha^{-2};\lambda) + K_b(\alpha^{-2};\lambda)}{K_{\frac{1}{2}}(\alpha^{-2})}\right], \quad \text{s} \in \mathbb{R},
\end{aligned}$$

em que $a = (2s+1)/2$, $b = (2s-1)/2$, $K_\nu(\omega;\lambda) = (1/2)\int_0^\infty \exp(-\omega\cosh(u) - \nu u)\Phi\left(\frac{2\lambda}{\alpha} senh(\frac{u}{2})\right)du$ e $K_\nu(\omega, \lambda = 0) = (1/2)K_\nu(\omega)$, com $K_\nu(.)$ é a função de Bessel modificada do terceiro tipo. Para mais detalhes veja Gradshteyn e Randzhik (2000), p. 907.

Apêndice C

Neste apêndice apresentamos de forma resumida o método de influência local proposto por Cook(1986).

Para um conjunto de dados observados, seja $l(\boldsymbol{\theta}) = \sum_{i=1}^{n} l_i(\boldsymbol{\theta})$ a função log-verossimilhança do modelo postulado, onde $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \alpha, \lambda)^\top$ é um vetor $(p+2) \times 1$ de parâmetros desconhecidos e $l_i(\boldsymbol{\theta})$ é a contribuição da i-ésima observação à log-verossimilhança. Introduziremos perturbações no modelo através de um vetor $\boldsymbol{\omega}$ de dimensão $q \times 1$, onde $\boldsymbol{\omega} \in \boldsymbol{\Omega} \subseteq \mathbb{R}^q$, $\boldsymbol{\Omega}$ aberto. Geralmente $\boldsymbol{\omega}$ pode refletir qualquer esquema de perturbação bem definida. Por exemplo, $\boldsymbol{\omega}$ pode ser usado para introduzir uma menor modificação nas variáveis explicativas ou para perturbar a matriz de covariância nos erros no modelo de regressão linear (veja, Galea, Paula e Bolfarine, 1997). Seja $l(\boldsymbol{\theta}/\boldsymbol{\omega})$ a função de log-verossimilhança do modelo perturbado. Assumiremos que existe um $\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}$ tal que $l(\boldsymbol{\theta}) = l(\boldsymbol{\theta}/\boldsymbol{\omega}_0)$, para todo $\boldsymbol{\theta}$ e que $l(\boldsymbol{\theta}/\boldsymbol{\omega})$ é duas vezes diferenciável.

Sejam $\widehat{\boldsymbol{\theta}}$ e $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ os EMV de $\boldsymbol{\theta}$ sob o modelo postulado e perturbado, respectivamente. O objetivo é comparar $\widehat{\boldsymbol{\theta}}$ e $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ quando $\boldsymbol{\omega}$ varia em $\boldsymbol{\Omega}$. Cook (1986) sugere que a comparação entre $\boldsymbol{\theta}$ e $\boldsymbol{\theta}_{\boldsymbol{\omega}}$ seja feita através do afastamento pela verossimilhança $LD(\boldsymbol{\omega})$, definida por

$$LD(\boldsymbol{\omega}) = 2\left[l(\widehat{\boldsymbol{\theta}}) - l(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})\right], \quad \boldsymbol{\omega} \in \boldsymbol{\Omega}. \tag{1}$$

O afastamento pela verossimilhança descrito acima é uma medida de influência que permite avaliar a perturbação $\boldsymbol{\omega}$ na estimação de máxima verossimilhança do vetor paramétrico completo $\boldsymbol{\theta}$.

O sentido da distância entre $\widehat{\boldsymbol{\theta}}$ e $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$, baseado na função de afastamento de verossimilhança $LD(\boldsymbol{\omega})$, pode depender da concavidade da função de log-verossimilhança $l(\boldsymbol{\theta})$. Se $l(\boldsymbol{\theta})$ é suficientemente achatada, pode-se dizer que $\widehat{\boldsymbol{\theta}}$ e $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ estão próximos entre si, enquanto que se $l(\boldsymbol{\theta})$ for suficientemente concentrada em torno de $\widehat{\boldsymbol{\theta}}$ estas estimativas podem estar distantes entre si (veja, Cook, 1986).

