Extensões do modelo α-potência

Guillermo Domingo Martínez Flórez

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

Área de Concentração: **Estatística**
Orientador: **Prof. Dr. Heleno Bolfarine**
Co-orientador: **Prof. Dr. Héctor W. Gómez**

Durante a elaboração deste trabalho o autor
recebeu apoio financeiro da CNPq.

– São Paulo, junho de 2011 –

Extensões do modelo α-Potência

Este exemplar corresponde à redação final
da tese devidamente corrigida e defendida
por Guillermo Domingo Martínez Flórez
e aprovada pela comissão julgadora.

São Paulo, 22 de junho de 2011.

Comissão Julgadora:

- Prof. Dr. Heleno Bolfarine (Presidente) - IME-USP
- Prof. Dr. Gilberto Alvarenga Paula - IME-USP
- Prof. Dr. Víctor Hugo Lachos Davila - UNICAMP
- Prof. Dr. Filidor Edilfonso Vilca Labra - UNICAMP
- Prof. Dr. Hugo Segundo Salinas Pérez - U. de Atacama

"O temor do Senhor é o princípio da sabedoria, têm bom entendimento todos os que cumprem os seus preceitos".

Salmo 111:10.

À meu pai, in memoriam,
minhas filhas, María José e Mailen
minha mãe, irmãos e sobrinhos.

Agradecimentos

- Acima de tudo, agradeço a Deus.
- A minha mãe, meus irmãos, meus sobrinhos e minhas filhas María José e Mailen pelo apoio incondicional e as forças que sempre me brindaram em todo momento.
- A meu orientador, Prof. Dr. Heleno Bolfarine, pela dedicação, compreensão e a confiança em mim depositada, pela orientação clara e objetiva que recebi ao longo do desenvolvimento deste trabalho.
- A meu co-orientador, Prof. Dr. Héctor Gómez, por aceitar ser meu co-orientador e pela orientação que recebi ao longo do desenvolvimento deste trabalho.
- Aos membros da banca examinadora pelas sugestões, aportes e revisão deste trabalho.
- Aos professores Barry Arnold e Hugo Salinas, por sua participação como co-autores no desenvolvimento e a escrita de alguns dos artigos deste trabalho.
- A Adriana por ter compartido comigo parte deste tempo, pela sua paciência, apoio e compreensão durante os dias e momentos de ausência.
- A meus companheiros do IME, em especial Artur, Alexandre, Rafhael e Amanda, pelas sugestões e troca de conhecimentos. Nubia, Luz Marina, Betsabe, Paola, Marcos, Germán, Luz Mery, Mary Luz, Diana, Eliane, Alexander, Glauce, Michel, Lisandra e Jalmar pelo companheirismo e amizade.
- A Milena por cuidar de meus dois mais grandes tesouros.
- À CNPq pelo apoio financeiro.

- A todos os meus professores do IME pelos seus ensinamentos durante meus estudos de doutorado.
- À Universidade de Córdoba por me dar a oportunidade de vir estudar neste fabuloso pais e nesta excelente Universidade.
- A todos os professores e funcionários do IME-USP que tenham colaborado na realização deste trabalho.
- Ao projeto FONDECIT 1090411(Chile) pelo apoio econômico na minha visita a Chile.

Resumo

Em análise de dados que apresentam certo grau de assimetria a suposição que as observações seguem uma distribuição normal, pode resultar ser uma suposição irreal e a aplicação deste modelo pode ocultar características importantes do modelo verdadeiro. Este tipo de situação deu força à aplicação de modelo assimétricos, destacando-se entre estes a família de distribuições *skew-symmetric*, desenvolvida por Azzalini (1985). Neste trabalho nós apresentamos uma segunda proposta para a análise de dados com presença importante de assimetria e/ou curtose, comparado com a distribuição normal. Nós apresentamos e estudamos algumas propriedades dos modelos α-potência e log-α-potência, onde também estudamos o problema de estimação, as matrizes de informação observada e esperada de Fisher e o grau do viés dos estimadores mediante alguns processos de simulação. Nós introduzimos um modelo mais flexível que o modelo α-potência do qual derivamos o caso bimodal desta distribuição e introduzimos os modelos bimodal simétrico e assimétrico α-potência. Posteriormente nós estendemos a distribuição α-potência para o caso do modelo Birnbaum-Saunders, estudamos as propriedades deste novo modelo, desenvolvemos estimadores para os parâmetros e propomos estimadores com viés corrigido. Também introduzimos o modelo de regressão α-potência para dados censurados e não censurados e para o modelo de regressão log-linear Birnbaum-Saunders; aqui nós derivamos os estimadores dos parâmetros e estudamos algumas técnicas de validação dos modelos. Por último nós fazemos a extensão multivariada do modelo α-potência e estudamos alguns processos de estimação dos parâmetros. Para todos os casos estudados apresentam-se ilustrações com dados já analisados previamente com outras suposições de distribuições.
Palavras Chaves: estimação de máxima verossimilhança, matriz de informação de Fisher, teste de razão de verossimilhança, distribuição assintótica.

Abstract

In data analysis where data present certain degree of asymmetry the assunption of normality can result in an unreal situation and the application of this model can hide important caracteristics of the true model. Situations of this type has given strength to the use of asymmetric models with special emphasis on the skew-symmetric distribution developed by Azzalini (1985). In this work we present an alternative for data analysis in the presence of significant asymmetry or kurtosis, when compared with the normal distribution, as well as other situations that involve such model. We present and study of the properties of the α-power and log-α-power distributions, where we also study the estimation problem, the observed and expected information matrices and the degree of bias in estimation using simulation procedures. A flexible model version is proposed for the α-power distribution, following an extension to a bimodal version. Follows next an extension of the Birnbaum-Saunders distribution using the α-power distribution, where some properties are studied, estimating approaches are developed as well as corrected bias estimator developed. We also develop censored and uncensored regression for the α-power model and for the log-linear Birnbaum-Saunders regression models, for which model validation techniques are studied. Finally a multivariate extension of the α-power model is proposed and some estimation procedures are investigated for the model. All the situations investigated were illustrated with data application using data sets previally analysed with other distributions.

Keywords: maximum likelihood estimator, Fisher information matrix, likelihood ratio test, asymptotic distribution.

Índice

Lista de Tabelas

Lista de Figuras

Capítulo 1

Introdução

1.0.1 Motivação

Em aplicações relacionadas com diversas áreas científicas, por exemplo, biologia, ciências médicas, economia, entre outras, os processos inferenciais têm sido considerados sob a suposição da distribuição normal. Quando os dados são provenientes de distribuições com caudas mais ou menos pesadas que da distribuição normal ou ainda assimétricas, estes processos inferenciais sob normalidades são inadequados, mesmo acontecendo quando se têm presença de valores atípicos. Ante este tipo de situações, muitos autores têm usado transformações de variáveis para alcançar normalidade ou pelo menos simetria dos dados, chegando muitas vezes a resultados satisfatórios. Esta técnica como uma alternativa, também apresenta algumas dificuldades, por exemplo a interpretação dos resultados com a variável transformada e ainda mais quando se têm uma transformação diferente para cada variável(veja Azzalini et al., 1999).

Ainda, quando a classe das distribuições elípticas é uma boa alternativa para o afastamento da distribuição normal, esta não é adequada quando a distribuição das observações é assimétrica. Portanto, é de interesse estudar distribuições que sejam menos sensíveis que a distribuição normal a certos desvios das suposições consideradas, mas que contenham esta como un caso especial, e que possam acomodar valores práticos de assimetria e cur-

tose.

Neste contexto, nas duas últimas décadas, tem sido observado um interesse crescente na literatura estatística pelas famílias de distribuições flexíveis capazes de modelar diferentes graus de assimetria e curtose. Diferentes trabalhos têm sidos publicados neste contexto, com resultados iniciais dados em Birnbaum (1950), Lehmann (1953), Roberts (1966) e O'Hagan e Leonard (1976) entre outros. Mais recentemente Azzalini (1985), Fernandez e Steel (1998), Mudholkar e Hutson (2000), Gupta et al. (2002), Arellano-Valle et al. (2004-2005), Gómez et al. (2007) também têm contribuções importantes.

Uma família mais geral que a normal e que pode acomodar valores práticos de assimetria é a família denominada normal assimétrica (*skew*-normal). Esta surgiu dos trabalhos de Roberts (1966) e O'Hagan et al., (1976), mais foi Azzalini (1985) quem introduziu formalmente esta distribuição e estudou suas propriedades. Além disso, esta distribuição é uma extensão da distribuição normal para modelar a estrutura assimétrica presente nos dados, no entanto esta apresenta problemas pois sua matriz de informação é singular, quando o parâmetro de assimetria é igual a zero, também observamos problemas na estimação do parâmetro que controla a assimetria.

Após Azzalini (1985), outros tipos de distribuições assimétricas foram introduzidas nesta área, entre estes temos: *skew*-t, *skew*-slash, *skew*-exponencial potência, etc. Durrans (1992) num contexto hidrológico introduz a distribuição de estatísticas de ordem fracionárias. Este modelo foi estudado por Gupta et al., (2008), e posteriormente, Pewsey et al., (2010) mostrou que sua matriz de informação é não singular quando o parâmetro que controla a assimetria é igual a um, caso contrario à normal assimétrica de Azzalini, com matriz de informação singular quando o parâmetro de assimetria é igual a zero.

Azzalini (1985), desenvolveu a estrutura geral para a função de densidade de probabilidade

assimétrica, *skew-symmetric*, a qual é dada por

$$\varphi(z;\lambda) = 2f(z)G(\lambda z), \qquad z, \lambda \in \mathbb{R}, \tag{1.1}$$

onde f é uma função de densidade de probabilidade (fdp) a qual é simétrica ao redor de zero, G é uma função de distribuição cumulativa (fda) absolutamente contínua também simétrica ao redor de zero e λ é um parâmetro de assimetria. Assim, para $f = \phi$ e $G = \Phi$, a função de densidade e de distribuição da normal padrão, respectivamente, se obtêm a chamada distribuição normal assimétrica, a qual denotamos por $Z \sim SN(\lambda)$, cuja fdp é

$$\varphi(z;\lambda) = 2\phi(z)\Phi(\lambda z), \qquad z \in \mathbb{R}. \tag{1.2}$$

A densidade (1.2) foi extensamente estudada por Azzalini (1985), Henze (1986), Chiogna (1997) e Pewsey (2000).

1.0.2 Modelo α-Potência

No entanto, Lehmann (1953) propôs um modelo probabilístico cuja distribuição cumulativa, $\mathcal{F}$, é a potência racional de outra função de distribuição cumulativa continua e diferenciável F; na literatura este é conhecido como modelo de alternativas de Lehmann. Este modelo é dado por

$$\mathcal{F}_F(z;\alpha) = \{F(z)\}^{\alpha}, \qquad z \in \mathbb{R}, \tag{1.3}$$

onde F é uma função de distribuição e α é um número racional. Em termos gerais esta distribuição é gerada a partir da distribuição do máximo na amostra. Este modelo tem sido discutido amplamente por Lehmann (1959), Gupta (1998) e Arnold (2004). Durrans (1992) num contexto hidrológico deriva este modelo como uma distribuição de estatística de ordem fracionária, que nós chamamos **de distribuição α-potência,** com função de densidade de probabilidade

$$\varphi_F(z;\alpha) = \alpha f(z)\{F(z)\}^{\alpha-1}, \qquad z \in \mathbb{R}, \quad \alpha \in \mathbb{R}^+, \tag{1.4}$$

onde $\alpha \in \mathbb{R}^+$ e F é uma função de distribuição absolutamente contínua com função de densidade de probabilidade $f = dF$. Nós a denotamos por $Z \sim AP(\alpha)$. Este novo modelo

dá origem à família de distribuições α-potência a qual é uma alternativa às famílias de distribuições que têm um alto grau de assimetria e/ou curtose.

Não entanto, Durrans (1992) considera o caso onde $F = \Phi$, a função de distribuição da normal padrão e que nós chamamos **a classe de distribuições α-potência normal assimétrica,** cuja função de densidade é dada por

$$\varphi(z;\alpha) = \alpha\phi(z)\{\Phi(z)\}^{\alpha-1}, \qquad z \in \mathbb{R}, \tag{1.5}$$

onde ϕ e Φ são como acima e α é um parâmetro de forma. Se denota por $Z \sim PN(\alpha)$. Nota-se que para $\alpha = 1$, este modelo contém à distribuição normal e para $\alpha = 2$ contém a função de densidade da distribuição normal assimétrica de Azzalini do modelo (1.2) para $\lambda = 1$.

A função de distribuição de Z no modelo (1.5) é dada por:

$$\mathcal{F}_F(z;\alpha) = \{\Phi(z)\}^{\alpha}, \qquad z \in \mathbb{R}. \tag{1.6}$$

O gráfico da Figura 1.1 mostra a distribuição PN para alguns valores do parâmetro α. Nota-se como o parâmetro α afeta a localização, a dispersão, a curtose e a assimetria da distribuição.

Uma generalização dos resultados em Durrans (1992) foram apresentados por Eugene et al. (2002) motivado pela distribuição de estatísticas de ordem. Jones (2004) também considerou este modelo como uma distribuição derivada a partir de estatística de ordem. Propriedades desta distribuição para potências inteiras são estudadas por Gupta e Gupta (2004) e Gupta et al. (2008).

Caso Localização-Escala

Se $Z \sim AP(\alpha)$, então a extensão de localização-escala de Z, $X = \xi + \eta Z$, onde $\xi \in \mathbb{R}$ e $\eta \in \mathbb{R}^+$, têm função de densidade de probabilidade dada por

$$\varphi(x;\xi,\eta,\alpha) = \frac{\alpha}{\eta} f\left(\frac{x-\xi}{\eta}\right)\left\{F\left(\frac{x-\xi}{\eta}\right)\right\}^{\alpha-1}, \qquad x \in \mathbb{R}. \tag{1.7}$$

Usamos a notação $X \sim AP(\xi,\eta,\alpha)$, onde ξ é o parâmetro de localização e η é o parâmetro

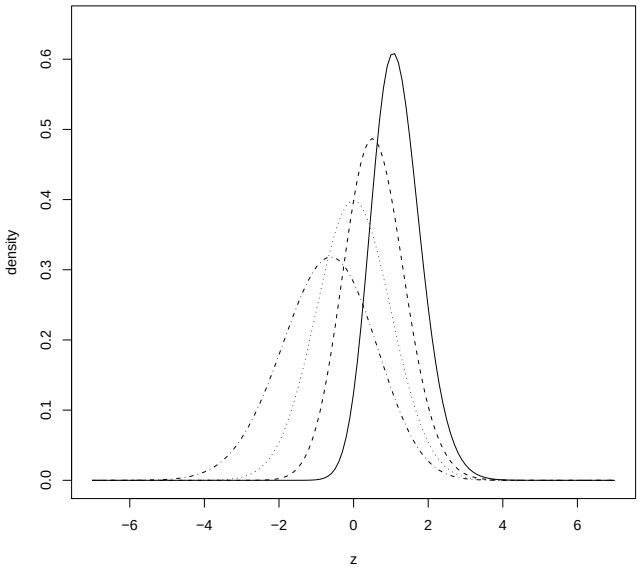


Figura 1.1 *Densidade* $\varphi(z;\alpha)$, *para valores* α *iguais a* 7 *(linha contínua),* 2 *(linha tracejada),* 1 *(linha de pontos) e* 0.5 *(linha de pontos e tracejada).*

de escala.

Para $F = \Phi$ temos a particularização de localização-escala da distribuição α-potência normal assimétrica, a qual denotamos por $X \sim PN(\xi, \eta, \alpha)$.

Neste trabalho nós apresentamos algumas extensões do modelo (1.4), para alguns casos ainda não estudados na literatura, o qual é o principal objetivo desta tese. Definimos os seguintes pontos como objetivos específicos do trabalho.

- Estender a distribuição log-normal para o caso da distribuição α-potência.
- Estender o modelo α-potência para o caso de dados com distribuição bimodal.
- Estender a distribuição generalizada de Birnbaum-Saundera para o caso da distribuição α-potência.
- Estender os modelos de regressão linear e não linear sob normalidade, assumindo que os erros seguem uma distribuição α-potência.

- Estender o modelo de regressão log-linear Birnbaum-Saunders, assumindo que os erros seguem uma distribuição senh-potência normal.

- Estender os modelos de regressão para dados censurados sob normalidade, assumindo que os erros seguem uma distribuição α-potência normal assimétrica.

- Discutir o modelo de mistura discreta-contínua para o caso Bernoulli - distribuição log-α-Potência normal assimétrica.

- Estender o modelo α-potência para o caso multivariado.

1.0.3 Organização da Tese

A presente tese está dividida em oito capítulos. No capítulo 2 apresentamos algumas das propriedades do modelo α-potência, também desenvolvemos a distribuição log-α-potência normal assimétrica e suas propriedades estatísticas. Nós estudamos alguns métodos de estimação, encontramos a função escore e as matrizes de informação observada e esperada de Fisher. Finalmente fizemos um pequeno estudo de simulação e uma ilustração para dados sobre um estudo de contaminação ambiental, em Santiago do Chile, considerados por Balakrishnan et al. (2009). Nós terminamos este capítulo introduzindo o modelo potência-*skew*-normal. No capítulo 3 apresentamos o modelo α-potência flexível e introduzimos o modelo bimodal α-potência para os casos simétrico e assimétrico. Estudamos os estimadores dos parâmetros por máxima verossimilhança e as matrizes de informação observada e esperada. No capítulo 4 apresentamos a extensão do modelo generalizado de Birnbaum-Saunders para o caso da distribuição α-potência. Nós estudamos suas principais propriedades estatísticas e alguns métodos de estimação; encontramos a função escore e as matrizes de informação observada e esperada de Fisher, também fazemos correção do viés dos estimadores. Por último, fizemos um estudo de simulação e uma ilustração para os dados analisados por Birnbaum e Saunders (1969b). No capítulo 5 apresentamos o modelo de regressão α-potência normal assimétrico e o modelo senh-normal α-potência. Nós desenvolvemos os estimadores de máxima verossimilhança e de momentos, encon-

tramos as matrizes de informação, apresentamos algumas simulações e uma ilustração para cada modelo estudado. No capítulo seis nós introduzimos as variáveis censuradas e truncadas para a distribuição α-potência normal assimétrica, estudamos os modelos para o caso de dados censurados, o modelo tobit e o modelo de mistura discreta-contínua. Nós desenvolvemos os estimadores de máxima verossimilhança, encontramos a matriz de informação observada, apresentamos algumas simulações e uma ilustração para cada modelo estudado. No capítulo sete nós consideramos a extensão do modelo α-potência ao caso multivariado a partir de distribuições condicionais; estudamos os estimadores de máxima verossimilhança e pseudo verossimilhança. Finalmente consideramos uma ilustração onde se mostra as vantagens do modelo α-potência normal assimétrico bivariado frente ao modelo normal bivariado. Por último o capítulo oito contém algumas conclusões gerais do trabalho e alguns tópicos de trabalhos futuros.

Capítulo 2

Distribuição Log-α-Potência. Modelo Potência-*Skew*-Normal.

2.1 Motivação

A distribuição log-normal tem sido muito utilizada em análise de dados de sobrevivência bem como em situações de tempo de falha de isolamento elétrico (Nelson e Hahn, 1972) e estudos de câncer de pulmão (Whittemore e Altschuler, 1976). Assim, uma variável aleatória Y é dita ter uma distribuição log-normal se a variável $Z = \log(Y)$ tem distribuição normal com média μ e variância σ^2.

Uma de suas aplicações mais importantes está na área da química pela lei fundamental da geoquímica, enunciada por Ahrens (1954), "a concentração de um elemento químico têm distribuição log-normal". Em muitas destes situações, a assimetria da distribuição assim como sua curtose estão acima o abaixo do esperado com a distribuição log-normal, razão pela qual precisa-se pensar em modelos mais flexíveis que modelem tais desvios, para estes dois parâmetros, em modelamento de dados positivos.

Como uma generalização da função de distribuição log-normal para o caso da família de distribuições normal assimétrica, Mateu-Figueras et al. (2003-2004) estudam a distribuição univariada log-normal assimétrica, de parâmetros ξ, η e λ, cuja função de densidade é dada por:

$$\varphi(y;\xi,\eta,\lambda) = \frac{2}{\eta y}\phi\left(\frac{\log(y)-\xi}{\eta}\right)\left\{\Phi\left(\lambda\frac{\log(y)-\xi}{\eta}\right)\right\}, \quad y \in \mathbb{R}^+, \tag{2.1}$$

Tabela 2.1 *Comportamento da moda da distribuição PN*

α	0.25	0.50	0.75	1.25	1.5	1.75	2.0	3.0	5.0
Moda	-1.3744	-0.6120	-0.2390	0.1728	0.3067	0.4162	0.5060	0.7652	1.0615

onde $\xi \in \mathbb{R}$, $\eta \in \mathbb{R}^+$, com ϕ e Φ, como definidas anteriormente. Esta distribuição igual que a normal assimétrica de Azzalini (1985), tem matriz de informação singular o qual é uma desvantagem em processos inferenciais.

Agora nós vamos estudar a distribuição log-α-potência como uma extensão do caso da distribuição AP, a qual é uma segunda alternativa para modelar dados positivos, mas primeiro vamos estudar algumas propriedades da distribuição AP.

2.2 Propriedades da Distribuição AP e PN

1. Se $Z \sim PN(\alpha)$ então a moda da distribição de Z é a solução da equação: $(\alpha-1)\Phi(z) - z\phi(z) = 0$ ou $(\alpha-1)h(z) - z = 0$, onde $h(z) = \frac{\Phi(z)}{\phi(z)}$ é a taxa de risco inversa da normal padrão. A Tabela 2.1 mostra que a única moda é uma função decrescente de α.

2. Se $Z \sim AP(\alpha)$ então, $Y = -Z \sim AP(\alpha)$

3. Se $Z \sim AP(\alpha)$, os momentos da variável Z não tem uma forma fechada. Uma mudança de variável permite escrever o n-ésimo momento da variável Z como

$$E(Z^n) = \alpha \int_0^1 [F^{-1}(u)]^n u^{\alpha-1} du, \tag{2.2}$$

onde F^{-1} é a função inversa de F. Esta esperança concorda com a esperança da expressão $[F^{-1}(z)]^n$ de uma variável aleatória com distribuição beta de parâmetros α e 1.