A idéia de Cook (1986) é estudar o comportamento da função $LD(\boldsymbol{\omega})$ numa vizinhança de $\boldsymbol{\omega}_0$, que é o ponto em que as duas verossimilhanças são iguais. Para isso considerou o gráfico de $LD(\boldsymbol{\omega})$ versus $\boldsymbol{\omega}$ que pode ser visto como a superfície geométrica formada pelos valores do vetor $(q+1)$-dimensional

$$\boldsymbol{\alpha}(\boldsymbol{\omega}) = \begin{pmatrix} \boldsymbol{\omega} \\ LD(\boldsymbol{\omega}) \end{pmatrix},$$

onde $\boldsymbol{\omega}$ varia em $\boldsymbol{\Omega}$. A função $\boldsymbol{\alpha}(\boldsymbol{\omega})$ é chamada de gráfico de influência local, esta é uma superfície em $\mathbb{R}^{q+1}$ e pode ser usada para avaliar a influência ao variar $\boldsymbol{\omega}$ através de $\boldsymbol{\Omega}$. O método de influência local utiliza a curvatura normal de $\boldsymbol{\alpha}(\boldsymbol{\omega})$ em $\boldsymbol{\omega}_0$ que é um ponto de mínimo local da função $LD(\boldsymbol{\omega})$.

O método de Cook (1986), baseado na superfície $\boldsymbol{\alpha}(\boldsymbol{\omega})$ em torno de $\boldsymbol{\omega}_0$, consiste em considerar o plano tangente ($\mathbf{T}_o$) à superfície $\boldsymbol{\alpha}(\boldsymbol{\omega})$ em $\boldsymbol{\omega}_0$. Como $LD(\boldsymbol{\omega})$ atinge o mínimo em $\boldsymbol{\omega}_0$, temos que $\mathbf{T}_o$ é paralelo a $\boldsymbol{\Omega} \subseteq \mathbb{R}^q$. Cada vetor $\mathbf{h}$ em $\boldsymbol{\Omega}$, de norma um, determina um plano que contém $\mathbf{h}$ e que é ortogonal a $\mathbf{T}_o$. A intersecção, chamada de seção normal, deste plano com a superfície é chamada de linha levantada (tradução livre de "lifted line") e pode ser obtida ao considerar o gráfico de $LD(\boldsymbol{\omega}_0 + a\mathbf{h})$ em função de a, onde $a \in \mathbb{R}$. A curvatura normal da linha levantada, denotada por $C_{\mathbf{h}}$, é agora definida como a curvatura da curva plana $(a, LD(\boldsymbol{\omega} + a\mathbf{h}))$ em $a = 0$ e pode ser visualizada como a inversa do raio do círculo que melhor se ajusta em $\boldsymbol{\omega}_0$. Valores "grandes" de $C_{\mathbf{h}}$ indicam sensibilidade em relação à perturbação considerada na direção $\mathbf{h}$.

Cook (1986) mostra que a curvatura normal na direção $\mathbf{h}$ pode ser expressa na seguinte forma:

$$C_{\mathbf{h}} = 2\left|\mathbf{h}^\top \ddot{\mathbf{F}}\mathbf{h}\right|, \tag{2}$$

onde $\ddot{\mathbf{F}} = \boldsymbol{\Delta}^\top(-\ddot{\mathbf{L}})^{-1}\boldsymbol{\Delta}$, com $\boldsymbol{\Delta}$ sendo uma matriz $q \times q$ com elementos $\boldsymbol{\Delta}_{ij} = \dfrac{\partial^2 l(\boldsymbol{\theta}/\boldsymbol{\omega})}{\partial\theta_i \partial\boldsymbol{\omega}_j}$, $i = 1, \ldots, p$, $j = 1, \ldots, q$ e $(-\ddot{\mathbf{L}})$ é a matriz de informação observada para o modelo postulado avaliada no ponto $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. O resultado na equação acima pode ser utilizado para avaliar a influência que o esquema de perturbações considerado exerce sobre as componentes do modelo, tais como estimativa dos parâmetros e outros resultados de análise.

Existem muitas formas na qual a função (2) pode ser usada para estudar $\boldsymbol{\alpha}(\boldsymbol{\omega})$, cada uma correspondendo a uma específica escolha do vetor unitário $\mathbf{h}$. Seja $C_{max} = \max_{\mathbf{h}:\|\mathbf{h}\|=1} C_{\mathbf{h}}$ que corresponde ao maior autovalor de $2\mathbf{F}$ e seja $\mathbf{h}_{max}$ o autovetor associado ao maior autovalor C_{max}. Neste caso, $\mathbf{h}_{max}$ é a direção que produz a maior mudança local nas estimativas dos parâmetros, sendo que os elementos mais influentes podem ser identificados através do "tamanho" relativo dos elementos em $\mathbf{h}_{max}$, enquanto que a sensibilidade em relação à perturbação causada pode ser indicada pelo valor de C_{max}.

A metodologia proposta por Cook (1986) vem sendo aplicada com sucesso em diversas áreas da estatística aplicada, porém observamos que dependendo da complexidade do modelo, a aplicação desta metodologia envolve extensas manipulações algébricas. Uma alternativa interessante à proposta de Cook (1986) e útil nos casos em que resulta difícil ou impossível aplicar diretamente os métodos apresentados por Cook (1986), foi proposta por Zhu e Lee (2001), que inspirados pela ideia básica do algoritmo EM propuseram um método unificado para análise de influência local em modelos estatísticos com dados faltantes, através da função de afastamento da verossimilhança completa.

Quando o interesse for um subconjunto de θ_1 de $\theta^\top = (\theta_1, \theta_2)^\top$, o equivalente a (1) é dado por

$$LD_s(\boldsymbol{\omega}) = 2\left[l(\widehat{\boldsymbol{\theta}}) - l(\widehat{\boldsymbol{\theta}}_{1\omega}, g(\widehat{\boldsymbol{\theta}}_{1\omega}))\right], \tag{3}$$

onde g é a função que maximiza $l(\theta_1, \theta_2)$ para cada θ_1 fixo de $\widehat{\theta}_{1\omega}$ e é determinada pela partição de $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}^\top = (\widehat{\theta}_{1\omega}^\top, \widehat{\theta}_{2\omega}^\top)^\top$. Neste caso, a curvatura normal na direção $\mathbf{h}$ é dada por

$$C_{\mathbf{h}}(\theta_1) = 2\left[\mathbf{h}^\top\boldsymbol{\Delta}^\top\left(\ddot{\mathbf{L}} - B_{22}\right)\boldsymbol{\Delta}\mathbf{h}\right], \tag{4}$$

onde $\|\mathbf{h}\| = 1$, $B_{22} = \begin{pmatrix} 0 & 0 \\ 0 & \ddot{L}_{22}^{-1} \end{pmatrix}$, e $\ddot{L}_{22}$ é determinada pela partição de $\ddot{\mathbf{L}} = \begin{pmatrix} \ddot{L}_{11} & \ddot{L}_{12} \\ \ddot{L}_{21} & \ddot{L}_{22} \end{pmatrix}$.

Apêndice C

Neste apêndice apresentamos os dados reais utilizados no Capítulo 4.

Tabela 5: Dados de fadiga.

Peça	**X** (mJ/m^3)	**Ciclos de vida**	**Peça**	**X** (mJ/m^3)	**Ciclos de vida**
1	11.5	3280	24	34.5	736
2	13.0	5046	25	40.1	750
3	14.3	1563	26	40.1	316
4	15.6	4707	27	43.0	456
5	16.0	977	28	44.1	552
6	17.3	2834	29	46.5	355
7	19.3	2266	30	47.3	242
8	21.1	2208	31	48.7	190
9	21.5	1040	32	52.9	127
10	22.6	700	33	56.6	185
11	22.6	1583	34	59.9	255
12	24.0	482	35	60.2	195
13	24.0	804	36	60.3	283
14	24.6	1093	37	60.5	212
15	25.2	1125	38	62.1	327
16	25.5	884	39	62.8	373
17	26.3	1300	40	66.5	125
18	27.9	852	41	67.0	187
19	28.3	580	42	67.1	135
20	28.4	1066	43	67.9	245
21	28.6	1114	44	68.8	137
22	30.9	386	45	75.4	200
23	31.9	745	46	100.5	190

BIBLIOGRAFIA

[1] Achcar, J. A. (1993). Inference for the Birnbaum-Saunders fatigue life model using Bayesian methods. *Computational Statistics* & *Data Analysis*, **15**, 367–380.

[2] Akaike, H. (1974). A new look at the statistical model identification. *IEEE. Transactions on Automatic Control*, **19**, 716–723.

[3] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal Statistics*, **12**, 171–178.

[4] Bagdonavicius, V. e Nikulin, M. (2002). *Accelerated Life Models: Modeling and Statistical Analysis*. Boca Raton, FL: Chapman and Hall/CRC.

[5] Barros, M., Paula, G. A. e Leiva, V. (2007). *Modelo de regressão t-Student log-Birnbaum-Saunders*. 10a Escola de Modelos de Regressão, Salvador. Programa e Resumos. São Paulo : ABE-Associação Brasileira de Estatística.

[6] Barros, M., Paula, G. A. e Leiva, V. (2009). An R implementation for generalized Birnbaum-Saunders distributions. *Computational Statistics* & *Data Analysis*, **53**, 1511–1528.

[7] Belsley, D.A. (1991). *Conditioning diagnostics: collinearity and weak data in regression*. Wiley, New York.