4. Se $Z \sim PN(\alpha)$ onde α é um número inteiro positivo maior que um, então os momentos de Z podem ser estudados a partir dos momentos da variável aleatória beta-normal (veja Gupta e Nadarajah, 2004) para o caso especial $\beta = 1$. Assim, para o caso de localização

- escala $X = \xi + \eta Z$, nós temos que

$$\mathbb{E}(X^n) = \alpha\xi^n \sum_{k=0}^{n} \binom{n}{k} \left(\frac{\eta}{\xi}\right)^k \left\{ \sum_{j=0}^{\alpha-1} (-1)^j \binom{\alpha-1}{j} I_{k,j} + (-1)^k I_{k,\alpha-1} \right\}.$$

onde, $I_{i,k} = \int_0^\infty z^i \phi(z)\{1-\phi(z)\}^k = \frac{1}{2^k\sqrt{2\pi}} \int_0^\infty z^i \exp\left(-\frac{z^2}{2}\right) erfc^k\left(\frac{z}{\sqrt{2}}\right) dz$, onde $erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-t^2)dt$, é a função de erro (veja Prudnikov et al., 1990). Se $n = 1$, temos que

$$\mathbb{E}(X) = \xi + \alpha\eta \left\{ \sum_{j=0}^{\alpha-2} \frac{(-1)^j}{j+1} \binom{\alpha-1}{j} \delta_{j+1} + \frac{(-1)^{\alpha-1}}{\alpha} \delta_\alpha \right\},$$

com $\delta_j = \int_0^\infty \{1-\Phi(z)\}^j dz = \frac{1}{2^j} \int_0^\infty erfc^j\left(\frac{x}{\sqrt{2}}\right)$. Portanto, de acordo com os resultados obtidos por Prudnikov et al. (1990) temos os seguintes casos especiais, veja Tabela 2.2:

Tabela 2.2 *Momentos da variável aleatória PN*

α	2	3	4	5
$\mathbb{E}(X)$	$\xi + \frac{\eta}{\sqrt{\pi}}$	$\xi + \frac{3\eta}{2\sqrt{\pi}}$	$\xi + \frac{6\eta \arctan\sqrt{2}}{\pi^{3/2}}$	$\xi + \frac{15\eta \arctan\sqrt{2}}{\pi^{3/2}} - \frac{5\eta}{2\sqrt{\pi}}$

Ruben(1954) desenvolveu os primeiros dez momentos da variável aleatória PN. Durrans (1992), apresenta uma reprodução parcial da Tabela destes momentos e onde os termos $\Phi_1(\alpha)$, $\Phi_2(\alpha)$ e $\Phi_3(\alpha)$ são a média, a variância e o viés da distribuição $PN(\alpha)$.

2.3 Distribuição Log-α-Potência

Definição 2.3.1. *A variável aleatória positiva Y, com suporte em R^+, tem uma distribuição univariada log-α-potência de parâmetro α, se a variável transformada $Z = \log(Y)$ tem uma distribuição AP de parâmetro α.*

Se Y tem distribuição univariada log-α-potência de parâmetro α, denotamos por $Y \sim LAP(\alpha)$. A função densidade de probabilidade de uma variável aleatória Y com distribuição LAP de parâmetro α é

$$\varphi(y;\alpha) = \frac{\alpha}{y} f(\log(y))\{F(\log(y))\}^{\alpha-1}, \qquad y \in \mathbb{R}^+, \tag{2.3}$$

onde F é uma função de distribuição absolutamente continua e diferenciável com função de densidade de probabilidade $f = dF$. A função de distribuição de Y fica

$$\mathcal{F}_F(y;\alpha) = \{F(\log y)\}^{\alpha}, \qquad y \in \mathbb{R}^+. \tag{2.4}$$

De acordo com (2.4), pode-se utilizar o método de inversão para gerar uma variável aleatória com distribuição $LAP(\alpha)$. Desta forma, se U é uma variável aleatória uniforme em (0,1) então a variável aleatória $Y = e^{F^{-1}(U^{1/\alpha})}$ tem distribuição LAP de parâmetro α.

Para $F = \Phi$ nós temos a distribuição log-α-potência normal assimétrica, a qual é denotada $LPN(\alpha)$. A Figura 2.1 mostra a forma da função de densidade da distribuição LPN para valores α de 0.75, 1, 2 e 7. Neste gráfico pode-se observar como o parâmetro α afeta a forma da distribuição. Assim, para $\alpha = 1$, caso log-normal, a curtose é maior que para $\alpha = 2$, caso log-normal assimétrica, mesmo acontecendo para $\alpha = 7$. Enquanto, que para $\alpha = 0.75$ a curtose é maior que para o caso $\alpha = 1$. Também podemos observar que a assimetria da distribuição muda quando α aumenta ou decresce. Portanto, α afeta a assimetria e a curtose da distribuição, além também a localização e da variância.

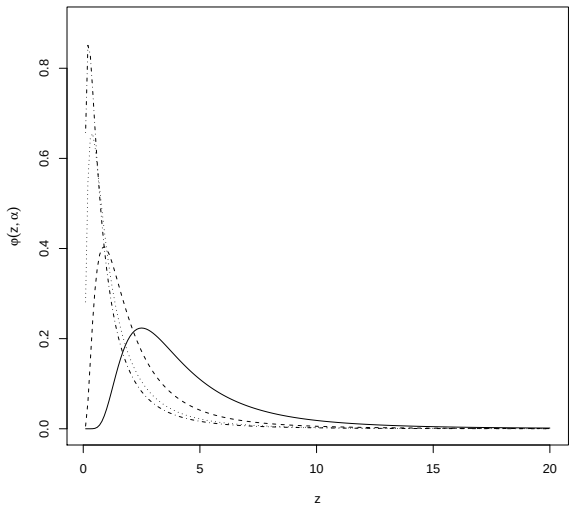


Figura 2.1 *Densidade LPN para $\alpha = 0.75$ (linha de pontos e tracejada), 1 (linha de pontos), 2 (linha tracejada), 7 (linha contínua).*

Uma das aplicações da distribuição log-normal é para explicar o comportamento de dados de tempos de sobrevivência. Entre as funções matemáticas mais usadas para explicar o tempo de sobrevivência estão a função de sobrevivência $S(t) = 1 - F(t)$, onde $F(t)$ é a função de distribuição cumulada, a função de densidade de probabilidade $f(t) = -\frac{dS(t)}{dt}$ e a função de risco $r(t) = \frac{f(t)}{S(t)}$. Estas três definições são matematicamente equivalentes. Para o modelo LPN as funções de sobrevivência e risco são, respectivamente,

$$S(t) = 1 - \{\Phi(\log(t))\}^{\alpha} \quad \text{e} \quad r(t) = r_{LN}(t)\frac{\alpha\{\Phi(\log(t))\}^{\alpha-1} - \{\Phi(\log(t))\}^{\alpha}}{1 - \{\Phi(\log(t))\}^{\alpha}},$$

onde $r_{LN}(t)$ é a função de risco da distribuição log-normal. Porem, de acordo com Gupta et al. (1997), para o caso log-normal, temos que $r(t)$ cresce inicialmente até atingir o máximo e então decresce até zero quando t atinge infinito. Este comportamento pode ser verificado na Figura 2.2, onde também podemos concluir que a função de risco é uma função não crescente do parâmetro α. Entretanto, a função de taxa de risco acumulado e o índice de risco invertido são respectivamente,

$$H(t, \alpha) = -\log[S(t)] \quad \text{e} \quad R(t) = \alpha R_{LN}(t),$$

onde $R_{LN}(t)$ é a função de risco inverso da distribuição log-normal; ou seja, a taxa de risco inversa de T é proporcional à taxa de risco inverso da distribuição log-normal.

2.3.1 Estimação por Máxima Verossimilhança (EMV)

Apresentamos agora a estimação do parâmetro da distribuição LPN pelo método de máxima verossimilhança. Assim, seja uma amostra aleatória de tamanho n, $\mathbf{Y} = (Y_1, \ldots, Y_n)'$, da distribuição $LPN(\alpha)$, a função de log-verossimilhança para α dado $\mathbf{Y}$ fica dada por:

$$\ell(\alpha; \mathbf{Y}) = n[\log(\alpha) - \sqrt{2\pi}] - \sum_{i=1}^{n} \log(y_i) - \frac{1}{2}\sum_{i=1}^{n}(\log(y_i))^2 + (\alpha - 1)\sum_{i=1}^{n} \log\{\Phi(\log(y_i))\}.$$

Então a função escore definida como a derivada de $\ell(\alpha; \mathbf{Y})$ com respeito α é:

$$U(\alpha) = \frac{\partial \ell(\alpha; \mathbf{Y})}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} \log\{\Phi(\log(y_i))\},$$

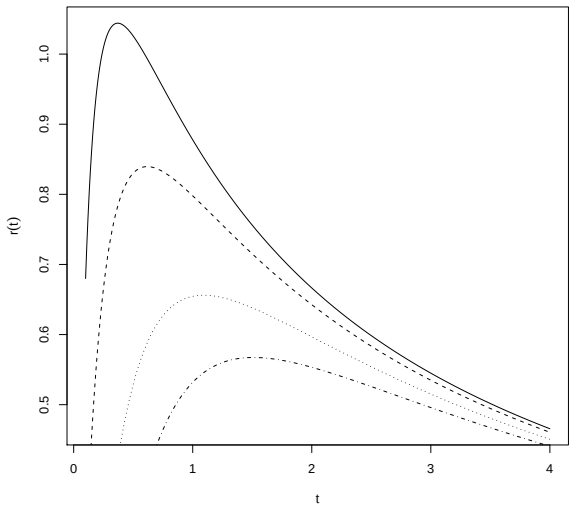


Figura 2.2 *A função de risco $r(t)$, para valores α de 0.75 (linha contínua), 1 (linha tracejada), 1.5 (linha de pontos) e 2 (linha de pontos e tracejada).*

portanto, o estimador de máxima verossimilhança de α fica dado por:

$$\hat{\alpha} = -n/\sum_{i=1}^{n} \log\{\Phi(\log(y_i))\}.$$

Nota-se que este estimador é função da estatística suficiente e completa para α, $T(\mathbf{Y}) = \sum_{i=1}^{n} \log\{\Phi(\log(y_i))\}$. Assim,

$$\frac{\partial^2 \ell(\alpha; \mathbf{Y})}{\partial \alpha^2} = -\frac{n}{\alpha^2},$$

e conclui-se que a variância assintótica de $\hat{\alpha}$ é α^2/n.

Agora, a variável aleatória $V = -\sum_{i=1}^{n} \log\{\Phi(\log(y_i))\} \sim Gamma(n, 1)$. Assim, $\hat{\alpha}$ tem distribuição como a da variável aleatória $W = \frac{n\alpha}{V}$ para $n > 2$. Logo,

$$\mathbb{E}(\hat{\alpha}) = n\alpha\mathbb{E}\left(\frac{1}{V}\right) = \frac{n}{n-1}\alpha.$$

Portanto, um estimador não viesado de α é

$$\hat{\alpha}_c = \frac{n-1}{n}\hat{\alpha}.$$

Da mesma forma temos que

$$Var(\hat{\alpha}) = \frac{n^2}{(n-1)^2(n-2)}\alpha^2 \quad \text{e} \quad EQM(\hat{\alpha}) = \frac{n+2}{(n-1)(n-2)}\alpha^2,$$

onde Var e EQM denotam a variância e o erro quadrático médio. Portanto, quando $n \to \infty$ então $\mathbb{E}(\hat{\alpha}) \to \alpha$ e $EQM(\hat{\alpha}) \to 0$, implicando a consistência assintótica do estimador. Por último nós temos que

$$Var(\hat{\alpha}_c) = \frac{\alpha^2}{n-2} > \frac{\alpha^2}{n},$$

onde $\frac{\alpha^2}{n}$ é a cota inferior de Cramer-Rao de $\hat{\alpha}$. Mesmo assim, $-2\alpha \sum_{i=1}^{n} \log\{\Phi(\log(y_i))\}$ é uma quantidade pivotal para α, porém um intervalo de $100(1-\delta)\%$ de confiança para α é

$$\left(\frac{\chi^2_{2n,\delta/2}}{-2\sum_{i=1}^{n} \log\{\Phi(\log(y_i))\}}, \frac{\chi^2_{2n,1-\delta/2}}{-2\sum_{i=1}^{n} \log\{\Phi(\log(y_i))\}} \right).$$

2.3.2 Momentos da Distribuição LPN

O r-ésimo momento da variável aleatória Y com distribuição $LPN(\alpha)$ é dado por:

$$E(Y^n) = \alpha \int_0^1 \left[\exp\left(n\Phi^{-1}(z)\right)\right] z^{\alpha-1} dz. \tag{2.5}$$

Entretanto, os primeiros quatro momentos centrais de Y, $\acute{\mu}_r = E(Y - E(Y))^r$, podem ser calculados das expressões

$$\acute{\mu}_2 = \mu_2 - \mu_1^2, \quad \acute{\mu}_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3 \;\; e \;\; \acute{\mu}_4 = \mu_4 - 4\mu_3\mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4.$$

portanto, a variância e os coeficientes de variação, assimetria e curtose são dados por:

$$\sigma^2 = Var(Y) = \acute{\mu}_2, \; CV = \sqrt{\sigma^2}/\mathbb{E}(Y), \; \sqrt{\beta_1} = \acute{\mu}_3/[\acute{\mu}_2]^{3/2} \text{ e } \beta_2 = \acute{\mu}_4/[\acute{\mu}_2]^2.$$

Um pequeno estudo de simulação leva à obtenção dos intervalos para os parâmetros definidos acima para α entre zero e 1000. Os resultados foram obtido no pacote estatístico R e as estatísticas foram obtidas numericamente mediante a função "integrate" a qual calcula integrais simples. Na Tabela 2.3 mostra-se o intervalo de valores da média, variância e dos coeficientes de variação, assimetria e curtose. Na Figura 2.3 nota-se como o grau da assimetria e da curtose decrescem quando α aumenta. Mesmo assim temos que o CV é uma função decrescente do parâmetro α, assim nós percebemos que para pequenos valores de α a variação relativa é muito grande. Nós também estudamos o comportamento da média, a qual foi totalmente crescente no intervalo de α, mesmo para a variância.

Tabela 2.3 *Descrição da assimetria e a curtose no Modelo LPN Para* $0 < \alpha \leq 1000$

Estatística	Média	Variância	Assimetria	Curtose	$CV(\%)$
Minimo	0.28	0.74	3.09	27.37	43.13
Máximo	27.37	139.40	13.18	517.00	305.96

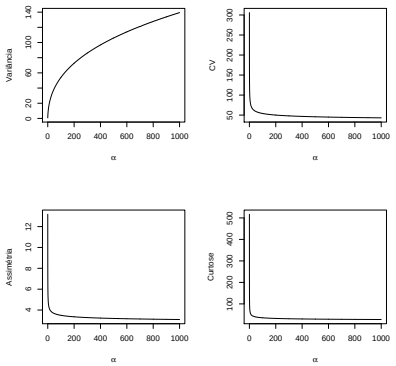


Figura 2.3 *Variância, CV(%), coeficientes de assimetria e curtose da variável* $LPN(\alpha)$.

2.3.3 Caso Localização-Escala

Seja $X \sim LPN(\xi, \eta, \alpha)$, onde $\xi \in \mathbb{R}$ é o parâmetro de localização e $\eta \in \mathbb{R}^+$ é o parâmetro de escala. Então, a transformação $X = \log(Y)$ leva ao caso de localização - escala para o modelo log-α-potência normal assimétrico, cuja função de densidade pode ser escrita como:

$$\varphi(y; \xi, \eta, \alpha) = \frac{\alpha}{\eta y}\phi\left(\frac{\log(y) - \xi}{\eta}\right)\left\{\Phi\left(\frac{\log(y) - \xi}{\eta}\right)\right\}^{\alpha-1}, \quad y \in \mathbb{R}^+. \tag{2.6}$$

Usamos a notação $Y \sim LPN(\xi, \eta, \alpha)$. De (2.6) conclui-se que $LPN(\alpha) = LPN(0, 1, \alpha)$.

Proposição 2.3.2. *Para qualquer* $\alpha \in \mathbb{R}^+$*, a variável aleatória* $Y \sim LPN(\xi, \eta, \alpha)$*, não tem função geradora de momentos.*

De fato: O caso $\alpha = 1$, corresponde ao caso da distribuição log-normal a qual não tem função geradora de momentos. Agora, fazendo $h_\alpha(y) = \frac{\alpha}{y}e^{ty}\phi(\log(y))\{\Phi(\log(y))\}^{\alpha-1}$, nós

temos que

$$M(t) = \int_0^\infty h_\alpha(y)dy, \qquad y \in \mathbb{R}^+.$$

Assim, pelos resultados dados em Lin e Stoyanov (2009) e as propriedades do ínfimo e do supremo, nós temos que $J_\alpha = \int_0^\infty h_\alpha(y)dy = \infty$ para todo $\alpha \in \mathbb{R}^+$.

2.4 Inferência no Modelo LPN(ξ, η, α)

2.4.1 Estimação pelo Método dos Momentos

Para o caso localização - escala temos que a média (μ), a variância (σ^2) e o coeficiente de assimetria ($\sqrt{\beta_1}$) são dados por

$$\mu = \xi + \eta\Phi_1(\alpha), \quad \sigma^2 = \eta^2\Phi_2(\alpha) \ \text{ e } \ \sqrt{\beta_1} = \frac{\mu_3}{\sigma^3} = \Phi_3(\alpha).$$

Portanto, pode-se obter estimadores de momentos para os parâmetros do modelo α, ξ e η substituindo nas expressões anteriores μ, σ^2 e $\sqrt{\beta_1}$ por seus respectivos momentos amostrais $\bar{y}$, s^2 e $\sqrt{b_1}$; enquanto que os valores de $\Phi_1(\alpha)$, $\Phi_2(\alpha)$ e $\Phi_3(\alpha)$ podem ser obtidos integrando numéricamente. Assim, inicialmente precisamos estimar o parâmetro α da equação não linear obtida ao igualar os coeficiente de assimetria amostral e populacional. Estimado este parâmetro, podemos obter por integração numérica $\Phi_1(\alpha)$, e $\Phi_2(\alpha)$; com as quais se pode obter os estimadores de momentos de ξ e η.

2.4.2 Estimadores de Máxima Verossimilhança (EMV)

Equações de Máxima Verossimilhança

Para uma amostra aleatória de tamanho n, $\mathbf{Y} = (Y_1, \ldots, Y_n)'$, da distribuição $LPN(\xi, \eta, \alpha)$, a função de log-verossimilhança para $\boldsymbol{\theta} = (\xi, \eta, \alpha)'$ dado $\mathbf{Y}$ fica dada por:

$$\ell(\boldsymbol{\theta}; \mathbf{Y}) = n\{\log(\alpha) - \log(\eta) - (1/2)\log(2\pi)\} - \sum_{i=1}^n \log(y_i) - \sum_{i=1}^n z_i^2/2 + \\ + (\alpha - 1)\sum_{i=1}^n \log\{\Phi(z_i)\}, \tag{2.7}$$

onde, $z_i = (\log(y_i) - \xi)/\eta$. Então, a função escore definida como a derivada parcial de primeira ordem de (2.7) com respeito ξ, η e α, pode ser escritas como $U(\boldsymbol{\theta}) = (U(\xi), U(\eta), U(\alpha))'$. Assim, usando a notação como em Pewsey et al. (2010), nós temos que:

$$U(\xi) = n\left\{\overline{z} - (\alpha - 1)\overline{w}\right\}/\eta, \quad U(\eta) = -n\{1 - \overline{z^2} + (\alpha - 1)\overline{zw}\}/\eta \ \text{ e } \ U(\alpha) = n\{\overline{u} + 1/\alpha\},$$

onde $w_i = \phi(z_i)/\Phi(z_i)$ e $u_i = \log\{\Phi(z_i)\}$.

Assim, os estimadores de máxima verossimilhança dos parâmetros ξ, η e α obtêm-se resolvendo o sistema de equações precedente. Logo, fazendo os escores iguais a zero obtemos que as soluções destes satisfazem as equações:

$$\overline{z} = (\alpha - 1)\overline{w}, \quad 1 - \overline{z^2} = (1 - \alpha)\overline{zw} \ \text{ e } \ \alpha = -1/\overline{u},$$

onde $\overline{z} = \sum_{i=1}^n z_i/n$, $\overline{z^2} = \sum_{i=1}^n z_i^2/n$, $\overline{w} = \sum_{i=1}^n w_i/n$, $\overline{zw} = \sum_{i=1}^n z_i w_i/n$ e $\overline{u} = \sum_{i=1}^n u_i/n$.

2.4.3 Método do Percentil *Elemental*

Um terceiro método de estimação para os parâmetros da distribuição LPN pode ser desenvolvido usando o método do percentil *elemental* proposto por Castillo e Hadi (1995) e adaptado para o caso de distribuições exponenciadas por Castillo et al. (2005). O método é uma aplicação da teoria de estimação de parâmetros e quantis de distribuições de variáveis aleatórias continuas. Assim, seja $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)'$ uma amostra aleatória da distribuição LPN, $\mathcal{F}_\Phi(y; \boldsymbol{\theta})$. O método obtém os estimadores em dois passos. O primeiro passo é como segue. Como a distribuição LPN tem três parâmetros, então os estimadores são baseados em três estatísticas de ordem distintas, então, seja $I = \{i, j, r\}$ com $i < j < r \in \{1, 2, ..., n\}$ os índices destas três estatísticas de ordem. Igualando os quanties teóricos e amostrais da distribuição LPN nós obtemos o sistema de equações

$$y_{i:n} = \exp\left[\xi + \eta\Phi^{-1}(p_{i:n}^{1/\alpha})\right], \ y_{j:n} = \exp\left[\xi + \eta\Phi^{-1}(p_{j:n}^{1/\alpha})\right] \ \text{ e } \ y_{r:n} = \exp\left[\xi + \eta\Phi^{-1}(p_{r:n}^{1/\alpha})\right],$$

onde $p_{i:n} = \frac{i}{n+1}$. Eliminando ξ e η deste sistema de equações, nós obtemos a seguinte equação para α

$$\omega(\alpha) = \left(\frac{y_{j:n}}{y_{r:n}}\right)^{\Phi^{-1}\left(p_{i:n}^{1/\alpha}\right)-\Phi^{-1}\left(p_{r:n}^{1/\alpha}\right)} - \left(\frac{y_{i:n}}{y_{r:n}}\right)^{\Phi^{-1}\left(p_{j:n}^{1/\alpha}\right)-\Phi^{-1}\left(p_{r:n}^{1/\alpha}\right)}.$$

A solução desta equação permite obter uma estimativa, $\hat{\alpha}$, de α baseada nas três observações $y_{i:n}$, $y_{j:n}$ e $y_{r:n}$ (conjunto *elemental*). Substituindo este estimador em duas das três equações anteriores, se obtêm os estimadores de ξ e η,

$$\hat{\eta}_m = \frac{\log\left(\frac{y_{i:n}}{y_{r:n}}\right)}{\Phi^{-1}\left(p_{i:n}^{1/\hat{\alpha}_m}\right) - \Phi^{-1}\left(p_{r:n}^{1/\hat{\alpha}_m}\right)} \quad \text{e} \quad \hat{\xi}_m = \log(y_{i:n}) - \hat{\eta}_m \Phi^{-1}\left(p_{i:n}^{1/\hat{\alpha}_m}\right).$$

Agora nós podemos obter estimadores eficientes e robustos para ξ, η e α usando estatísticas de ordem como se mostra a continuação. Selecionamos um número pre-especificado N de subconjuntos *elementais*, cada uns de tamanho três, selecionados aleatoriamente. Para cada uns destes subconjuntos *elementais* nós obtemos uma estimativa *elemental* do parâmetro $\theta = (\xi, \eta, \alpha)$, que nós denotamos por $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_N$. Estes estimativas *elementais* podem ser combinadas a través de alguma função robusta para obter uma estimativa final de θ. Exemplos de funções robustas inclui a mediana (MED) ou também uma média podada (Rousseeuw, 1984).

Geralmente a variância destes estimadores é obtida por meio de métodos de reamostragem como jackknife ou também bootstrap (Efron, 1979; Diaconis e Efron, 1983).