[8] Birbnaum, Z. W. e Saunders, S. C. (1969a). A new family of life distributions. *Journal of Applied Probability*, **6**, 319–327.

[9] Birnbaum, Z. W. e Saunders, S. C. (1969b). Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability*, **6**, 328–347.

[10] Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society*, B, **48**, 133–169.

[11] Christensen, R. (1997). *Log-Linear Models ans Logistic Regression*. 2ed edition. New York, Springer.

[12] 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*, **39**, 1–22.

[13] Desmonde, A. F. (1985). Stochastic models of failure in random environments. *Canadian Journal of Statistics*, **13**, 171–183.

[14] Desmonde, A. F. (1986). On the relationship between two fatigue-life models. *IEEE Transactions on Reliability*, **35**, 167–169.

[15] Díaz–García, J. A. e Leiva, V. (2005). A new family of life distributions based on the contoured elliptically distributions. *Journal of Statistical Planning and Inference*, **128**, 88–95.

[16] Dupuis, D. J. e Mills, J. E. (1998). Robust estimation of the Birnbaum-Saunders distribution. *IEEE Transactions on Reliability*, **47**, 88–95.

[17] Engelhardt, M., Bain, L. J. e Wright, F. T. (1981). Inferences on the parameters of the Birnbaum-Saunders fatigue life distribution based on maximum likelihood estimation. *Technometrics*, **23**, 251–255.

[18] Freitas, M. A. e Colosimo, E. A. (1997). *Confiabilidade: análise de tempo de falha e testes de vida acelerados*. Belo Horizonte: Fundação Christiano Ottoni, Universidade Federal de Minas Gerais.

[19] Galea, M., Paula, G.A. e Bolfarine, H. (1997). Local influence in elliptical linear regression models. *The Statistician*, **46**, 71–79.

[20] Galea, M., Leiva, V. e Paula, G. A. (2004). Influence diagnostics in log-Birnbaum-Saunders regression models. *Journal of Applied Statistics*, **31**, 1049–1064.

[21] Gokhale, S. e Khare, M. (2004) A review of deterministic, stochastic and hybrid vehicular exhaust emission models International. *J. Transp. Manag.*, **2**, 59–74.

[22] Gradshteyn, I. S. e Randzhik, I. M. (2000). *Table of Integrals, Series, and Products.* Academic Press, New York.

[23] Gupta, R. C. e Kirmani, S. N. U. A. (1998). Residual life function in reliability studies. *Frontiers in Reliability*, **4**, 175–190.

[24] Gupta, R. C. e Brown, N. (2001). Reliability studies of the skew-normal distribution and its application to a strength-stress model. *Comm. Stat. Theoret. Methods*, **30 (11)**, 2427–2445.

[25] Hill, M.A. e Dixon, W.J. (1982). Robustness in real life: a study of clinical laboratory data. *Biometrics*, **38**, 377–396.

[26] Jeng, S. L. (2003). Inferences for the fatigue life model based on the Birnbaum-Saunders distribution. *Commun. Statist. Simul. Comp.*, **32 (1)**, 43–60.

[27] Johnson, N. L., Kotz, S. e Kemp, A. (1993). *Univariate Discrete Distributions.* Vol. 2. New York: Wiley & Sons.

[28] Johnson, N., Kotz, S. e Balakrishnan, N. (1995). *Continuous Univariate Distributions.* Volume 2 , 2nd ed. John Wiley & Sons, New York.

[29] Kass, R. E. e Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, **90**, 773–795.

[30] Lachos, V. H. (2004). *Modelos lineares mistos assimétricos.* Tese de Doutorado, Departamento de Estatística, IME-USP. São Paulo.

[31] Lee, S. e Xu, L. (2004). Influence analysis of nonlinear mixed-efects models. *Computational Statistics and Data Analysis*, **45**, 321–341.

[32] Leiva, V., Hernández, H. e Riquelme, M. (2006). A new package for the Birnbaum-Saunders distribution. *R News*, **6**, 35–40.

[33] Leiva, V., Barros, M., Paula, G. A. e Galea, M. (2007a). Influence Diagnostics in Log-Birnbaum-Saunders Regression Models with Censored Data. *Computational Statistics & Data Analysis*, **51**, 5694–5707.

[34] Leiva, V., Barros, M., Paula, G. e Sanhueza, A. (2007b). Generalized Birnbaum-Saunders distribution applied to air pollutant concentration. *Environmetrics*, **19**, 235–249.