2.4.4 Matriz de Informação Observada

Os elementos da matriz de informação observada, denotados por $j_{\xi\xi}, j_{\eta\xi}, \ldots, j_{\alpha\alpha}$, são menos a segunda derivada parcial da função de log-verossimilhança com respeito cada uns dos parâmetros. Então, depois de algumas simplificações algébricas tem-se que:

$$j_{\xi\xi} = n\left\{1 + (\alpha - 1)\left[\overline{w^2} + \overline{zw}\right]\right\}/\eta^2, \quad j_{\eta\xi} = n\left\{2\overline{z} + (\alpha - 1)\left[\overline{zw^2} + \overline{z^2w} - \overline{w}\right]\right\}/\eta^2,$$

$$j_{\alpha\xi} = n\overline{w}/\eta, \quad j_{\eta\eta} = n\left\{3\overline{z^2} - 1 - (\alpha - 1)\left[2\overline{zw} - \overline{z^3w} - \overline{z^2w^2}\right]\right\}/\eta^2,$$

$$j_{\alpha\eta} = n\overline{zw}/\eta \quad \text{e} \quad j_{\alpha\alpha} = n/\alpha^2.$$

onde $\overline{z^k w^j} = \sum_{i=1}^{n} z_i^k w_i^j / n$. Então, a partir das funções escores e a matriz de informação observada podem-se obter as estimativas dos parâmetros por métodos numéricos iterativos. Se os EMV existem, dadas as relações $\overline{z} = (\alpha - 1)\overline{w}$, $1 - \overline{z^2} = (1-\alpha)\overline{zw}$ e $\alpha = -1/\overline{u}$, os elementos da matriz de informação observada avaliada nos EMV ficam dados por:

$$j_{\hat{\xi}\hat{\xi}} = n\left\{2 - \overline{z^2} + (\hat{\alpha} - 1)\overline{w^2}\right\}/\hat{\eta}^2, \quad j_{\hat{\eta}\hat{\xi}} = n\left\{\overline{z} + (\hat{\alpha} - 1)\left[\overline{zw^2} + \overline{z^2 w}\right]\right\}/\hat{\eta}^2,$$

$$j_{\hat{\alpha}\hat{\xi}} = n\overline{z}/\left[(\hat{\alpha} - 1)\hat{\eta}\right], j_{\hat{\eta}\hat{\eta}} = n\left\{1 + \overline{z^2} + (\hat{\alpha} - 1)\left[\overline{z^3 w} + \overline{z^2 w^2}\right]\right\}/\hat{\eta}^2,$$

$$j_{\hat{\alpha}\hat{\eta}} = n\left(1 - \overline{z^2}\right)/\left[(\hat{\alpha} - 1)\hat{\eta}\right], \quad j_{\hat{\alpha}\hat{\alpha}} = n/\hat{\alpha}^2.$$

2.4.5 Matriz de Informação Esperada

Os elementos da matriz de informação esperada são os valores esperados dos correspondentes elementos da matriz de informação observada. Denotam-se n^{-1} vezes seus valores por $i_{\xi\xi}, i_{\eta\xi}, \ldots, i_{\alpha\alpha}$. Assim, fazendo $a_{kj}^* = E\left(Z^k W^j\right)$ para $k = 0, 1, 2, 3$ e $j = 1, 2$ obtêm-se:

$$i_{\xi\xi} = \{1 + (\alpha - 1)\left[a_{02}^* + a_{11}^*\right]\}/\eta^2, \quad i_{\eta\xi} = 2a_{10}^* + (\alpha - 1)\left[a_{12}^* + a_{21}^* - a_{01}^*\right]/\eta^2,$$

$$i_{\alpha\xi} = a_{01}^*/\eta, \quad i_{\eta\eta} = \{3a_{20}^* - 1 + (\alpha - 1)\left[a_{22}^* + a_{31}^* - 2a_{11}^*\right]\}/\eta^2, \quad i_{\alpha\eta} = a_{11}^*/\eta, \quad i_{\alpha\alpha} = 1/\alpha^2.$$

Estas expressões são calculadas numéricamente pois não têm forma fechadas. Então, para estudar a singularidade ou não singularidade da matriz de informação esperada, é suficiente verificar a existência da matriz inversa na fronteira de $\alpha = 1$. Logo, para $\alpha = 1$ se obtém o modelo log-normal com matriz de informação esperada dada por:

$$\begin{pmatrix} 1/\eta^2 & 0 & a_{01}^*/\eta \\ 0 & 2/\eta^2 & a_{11}^*/\eta \\ a_{01}^*/\eta & a_{11}^*/\eta & 1 \end{pmatrix}$$

cujo determinante é $[2 - [(a_{11}^*) + 2(a_{01}^*)]]/\eta^4 \neq 0$, e então a matriz de informação de Fisher é não-singular. Portanto, pela distribuição assintótica do estimador de máxima verossimilhança temos que, $\hat{\boldsymbol{\theta}} \overset{A}{\sim} N_3\left(\boldsymbol{\theta}, I_F^{-1}(\boldsymbol{\theta})\right)$. Assim, concluímos que o estimador de máxima verossimilhança é assintoticament consistente e com distribuição normal, onde a matriz de covariâncias é a inversa da matriz de informação de Fisher.

2.4.6 Estimação por Verossimilhança Perfilada

As vezes as inferências sobre um determinado modelo envolvem alguns mas não todos os parâmetros do modelo. Os parâmetros envolvidos são chamados parâmetros de interesses, enquanto que os demais se dizem são parâmetros de incômodo. Suponha que nossos parâmetros de interesses sejam descritos pelo vetor $\boldsymbol{\tau}$ e aqueles de incômodo, sejam expressos pelo vetor ϕ. Muitas vezes é possível fazer a inferência sobre $\boldsymbol{\tau}$ baseados numa função de verossimilhança marginal ou de verossimilhança condicional. Mas só em certos casos é possível ter tais funções. Considere-se uma pseudo-verossimilhança que não requer tais verossimilhanças e pode ser utilizada quando o modelo envolve parâmetros de incômodo, ϕ. Esta verossimilhança pode ser obtida substituindo o vetor paramétrico ϕ por uma estimativa consistente na verossimilhança original, comumente a estimativa de máxima verossimilhança, para valores fixados de $\boldsymbol{\tau}$. A função resultante é chamada de função de verossimilhança perfilada e depende somente dos parâmetros de interesse. Essa função de verossimilhança não é uma verossimilhança genuína e algumas propriedades básicas da função de verossimilhança podem não ser validas para a verossimilhança perfilada. Entre estas temos que a função escore pode não ter média zero e a informação observada pode apresentar vício, mas também possui propriedades interessantes que a fazem parecer como una verossimilhança verdadeira. (para outras propriedades vide Severini et al., 2000). Nosso objetivo agora é encontrar os estimadores do vetor de parâmetros $\boldsymbol{\theta} = (\xi, \eta, \alpha)' = (\boldsymbol{\tau}', \phi)'$, do modelo LPN, por máxima verossimilhança perfilada, onde o parâmetro de interesse é $\boldsymbol{\tau} = (\xi, \eta)'$, da mesma forma como Gupta et al. (2008) fizeram com a distribuição normal generalizada.

Sendo $\boldsymbol{\tau} = (\xi, \eta)'$ o parâmetro de interesse e $\phi = \alpha$ o parâmetro de incômodo, a função de verossimilhança perfilada para $\boldsymbol{\tau}$ é definida por

$$L_p(\boldsymbol{\tau}) = \underbrace{sup}_{\boldsymbol{\theta}|\boldsymbol{\tau}} L(\boldsymbol{\theta}),$$

onde $\boldsymbol{\theta}|\boldsymbol{\tau}$ denota todos os valores de $\boldsymbol{\theta}$ tais que $\boldsymbol{\tau}(\boldsymbol{\theta}) = \boldsymbol{\tau}$. Denotamos o logaritmo da função de verossimilhança perfilada por $l_p(\boldsymbol{\tau})$ e a função escore perfilada, por $u_p(\boldsymbol{\tau}) = \frac{\partial l_p(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}}$. O

ponto do máximo de $L_p(\boldsymbol{\tau})$ é denotado por $\boldsymbol{\tau}_p$. A matriz de informação observada perfilada é definida da mesma forma que a matriz de informação. Definimos,

$$L_p(\boldsymbol{\tau}) = L(\boldsymbol{\tau}, \hat{\phi}_{\boldsymbol{\tau}}),$$

onde $\hat{\phi}_{\boldsymbol{\tau}}$ é a estimativa de máxima verossimilhança de ϕ para $\boldsymbol{\tau}$ fixado. Agora apresentamos os resultados obtidos para a distribuição LPN. Assim, fixados ξ e η, a estimação de máxima verossimilhança para o parâmetro α resulta no estimador, como função de ξ e η, dado por

$$\hat{\alpha} = \hat{\alpha}(\xi, \eta) = -\frac{n}{\sum_{i=1}^{n} \log\left[\Phi\left(\frac{\log(y_i)-\xi}{\eta}\right)\right]},$$

então, substituindo $\hat{\alpha}(\xi, \eta)$ em (2.7) temos que

$$\begin{aligned}\ell_p(\xi, \eta) = -n\left\{\log\left(-\sum_{i=1}^{n}\log\left[\Phi\left(\frac{\log(y_i)-\xi}{\eta}\right)\right]\right) - \frac{1}{2}\log(2\pi)\right\} - n\log(\eta)\\ -\sum_{i=1}^{n}\log(y_i) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\log(y_i)-\xi}{\eta}\right)^2 - \sum_{i=1}^{n}\log\left[\Phi\left(\frac{\log(y_i)-\xi}{\eta}\right)\right].\end{aligned} \tag{2.8}$$

Conseqüentemente, as estimativas de máxima verossimilhança perfilada para ξ e η, $\hat{\xi}_p$ e $\hat{\eta}_p$, são as soluções, geralmente por métodos numéricos, das equações não lineais $u_p(\xi) = 0$ e $u_p(\eta) = 0$, onde

$$U_p(\xi) = \frac{n}{\eta}\frac{\sum_{i=1}^{n} w_i}{\sum_{i=1}^{n}\log\left[\Phi\left(\frac{\log(y_i)-\xi}{\eta}\right)\right]} + \frac{1}{\eta}\sum_{i=1}^{n} w_i + \frac{1}{\eta}\sum_{i=1}^{n} z_i$$

e

$$U_p(\eta) = \frac{n}{\eta}\frac{\sum_{i=1}^{n} z_i w_i}{\sum_{i=1}^{n}\log\left[\Phi\left(\frac{\log(y_i)-\xi}{\eta}\right)\right]} + \frac{1}{\eta}\sum_{i=1}^{n} z_i w_i + \frac{1}{\eta}\sum_{i=1}^{n} z_i^2 - \frac{n}{\eta}.$$

O fraco desempenho de $\ell_p(\boldsymbol{\theta})$ que leva a estimativas inconsistentes e ineficientes, em alguns casos pode ser explicado pelo fato desta função não constituir uma tentativa de aproximação para verossimilhanças marginal e condicional genuínas. Barndorff-Nielsen (1983) propõe uma verossimilhança perfilada modificada a qual é invariante sob reparametrizações da forma $(\boldsymbol{\tau}, \phi) \to (\boldsymbol{\tau}, \varsigma(\boldsymbol{\tau}, \phi))$ onde ς é uma função de $\boldsymbol{\tau}$ e ϕ. (Para maiores detalhes sobre esta proposta veja Severini, 1998). Assim, a função de verossimilhança perfilada

modificada de Barndorff-Nielsen é dada por

$$L_{BN}(\boldsymbol{\tau}) = \left|\frac{\partial\hat{\phi}_{\boldsymbol{\tau}}}{\partial\hat{\phi}}\right|^{-1} \left|j_{\phi\phi}(\boldsymbol{\tau}, \hat{\phi}_{\boldsymbol{\tau}})\right|^{-1/2} L_p(\boldsymbol{\tau}),$$

onde $\left|\frac{\partial\hat{\phi}_{\boldsymbol{\tau}}}{\partial\hat{\phi}}\right|$ é a matriz de derivadas parciais de $\hat{\phi}_{\boldsymbol{\tau}}$ com respeito a $\hat{\phi}$, $j_{\phi\phi}(\boldsymbol{\tau}, \phi) = -\partial^2 l/\partial\phi\partial\phi'$ é a matriz de informação observada para ϕ quando $\boldsymbol{\tau}$ é fixo e $L_p(\boldsymbol{\tau})$ é a função de verossimilhança perfilada para $\boldsymbol{\tau}$. Pela estrutura matemática da quantidade $\frac{\partial\hat{\phi}_{\boldsymbol{\tau}}}{\partial\hat{\phi}}$, esta na maioria das situações é de dificil obtenção ou até mesmo impossível. Assim, Severini (1998) obtém uma aproximação para a função de verossimilhança perfilada modificada de Barndorff-Nielsen. Esta expressão é dada por

$$\bar{\ell}_{BN}(\boldsymbol{\tau}) = \ell_p(\boldsymbol{\tau}) + \frac{1}{2}\log\left|j_{\phi\phi}(\boldsymbol{\tau}, \hat{\phi}_{\boldsymbol{\tau}})\right| - \log\left|I_\phi(\boldsymbol{\tau}, \hat{\phi}_{\boldsymbol{\tau}}; \hat{\boldsymbol{\tau}}, \hat{\phi})\right|$$

onde $I_\phi(\boldsymbol{\tau}, \phi; \boldsymbol{\tau}_0, \phi_0) = \mathbb{E}_{\boldsymbol{\tau}_0,\phi_0}\left\{\ell_\phi(\boldsymbol{\tau}, \phi)\ell_\phi(\boldsymbol{\tau}_0, \phi_0)'\right\}$ é a matriz de covariâncias da derivada da log-verossimilhança e $\ell_\phi(\boldsymbol{\tau}, \phi) = \partial\ell/\partial\phi$. Dado o complexo que pode resultar o calculo da esperança anterior, uma aproximação alternativa, a qual pode ser facilmente calculada, com o mesmo erro de aproximação foi proposta por Severini (1999) e é dada por

$$\breve{\ell}_{BN}(\boldsymbol{\tau}) = \ell_p(\boldsymbol{\tau}) + \frac{1}{2}\log\left|j_{\phi\phi}(\boldsymbol{\tau}, \hat{\phi}_{\boldsymbol{\tau}})\right| - \log\left|\breve{I}_\phi(\boldsymbol{\tau}, \hat{\phi}_{\boldsymbol{\tau}}; \hat{\boldsymbol{\tau}}, \hat{\phi})\right|,$$

onde $\breve{I}_\phi(\boldsymbol{\tau}, \phi; \boldsymbol{\tau}_0, \phi_0) = \sum_{j=1}^n \left\{\ell_\phi^{(j)}(\boldsymbol{\tau}, \phi)\ell_\phi^{(j)}(\boldsymbol{\tau}_0, \phi_0)'\right\}$ e $\ell_{\boldsymbol{\theta}}^{(j)}(\boldsymbol{\theta}) = \left(\ell_{\boldsymbol{\tau}}^{(j)}(\boldsymbol{\theta}), \ell_\phi^{(j)}(\boldsymbol{\theta})\right)$ sendo a função escore da j-ésima observação (vide Cysneiros, Cribari-Neto e Araújo, 2008). Para a distribuição LPN a log verossimilhança perfilada modificada de Barndorff-Nielsen é

$$\ell_{BN}(\xi, \eta) = \ell_p(\xi, \eta) + \log\left(\frac{\sqrt{n}}{\hat{\alpha}_{\xi,\eta}}\frac{1}{\left|\breve{I}(\hat{\alpha}_{\xi,\eta}, \xi, \eta; \hat{\alpha}, \hat{\xi}, \hat{\eta})\right|}\right) \tag{2.9}$$

onde

$$\begin{aligned}\breve{I}(\hat{\alpha}_{\xi,\eta}, \xi, \eta; \hat{\alpha}, \hat{\xi}, \hat{\eta}) = {} & \frac{n}{\hat{\alpha}_{\xi,\eta}\hat{\alpha}} + \frac{1}{\hat{\alpha}_{\xi,\eta}\hat{\alpha}}\sum_{i=1}^n \log\left[\Phi\left(\frac{\log(y_i) - \xi}{\eta}\right)\right]\log\left[\Phi\left(\frac{\log(y_i) - \hat{\xi}}{\hat{\eta}}\right)\right] \\ & + \frac{1}{\hat{\alpha}_{\xi,\eta}}\sum_{i=1}^n \log\left[\Phi\left(\frac{\log(y_i) - \xi}{\eta}\right)\right] + \frac{1}{\hat{\alpha}}\sum_{i=1}^n \log\left[\Phi\left(\frac{\log(y_i) - \hat{\xi}}{\hat{\eta}}\right)\right].\end{aligned}$$

com $\hat{\alpha}_{\xi,\eta}$ o estimador de α sob a log verossimilhança perfilada e $\hat{\xi}, \hat{\eta}$ e $\hat{\alpha}$ são os estimadores de máxima verossimilhança de ξ, η e α.

Outra aplicação importante da verossimilhança perfilada são os testes de hipóteses. Assim, para testar $H_0 : \boldsymbol{\tau} = \mathbf{0}$ contra $\boldsymbol{\tau} \neq \mathbf{0}$ nós podemos usar as estatísticas de razão de verossimilhança perfilada ou de máxima verossimilhança perfilada modificada,

$$LR_p = 2\{\ell_p(\hat{\boldsymbol{\tau}}) - \ell_p(\boldsymbol{\tau})\} \qquad \text{ou} \qquad LR_{BN} = 2\{\ell_{BN}(\hat{\boldsymbol{\tau}}) - \ell_{BN}(\boldsymbol{\tau})\}.$$

as quais sob condições de regularidades satisfeitas seguem uma distribuição χ^2 com k graus de libertade, onde k é a dimensão do vetor $\boldsymbol{\tau}$.

2.5 Resultados Numéricos: estimação

O objetivo é comparar, através de simulação, o desempenho dos estimadores de máxima verossimilhança, $\hat{\boldsymbol{\theta}}$, máxima verossimilhança perfilada, $\widetilde{\boldsymbol{\theta}}$, e máxima verossimilhança perfilada modificada, $\widetilde{\widetilde{\boldsymbol{\theta}}}$, dos parâmetros que indexam a distribuição LPN. Esta verossimilhança perfilada é obtida para o casso onde o parâmetro de interesse é $\boldsymbol{\tau} = (\xi, \eta)'$. Para isso foram calculados medidas de qualidades para a estimação pontual como viés e a raiz do erro quadrático médio ($\sqrt{EQM}$). Para a geração das sequências de números pseudo-aleatórios com distribuição LPN foi usado o método de inversão. As simulações foram feitas usando a função optim("L-BFGS-B") do pacote estatístico R.

Se formula um experimento de simulação de Monte Carlo baseado em 5.000 replicas considerando os tamanhos amostrais $n = 20, 60, 100$. Os valores verdadeiros considerados para os parâmetros foram: $\alpha = 0.75, 1.25, 1.75, 2.0$, $\xi = 2.0, 4.0, 8.0$ e para o parâmetro de escala atribui-se, sem perda de generalidade, um único valor $\eta = 1.0$.

Os resultados mostram, veja as Tabelas $A.1$ e $A.2$, que o viés e $\sqrt{EQM}$ para cada um dos estimadores estudados decrescem quando o tamanho de amostra é maior, ou seja que os estimadores são assintoticamente consistentes. Geralmente para $\alpha > 1$ o estimador de máxima verossimilhança de ξ apresenta um viés positivo ou seja que este superestima o valor do verdadeiro parâmetro, enquanto que o estimador pelo método de máxima verossimilhança perfilada apresenta um viés negativo, subestimando o valor do parâmetro. Mesmo assim, os estimadores de máxima verossimilhança e máxima verossimilhança perfilada modificada também apresentam viés negativo para o parâmetro η. Em

geral, o viés dos estimadores de máxima verossimilhança perfilada e modificada apresentam modulos menores ao do estimador de máxima verossimilhança para o parâmetro η. Quanto ao parâmetro α, nós concluímos que o estimador de máxima verossimilhança é menos viésado e mais eficiente que os estimadores de máxima verossimilhança perfilada e modificada; além disso, o viés do estimador de máxima verossimilhança é negativo para os valores maiores de α. Nós também concluímos que os estimadores de máxima verossimilhança perfilada e modificada do parâmetro ξ é menos eficientes, maior $\sqrt{EQM}$, que os estimadores de máxima verossimilhança.

2.5.1 Ilustração

Os dados na seguinte ilustração foram considerados por Balakrishnan, Leiva, Sanhueza e Cabrera (2009), num estudo de contaminação ambiental, em Santiago de Chile, uma das cidades mais contaminadas ambientalmente no mundo. Os dados correspondem à concentração de dióxido de enxofre no ar em certo horário do dia, variável Y, observada em uma estação de monitoramento localizada na cidade de Santiago. Uma analise descritiva de Y e $\log(Y)$ são dadas na tabela 2.4. Os altos valores dos coeficientes de assimetria e

Tabela 2.4 *Estatísticas descritivas para as variáveis Y e $\log(Y)$.*

Variável	n	media	variância	$\sqrt{b_1}$	b_2
Y	744	2.9260	4.0604	4.3213	37.5707
$\log(Y)$	744	0.9226	0.2762	0.4099	3.7611

curtose para o $\log(Y)$, com respeito ao da distribuição normal (0 e 3 respectivamente), são uma forte justificativa para acreditar que uma distribuição LPN pode apresentar melhor ajuste. Mesmo assim, os coeficientes de assimetria e curtose da variável aleatória Y são muito distintos que os esperados para o modelo log-normal (1.8714 e 39.4568, respectivamente). Para uma maior justificativa, nós levamos a cabo o teste de hipóteses paramétrico de log-normalidade do tipo

$$H_0 : \alpha = 1 \quad \text{contra} \quad H_1 : \alpha \neq 1,$$

Tabela 2.5 *Estimativas (erro padrão) dos parâmetros das distribuições LN e LPN.*

Modelo log-Nomal		Modelo LPN	
Parâmetro	**estimativa(Erro padrão)**	**Parâmetro**	**estimativa(Erro padrão)**
Loglik	-2753.666	*Loglik*	-2751.538
μ	0.9226(0.0192)	ξ	0.5171(0.0553)
σ^2	0.2759(0.0143)	η	0.6867(0.0466)
		α	2.0891(0.1039)

mediante a estatística de prova

$$\lambda = \frac{\ell_{LN}(\hat{\boldsymbol{\theta}})}{\ell_{LPN}(\hat{\boldsymbol{\theta}})}.$$

Os resultados obtidos levam a $-2\log(\Lambda) = 4.256$, o qual é um valor maior que o valor critico da distribuição qui-quadrado, com um grau de liberdade, ao nível 5%, $\chi^2_{1,5\%} = 3.8414$. Portanto, o modelo LPN apresenta um melhor ajuste para modelar os dados de concentração de dióxido de enxofre na cidade de Santiago do Chile. As estimativas de máxima verossimilhança dos parâmetros para os modelos log-normal(LN) e LPN, com erros padrão em parenteses obtidos a partir da matriz de informação observada, são dados na Tabela 2.5, mesmo assim, um gráfico da situação é dado pela Figura 2.4.

2.5.2 Modelo Potência-*Skew*-Normal

Diversas generalizações do modelo (1.3) foram dadas. Balakrishnan (2002), propõe a família de funções de densidade de probabilidade ($a > 0, \quad \lambda \in \mathbb{R}$)

$$\varphi(z;\lambda,a) = k(\lambda,a)\,[\Phi(\lambda z)]^{a-1}\,\phi(z), \quad z \in \mathbb{R}, \tag{2.10}$$

a qual para $\lambda = 1$ é a função de densidade de Durran e para $a = 2$ se obtém a normal assimétrica de Azzalini.

Gupta e Gupta (2004), definem a função de densidade de probabilidade da distribuição

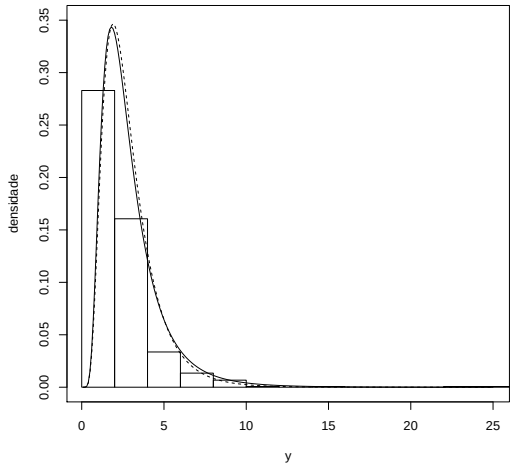


Figura 2.4 *Histograma da variável Y e modelos ajustados* $LN(0.9226, 0.2759)$ *(linha tracejada) e* $LPN(0.5171, 0.6867, 2.0891)$ *(linha contínua).*

normal assimétrica generalizada por

$$\varphi(z;\lambda) = \frac{\phi(z)\{\Phi(\lambda z)\}^n}{C_n(\lambda)}, \quad z \in \mathbb{R}, \tag{2.11}$$

onde $n \in \mathbb{Z}^+$, $\lambda \in \mathbb{R}$ e $C_n(\lambda) = \int_{-\infty}^{\infty} \phi(z)\{\Phi(\lambda z)\}^n dz$. Assim, para $n = 1$ se tem a distribuição (1.2) de Azzalini e para $\lambda = 1$ temos a distribuição (1.5) de Durrans.

Uma desvantagem dos modelos (2.10) e (2.11) é a constante de normalização a qual na maioria dos casos apresenta problemas nos processos de maximização para a estimação dos parâmetros do modelo os quais são maiores para o caso de localização e escala. Baseados nas características dos parâmetros de forma α e λ de modelar assimetria e curtose nos modelos de Durrans e Azzalini, respectivamente, nós propomos agora uma terceira alternativa para a generalização dos dois modelos. Assim, nós definimos o modelo potência-*skew*-normal, cuja função de densidade é dada por

$$\phi_{\alpha,\lambda}(z) = \alpha\phi_\lambda(z)\left\{\Phi_\lambda(z)\right\}^{\alpha-1}, \quad z \in \mathbb{R}, \ \ \lambda \in \mathbb{R} \ \ e \ \ \alpha \in \mathbb{R}^+, \tag{2.12}$$

onde

$$\phi_\lambda(z) = 2\phi(z)\Phi(\lambda z), \qquad \Phi_\lambda(z) = \int_{-\infty}^{z} \phi_\lambda(t)dt.$$

Nós a denotamos $PSN(\lambda, \alpha)$. No modelo (2.12), se $\alpha = 1$, então $\phi_{\alpha=1,\lambda}(z) = \varphi(z;\lambda)$,

a normal assimétrica de Azzalini, enquanto que se $\lambda = 0$ então $\phi_{\alpha,\lambda=0}(z) = \varphi(z;\alpha)$, a normal generalizada de Durrans. Mesmo assim este novo modelo contém a distribuição normal para o caso especial $\alpha = 1$ e $\lambda = 0$, isto é $\phi_{\alpha=1,\lambda=0}(z) = \phi(z)$. Se $\alpha = 2$ então temos um tipo de distribuição *skew-skew*, enquanto se $\lambda = 1$, então temos um modelo potência-potência. Os gráficos (a) e (b) na Figura 2.5 mostram que simultaneamente os parâmetros α e λ afetam a assimetria e a curtose da distribuição, com o qual se obtêm um modelo mais flexível que os obtidos por Azzalini (1985) e Durrans (1992).

(a) (b)

Figura 2.5 *Gráfico da distribuição PSN. (a)* $\lambda = 0.70$ *e* $\alpha =$ *0.50 (linha de pontos e tracejada), 1.0 (linha de pontos), 2.0 (linha tracejada) e 5.0 (linha continua). (b)* $\alpha = 0.5$ *e* $\lambda =$*-0.70 (linha de pontos e tracejada), 0 (linha de pontos), 1 (linha tracejada) e 1.75 (linha continua).*

Se $Z \sim PSN(\lambda, \alpha)$, os momentos da variável Z não tem uma forma fechada, mas através de uma mudança de variável o r-ésimo momento da variável Z pode escrever-se como

$$E(Z^r) = \alpha \int_0^1 [\Phi_\lambda^{-1}(z)]^r z^{\alpha-1} dz, \tag{2.13}$$

onde Φ_λ^{-1} é a função inversa de Φ_λ. Esta esperança é igual à esperança da expressão $[\Phi_\lambda^{-1}(z)]^r$ de uma variável aleatória com distribuição beta de parâmetros α e 1.

Para alguns valores de λ e valores de α entre 0.1 e 100, temos que os coeficientes de assimetria e curtose, $\sqrt{\beta_1}$, β_2, para $Z \sim PSN(\lambda, \alpha)$ estão no intervalo [-1.4676,0.9953) e [1.4672,5.4386] respectivamente, os quais contem os intervalos para os modelos *skew* -

normal, (-0.9953,0.9953) e [3,3.8692) respectivamente, e PN, [-0.6115,0.9007] e [1.7170,4.3556], respectivamente. Isto mostra que o modelo potência-*skew*-normal contém maior assimetria negativa e além disso contem distribuições que são mais platicúrticas e/ou leptocúrticas, que os modelos de Azzalini e Durrans. Estas duas propriedades são uma grande vantagem do modelo PSN frente aos modelos *skew*-normal e PN.

2.6 Caso Localização-Escala

Se $Z \sim PSN(\lambda, \alpha)$, a extensão ao caso localização-escala segue da transformação $X = \xi + \eta Z$, onde $\xi \in \mathbb{R}$ e $\eta \in \mathbb{R}^+$. Então temos que a densidade de X fica:

$$\phi_{\alpha,\lambda}(x;\xi,\eta) = \alpha\phi_\lambda(x;\xi,\eta,\lambda,\alpha)\left\{\Phi_\lambda(x;\xi,\eta)\right\}^{\alpha-1}, \qquad x \in \mathbb{R}. \tag{2.14}$$

onde

$$\phi_\lambda(x;\xi,\eta) = \frac{2}{\eta}\phi\left(\frac{x-\xi}{\eta}\right)\Phi\left(\lambda\frac{x-\xi}{\eta}\right), \qquad \Phi_\lambda(x;\xi,\eta) = \int_{-\infty}^{x}\phi_\lambda(t;\xi,\eta)dt.$$

Usamos a notação $X \sim PSN(\xi,\eta,\lambda,\alpha)$, onde ξ é o parâmetro de localização e η é o parâmetro de escala. Assim nós temos que $PSN(\lambda,\alpha) = PSN(0,1,\lambda,\alpha)$.

Para uma amostra aleatória $\mathbf{X} = (X_1, X_2, ..., X_n)'$ com $X_i \sim PSN(\xi,\eta,\lambda,\alpha)$, a função de log-verossimilhança de $\boldsymbol{\theta} = (\xi,\eta,\lambda,\alpha)'$ é dada por:

$$\ell(\theta,\mathbf{X}) = n\log(\alpha) - n\log(\eta) - \frac{1}{2}\sum_{i=1}^{n} z_i^2 + \sum_{i=1}^{n}\log\left\{\Phi(\lambda z_i)\right\} + (\alpha-1)\sum_{i=1}^{n}\log\left\{\Phi_\lambda(z_i)\right\},$$

onde $z_i = \frac{x_i-\xi}{\eta}$. Os elementos da função escore são:

$$U(\xi) = \frac{1}{\eta}\sum_{i=1}^{n} z_i - \frac{\lambda}{\eta}\sum_{i=1}^{n}\frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)} - n\frac{\alpha-1}{\eta},$$

$$U(\eta) = -\frac{n}{\eta} + \frac{1}{\eta}\sum_{i=1}^{n} z_i^2 - \frac{\lambda}{\eta}\sum_{i=1}^{n} z_i\frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)} - \frac{\alpha-1}{\eta}\sum_{i=1}^{n} z_i,$$

$$U(\lambda) = \sum_{i=1}^{n} x_i\frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)} - \sqrt{\frac{2}{\pi}}\frac{(\alpha-1)}{1+\lambda^2}\sum_{i=1}^{n}\frac{\phi\left(\sqrt{1+\lambda^2}z_i\right)}{\Phi_\lambda(z_i)},$$

$$U(\alpha) = \frac{n}{\alpha} + \sum_{i=1}^{n}\log\left\{\Phi_\lambda(z_i)\right\}.$$

Os estimadores de máxima verossimilhança do vetor $\boldsymbol{\theta} = (\xi, \eta, \lambda, \alpha)$ podem ser obtidos a partir de métodos numéricos iterativos. Para $w = \frac{\phi(\lambda z)}{\Phi(\lambda z)}$ e $w_\lambda = \frac{\phi\left(\sqrt{1+\lambda^2}z\right)}{\Phi_\lambda(z)}$, obtemos os elementos da matriz hessiana (H) dados por:

$$\frac{\partial^2 \ell(\boldsymbol{\theta}, \mathbf{X})}{\partial \xi^2} = -\frac{n}{\eta^2} - \frac{\lambda^2}{\eta^2}\sum_{i=1}^{n} \lambda z_i w_i - \frac{\lambda^2}{\eta^2}\sum_{i=1}^{n} w_i^2,$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta}, \mathbf{X})}{\partial \eta \partial \xi} = -\frac{2}{\eta^2}\sum_{i=1}^{n} z_i - \frac{\lambda^3}{\eta^2}\sum_{i=1}^{n} z_i^2 w_i - \frac{\lambda^2}{\eta^2}\sum_{i=1}^{n} z_i w_i^2 + \frac{\lambda}{\eta^2}\sum_{i=1}^{n} w_i + n\frac{\alpha-1}{\eta^2},$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta}, \mathbf{X})}{\partial \lambda \partial \xi} = -\frac{1}{\eta}\sum_{i=1}^{n}\left[w_i - \lambda^2 z_i^2 w_i - \lambda z_i w_i^2\right], \qquad \frac{\partial^2 \ell(\boldsymbol{\theta}, \mathbf{X})}{\partial \alpha \partial \xi} = -\frac{n}{\eta},$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta}, \mathbf{X})}{\partial \eta^2} = \frac{n}{\eta^2} - \frac{3}{\eta^2}\sum_{i=1}^{n} z_i^2 + \frac{2\lambda}{\eta^2}\sum_{i=1}^{n} z_i w_i - \frac{\lambda^3}{\eta^2}\sum_{i=1}^{n} z_i^3 w_i - \frac{\lambda^2}{\eta^2}\sum_{i=1}^{n} z_i^2 w_i^2 + 2\frac{\alpha-1}{\eta^2}\sum_{i=1}^{n} z_i,$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta}, \mathbf{X})}{\partial \lambda \partial \eta} = -\frac{1}{\eta}\sum_{i=1}^{n} z_i w_i + \frac{\lambda^2}{\eta}\sum_{i=1}^{n} z_i^3 w_i + \frac{\lambda}{\eta}\sum_{i=1}^{n} z_i^2 w_i^2,$$

$$\begin{aligned}\frac{\partial^2 \ell(\boldsymbol{\theta}, \mathbf{X})}{\partial \lambda^2} = &-\sum_{i=1}^{n} z_i^2(\lambda z_i w_i + w_i^2) + \sqrt{\frac{2}{\pi}}\frac{2\lambda(\alpha-1)}{(1+\lambda^2)^2}\sum_{i=1}^{n} w_{\lambda i} \\ &- 2(\alpha-1)\sum_{i=1}^{n}\left[-\sqrt{\frac{1}{2\pi}}\frac{\lambda}{1+\lambda^2} z_i^2 w_{\lambda i} + \frac{1}{\pi}\frac{1}{(1+\lambda^2)^2} w_{\lambda i}^2\right],\end{aligned}$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta}, \mathbf{X})}{\partial \alpha \partial \eta} = -\frac{1}{\eta}\sum_{i=1}^{n} z_i, \qquad \frac{\partial^2 \ell(\boldsymbol{\theta}, \mathbf{X})}{\partial \alpha \partial \lambda} = -\sqrt{\frac{2}{\pi}}\frac{1}{1+\lambda^2}\sum_{i=1}^{n} w_{\lambda i}, \qquad \frac{\partial^2 \ell(\boldsymbol{\theta}, \mathbf{X})}{\partial \alpha^2} = -\frac{n}{\alpha^2}.$$

Os elementos da matriz de informação esperada($I_O(\theta)$) podem ser encontrados a partir da matriz hessiana como $I_O(\theta) = -H$. Chamando $a_{jk} = \mathbb{E}(Z^j W^k)$ e $a_{jk,\lambda} = \mathbb{E}(Z^j W_\lambda^k)$ nós escrevemos os elementos da matriz de informação esperada de Fisher, definidos e denotados da mesma forma que no modelo LPN, como

$$i_{\xi\xi} = \frac{1}{\eta^2} + \frac{\lambda^3}{\eta^2}a_{11} + \frac{\lambda^2}{\eta^2}a_{02}, \qquad i_{\eta\xi} = \frac{2}{\eta^2}a_{10} + \frac{\lambda^3}{\eta^2}a_{21} + \frac{\lambda^2}{\eta^2}a_{12} - \frac{\lambda}{\eta^2}a_{10} - \frac{\alpha-1}{\eta^2},$$

$$_{\lambda\xi} = \frac{1}{\eta}\left[a_{01} - \lambda^2 a_{21} - \lambda a_{12}\right], \qquad i_{\alpha\xi} = \frac{1}{\eta}, \qquad i_{\lambda\eta} = \frac{1}{\eta}\left[a_{11} - \lambda^2 a_{31} - \lambda a_{22}\right],$$

$$i_{\eta\eta} = -\frac{1}{\eta^2} + \frac{3}{\eta}a_{20} - \frac{2\lambda}{\eta^2}a_{11} + \frac{\lambda^3}{\eta}a_{31} + \frac{\lambda^2}{\eta^2}a_{22} - 2\frac{\alpha-1}{\eta^2}a_{10}, \qquad i_{\alpha\eta} = \frac{1}{\eta}a_{10},$$

$$i_{\lambda\lambda} = \lambda a_{31} + a_{22} - \sqrt{\frac{2}{\pi}}\frac{2\lambda(\alpha-1)}{(1+\lambda^2)^2}a_{01,\lambda} + \sqrt{\frac{2}{\pi}}(\alpha-1)\left[-\frac{\lambda}{1+\lambda^2}a_{21,\lambda} + \sqrt{\frac{2}{\pi}}\frac{1}{(1+\lambda^2)^2}a_{02,\lambda}\right],$$

$$i_{\alpha\lambda} = \sqrt{\frac{2}{\pi}}\frac{1}{1+\lambda^2}a_{01,\lambda}, \qquad i_{\alpha\alpha} = \frac{1}{\alpha^2},$$

Para $\alpha = 1$ e $\lambda = 0$ e com base na aproximação de funções obtidas por Chaibub-Neto e Branco (2003), $\frac{1}{\pi}\frac{\phi(z)}{\sqrt{\Phi(z)[1-\Phi(z)]}} \approx \frac{1}{\sqrt{2\pi}(\pi/2)}\exp\left(-\frac{z^2}{2(\pi^2/4)}\right)$ nós obtemos a seguinte matriz de informação esperada

$$I_F(\theta) = \begin{pmatrix} \frac{1}{\eta^2} & 0 & \sqrt{\frac{2}{\pi}}\frac{1}{\eta} & \frac{1}{\eta} \\ 0 & \frac{2}{\eta^2} & 0 & 0 \\ \sqrt{\frac{2}{\pi}}\frac{1}{\eta} & 0 & \frac{2}{\pi} & \sqrt{\frac{1}{2}} \\ \frac{1}{\eta} & 0 & \sqrt{\frac{1}{2}} & 1 \end{pmatrix},$$

cujo determinante é $|I_F(\theta)| = \frac{8-\pi}{2\pi}\frac{1}{\eta^4}$, portanto a matriz de informação esperada é não singular. Então, pela distribuição assintótica do estimador de máxima verossimilhança temos que, $\hat{\boldsymbol{\theta}} \overset{A}{\sim} N_4\left(\boldsymbol{\theta}, I_F^{-1}(\boldsymbol{\theta})\right)$. Assim, concluímos que o estimador de máxima verossimilhança é assintoticamente consistente, com distribuição normal onde a matriz de covariâncias é a inversa da matriz de informação de Fisher.

2.7 Ilustração

Na seguinte aplicação nós usamos os dados relacionados com as 1150 medições da altura em intervalos de uma micra do tambor dum rolo, na ampla pesquisa sobre a rugosidade da superfície dos rolos. Os dados estão disponíveis em http://lib.stat.cmu.edu/jasadata/laslett. A Tabela 2.6 mostra algumas estatísticas descritiva dos dados, onde $\sqrt{b_1}$ e b_2 são os coeficientes de assimetria e curtose amostrais.

Tabela 2.6 *Estatísticas descritivas para altura de rolo*

n	Média	Variância	$\sqrt{b_1}$	b_2
1150	3.535	0.650	-0.986	4.855

De acordo aos valores de assimetria e curtose nós propomos um modelo assimétrico. Assim, ajustamos os modelos PN, *skew*-normal e PSN. As estimativas de máxima verossimilhança

para os modelos em questão são dados na Tabela 2.7, com erros padrão (entre parênteses) calculados a partir da matriz de informação observada para cada modelo, respectivamente. Assim, temos que o modelo PSN se ajusta melhor aos dados de altura do rolo. Um gráfico dos modelos ajustados é apresentado na figura 2.6. Mesmo assim, o gráfico Q-Q "plot" para os três modelos é apresentado na Figura 2.7; este gráfico foi obtido dos quantis amostrais e os teóricos obtidos a partir da função de distribuição com as estimativas dos parâmetros. Para uma justificativa mais forte nós fazemos os dois testes de hipóteses paramétricos de não diferença dos modelo PN e *skew*-normal contra o modelo PSN. Assim, nós temos

$$H_{01} : \lambda = 0 \quad \text{contra} \quad H_{11} : \lambda \neq 0 \qquad \text{e} \qquad H_{02} : \alpha = 1 \quad \text{contra} \quad H_{12} : \alpha \neq 1$$

mediante o uso das estatísticas de teste

$$\Lambda_1 = \frac{\ell_{PN}(\hat{\boldsymbol{\theta}})}{\ell_{PSN}(\hat{\boldsymbol{\theta}})} \qquad \text{e} \qquad \Lambda_2 = \frac{\ell_{SN}(\hat{\boldsymbol{\theta}})}{\ell_{PSN}(\hat{\boldsymbol{\theta}})}.$$

Os valores calculados destas estatísticas levam que: $-2\log(\Lambda_1) = 54.442$ e $-2\log(\Lambda_2) = 26.324$, os quais são valores muito maiores que o valor da distribuição qui-quadrado, com um grau de liberdade, ao nível 5%.

Tabela 2.7 *Estimativas (erro padrão) dos parâmetros das distribuições PN, SN e PSN.*

	Modelo PN	Modelo SN	Modelo PSN
Parâmetro	**estimativa**	**estimativa**	**estimativa**
Loglik	-1085.241	-1071.362	-1058.200
ξ	4.5495(0.0572)	4.2503(0.0284)	3.9939(0.0522)
η	0.1982(0.0279)	0.9694(0.0304)	5.0493(0.1417)
λ	–	-2.7864(0.2529)	-12.4238(0.3567)
α	0.0479(0.0156)	–	10.2356(0.4016)

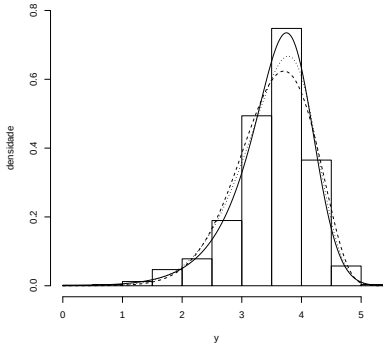
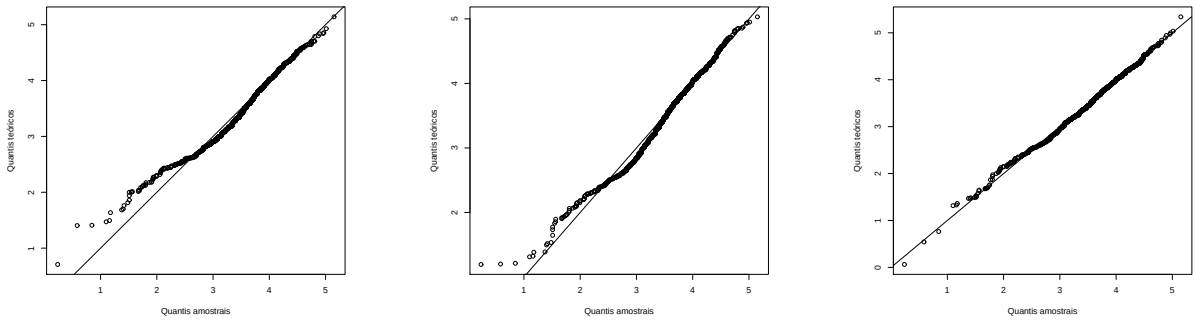


Figura 2.6 *Densidades:* $PN(4.5495, 0.1982, 0.0479)$ *(linha tracejada),* $SN(4.2503, 0.9694, -2.7864)$ *(linha de pontos) e* $PSN(3.9939, 5.0493, -12.4238, 10.2356)$ *(linha continua).*



(a) (b) (c)

Figura 2.7 *Gráfico Q-Q "plot"(a)* PN*, (b) SN e (c) PSN.*

Capítulo 3

Distribuições PN Flexível e Bimodal

3.1 Motivação

No capítulo (1) nós introducimos os modelos *skew-symmetric* e α-potência e como casos especiais os modelos *skew-normal* (SN) e potência normal assimétrico (PN). Estes duas famílias de densidades tem a desvantagem que as distribuições são unimodal. As vezes a distribuição dos dados apresentam um comportamento bimodal e estes modelos não são óptimos para os respectivos ajustes. Alguns modelos tem sido implemetados neste contexto, entre estes, a mistura de distribuições normais ou normais assimétricas, más alguns destes modelos apresentam complicadas implementações computacionais e para alguns casos os processos de estimação dos parâmetros do modelos apresentam problemas de identificabilidade. Portanto, é de interesse estudar distribuições assimétricas que sejam mais flexíveis que os modelos SN e PN, para modelar comportamentos bimodais dos dados, além disso que não sejam misturas de distribuições e que seus processos de estimação sejam simples de implementar.

O seguinte lema é dado por Azzalini (1985), o qual apresenta um resultado fundamental para o desenvolvimento de modelos assimétricos e simétricos unimodal e bimodal.