[35] Leiva, V., Vilca-Labra, F. E., Balakrishnan, N. e Sanhueza, A. (2008). A skewed sinh-normal distribution and its properties and application to air pollution. *Comm. Stat. Theoret. Methods. Submetido.*

[36] Leiva, V., Sanhueza, A. e Angulo, J. M. (2009a). A length-biased version of the Birnbaum-Saunders distribution with application in water quality. *Stoch Environ Res Risk Assess*, **23**, 299–307.

[37] Leiva, V., Rojas, E., Galea, M. e Sanhueza, A. (2009b) Influence diagnostics in a Birnbaum-Saunders accelerated life model. *Submetido.*

[38] Lesafre, E. e Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, **54**, 570–582.

[39] Lu, M. e Chang, D. S. (1997). Bootstrap prediction intervals for the Birnbaum-Saunders distribution. *Microelectron Reliability*, **37**, 1213–1216.

[40] Mann, N. R., Schafer, R. E. e Singpurwalla, N. (1974). *Methods for Statistical Analysis of Reliability and Life Data*. John Wiley & Sons, New York.

[41] Marshall, A. W. e Olkin, I. (2007) *Life Distributions-Structure of Nonparametric, Semiparametric, and Parametric Families*. Springer, New York.

[42] McLachlan, G. J. e Krishnan, T. (1997). *The EM Algorithm and Extensions*. New York: John Wiley & Sons.

[43] Nadarajah, S. (2007). A truncated inverted beta distribution with application to air pollution data. *Stoch. Environ. Res. Risk. Assess.*, **22**, 285–289.

[44] Nelson, W. B. (2004). *Accelerated Testing : Statistical Models, Test Plans, and Data Analysis*. New York: Wiley.

[45] Ng, H. K., Kundu, D. e Balakrisnan, N. (2003). Modified moment estimation for the two-parameter Birnbaum-Saunders distribution. *Computational Statistics* & *Data Analysis*, **43**, 283–298.

[46] O'Hagan, A. e Leonard, T. (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, **62**, 607–614.

[47] Ortega, E. M. M., Bolfarine, H. e Paula, G. A. (2003). Influence diagnostics in generalized log-gamma regression models. *Computational Statistics* & *Data Analysis*, **42**, 165–186.

[48] Owen, W.J. e Padgett, W.J. (1999). Accelerated test models for system strength based on Birnbaum-Saunders distribution. *Lifetime Data Anal.*, **5**, 133–147.

[49] Poon, W.Y. e Poon, Y.S. (1999). Conformal normal curvature e assessment of local influence. *Journal of the Royal Statistical Society*, B, **61**, 51–61.

[50] Raftery, A. E. (1995). Bayesian Model Seletion in Social Research. *Sociological Methodology*, **25**, 111–163.

[51] Rieck, J. R. (1989). *Statistical Analysis for the Birnbaum-Saunders Fatigue Life Distribution*. Ph. D. Thesis, Clemson University, Department of Mathematical Sciences.

[52] Rieck, J. R. e Nedelman, J. R. (1991). A log-linear model for the Birnbaum-Saunders distribution. *Technometrics*, **33**, 51–60.

[53] Rieck, J. R. (1999). A moment–generating function with application to the Birnbaum-Saunders distribution. *Communications in Statistics – Theory and Methods*, **28**, 2213–2222.

[54] Saunders, S. C. (1974). A family of random variables closed under reciprocation. *Journal of the American Statistical Association*, **69**, 533–539.

[55] Schwarz, G. (1978). Estimating the dimension of a mode. *Annals of Statistics*, **6**, 461–464.

[56] Shaked, M. e Shanthikumar, J. G. (1994) *Stochastic Orders and Their Applications*. Academic Press: New York.

[57] Spieglhaiter, D. J., Best, N. G., Carlin, B. P. e van der Linde, A. (2002). Bayesian measures of complexity and fit. *Journal of the Royal Statistical Society Series B*, **64**, 1–34.

[58] Vilca-Labra, F. E. e Leiva, V. (2006). A new fatigue life model based on the family of skew–elliptical distributions. *Communications in Statistics–Theory and Methods*, **35**, 229–244.

[59] Xie, F. C. e Wei, B. C. (2007). Diagnostics analysis for log-Birnbaum-Saunders regression models. *Computational Statistics* & *Data Analysis*, **51**, 4692–4706.

[60] Wei, B.C., Qu, Y.Q. e Fung, W.K. (1998). Generalized leverage and its applications. *Scandinavian Journal of Statistics*, **25**, 25–37.

[61] Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics*, **11**, 95–103.

[62] Zhu, H. e Lee, S. (2001). Local influence for incomplete-data models. *Journal of the Royal Statistical Society*, B, **63**, 111–126.