Lema 3.1.1. *Sejam f_0 uma função de densidade de probabilidade simétrica ao redor de zero e G uma função de distribuição tal que G' existe e é uma função de densidade*

simétrica ao redor de zero, então

$$f_Z(z;\lambda) = 2f_0(z)G(\lambda z), \qquad z \in \mathbb{R}, \tag{3.1}$$

é uma função de densidade para qualquer $\lambda \in \mathbb{R}$.

Lema 3.1.2. *Seja g uma função de densidade de probabilidade simétrica ao redor de zero e G sua função de distribuição cumulativa tal que G' existe e é uma função de densidade simétrica ao redor de zero, então*

$$f_Z(z;\lambda) = 2g(z)G(w(z)), \qquad z \in \mathbb{R}, \tag{3.2}$$

é uma função de densidade para qualquer função impar $w(z)$.

Lema 3.1.3. *Seja f uma função densidade de probabilidade simétrica ao redor de zero, F sua respectiva função de distribuição cumulativa e G uma função de distribuição absolutamente contínua tal que G' é simétrica ao redor de zero. Então*

$$g(z;\lambda,\delta) = c_\delta f(|z|+\delta)G(\lambda z), \qquad z,\lambda,\delta \in \mathbb{R}, \tag{3.3}$$

é a função densidade da variável aleatória Z e $c_\delta^{-1} = (1 - F(\delta))$.

Demostração: vide Gómez et al. (2009).

Gómez et al. (2009), estuda a distribuição skew-normal-flexible com função de densidade de probabilidade dada por

$$f(z;\lambda,\delta) = c_\delta \phi(|z|+\delta)\Phi(\lambda z), \tag{3.4}$$

onde δ é um número real e c_δ é uma constante de normalização. Eles demostram que para $\delta < 0$ o modelo (3.4) é bimodal. Este é denotado por $SNF(\lambda,\delta)$.

Neste capítulo considera-se uma extensão ao modelo α-potência normal assimétrico flexível com a inclusão de um parâmetro adicional no modelo PN, com o qual pode-se obter um modelo unimodal ou bimodal assimétrico. Além disso, também se definem os modelos bimodal simétrico e o assimétrico para o caso da distribuição α-potência, baseados nos trabalhos de Kim (2005) e Arnold (2009), respectivamente.

3.2 Distribuição PNF

Um dos resultados mais importante desta seção é que a partir da família de densidades PNF, pode-se obter densidades bimodais.

Definição 3.2.1. *A variável aleatória Z segue a distribuição α-potência normal assimétrica flexível de parâmetros α e δ, se sua função de densidade de probabilidade é dada por*

$$\varphi(z;\alpha,\delta) = k(\alpha,\delta)\phi(|z|+\delta)\{\Phi(z)\}^{\alpha-1}, \qquad z,\delta \in \mathbb{R}, \quad \alpha \in \mathbb{R}^+, \tag{3.5}$$

onde

$$k(\alpha,\delta) = k_{\alpha,\delta} = \left[\sqrt{2\pi}\phi(\delta)\int_0^1 u^{\alpha-1}\exp\{-\delta\left|\Phi^{-1}(u)\right|\}du\right]^{-1},$$

é uma constante de normalização.

Se a variável Z segue o modelo α-potência normal assimétrica flexível, usamos a notação $Z \sim PNF(\alpha,\delta)$.

3.2.1 Propriedades

Para a família de distribuições PNF temos as seguintes propriedades.

Proposição 3.2.2. *Se $Z \sim PNF(\alpha,\delta)$, então*

1. $\varphi(z;\alpha=1,\delta=0) = \phi(z)$.

2. $\varphi(z;\alpha,\delta=0) = \alpha\phi(z)\{\Phi(z)\}^{\alpha-1} \to PN(\alpha)$.

3. $\varphi(z;\alpha=2,\delta=0) = 2\phi(z)\{\Phi(z)\} \to SN(\lambda=1)$, *a normal assimétrica com* $\lambda=1$.

4. $\varphi(z;\alpha=2,\delta) = c_\delta\phi(|z|+\delta)\{\Phi(z)\} \to SNF(1,\delta)$, *SNF de Gómez et al. (2009).*

3.3 Distribuição PN Bimodal

Proposição 3.3.1. *Seja $Z \sim PNF(\alpha,\delta)$, se $\delta \neq 0$ então a função de densidade PNF não é diferenciável em Z=0.*

Demostração: Note que

$$Lim_{z\to 0^+}\varphi(z;\acute{\alpha},\delta) = (1/2)^{\alpha-2}k_{\alpha,\delta}\phi(\delta)\left[\frac{\alpha-1}{\sqrt{2\pi}} - \frac{\delta}{2}\right]$$

e

$$Lim_{z\to 0^-}\varphi(z;\acute{\alpha},\delta) = (1/2)^{\alpha-2}k_{\alpha,\delta}\phi(\delta)\left[\frac{\alpha-1}{\sqrt{2\pi}} + \frac{\delta}{2}\right],$$

de onde conclui-se que a função de densidade da distribuição PNF não é diferenciável em Z=0 quando $\delta \neq 0$.

Proposição 3.3.2. *Seja $Z \sim PNF(\alpha,\delta)$, se $\delta < 0$ então a variável aleatória Z é bimodal.*

Demostração: Igualando a zero a derivada com respeito a Z da função de densidade de probabilidade com distribuição α-potência normal assimétrica flexível, obtemos:

1. Se $Z < 0$ então, segue a solução

$$z_1 = (\alpha-1)\frac{\phi(z_1)}{\Phi(z_1)} + \delta.$$

2. Se $Z \geq 0$ então, segue a solução

$$z_2 = (\alpha-1)\frac{\phi(z_2)}{\Phi(z_2)} - \delta.$$

Agora, se $\alpha \to 1^-$, $z_1 \in \mathbb{R}^-$ e se $\alpha \to 1^+$, $z_2 \in \mathbb{R}^+$. Logo, z_1 e z_2 são pontos modais distintos. Portanto, Z é uma variável aleatória bimodal.

Corolário 3.3.3. *Seja $Z \sim PNF(\alpha,\delta)$. Se $\alpha = 1$ e $\delta < 0$ então a variável aleatória Z é simétrica bimodal.*

Demostração: Se $\alpha = 1$ então $\varphi(z;\alpha,\delta) = k_{\alpha,\delta}\phi(|z|+\delta)$ e como ϕ é uma funcão simétrica, segue-se que Z é uma variável aleatória simétrica bimodal.

A Figura 3.1 mostra a forma da função de densidades PNF para valores α de 0.75, 1.0, 1.25, 1.75 e para δ de -0.75 e 0.75. Neste gráfico pode-se observar que para $\delta = -0.75$, Figura (a), a densidade é bimodal e com diferentes graus de assimetria em cada caso. Para a Figura (b) observa-se que se $\delta = 0.75 > 0$, as densidades somente apresentam distintos grau de assimetria de acordo ao valor do parâmetro α.

(a) (b)

Figura 3.1 *Densidade $\varphi(z;\alpha,\delta)$ para $\alpha = 0.75$ (linha de pontos e tracejada), 1 (linha de pontos), 1.25 (linha tracejada), 1.75 (linha contínua). (a) $\delta = -0.75$ e (b) $\delta = 0.75$.*

3.4 Momentos da Distribuição PNF

Definição 3.4.1. *Seja $Z \sim PNF(\alpha,\delta)$. Então, para $r = 0,1,2,3,\ldots,n$. define-se os r-ésimos momentos incompletos de Z por:*

$$\mu_r^-(Z) = \int_{-\infty}^{z} y^r \varphi(y;\alpha,\delta)dy, \quad e \quad \mu_r^+(Z) = \int_{z}^{\infty} y^r \varphi(y;\alpha,\delta)dy$$

e o r-ésimo momento de Z por:

$$\mu_r = E(Z^r) = \mu_r^-(Z) + \mu_r^+(Z).$$

O próximo resultado apresenta uma expressão para o cálculo dos momentos.

Proposição 3.4.2. *Seja $Z \sim PNF(\alpha,\delta)$, o r-ésimo momento da variável aleatória Z, $\mu_r = E(Z^r)$ pode-se escrever como:*

$$\begin{aligned}\mu_r = (r-1)\mu_{r-2}(0) + \delta k_{\alpha,\delta}\left[\mu_{r-1}^-(0) - \mu_{r-1}^+(0)\right] + \\ + (\alpha-1)k_{\alpha,\delta}\int_0^1 \left[\Phi^{-1}(v)\right]^{r-1} \phi\left[\left|\Phi^{-1}(v)\right| + \delta\right] v^{\alpha-2}dv. \end{aligned} \tag{3.6}$$

Os momentos centrais, $\acute{\mu}_r = E(Z - E(Z))^r$, para $r = 2,3,4$ podem ser calculados das expressões:

$\acute{\mu}_2 = \mu_2 - \mu_1^2, \quad \acute{\mu}_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3 \quad$ e $\quad \acute{\mu}_4 = \mu_4 - 4\mu_3\mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4.$

Consequentemente, a variância e os coeficientes de assimetria e curtose são dados por: $\sigma^2 = Var(Z) = \acute{\mu}_2$, $\sqrt{\beta_1} = \acute{\mu}_3/[\acute{\mu}_2]^{3/2}$ e $\beta_2 = \acute{\mu}_4/[\acute{\mu}_2]^2$.

3.4.1 Caso Localização - Escala

Considera-se agora a extensão ao caso de localização-escala para a distibuição PNF.

Definição 3.4.3. *Seja $Z \sim PNF(\alpha, \delta)$, com $\alpha \in \mathbb{R}^+$ e $\delta \in \mathbb{R}$. A família de distribuições PNF, com parâmetro de localização $\xi \in \mathbb{R}$ e parâmetro de escala $\eta \in \mathbb{R}^+$, é definida como a distribuição da variável aleatória $X = \xi + \eta Z$ cuja função de densidade é:*

$$\varphi(x; \xi, \eta, \alpha, \delta) = \frac{k_{\alpha,\delta}}{\eta}\phi\left(\frac{|x-\xi|}{\eta} + \delta\right)\left\{\Phi\left(\frac{x-\xi}{\eta}\right)\right\}^{\alpha-1}, \qquad x \in \mathbb{R}. \tag{3.7}$$

Se X segue o modelo (3.7), usamos a notação $X \sim PNF(\xi, \eta, \alpha, \delta)$.

3.5 Inferência no Modelo PNF

Agora apresentamos os estimadores de máxima verossimilhança, a matriz de informação observada e a matriz de informação esperada para α e δ no caso padrão e do vetor $\theta = (\xi, \eta, \alpha, \delta)'$ no caso de localização-escala.

3.5.1 Caso Padrão

Para uma amostra aleatória de tamanho n, $\mathbf{Z} = (Z_1, \ldots, Z_n)'$, da distribuição $PNF(\alpha, \delta)$, a função de log-verossimilhança para $\boldsymbol{\theta} = (\alpha, \delta)'$ dado $\mathbf{Z}$ fica dada por:

$$\ell(\theta; \mathbf{Z}) = n\left\{\log(k_{\alpha,\delta}) - \log(\sqrt{2\pi})\right\} - \frac{1}{2}\sum_{i=1}^{n}(|z_i| + \delta)^2 + (\alpha - 1)\sum_{i=1}^{n}\log\left\{\Phi(z_i)\right\}, \tag{3.8}$$

então os elementos da função escore são

$$U(\alpha) = n\left\{-\sqrt{2\pi}h(\alpha, \delta) + \overline{u}\right\}, \qquad U(\delta) = n\left\{\sqrt{2\pi}g(\alpha, \delta) - \overline{|z|} - \delta\right\}, \tag{3.9}$$

onde,

$$h(\alpha, \delta) = \phi(\delta)k(\alpha, \delta)\int_0^1 z^{\alpha-1}\log z \exp\left\{-\delta\left|\Phi^{-1}(z)\right|\right\}dz, \quad \overline{u} = \sum_{i=1}^{n} u_i/n,$$

$$g(\alpha,\delta)=\phi(\delta)k(\alpha,\delta)\int_0^1\left(\left|\Phi^{-1}(z)\right|+\delta\right)z^{\alpha-1}\exp\left\{-\delta\left|\Phi^{-1}(z)\right|\right\}dz \quad \text{e} \quad \overline{|z|}=\sum_{i=1}^{n}|z_i|/n.$$

Fazendo as derivadas descritas acima iguais zero, obtemos que a solução da função escore satisfaz as equações:

$$\overline{u}=\sqrt{2\pi}h(\alpha,\delta) \quad \text{e} \quad \delta=\sqrt{2\pi}g(\alpha,\delta)-\overline{|z|}.$$

Aqui com um processo numérico iterativo pode-se resolver estas equacões. O nosso próximo passo é obter a matriz de informação de Fisher para (α,δ). Os elementos da matriz de informação observada (j_{xy}) são:

$$j_{\alpha\alpha}=n\left[-\Delta_{1\alpha}(\alpha,\delta)+\Delta_{2\alpha}(\alpha,\delta)\right], \qquad j_{\delta\alpha}=n[-\sqrt{2\pi}\Delta_{1\alpha\delta}(\alpha,\delta)+\Delta_{2\alpha\delta}(\alpha,\delta)],$$

$$j_{\delta\delta}=n\left[-\Delta_{1\delta}(\alpha,\delta)+\Delta_{2\delta}(\alpha,\delta)+1\right],$$

onde

$$\Delta_{1\alpha}(\alpha,\delta)=\left(\frac{\partial\log k(\alpha,\delta)}{\partial\alpha}\right)^2, \quad \Delta_{1\delta}(\alpha,\delta)=\left(\frac{\partial\log k(\alpha,\delta)}{\partial\delta}\right)^2,$$

$$\Delta_{2\alpha}(\alpha,\delta)=k(\alpha,\delta)\frac{\partial^2k_{\alpha,\delta}^{-1}}{\partial\alpha^2}, \quad \Delta_{2\delta}(\alpha,\delta)=k(\alpha,\delta)\frac{\partial^2k_{\alpha,\delta}^{-1}}{\partial\delta^2},$$

$$\Delta_{1\alpha\delta}(\alpha,\delta)=g(\alpha,\delta)\left(\frac{\partial\log(k(\alpha,\delta))}{\partial\alpha}\right), \quad \Delta_{2\alpha\delta}(\alpha,\delta)=k(\alpha,\delta)\frac{\partial^2k_{\alpha,\delta}^{-1}}{\partial\alpha\partial\delta}.$$

Neste caso os elementos da matriz de informação esperada $i_{\alpha\alpha}, i_{\alpha\delta}, i_{\delta\alpha}, i_{\delta\delta}$ são n^{-1} vezes os valores esperados dos correspondentes elementos da matriz de informação observada.

3.5.2 Caso Localização-Escala

Para uma amostra aleatória de tamanho n, $\mathbf{X}=(X_1,\ldots,X_n)$, da distribuição $PNF(\xi,\eta,\alpha,\delta)$, a função de log-verossimilhança para $\boldsymbol{\theta}=(\xi,\eta,\alpha,\delta)'$ dado $\mathbf{X}$ fica dada por:

$$\ell(\boldsymbol{\theta};\mathbf{X})=n\{\log(k_{\alpha,\delta})-\log(\eta)-\log(\sqrt{2\pi})\}-\frac{1}{2}\sum_{i=1}^{n}z_i^2-\frac{n}{2}\delta^2-\delta\sum_{i=1}^{n}|z_i|+$$
$$+(\alpha-1)\sum_{i=1}^{n}\log\left\{\Phi\left(z_i\right)\right\}, \tag{3.10}$$

onde $z_i=\frac{x_i-\xi}{\eta}$ e $k_{\alpha,\delta}=\left[\sqrt{2\pi}\phi(\delta)\int_0^1 z^{\alpha-1}\exp\{-\delta\left|\Phi^{-1}(z)\right|\}dz\right]^{-1}$.

Então, os elementos da função escore ficam dados por

$$U(\xi) = \frac{1}{\eta}\sum_{i=1}^{n} z_i + \frac{\delta}{\eta}\sum_{i=1}^{n} sinal(z_i) - \frac{\alpha-1}{\eta}\sum_{i=1}^{n} w_i,$$

$$U(\eta) = -\frac{n}{\eta} + \frac{1}{\eta}\sum_{i=1}^{n} z_i^2 + \frac{\delta}{\eta}\sum_{i=1}^{n} z_i sinal(z_i) - \frac{\alpha-1}{\eta}\sum_{i=1}^{n} z_i w_i,$$

$$U(\alpha) = -\sqrt{2\pi} n h(\alpha,\delta) + \sum_{i=1}^{n} \log\{\Phi(z_i)\} = n\left\{-\sqrt{2\pi}h(\alpha,\delta) + \overline{u}\right\}$$

e

$$U(\delta) = \sqrt{2\pi} n g(\alpha,\delta) - \sum_{i=1}^{n} |z_i| - n\delta = n\left\{\sqrt{2\pi}g(\alpha,\delta) - \overline{|z|} - \delta\right\},$$

onde $u_i = \log\{\Phi(z_i)\}$, $w_i = \frac{\phi(z_i)}{\Phi(z_i)}$ e $\overline{u} = \sum_{i=1}^{n} u_i/n$, sinal é a função definida por: $sinal(z) = -1$ se $z < 0$, $sinal(z) = 0$ se $z = 0$ e $sinal(z) = 1$ se $z > 0$. As funções g e h são como foram definidas acima. Segue-se que as soluções das equações dos escores satisfazem:

$$\delta = \sqrt{2\pi}g(\alpha,\delta) - \overline{|z|}, \qquad (\alpha-1)\overline{(zw)} + 1 = \overline{z^2} + \delta\overline{z sinal(z)},$$

$$\overline{u} = \sqrt{2\pi}h(\alpha,\delta), \qquad \overline{z} + \delta\overline{sinal(z)} = (\alpha-1)\overline{w},$$

onde,

$$\overline{w} = \sum_{i=1}^{n} w_i/n, \quad \overline{z} = \sum_{i=1}^{n} z_i/n, \quad \overline{|z|} = \sum_{i=1}^{n} |z_i|/n, \quad \overline{z^2} = \sum_{i=1}^{n} z_i^2/n, \quad \overline{zw} = \sum_{i=1}^{n} z_i w_i/n.$$

3.5.3 Matriz de Informação Esperada

Para uma observação ($z_i = z$) temos que:

$$\ell_i(\boldsymbol{\theta}; z) = \log(k(\alpha,\delta)) - \log(\eta) - \log\sqrt{2\pi}\} - \frac{1}{2}(|z|+\delta)^2 + (\alpha-1)\log\{\Phi(z)\}, \quad (3.11)$$

onde $z = (x-\xi)/\eta$. Então, menos as segundas derivadas da função (3.11) com respeito aos parâmetros são

$$j^i_{\xi\xi} = \frac{1}{\eta^2}\left[1 + (\alpha-1)(zw + w^2)\right], \qquad j^i_{\alpha\xi} = \frac{w}{\eta}, \qquad j^i_{\delta\xi} = -\frac{1}{\eta}sinal(z),$$

$$j^i_{\eta\xi} = \frac{1}{\eta^2}\left\{sinal(z)\left(|z|+\delta\right)+z+(\alpha-1)\left[z^2w+zw^2-w\right]\right\},$$

$$j^i_{\eta\eta} = \frac{1}{\eta^2}\left[-1+2sinal(z)\left(|z|+\delta\right)z+z^2-(\alpha-1)\left(2zw-z^2w^2-z^3w\right)\right],$$

$$j^i_{\alpha\eta} = \frac{zw}{\eta}, \qquad j^i_{\delta\eta} = -\frac{z}{\eta}sinal(z), \qquad j^i_{\alpha\alpha} = -\Delta_{1\alpha}(\alpha,\delta)+\Delta_{2\alpha}(\alpha,\delta),$$

$$j^i_{\delta\alpha} = [-\sqrt{2\pi}\Delta_{1\alpha\delta}(\alpha,\delta)+\Delta_{2\alpha\delta}(\alpha,\delta)], \qquad j^i_{\delta\delta} = -\Delta_{1\delta}(\alpha,\delta)+\Delta_{2\delta}(\alpha,\delta)+1,$$

onde $\Delta_{1\alpha}$, $\Delta_{2\alpha}$, $\Delta_{1\delta}$, $\Delta_{2\delta}$, $\Delta_{1\alpha\delta}$ e $\Delta_{2\alpha\delta}$ são dados como na seção (3.5.1). Logo, dado que $zsinal(z) = |z|$ e $|z|\,sinal(z) = z$, além disso, $E(sinal(Z)) = \mu_0^+(\xi) - \mu_0^-(\xi)$ e $E(|Z|) = \mu_1^+(\xi) - \mu_1^-(\xi)$ então, chamando $i_{\xi\xi}, i_{\xi\eta}, \ldots, i_{\delta\delta}$ os valores esperados da matriz de informação observada e fazendo $a_{kj} = \mathbb{E}\{Z^kW^j\}$ para $k = 0,1,2,3$ e $j = 0,1,2$ se obtêm:

$$i_{\xi\xi} = \frac{1}{\eta^2}\left[1+(\alpha-1)(a_{11}+a_{02})\right], \qquad i_{\alpha\xi} = \frac{a_{01}}{\eta}, \qquad i_{\delta\xi} = -\frac{\mu_0^+(\xi)-\mu_0^-(\xi)}{\eta},$$

$$i_{\eta\xi} = \frac{1}{\eta^2}\left\{\delta(\mu_0^+(\xi)-\mu_0^-(\xi))+2a_{10}+(\alpha-1)\left(a_{21}+a_{12}-a_{01}\right)\right\},$$

$$i_{\eta\eta} = \frac{1}{\eta^2}\left[-1+2\delta(\mu_1^+(\xi)-\mu_1^-(\xi))+3a_{20}-(\alpha-1)\left(2a_{11}-a_{22}-a_{31}\right)\right],$$

$$i_{\alpha\eta} = \frac{a_{11}}{\eta}, \qquad i_{\delta\eta} = -\frac{\mu_1^+(\xi)-\mu_1^-(\xi)}{\eta}, \qquad i_{\alpha\alpha} = -\mathbb{E}(\Delta_{1\alpha}(\alpha,\delta))+\mathbb{E}(\Delta_{2\alpha}(\alpha,\delta)),$$

$$i_{\delta\alpha} = [-\sqrt{2\pi}\mathbb{E}(\Delta_{1\alpha\delta}(\alpha,\delta))+\mathbb{E}(\Delta_{2\alpha\delta}(\alpha,\delta))], \qquad i_{\delta\delta} = -\mathbb{E}(\Delta_{1\delta}(\alpha,\delta))+\mathbb{E}(\Delta_{2\delta}(\alpha,\delta))+1,$$

onde os a_{kj} podem ser calculados numéricamente pois não tem forma fechada.

3.6 Ilustrações

Caso 1. A seguinte aplicação correspondem 70 observações de precipitação em cidades dos EUA. Os dados encontram-se na base zCO2 da biblioteca base/R-ex do pacote R. Uma análise descritiva dos dados se encontra na tabela 3.1.

Os valores dos coeficientes de assimetria e curtose justificam um modelo do tipo PN,

Tabela 3.1 *Estatísticas descritivas para a variável precipitação.*

n	média	variância	$\sqrt{b_1}$	b_2
70	34.8799	185.1023	-0.2846	2.6166

além disso, o gráfico do histograma dado na figura 3.2 também suporta um modelo bimodal. Assim, um teste de hipótese para

$$H_0 : \delta = 0 \quad \text{contra} \quad H_1 : \delta \neq 0,$$

mediante o uso da estatística

$$\Lambda = \frac{\ell_{PN}(\theta)}{\ell_{FPN}(\theta)},$$

leva a

$$-2\log(\Lambda) = 458.9966,$$

o qual é um valor maior que o valor crítico da distribuição qui-quadrado, com um grau de liberdade, ao nível 5%, $\chi^2_{1,5\%} = 3.8414$. Portanto, o modelo PNF bimodal pode ser usado para modelar os dados de precipitação em cidades dos EUA. A tabela 3.2 apresenta as estimativas dos parâmetros, com erros padrão em parenteses, para os dois modelos. Finalmente, um gráfico do modelo estimado, junto com os modelos normal e PNF são apresentados na figura 3.2.

Caso 2. A seguinte aplicação corresponde às 1150 observações do conjunto de dados "roller"estudados anteriormente no modelo PSN.

Para os modelos PN e PNF, temos que este último modelo se ajusta melhor aos dados de altura do rolo (veja tabela 3.3).

3.7 Extensões Bimodal α-Potência

Nesta seção nós estudamos duas novas famílias de distribuições com base na distribuição α-potência. O primeiro modelo corresponde a uma extensão direita do modelo de Durrans(1992), com o qual se logra modelar dados com distribuição simétrica bi-modal,

Tabela 3.2 *Estimativas (erro padrão) dos parâmetros das distribuições PN e PNF bimodal.*

Modelo $PN(\boldsymbol{\theta})$		Modelo $PNF(\boldsymbol{\theta})$	
Parâmetro	**estimativa(Erro padrão)**	**Parâmetro**	**estimativa(Erro padrão)**
Loglik	-281.4556	*Loglik*	-51.2172
ξ	48.8494(9.4578)	ξ	24.1084(1.7386)
η	8.3079(3.8857)	η	9.5079(0.6786)
α	0.2376(0.2964)	α	1.4576(0.1519)
		δ	-1.5255(0.1363)

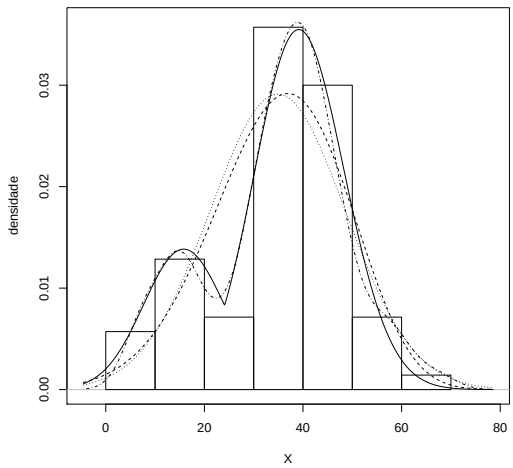


Figura 3.2 *Variável precipitação (linha de pontos e tracejada) e Distribuições:* $N(34.8799, 185.1023)$ *(linha de pontos),* $PN(48.8494, 8.3079, 0.2376)$ *(linha tracejada) e* $PNF(24.1084, 9.5079, 1.4576, -1.5255)$ *(linha contínua).*

enquanto que o segundo modelo contém um novo parâmetro que o faz mais flexível que o primeiro modelo pois se obtém distribuições bimodais assimétricas. Os dois modelos estudados contém a distribuição normal como um caso particular.

Diversas extensões para o modelo skew-normal de Azzalini(1985) foram estudadas para o caso bimodal. Kim (2005) introduz o modelo *two-pieces skew-normal model*, com função

Tabela 3.3 *Estimativas (erro padrão) dos parâmetros das distribuições PN e PNF.*

Modelo PN		Modelo PNF	
Parâmetro	**estimativa(Erro padrão)**	**Parâmetro**	**estimativa(Erro padrão)**
Loglik	-1085.241	*Loglik*	763.193
ξ	4.5495(0.0847)	ξ	3.6880(0.0096)
η	0.1982(0.0432)	η	2.0510(0.0883)
α	0.0479(0.0240)	α	0.6284(0.1008)
		δ	2.9219(0.1547)

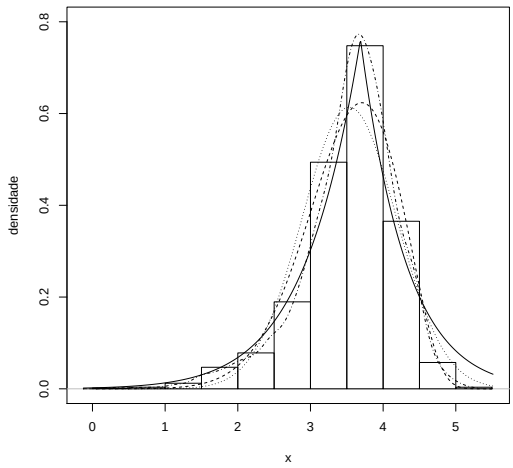


Figura 3.3 *Densidade da variável altura do rolo (linha de pontos e tracejada) e Distribuições:* $N(3.5347, 0.4221)$ *(linha de pontos),* $PN(4.5495, 0.1982, 0.0479)$ *(linha tracejada) e* $PNF(3.6880, 2.0510, 0.6284, 2.9219)$ *(linha contínua).*

de densidade

$$f(z;\lambda) = c_\lambda \phi(z)\Phi(\lambda|z|), \tag{3.12}$$

onde λ é um número real e c_λ é uma constante de normalização. Para $\lambda > 0$, Kim (2005) demostra que o modelo (3.12) é bimodal. Usamos a notação $TN(\lambda)$.

Arnold et al. (2009) estudam um modelo que chamam *the extended two-pieces skew-normal*

model, o qual tem função de densidade de probabilidade dada por

$$f(z;\lambda,\beta) = 2c_\lambda \phi(z)\Phi(\lambda|z|)\Phi(\beta z), \tag{3.13}$$

onde β e λ são números reais e c_λ é uma constante de normalização. Este é um modelo bimodal assimétrico para certos valores de λ e β. Arnold et al. (2009) demostram que a matriz de informação de Fisher do modelo é singular para $\lambda = \beta = 0$, isto é, para o caso da distribuição normal.

3.7.1 Modelo Bimodal Simétrico α-Potência

Com base nos modelos de Durrans (1992) e de Kim (2005), nós definimos a função de densidade de probabilidade para uma nova família de distribuições bimodais, a qual chamamos de família bimodal α-potência. A função de densidade é dada por:

$$\varphi(z;\alpha) = \alpha c_\alpha f(z)\left\{F(|z|)\right\}^{\alpha-1}, \qquad z \in \mathbb{R}, \tag{3.14}$$

onde $\alpha \in \mathbb{R}^+$, F é uma função de distribuição absolutamente contínua com função de densidade de probabilidade $f = dF$ simétrica ao redor de zero e $c_\alpha = \frac{2^{\alpha-1}}{2^\alpha-1}$ é a constante de normalização. Usamos a notação $BSP(\alpha)$.

Resultado 1. Se $Z \sim BSP(\alpha)$, então o modelo (3.14) é simétrico.

Resultado 2. Se $Z \sim BSP(\alpha)$, então $-Z \sim BSP(\alpha)$.

Resultado 3. Se $Z \sim BSP(\alpha)$, então a função de densidade de Z não é diferenciável em Z=0.

Prova: de fato, os dois primeiros resultados são óbvios. Para o resultado 3, temos que $\frac{d}{dz}\varphi(z;\alpha) = -z\phi(z)\Phi(|z|) + (\alpha-1)\frac{z}{|z|}\phi^2(z)$ a qual não existe em z=0, de modo que a densidade (3.14) não é diferenciável em z=0.

Resultado 4. Se $Z \sim BSP(\alpha)$, então a função de distribuição de Z é dada por:

$$\mathcal{F}_F(z;\alpha) = \begin{cases} \frac{2^{\alpha-1}}{2^\alpha-1}\left\{1-\left\{F(-z)\right\}^\alpha\right\}, & \text{se } z<0, \\ \frac{1}{2} + \frac{2^{\alpha-1}}{2^\alpha-1}\left\{\left\{F(z)\right\}^\alpha - \frac{1}{2^\alpha}\right\}, & \text{se } z \geq 0. \end{cases}$$

O método de inversão pode ser usado para gerar amostras com distribuição BSP de parâmetro α. Assim, se $U \sim u(0,1)$, a distribuição da variável $Z \sim BSP(\alpha)$ é

$$Z = \begin{cases} -F^{-1}\left(1 - 2(1-2^{-\alpha})U\right)^{1/\alpha}, & \text{para } U < 1/2, \\ F^{-1}\left(2^{-\alpha}(2(2^{\alpha}-1)(U-2^{-1}))+1\right)^{1/\alpha}, & \text{para } U \geq 1/2, \end{cases}$$

onde F^{-1} é a função inversa de F.

Para uma amostra aleátoria $\mathbf{Z} = (Z_1, Z_2, ..., Z_n)'$ de tamanho n da distribuição BSP, a função de log-verosimilhança de α dada $\mathbf{Z}$ é dada por

$$\ell(\alpha, \mathbf{Z}) = n\log(\alpha) + n\log(c_\alpha) + \sum_{i=1}^{n} \log(f(z_i)) + (\alpha - 1)\sum_{i=1}^{n} \log(F(|z_i|)).$$

Assim, a função escore é dada por:

$$U(\alpha) = \frac{n}{\alpha} - \frac{\log 2}{1-2^{-\alpha}} + \sum_{i=1}^{n} 2F(|z_i|),$$

de modo que o estimador de máxima verossimilhança é a solução da equação não linear $U(\alpha) = 0$. Além disso, para $n \to \infty$,

$$\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{D} N(0, \theta),$$

onde

$$\theta = \frac{1}{n\left[\alpha^{-2} + 2^{\alpha}(2^{\alpha}-1)^{-2}(\log 2)^2\right]}.$$

Momentos da Distribuição BSP

Os momentos de uma variável aleatória com distribuição BSP são funções do momento incompleto

$$\mu_r(Z) = \alpha c_\alpha \int_z^{\infty} y^r f(y) \left\{F(|y|)\right\}^{\alpha-1} dy, \qquad r = 0, 1, 2, ... \tag{3.15}$$

Resultado 5.

Se $Z \sim BSP(\alpha)$ então o r-ésimo momento da variável Z é dado por

$$\mathbb{E}(Z^r) = \begin{cases} 0, & \text{se r é impar}, \\ 2\mu_r(0), & \text{se r é par}. \end{cases}$$

3.7.2 Caso Localização-Escala

Seja $Z \sim BSP(\alpha)$, com $\alpha \in \mathbb{R}^+$. A família de distribuições com parâmetros de localização-escala é definida como a distribuição de $X = \xi + \eta Z$ para $\xi \in \mathbb{R}$ e $\eta > 0$. A correspondente função de densidade é dada por

$$\varphi(x;\xi,\eta,\alpha) = \frac{\alpha}{\eta}\frac{2^{\alpha-1}}{2^{\alpha}-1} f\left(\frac{x-\xi}{\eta}\right)\left\{F\left(\left|\frac{x-\xi}{\eta}\right|\right)\right\}^{\alpha-1}, \quad x \in \mathbb{R}, \tag{3.16}$$

onde ξ é o parâmetro de localização e η é o parâmetro de escala. Usamos a notação $BSP(\xi,\eta,\alpha)$, e em particular, $BSP(\alpha) = BSP(0,1,\alpha)$.

Inferência

Para uma amostra aleatória de tamanho n, $\mathbf{X} = (X_1, X_2, ..., X_n)'$, de uma distribuição $BSP(\xi,\eta,\alpha)$, a função de log-verosimilhança de $\boldsymbol{\theta} = (\xi,\eta,\alpha)'$ dada $\mathbf{X}$ é dada por:

$$\ell(\boldsymbol{\theta};\mathbf{X}) = n\log(\alpha) + n\log(c_\alpha) - n\log(\eta) + \sum_{i=1}^{n}\log(f(z_i)) + (\alpha-1)\sum_{i=1}^{n}\log(F(|z_i|)),$$

onde $z_i = \frac{x_i-\xi}{\eta}$. Assim, sob a suposição de que a derivada de f (f') existe, a função escore é dada por:

$$U(\xi) = -\frac{1}{\eta}\sum_{i=1}^{n}\frac{f'(z_i)}{f(z_i)} + \frac{\alpha-1}{\eta}\sum_{i=1}^{n} sinal(z_i)\frac{f(|z_i|)}{F(|z_i|)},$$

$$U(\eta) = -\frac{n}{\eta} - \frac{1}{\eta}\sum_{i=1}^{n} z_i\frac{f'(z_i)}{f(z_i)} - \frac{\alpha-1}{\eta}\sum_{i=1}^{n}|z_i|\frac{f(|z_i|)}{F(|z_i|)},$$

$$U(\alpha) = \frac{n}{\alpha} - \frac{\log 2}{1-2^{-\alpha}} + \sum_{i=1}^{n}\log[2F(|z_i|)].$$

Os estimadores de máxima verosimilhança são obtidos por métodos numéricos iterativos.

3.7.3 Matriz de Informação Observada

Assumindo que a segunda derivada de f (f'') existe e fazendo $w_i = \frac{f(|z_i|)}{F(|z_i|)}$, $s_i = \frac{f'(|z_i|)}{F(|z_i|)}$, $t_i = \frac{f''(z_i)}{f(z_i)}$ e $v_i = \frac{f'(z_i)}{f(z_i)}$, obtemos que os elementos da matriz de informação observada, que denotamos por $j_{\xi\xi}, j_{\eta\xi}, ..., j_{\alpha\alpha}$, são dados por:

$$j_{\xi\xi} = \frac{n}{\eta^2}\left\{(\overline{v^2} - \overline{t}) + (\alpha-1)\left[\overline{w^2} - \overline{s}\right]\right\},$$

$$j_{\eta\xi} = -\frac{n}{\eta^2}(\overline{v} + \overline{t} - \overline{v^2}) - n\frac{\alpha-1}{\eta^2}\left[\overline{sinal(z)|z|w^2} - \overline{sinal(z)|z|s} - \overline{sinal(z)w}\right],$$

$$j_{\eta\eta} = -\frac{n}{\eta^2} - \frac{n}{\eta^2}\left[2\overline{zv} + \overline{z^2t} - \overline{z^2v^2}\right] - n\frac{\alpha-1}{\eta^2}\left[2\overline{|z|w} + \overline{z^2s} - \overline{z^2w^2}\right],$$

$$j_{\alpha\xi} = -\frac{n}{\eta}\overline{sinal(z)w}, \qquad j_{\alpha\eta} = \frac{n}{\eta}\overline{|z|w}, \qquad j_{\alpha\alpha} = n\left[\alpha^{-2} + 2^{\alpha}(2^{\alpha}-1)^{-2}(\log 2)^2\right],$$

onde $\overline{t} = \frac{1}{n}\sum_{i=1}^{n} t_i$, $\overline{v^2} = \frac{1}{n}\sum_{i=1}^{n} v_i^2$,...,$\overline{z^2w^2} = \frac{1}{n}\sum_{i=1}^{n} z_i^2 w_i^2$.

Matriz de Informação Esperada

Considerando:

$$a_{kj} = \mathbb{E}\{Z^k(f(|Z|)/F(|Z|))^j\}, \qquad a^*_{kj} = \mathbb{E}\{|Z|^k(f(|Z|)/F(|Z|))^j\},$$

$$a^{**}_{kj} = \mathbb{E}\{sinal(Z)|Z|^k(f(|Z|)/F(|Z|))^j\}, \qquad a^{***}_{j} = \mathbb{E}\{sinal(Z)(f(|Z|)/F(|Z|))^j\},$$

$$b_{kj} = \mathbb{E}\{Z^k(f'(|Z|)/F(|Z|))^j\}, \qquad b^*_{kj} = \mathbb{E}\{sinal(Z)|Z|^k(f'(|Z|)/F(|Z|))^j\},$$

$$c_{kj} = \mathbb{E}\{Z^k(f'(Z)/f(Z))^j\}, \qquad d_k = \mathbb{E}\{z^Z(f''(|Z|)/f(|Z|))^j\},$$

para $k = 0, 1, 2, 3$ e $j = 1, 2$, temos que os elementos da matriz de informação esperada são dados por:

$$i_{\xi\xi} = \frac{1}{\eta^2}\left[(c_{02} - d_0) + (\alpha-1)(a_{02} - b_0)\right], \qquad i_{\eta\xi} = -\frac{1}{\eta^2}\left[c_{01} + d_1 - c_{02} + \frac{\alpha-1}{\eta^2}\left[a^{**}_{01} + b^*_1 - a^*_{12}\right]\right],$$

$$i_{\eta\eta} = -\frac{1}{\eta^2} - \frac{1}{\eta^2}\left[2c_{11} + d_2 - c_{22}\right] - \frac{\alpha-1}{\eta^2}\left[2a^*_{11} + b_2 - a_{22}\right],$$

$$i_{\alpha\xi} = -\frac{1}{\eta}a^{***}_1, \qquad i_{\alpha\eta} = \frac{1}{\eta}a^*_{11}, \qquad i_{\alpha\alpha} = \alpha^{-2} + 2^{\alpha}(2^{\alpha}-1)^{-2}(\log 2)^2.$$

Em geral as esperanças nas expressões anteriores se calculam usando métodos de integração numérica. Quando $\alpha = 1$, então $\varphi(x;\xi,\eta,1) = \frac{1}{\eta}f(x)$, a função de densidade do modelo de localização-escala, assim, a matriz de informação se reduz a

$$I(\theta) = \begin{pmatrix} (c_{02} - d_0)/\eta^2 & (c_{01} + d_1 - c_{02})/\eta^2 & a^{***}_1/\eta \\ (c_{01} + d_1 - c_{02})/\eta^2 & -(1 + 2c_{11} + d_2 - c_{22})/\eta^2 & a^*_{11}/\eta \\ a^{***}_1/\eta & a^*_{11}/\eta & 1 + 2(\log 2)^2 \end{pmatrix}$$

onde as características desta matriz dependem da função f. Esta matriz de informação é importante no processo de maximização numérica da função de log-verosimilhança pelos métodos de Fisher e Fisher escore. Também esta matriz será importante para a obtenção da distribuição assintótica (normalidade) do estimador de máxima verosimilhança do vetor de parâmetros do modelo.

3.8 Modelo Bimodal PN

Para $F = \Phi$, a função de distribuição da normal padrão, em (3.14) obtemos a função de densidade de probabilidade

$$\varphi(z;\alpha) = \alpha c_\alpha \phi(z) \left\{\Phi(|z|)\right\}^{\alpha-1}, \qquad z \in \mathbb{R}, \tag{3.17}$$

onde $\alpha \in \mathbb{R}^+$, a qual chamamos de função de densidade bimodal potência normal e a denotamos por $BPN(\alpha)$. Para $\alpha = 1$ temos $c_\alpha = 1$ e assim, chegamos ao caso da distribuição normal.

Resultado 6. Se $Z \sim BPN(\alpha)$ então sua função de densidade é unimodal para $\alpha \leq 1$ e bimodal para $\alpha > 1$.

Prova:

De fato, derivando $\varphi(z;\alpha)$ com respeito z e igualando a zero, obtemos que os pontos onde acontecem os máximos e mínimos são as soluções das equações

$$\alpha c_\alpha z \phi(z) \left\{\Phi(|z|)\right\}^{\alpha-1} = 0 \quad \text{ou} \quad \alpha - 1 = \frac{|z|\Phi(|z|)}{\phi(z)}.$$

Assim, para qualquer $\alpha \in \mathbb{R}^+$ a primeira equação tem uma única solução z=0, enquanto que a segunda equação não tem nenhuma solução quando $\alpha < 1$; tem a solução $z = 0$ para $\alpha = 1$, o caso da normal, e, dado o valor absoluto, têm duas soluções quando $\alpha > 1$. Portanto, concluímos que a densidade (3.17) é bimodal. A Figura 3.5 mostra o comportamento da função de densidade (3.17) para distintos valores de α.

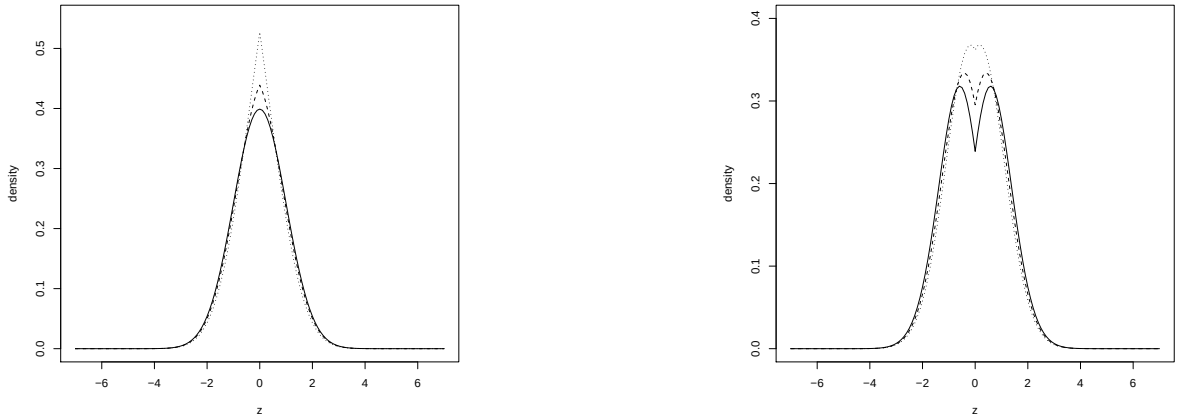
3.8.1 Momentos

De acordo com o caso geral, os momentos de uma variável aleatória com distribuição BPN são funções do momento incompleto

$$\mu_r(Z) = \alpha c_\alpha \int_z^\infty y^r \phi(y) \left\{\Phi(y)\right\}^{\alpha-1} dy, \qquad r = 0, 1, 2, ... \tag{3.18}$$

e o r-ésimo momento da variável Z é dado por

$$\mathbb{E}(Z^r) = \begin{cases} 0, & \text{se r é impar,} \\ 2\mu_r(0), & \text{se r é par.} \end{cases}$$



(a) (b)

Figura 3.4 *Gráfico da distribuição BPN (a) $\alpha = 0.25$ (linha de pontos), 0.75 (linha descontinua) e 1.0 (linha continua) (b) $\alpha = 1.25$ (linha de pontos), 1.75 (linha descontinua) e 2.25 (linha continua).*

Assim, $\mathbb{E}(Z) = \mathbb{E}(Z^3) = 0$. Para $r \geq 2$ temos a seguinte fórmula recursiva para o cálculo dos momentos pares de Z:

$$\mu_r(0) = (r-1)\mu_{r-2}(0) + \alpha(\alpha-1)c_\alpha \int_0^\infty t^{r-1}\phi(\sqrt{2}t)\left\{\Phi(t)\right\}^{\alpha-2}dt.$$

Depois de intensos cálculos algébricos temos

$$\mu_1(0) = \alpha c_\alpha \left[\frac{\alpha-1}{\sqrt{2\pi}}I_1 + \frac{1}{\sqrt{2\pi}}\left(\frac{1}{2}\right)^{\alpha-1}\right],$$

$$\mu_2(0) = \alpha c_\alpha \left[\frac{(\alpha-1)(\alpha-2)}{4\pi}I_2 + \alpha^{-1} - \left(\frac{1}{2}\right)^{\alpha}\frac{\alpha^2-\alpha-4\pi}{4\pi\alpha}\right],$$

$$\begin{aligned}\mu_3(0) = \alpha c_\alpha &\left[\sqrt{\frac{2}{\pi}}\left(\frac{1}{2}\right)^{\alpha-1} + \frac{(\alpha-1)(\alpha-2)}{12\pi\sqrt{\pi}}\left(\frac{1}{2}\right)^{\alpha-3} + \sqrt{\frac{1}{2\pi}}\frac{5(\alpha-1)}{2}I_1\right] \\ &+ \frac{c_\alpha\alpha(\alpha-1)(\alpha-2)(\alpha-3)}{12\pi\sqrt{2\pi}}I_3,\end{aligned}$$

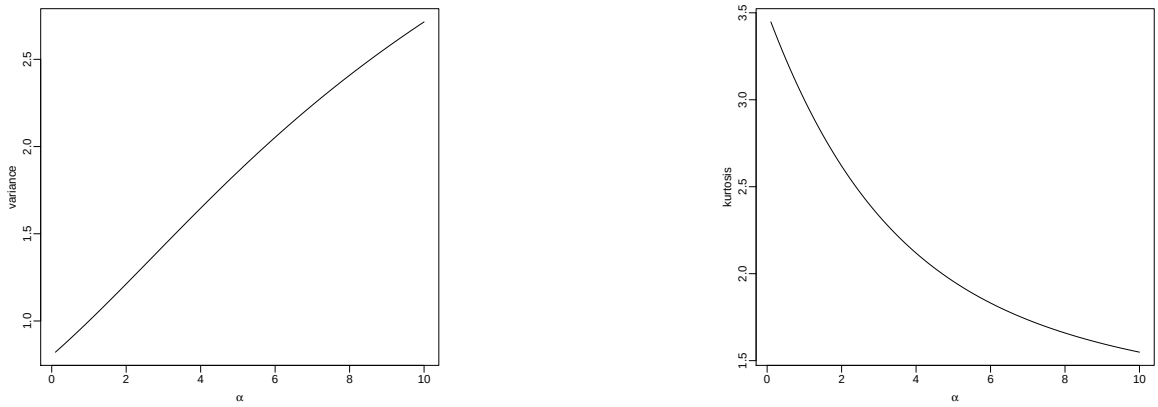
$$\begin{aligned}\mu_4(0) = 3\mu_2(0) &+ \frac{c_\alpha\alpha(\alpha-1)(\alpha-2)}{4\pi}\left[\frac{5}{6}I_2 + \frac{(\alpha-3)(\alpha-4)}{8}I_4\right] \\ &+ \frac{c_\alpha\alpha(\alpha-1)}{4\pi}\left(\frac{1}{2}\right)^{\alpha-4}\left[\left(\frac{1}{2}\right)^4 + \frac{(\alpha-2)(\alpha-3)}{4\pi}\right],\end{aligned}$$

onde
$I_1 = \int_0^\infty \phi(\sqrt{2}z)\,\{\Phi(z)\}^{\alpha-2}dz$, $I_2 = \int_0^\infty \phi(\sqrt{3}z)\,\{\Phi(z)\}^{\alpha-3}dz$, $I_3 = \int_0^\infty \phi(\sqrt{2}z)\,\{\Phi(z)\}^{\alpha-4}dz$
e $I_4 = \int_0^\infty \phi(\sqrt{5}z)\,\{\Phi(z)\}^{\alpha-5}dz$.

Portanto, a variância, o coeficiente de assimetria e o coeficiente de curtose são dados, respectivamente, por:

$$\sigma^2 = Var(Z) = 2\mu_2(0), \quad \sqrt{\beta_1} = 0 \;\; e \;\; \beta_2 = \frac{\mu_4(0)}{2[\mu_2(0)]^2}.$$

Para $\alpha = 1$, temos $\mu_2(0) = 1/2$, $\sigma^2 = Var(Z) = 1$ e $\beta_2 = 3$, que coincidem com o caso da distribuição normal padrão. A Figura seguinte mostra o comportamento da variância e o coeficiente de curtose da distribuição BPN para valores de α entre 0.1 e 10. Como podemos ver, a variância é uma função crescente do parâmetro de forma α enquanto que a curtose é função decrescente deste parâmetro.



(a) (b)

Figura 3.5 *Comportamento da: (a) variância e (b) curtose, da distribuição $BPN(\alpha)$.*

3.8.2 Caso Localização-Escala

Seja $Z \sim BPN(\alpha)$, com $\alpha \in \mathbb{R}^+$. A família de distribuições com parâmetros de localização-escala é definida como a distribuição de $X = \xi + \eta Z$ para $\xi \in \mathbb{R}$ e $\eta > 0$. A correspondente

função de densidade é dada por

$$\varphi(x;\xi,\eta,\alpha) = \frac{\alpha}{\eta}\frac{2^{\alpha-1}}{2^{\alpha}-1}\phi\left(\frac{x-\xi}{\eta}\right)\left\{\Phi\left(\left|\frac{x-\xi}{\eta}\right|\right)\right\}^{\alpha-1}, \quad x \in \mathbb{R}, \tag{3.19}$$

onde ξ é o parâmetro de localização e η é o parâmetro de escala. Usamos a notação $BPN(\xi,\eta,\alpha)$. Novamente segue que $BPN(\alpha) = BPN(0,1,\alpha)$.

3.8.3 Inferência

Assim como no caso geral, encontramos agora os estimadores de máxima verosimilhança para os parâmetros do modelo BPN. Para uma amostra aleatória $\mathbf{X} = (X_1, X_2, ..., X_n)'$ da distribuição $BPN(\xi,\eta,\alpha)$, a função de log-verosimilhança de $\boldsymbol{\theta} = (\xi,\eta,\alpha)'$ dado $\mathbf{X}$ é dada por:

$$\ell(\boldsymbol{\theta};\mathbf{X}) = n\log(\alpha) + n\log(c_\alpha) - n\log(\eta) + \sum_{i=1}^{n}\log(\phi(z_i)) + (\alpha-1)\sum_{i=1}^{n}\log(\Phi(|z_i|)),$$

onde $z_i = \frac{x_i-\xi}{\eta}$. Assim, a função escore é dada por:

$$U(\xi) = \frac{1}{\eta}\sum_{i=1}^{n} z_i + \frac{\alpha-1}{\eta}\sum_{i=1}^{n} sinal(z_i)\frac{\phi(|z_i|)}{\Phi(|z_i|)},$$

$$U(\eta) = -\frac{n}{\eta} + \frac{1}{\eta}\sum_{i=1}^{n} z_i^2 - \frac{\alpha-1}{\eta}\sum_{i=1}^{n}|z_i|\frac{\phi(|z_i|)}{\Phi(|z_i|)},$$

$$U(\alpha) = \frac{n}{\alpha} - \frac{\log 2}{1-2^{-\alpha}} + \sum_{i=1}^{n}\log[2\Phi(|z_i|)].$$

Mesmo como no caso geral, os estimadores de máxima verosimilhança são obtidos por métodos numéricos iterativos.

3.8.4 Matriz de Informação Observada

Assim como no caso geral, para $F = \Phi$ temos que os elementos da matriz de informação observada são dados por:

$$j_{\xi\xi} = \frac{n}{\eta^2} + n\frac{\alpha-1}{\eta^2}\left[\overline{w^2} - \overline{sinal(z)zw}\right],$$

$$j_{\eta\xi} = \frac{2n}{\eta^2}\overline{z} + n\frac{\alpha-1}{\eta^2}\left[-\overline{zw^2} - \overline{sinal(z)z^2w} + \overline{sinal(z)w}\right],$$

$$j_{\eta\eta} = -\frac{n}{\eta^2} + \frac{3n}{\eta^2}\overline{z^2} + n\frac{\alpha-1}{\eta^2}\left[-2\overline{|z|w} + \overline{z^2w^2} + \overline{|z|^3w}\right],$$

$$j_{\alpha\xi} = -\frac{n}{\eta}\overline{sinal(z)w}, \qquad j_{\alpha\eta} = \frac{n}{\eta}\overline{|z|w}, \qquad j_{\alpha\alpha} = n\left[\alpha^{-2} + 2^{\alpha}(2^{\alpha}-1)^{-2}(\log 2)^2\right],$$

onde $\overline{w} = \frac{1}{n}\sum_{i=1}^{n} w_i$, $\overline{w^2} = \frac{1}{n}\sum_{i=1}^{n} w_i^2$, $\overline{zw} = \frac{1}{n}\sum_{i=1}^{n} z_iw_i$, $\overline{sign(z)zw} = \frac{1}{n}\sum_{i=1}^{n} sinal(z_i)z_iw_i$,..., $\overline{z^2w^2} = \frac{1}{n}\sum_{i=1}^{n} z_i^2w_i^2$.

Matriz de Informação Esperada

Considerando:

$$a_{kj} = \mathbb{E}\{Z^k(\phi(Z)/\Phi(|Z|))^j\}, \qquad a^*_{kj} = \mathbb{E}\{|Z|^k(\phi(Z)/\Phi(|Z|))^j\}$$

e

$$a^{**}_{kj} = \mathbb{E}\{sinal(Z)Z^k(\phi(Z)/\Phi(|Z|))^j\},$$

temos que os elementos da matriz de informação esperada são dados por:

$$i_{\xi\xi} = \frac{1}{\eta^2}\left[1 + (\alpha-1)(a_{02} - a^{**}_{11})\right], \qquad i_{\eta\xi} = \frac{2}{\eta^2}a_{10} + \frac{\alpha-1}{\eta^2}\left[a^{**}_{01} - a^{**}_{21} - a_{12}\right],$$

$$i_{\eta\eta} = -\frac{1}{\eta^2} + \frac{3}{\eta^2}a_{20} + \frac{\alpha-1}{\eta^2}\left[a^*_{31} + a_{22} - 2a^*_{11}\right],$$

$$i_{\alpha\xi} = -\frac{1}{\eta}a^{**}_1, \qquad i_{\alpha\eta} = \frac{1}{\eta}a^*_{11}, \qquad i_{\alpha\alpha} = \alpha^{-2} + 2^{\alpha}(2^{\alpha}-1)^{-2}(\log 2)^2.$$

Quando $\alpha = 1$, então $\varphi(x;\xi,\eta,1) = \frac{1}{\eta}\phi(z)$, a função de densidade de localização-escala da normal, assim, a matriz de informação esperada se reduz a

$$I(\boldsymbol{\theta}) = \begin{pmatrix} 1/\eta^2 & 0 & a^{**}_{01}/\eta \\ 0 & 2/\eta^2 & a^*_{11}/\eta \\ a^{**}_{01}/\eta & a^*_{11}/\eta & 1 + 2(\log 2)^2 \end{pmatrix}$$

cujo determinante é $|I(\boldsymbol{\theta})| = \frac{1}{\eta^4}[2 + 4(\log 2) - a^{*2}_{11} - 2a^{**2}_{01}] = \frac{2.8088}{\eta^4}$, então concluímos que a matriz de informação de Fisher do modelo é não singular, para o caso especial de uma

distribuição normal. A submatriz superior de tamanho 2x2 é a matriz de informação da distribuição normal. Para n grande temos que,

$$\hat{\boldsymbol{\theta}} \xrightarrow{A} N_3(\boldsymbol{\theta}, I(\boldsymbol{\theta})^{-1}),$$

concluímos que $\hat{\boldsymbol{\theta}}$ é assintoticamente consistente e com distribuição normal onde $I(\boldsymbol{\theta})^{-1}$ é a matriz de covariâncias para grandes amostras.

3.9 Modelo Bimodal Assimétrico

O modelo (3.14) proposto anteriormente representa uma nova alternativa para dados que apresentam bi-modalidade, mas este modelo só pode ser aplicado em situações com bi-modalidade simétrica. Um segundo modelo mais flexível que o modelo (3.14), pois modela dados com bi-modalidade assimétrica é apresentado agora com base no modelo de Arnold et al. (2009), para a distribuição *skew*-normal e que é conhecido como *the extended two-pieces skew-normal model* (ETN). Assim, nós agora introduzimos a versão do modelo de Arnold et al. (2009), para a distribuição α-potência, como a função de densidade dada pela expressão:

$$\varphi(z;\alpha) = 2\alpha c_\alpha f(z)\left\{F(|z|)\right\}^{\alpha-1} G(\beta z), \qquad z \in \mathbb{R}, \tag{3.20}$$

onde $\alpha \in \mathbb{R}^+$, $\beta \in \mathbb{R}$, F é uma função de distribuição absolutamente contínua com função de densidade de probabilidade $f = dF$ simétrica ao redor de zero, G é uma função de distribuição absolutamente contínua e simétrica ao redor de zero e $c_\alpha = \frac{2^{\alpha-1}}{2^\alpha-1}$ é uma constante de normalização. Nós denotamos por $TPf(\beta, \alpha)$.

Evidentemente esta função é uma densidade pois dado que $f_0 = \alpha c_\alpha f(z)\left\{F(|z|)\right\}^{\alpha-1}$ é uma função contínua e simétrica ao redor de zero, então pelo lema (3.1.1) podemos concluir que (3.20) é uma função de densidade de probabilidade.

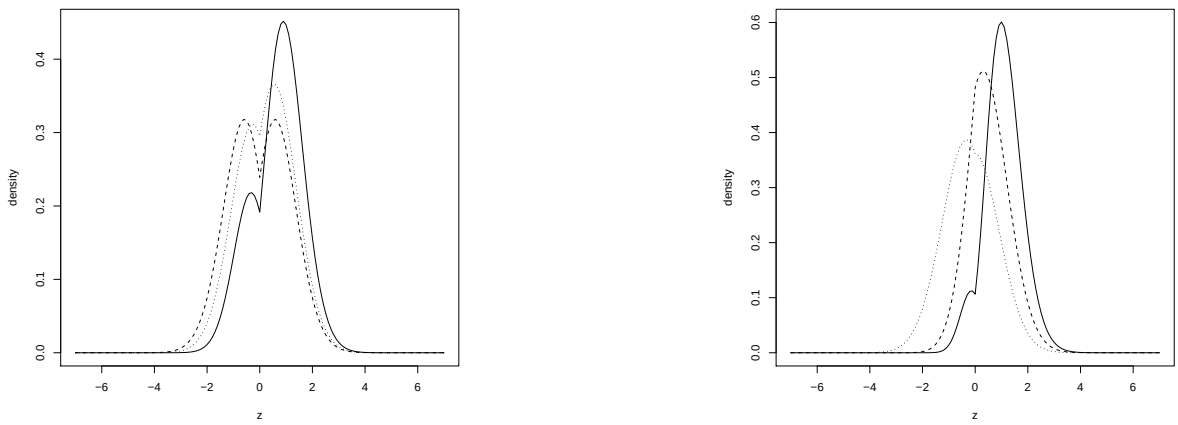
Também é claro que o modelo (3.14) é um caso particular da densidade (3.20) quando $\beta = 0$. Neste trabalho nós vamos dar ênfase no caso $F = G = \Phi$, com $\beta \neq 0$, o qual vamos denominar "'modelo bimodal potência-normal assimétrico"'.

3.9.1 Modelo bimodal Potência-Normal Assimétrico

O modelo bimodal potência-normal assimétrico é um caso especial do modelo (3.20) tomando $F = G = \Phi$. Portanto, sua função de densidade de probabilidade é dada por

$$\varphi(z;\beta,\alpha) = 2\alpha c_\alpha \phi(z)\left\{\Phi(|z|)\right\}^{\alpha-1}\Phi(\beta z), \qquad z \in \mathbb{R}, \tag{3.21}$$

onde $\alpha \in \mathbb{R}^+$ e $\beta \in \mathbb{R}$. Usamos a notação $BPNA(\beta,\alpha)$. A figura (3.6) mostra o comportamento da função de densidade (3.21) para distintos valores de α e β.



(a) (b)

Figura 3.6 *Gráfico da distribuição BPNA (a)* $\alpha = 1.75$ *e* $\beta = 0.25$ *(linha de pontos),* $\alpha = 2.25$ *e* $\beta = 0$, *(linha descontinua)* $\alpha = 2.75$ *e* $\beta = 0.75$ *(linha continua) (b)* $\alpha = 1.25$ *e* $\beta = -0.25$ *(linha de pontos),* $\alpha = 0.5$ *e* $\beta = 1$ *(linha descontinua)* $\alpha = 4$ *e* $\beta = 2$ *(linha continua).*

Derivando a função de densidade com respeito à variável Z, temos que a derivada de $\varphi(z;\beta,\alpha)$ é igual a zero quando

$$2\alpha c_\alpha \phi(z)\left\{\Phi(|z|)\right\}^{\alpha-2} = 0 \quad \text{ou} \quad \alpha - 1 = |z|\frac{\Phi(|z|)}{\phi(z)}\left[1 - \frac{\beta}{z}\frac{\phi(\beta z)}{\Phi(\beta z)}\right].$$

Assim, quando $\alpha > 1$ e para valores de β tais que $\left[1 - \frac{\beta}{z}\frac{\phi(\beta z)}{\Phi(\beta z)}\right] > 0$, o modelo é bimodal. Este modelo contém uma grande variedade de famílias ja conhecidas na literatura dos modelos simétricos e assimétricos uni ou bimodais, como enunciado nos seguintes resultados.

Resultado 7. Se $Z \sim BPNA(\beta, \alpha)$ então: 1) $BPNA(0, \alpha) = BPN(\alpha)$, 2) $BPNA(0, 1) = N(0, 1)$, 3) $BPNA(\beta, 1) = SN(\beta)$ e 4) $BPNA(\beta, 2) = ETN(1, \beta)$.

3.9.2 Momentos

Não existem expressões explícitas para os momentos de uma variável aleatória com distribuição BPNA, assim, nós temos que o r-ésimo momento da variável $Z \sim BPNA(\beta, \alpha)$ se pode expressar como

$$\mathbb{E}(Z^r) = \begin{cases} \mu_r(\alpha), & \text{se r é par,} \\ 2\mu_r(\beta, \alpha) - \mu_r(\alpha), & \text{se r é impar,} \end{cases}$$

onde

$$\mu_r(\alpha) = 2\alpha c_\alpha \int_0^\infty z^r \phi(z) \left\{\Phi(z)\right\}^{\alpha-1} dz \;\text{ e }\; \mu_r(\beta, \alpha) = 2\alpha c_\alpha \int_0^\infty z^r \phi(z) \left\{\Phi(z)\right\}^{\alpha-1} \Phi(\beta z) dz.$$

Assim, se obtém que os momentos de Z são a soma de $2c_\alpha$ vezes os momentos incompletos positivos de uma variável PN e de uma variável BPNA. Mediante uma mudança de variável nós temos que

$$\mu_r(\alpha) = 2\alpha c_\alpha \int_{1/2}^1 [\Phi^{-1}(t)]^r t^{\alpha-1} dt \;\text{ e }\; \mu_r(\beta, \alpha) = 2\alpha c_\alpha \int_{1/2}^1 [\Phi^{-1}(t)]^r \Phi(\beta \Phi^{-1}(t)) t^{\alpha-1} dt.$$

3.9.3 Caso Localização-Escala

Se $Z \sim BPNA(\beta, \alpha)$, então o caso de localização-escala do modelo BPNA é obtido a partir da transformação $X = \xi + \eta Z$, onde $\xi \in \mathbb{R}$ é um parâmetro de localização e $\eta > 0$ é um parâmetro de escala. Finalmente a família de localização-escala da distribuição BPNA é dada pela densidade

$$\varphi(x; \xi, \eta, \beta, \alpha) = \frac{2\alpha c_\alpha}{\eta} \phi\left(\frac{x-\xi}{\eta}\right) \left\{\Phi\left(\left|\frac{x-\xi}{\eta}\right|\right)\right\}^{\alpha-1} \Phi\left(\beta \frac{x-\xi}{\eta}\right), \qquad x \in \mathbb{R}. \quad (3.22)$$

3.9.4 Inferência

Suponha que temos uma amostra aleatória de tamanho n, $\mathbf{X} = (X_1, X_2, ..., X_n)'$ da distribuição $BPNA(\xi, \eta, \beta, \alpha)$, a função de log-verosimilhança de $\boldsymbol{\theta} = (\xi, \eta, \beta, \alpha)'$ dada $\mathbf{X}$

pode ser escrita como

$$\ell(\boldsymbol{\theta};\mathbf{X}) = n\log(\alpha)+n\log(c_\alpha)-n\log(\eta)+\sum_{i=1}^{n}\log(\phi(z_i))+(\alpha-1)\sum_{i=1}^{n}\log(\Phi(|z_i|))+\sum_{i=1}^{n}\log(\Phi(\beta z_i)),$$

onde $z_i = \frac{x_i-\xi}{\eta}$. As correspondentes equações de verosimilhança são dadas por

$$U(\xi) = \frac{1}{\eta}\sum_{i=1}^{n} z_i + \frac{\alpha-1}{\eta}\sum_{i=1}^{n} sinal(z_i)\frac{\phi(|z_i|)}{\Phi(|z_i|)} - \frac{\beta}{\eta}\sum_{i=1}^{n}\frac{\phi(\beta z_i)}{\Phi(\beta z_i)} = 0,$$

$$U(\eta) = -\frac{n}{\eta} + \frac{1}{\eta}\sum_{i=1}^{n} z_i^2 - \frac{\alpha-1}{\eta}\sum_{i=1}^{n}|z_i|\frac{\phi(|z_i|)}{\Phi(|z_i|)} - \frac{\beta}{\eta}\sum_{i=1}^{n} z_i\frac{\phi(\beta z_i)}{\Phi(\beta z_i)} = 0,$$

$$U(\beta) = \sum_{i=1}^{n} z_i\frac{\phi(\beta z_i)}{\Phi(\beta z_i)} = 0 \;\text{ e }\; U(\alpha) = \frac{n}{\alpha} - \frac{\log 2}{1-2^{-\alpha}} + \sum_{i=1}^{n}\log[2\Phi(|z_i|)] = 0.$$

Métodos numéricos iterativos, tipo Newton-Raphson, têm que ser usados para encontrar as soluções do sistema de equações anterior.

3.9.5 Matriz de Informação Observada

Os elementos da matriz de informação observada são dados por:

$$j_{\xi\xi} = \frac{n}{\eta^2} + n\frac{\alpha-1}{\eta^2}\left[\overline{w^2} - \overline{sinal(z)zw}\right] + n\frac{\beta^2}{\eta^2}\left[\beta\overline{zw_1} + \overline{w_1^2}\right],$$

$$j_{\eta\xi} = \frac{2n}{\eta^2}\overline{z} + n\frac{\alpha-1}{\eta^2}\left[-\overline{zw^2} - \overline{sinal(z)z^2w} + \overline{sinal(z)w}\right] + n\frac{\beta}{\eta^2}\left[\beta^2\overline{z^2w_1} + \beta\overline{zw_1^2} - \overline{w_1}\right],$$

$$j_{\beta\xi} = \frac{n}{\eta}\overline{w_1} - n\frac{\beta}{\eta^2}\left[\beta\overline{z^2w_1} + \overline{zw_1^2}\right], \qquad j_{\alpha\xi} = -\frac{n}{\eta}\overline{sign(z)w},$$

$$j_{\eta\eta} = -\frac{n}{\eta^2} + \frac{3n}{\eta^2}\overline{z^2} + n\frac{\alpha-1}{\eta^2}\left[-2\overline{|z|w} + \overline{z^2w^2} + \overline{|z|^3w}\right] - \frac{\beta}{\eta}\overline{zw_1}$$
$$+ n\frac{\beta}{\eta^2}\left[\beta^2\overline{z^3w_1} + \beta\overline{z^2w_1^2} - 2\overline{zw_1}\right],$$

$$j_{\beta\eta} = \frac{n}{\eta}[\overline{zw_1} - \beta^2\overline{z^3w_1} - \beta\overline{z^2w_1^2}], \qquad j_{\alpha\eta} = \frac{n}{\eta}\overline{|z|w}, \qquad j_{\beta\beta} = n[\beta\overline{z^3w} + \overline{z^2w_1^2}],$$

$$j_{\alpha\beta} = 0, \qquad j_{\alpha\alpha} = n\left[\alpha^{-2} + 2^\alpha(2^\alpha-1)^{-2}(\log 2)^2\right],$$

onde

$w_{1i} = \phi(\beta z_i)/\Phi(\beta z_i)$, $\overline{w_1} = \frac{1}{n}\sum_{i=1}^{n} w_{1i}$, $\overline{w_1^2} = \frac{1}{n}\sum_{i=1}^{n} w_{1i}^2$, $\overline{w} = \frac{1}{n}\sum_{i=1}^{n} w_i$, $\overline{w^2} = \frac{1}{n}\sum_{i=1}^{n} w_i^2$, $\overline{zw} = \frac{1}{n}\sum_{i=1}^{n} z_iw_i$, $\overline{sign(z)zw} = \frac{1}{n}\sum_{i=1}^{n} sign(z_i)z_iw_i$,..., $\overline{z^2w^2} = \frac{1}{n}\sum_{i=1}^{n} z_i^2w_i^2$.

Matriz de Informação Esperada

Considerando:

$$a_{kj} = \mathbb{E}\{Z^k(\phi(Z)/\Phi(|Z|))^j\}, \qquad a^*_{kj} = \mathbb{E}\{|Z|^k(\phi(Z)/\Phi(|Z|))^j\},$$

$$a^{**}_{kj} = \mathbb{E}\{sinal(Z)Z^k(\phi(Z)/\Phi(|Z|))^j\}, \qquad a_{1kj} = \mathbb{E}\{Z^k(\phi(\beta Z)/\Phi(\beta Z))^j\},$$

os elementos da matriz de informação esperada podem ser expressados como:

$$i_{\xi\xi} = \frac{1}{\eta^2}\left[1 + (\alpha-1)(a_{02} - a^{**}_{11})\right] + \frac{\beta^2}{\eta^2}[\beta a_{111} + a_{102}], \qquad i_{\beta\xi} = \frac{1}{\eta}a_{101} - \frac{\beta}{\eta^2}[\beta a_{121} + a_{112}],$$

$$i_{\eta\xi} = \frac{2}{\eta^2}a_{10} + \frac{\alpha-1}{\eta^2}\left[a^{**}_{01} - a^{**}_{21} - a_{12}\right] + \frac{\beta}{\eta^2}[\beta^2 a_{121} + \beta a_{112} - a_{101}],$$

$$i_{\alpha\xi} = -\frac{1}{\eta}a^{**}_1, \qquad i_{\beta\eta} = \frac{1}{\eta}a_{111} - \frac{\beta}{\eta^2}[\beta a_{131} + \beta a_{122}], \qquad i_{\alpha\eta} = \frac{1}{\eta}a^*_{11},$$

$$i_{\eta\eta} = -\frac{1}{\eta^2} + \frac{3}{\eta^2}a_{20} + \frac{\alpha-1}{\eta^2}\left[a^*_{31} + a_{22} - 2a^*_{11}\right] + \frac{\beta}{\eta^2}[\beta^2 a_{131} + \beta a_{122} - 2a_{111}],$$

$$i_{\beta\beta} = \beta a_{131} + a_{122}, \qquad i_{\alpha\beta} = 0, \qquad i_{\alpha\alpha} = \alpha^{-2} + 2^\alpha(2^\alpha - 1)^{-2}(\log 2)^2.$$

Novamente, as esperanças nas expressões anteriores são calculadas usando métodos de integração numérica. Quando $\alpha = 1$ e $\beta = 0$ então $\varphi(x;\xi,\eta,0,1) = \frac{1}{\eta}\phi(x)$, a função de densidade de localização-escala da normal e a matriz de informação se reduz a

$$I(\boldsymbol{\theta}) = \begin{pmatrix} 1/\eta^2 & 0 & \sqrt{\frac{2}{\pi}}/\eta & a^{**}_{01}/\eta \\ 0 & 2/\eta^2 & 0 & a^*_{11}/\eta \\ \sqrt{\frac{2}{\pi}}/\eta & 0 & 2/\pi & 0 \\ a^{**}_{01}/\eta & a^*_{11}/\eta & 0 & 1 + 2(\log 2)^2 \end{pmatrix}$$

com $|I(\boldsymbol{\theta})| = -\frac{4a^{**2}_{01}}{\pi\eta^4} = -\frac{0.2990}{\eta^4} \neq 0$. Portanto, a matriz de informação do modelo é não singular, para o caso especial da distribuição normal, isto é, as filas da matriz de informação são linearmente independentes, contrário ao caso da matriz de informação do modelo em Arnold et al. (2009), onde as filas da matriz de informação são dependentes, para o caso especial da distribuição normal. Novamente podemos notar que a submatriz

superior de tamanho 2×2 é a matriz de informação da distribuição normal. Assim, para $n \to \infty$ temos que,

$$\hat{\boldsymbol{\theta}} \xrightarrow{A} N_4(\boldsymbol{\theta}, I(\boldsymbol{\theta})^{-1}),$$

e concluímos que $\hat{\boldsymbol{\theta}}$ é assintoticamente consistente e com distribuição normal onde $I(\boldsymbol{\theta})^{-1}$ é a matriz de covariâncias dada pela inversa de I_F.

3.9.6 Estudo de Simulação

Agora nós apresentamos alguns resultados sobre o parâmetro α, usando o método de máxima verosimilhança, baseados com as amostras aleatórias da distribuição $BPN(\alpha)$ geradas usando o algoritmo descrito a seguir
Geração de variáveis aleatórias $Z \sim BPN(\alpha)$:

1. Seja $U \sim u(0,1)$.

2. Se $U < \frac{1}{2}$ então $Z = -\Phi^{-1}\left(1 - 2(1-2^{-\alpha})U\right)^{1/\alpha}$.

3. Caso contrário, $Z = \Phi^{-1}\left(2^{-\alpha}(2(2^{\alpha}-1)(U-2^{-1}))+1\right)^{1/\alpha}$,

onde Φ^{-1} corresponde à inversa da distribuição acumulada da normal.

Geram-se diferentes amostras aleatórias BPN, para diferentes valores de α. Todas as simulações foram realizadas para cinco diferentes tamanhos amostrais, $n = 100$, 200, 500, 1000 e 1500. Na Tabela 3.5, descreve-se a média e a $\sqrt{EQM}$ das estimativas usando o método de máxima verosimilhança. As estimativas estão baseadas em 1000 repetições. Os resultados mostram que gerando com $\alpha < 1$ as estimativas melhoram para tamanhos amostrais grandes, enquanto que para valores de $\alpha > 1.5$, de acordo com a tabela 3.5, as estimativas foram muito boas mesmo para tamanhos amostrais pequenos. Notamos que quando o tamanho da amostra aumenta, a média atinge ao valor do parâmetro, enquanto que a $\sqrt{EQM}$ diminui, o qual garante a consistência do estimador.

3.10 Ilustração

Para ilustrar a flexibilidade do modelo BPN usamos uma base de 500 dados com a variável $Z = PesoECO$ o qual é o peso ecográfico (antes de nascer) medido em gramo. Os estatísticos descritivos desta variável são: $\overline{Z} = 3210.356$, $S_Z = 834.092$, $\sqrt{b_1} = 0.071$ e $b_2 = 2.068$. Usando a livraría `bbmle` do pacote `R` se obtém as seguintes estimativas para os modelos normal, TN e BPN, respectivamente. Para comparar os modelos normal, TN e BPN, usamos o critério de Informação de Akaike (AIC) (Akaike, 1974). De acordo com este critério (veja Tabela 3.4 e figura 3.7), o modelo BPN apresenta melhor ajuste que os modelos normal e TN à distribuição da variável Peso ecográfico.

Tabela 3.4 *Estimativas (erro padrão) para os parâmetros dos modelo Normal, TN e BPN.*

Parâmetro	Normal	TN	BPN
ξ	3205.486(37.459)	3207.581(26.083)	3208.728(21.138)
η	837.578(26.694)	772.907(25.782)	662.196(21.733)
α			3.731(0.354)
λ		1.773(0.540)	
AIC	8148.282	8109.672	8088.22

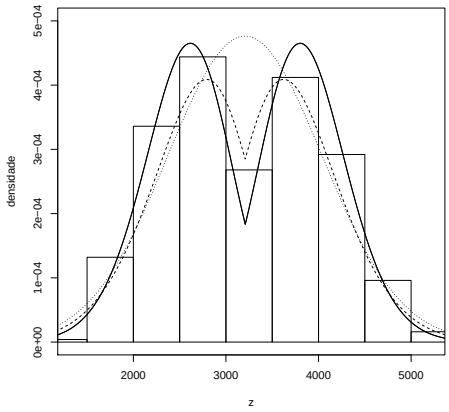


Figura 3.7 *Histograma para a variável Peso ecográfico, (em gramo) antes de nascer. Densidades ajustadas por máxima verosimilhança :* $N(\widehat{\xi}, \widehat{\eta})$ *(linha de pontos),* $TN(\widehat{\xi}, \widehat{\eta}, \widehat{\lambda})$ *(linha tracejada) e* $BSP(\widehat{\xi}, \widehat{\eta}, \widehat{\alpha})$ *(linha contínua).*

Tabela 3.5 *EMV e* $\sqrt{EQM}$ *para o parâmetro* α *no modelo BPN.*

	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 1500$
α	$\widehat{\alpha}(\sqrt{EQM})$	$\widehat{\alpha}(\sqrt{EQM})$	$\widehat{\alpha}(\sqrt{EQM})$	$\widehat{\alpha}(\sqrt{EQM})$	$\widehat{\alpha}(\sqrt{EQM})$
0.2	0.5744(0.3103)	0.4672(0.2165)	0.3764(0.1401)	0.3174(0.0960)	0.3032(0.0738)
0.5	0.8172(0.3294)	0.7496(0.2313)	0.6708(0.1419)	0.6209(0.1001)	0.6016(0.0815)
0.8	1.1061(0.3572)	1.0250(0.2396)	0.9537(0.1421)	0.9106(0.0973)	0.8956(0.0795)
1.0	1.2658(0.3563)	1.2154(0.2457)	1.1398(0.1447)	1.1118(0.1008)	1.0920(0.0833)
1.5	1.5102(0.5166)	1.5052(0.3608)	1.5045(0.2375)	1.5078(0.1582)	1.4977(0.1320)
2.0	2.0307(0.5250)	2.0121(0.3558)	2.0033(0.2337)	1.9968(0.1620)	1.9996(0.1379)
2.5	2.5232(0.5437)	2.5059(0.3786)	2.5016(0.2440)	2.5007(0.1688)	2.5056(0.1349)
3.0	3.0316(0.5507)	2.9885(0.4011)	3.0034(0.2417)	3.0013(0.1706)	2.9978(0.1429)
3.5	3.5175(0.5855)	3.4920(0.3985)	3.5056(0.2558)	3.4978(0.1801)	3.5029(0.1457)
4.0	4.0228(0.6141)	4.0203(0.4156)	4.0025(0.2691)	4.0074(0.1822)	4.0071(0.1523)
4.5	4.5549(0.6290)	4.5172(0.4347)	4.5145(0.2695)	4.4983(0.1921)	4.5006(0.1567)
5.0	5.0526(0.6525)	5.0131(0.4464)	5.0164(0.2984)	4.9991(0.2057)	5.0028(0.1684)
5.5	5.5376(0.6626)	5.5098(0.4831)	5.5022(0.3057)	5.4938(0.2131)	5.5101(0.1716)
6.0	6.0447(0.7334)	6.0015(0.5073)	5.9992(0.3183)	6.0004(0.2243)	5.9950(0.1842)
6.5	6.5461(0.7465)	6.5482(0.5366)	6.4957(0.3395)	6.5137(0.2300)	6.5123(0.1988)
7.0	7.0782(0.7687)	7.0330(0.5788)	6.9889(0.3453)	7.0053(0.2431)	6.9824(0.1990)
7.5	7.5359(0.7971)	7.5323(0.5734)	7.5019(0.3557)	7.5181(0.2585)	7.5150(0.2086)
8.0	8.0555(0.8678)	7.9986(0.5838)	8.0169(0.3592)	8.0106(0.2680)	8.0082(0.2213)
8.5	8.6065(0.9055)	8.4998(0.6101)	8.4918(0.4085)	8.4993(0.2840)	8.5066(0.2360)
9.0	9.0368(0.9414)	9.0030(0.6668)	9.0085(0.4281)	9.0125(0.2927)	9.0129(0.2474)
9.5	9.6064(0.9913)	9.5436(0.6804)	9.5325(0.4283)	9.5089(0.3131)	9.5116(0.2528)
10.0	10.0528(1.0801)	10.0404(0.7128)	10.0184(0.4662)	10.0090(0.3283)	10.0014(0.2562)
10.5	10.5481(1.0780)	10.5654(0.7582)	10.5266(0.4698)	10.5032(0.3487)	10.5111(0.2773)
11.0	11.1058(1.1754)	11.0372(0.7942)	10.9927(0.5025)	11.0129(0.3605)	11.0076(0.2710)

Capítulo 4

Modelo Birnbaum-Saunders α-Potência

4.1 Motivação

Nós agora apresentamos uma generalização da distribuição Birnbaum-Saunders para o caso de distribuições assimétricas. Esta generalização esta baseada na incorporação de um novo parâmetro α, o que torna mais flexível a curtose e a assimetria da distribução. Com esta generalização obtemos densidades que têm caudas mais o menos pesadas que os casos clássicos e por outro lado este novo modelo têm densidades cuja assimetria é mais ou menos pronunciada que a da densidade da BS obtida a partir de uma distribuição normal. Este novo modelo é motivado primeiro porque contém a distribuição BS obtida de um modelo normal, além disso também a BS generalizada estudada por Díaz-García e Leiva-Sánchez e um caso particular da BS elíptica assimétrica de Vilca-Labra e Leiva-Sánchez. Segundo, a partir de um ponto de vista prático, esta nova generalização é baseada na procura de distribuições mais flexível de forma que estas se adaptem melhor a dados de fadiga de vida e também pode ser usada em outras situações de grande interesse prático. Além disso, algumas das propriedades da BS clássica podem ser estendidas ao novo modelo proposto.

A distribuição de Birnbaum-Saunders (BS) foi proposta por Birnbaum e Saunders

(1969), e tem sido bastante utilizada para modelar processos de fadiga, a qual é um dano estrutural que ocorre quando um material é exposto a flutuações de estresse e tensão. Varias generalizações e extensões da distribuição BS foram proposta nos últimos anos.

A distribuição generalizada de Birnbaum-Saunders (GBS) foi proposta por Díaz-García e Leiva-Sanchéz (2005) a partir de distribuições de contornos elípticos e é muito utilizada para modelar distribuições flexíveis de vida com diferentes graus de assimetria e curtose em processos uni e bimodais. O objetivo dessa proposta foi a busca de distribuições de vida com altos graus de assimetria ou que cresçam mais rapidamente e que possuam caudas com menor ou maior curtose do que a distribuição de BS no caso normal. Uma revisão mais extensa da GBS foi feita por Sanhueza, Leiva e Balakrishnan (2008). Recentemente, Vilca-Labra e Leiva-Sánchez (2006), obtiveram uma maior generalização desenvolvendo a distribuição BS mediante o uso de distribuições elípticas assimétricas. Barros, Paula e Leiva-Sánchez (2007) têm alguns desenvolvimentos no caso de modelos de regressão log-Birnbaum-Saunders no caso generalizado. Outra característica muito importante desta distribuição é a robustez na estimação de seus parâmetros, o que foi estudado em Barros, Paula e Leiva-Sánchez (2008). Considera-se agora a extensão do modelo PN ao caso da distribuição BS.

Definição 4.1.1. *Uma variável aleatória T segue a distribuição de Birnbaum-Saunders se T é da forma*

$$T = \frac{\beta}{4}\left[\lambda Z + \sqrt{\lambda^2 Z^2 + 4}\right]^2, \tag{4.1}$$

onde $Z \sim N(0,1)$, $\lambda > 0$ é o parâmetro que controla a forma da distribuição e $\beta > 0$ é o parâmetro de escala e a mediana da distribuição, que denotamos por $T \sim BS(\lambda, \beta)$.

Proposição 4.1.2. *Se $T \sim BS(\lambda, \beta)$, então a função de densidade de probabilidade de T é dada por*

$$f_T(t) = \frac{1}{\sqrt{2\pi}} exp\left[-\frac{1}{2\lambda^2}\left(\frac{t}{\beta} + \frac{\beta}{t} - 2\right)\right]\frac{t^{-3/2}(t+\beta)}{2\lambda\sqrt{\beta}}, \qquad t > 0. \tag{4.2}$$

Esta distribuição torna-se assimétrica quando o parâmetro de forma λ cresce e simétrica,

em torno de β, quando λ está próximo de zero. A função de distribuição cumulativa de $T \sim BS(\lambda, \beta)$, é dada pela expressão

$$F_T(t) = \Phi(a_t),$$

onde

$$a_t = a_t(\lambda, \beta) = \frac{1}{\lambda}\left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right). \tag{4.3}$$

A média e a variância da distribuição são, respectivamente, dadas por

$$\mathbb{E}(T) = \beta\left(1 + \frac{\lambda^2}{2}\right) \quad \text{e} \quad Var(T) = (\lambda\beta)^2\left(1 + \frac{5}{4}\lambda^2\right).$$

Proposição 4.1.3. *Se $T \sim BS(\lambda, \beta)$, então*

$$Z = \lambda^{-1}\left[\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}}\right] \sim N(0, 1). \tag{4.4}$$

Demostração: Vide Johnson et al. (1995).

Seja $\mathbf{T} = (T_1, T_2, ..., T_n)^T$ uma amostra aleatória da distribuição Birnbaum-Saunders. A função de log-verossimilhança é

$$\ell(\lambda, \beta; \mathbf{T}) \propto -n\log(\lambda\beta) + \sum_{i=1}^{n}\log\left[\left(\frac{\beta}{t_i}\right)^{1/2} + \left(\frac{\beta}{t_i}\right)^{3/2}\right] - \frac{1}{2\lambda^2}\sum_{i=1}^{n}\left(\frac{t_i}{\beta} + \frac{\beta}{t_i} - 2\right).$$

Birnbaum-saunders (1969) mostram que o estimador de máxima verossimilhança de β, $\hat{\beta}$, é a raíz positiva da equação $\beta^2 - \beta[2r + k(\beta)] + r[s + k(\beta)] = 0$, onde $s = \frac{1}{n}\sum_{i=1}^{n} t_i$, $r = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{t_i}\right)^{-1}$ e $k(\beta) = \left[\frac{1}{n}\sum_{i=1}^{n}(\beta + t_i)^{-1}\right]^{-1}$. Além disso, $\hat{\lambda}$ pode ser escrita como $\hat{\lambda} = \left(\frac{s}{\hat{\beta}} + \frac{\hat{\beta}}{r} - 2\right)^{1/2}$. A matriz de informação de Fisher é dada por

$$K(\lambda, \beta) = \begin{pmatrix} \frac{2n}{\lambda^2} & 0 \\ 0 & \frac{n[\lambda(2\pi)^{-1/2}h(\lambda)+1]}{\lambda^2\beta^2} \end{pmatrix},$$

onde $h(\lambda) = \lambda\sqrt{\pi/2} - \pi e^{2/\lambda^2}[1 - \Phi(2/\lambda)]$. Assim, λ e β são ortogonais e portanto, os estimadores de máxima verossimilhança são assintoticamente independentes.

Para uma amostra de variáveis aleatórias independentes $T = (T_1, T_2, ..., T_n)'$, com $T_i \sim$

$BS(\lambda, \beta)$, a distribuição da estatística $T_{(n)} = Max[T_1, T_2, ..., T_n]$, onde Max denota a função máximo, é dada por

$$F_{T_{(n)}}(t) = P[T_{(n)} \leq t] = P[T_1 \leq t, T_2 \leq t, ..., T_n \leq t] = \prod_{i=1}^{n} \Phi(a_{t_i}) = \{\Phi(a_t)\}^n .$$

Então, a fdp de $T_{(n)}$ é dada pela expressão

$$\varphi_{T_{(n)}}(t; \lambda, \beta) = n\phi(a_t)\{\Phi(a_t)\}^{n-1} \frac{t^{-3/2}[t+\beta]}{2\lambda\beta^{1/2}}, \tag{4.5}$$

a qual é uma expressão é similar à fdp de uma variável aleatória $PN(n)$. Assim, tomando a ideia de Durrans (1992), nós vamos generalizar a distribuição do máximo no modelo BS para uma estatística de ordem fracionária α e vamos supor que $\alpha \in \mathbb{R}^+$. Portanto, mesmo como acontece no modelo PN, este novo modelo é uma alternativa para aquelas famílias de distribuições, em processos de fadiga, que apresentem um alto grau de assimetria e curtose.

4.2 Distribuição BSPN

Nós agora introduzimos o conceito de uma variável aleatória com distribuição Birnbaum-Saunders α-potência normal assimétrica.

Definição 4.2.1. *Uma variável aleatória T segue a distribuição de Birnbaum-Saunders α-potência normal assimétrica, que denotamos por $T \sim BSPN(\lambda, \beta, \alpha)$, se T é da forma*

$$T = \beta \left[\frac{\lambda}{2} Z + \sqrt{\left(\frac{\lambda}{2} Z\right)^2 + 1} \right]^2 , \tag{4.6}$$

onde $Z = \frac{1}{\lambda}\left(\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}}\right) \sim PN(\alpha)$, com $\lambda,\ \alpha > 0$ parâmetros de forma e $\beta > 0$ é um parâmetro de escala.

4.2.1 Função de Densidade e Propriedades

Seja $T \sim BSPN(\lambda, \beta, \alpha)$. Então, a função de densidade de probabilidade de T é dada por

$$\varphi_T(t; \lambda, \beta, \alpha) = \alpha\phi(a_t)\{\Phi(a_t)\}^{\alpha-1} \frac{t^{-3/2}[t+\beta]}{2\lambda\beta^{1/2}}, \tag{4.7}$$

Assim, para $\alpha = 1$ temos que $T \sim BS(\lambda, \beta)$, a BS padrão sob normalidade, se $\alpha = 2$, então temos o modelo BS normal assimétrico, estudado por Vilca-Labra e Leiva-Sánchez (2006), com parâmetro de assimetria $\lambda = 1$. A Figura 4.23 mostra a densidade da BSPN, onde (a) pode-se observar que quando λ aumenta a curtose é menor e a distribuição apresenta uma assimetria maior e positiva. Mesmo assim, o gráfico (b) mostra que quando α aumenta, a curtose é menor, enquanto que a assimetria da distribuição é sempre positiva e também diminui. Em geral nota-se como o incremento dos parâmetros λ e α afetam a curtose e a assimetria da distribuição.

(a) (b)

Figura 4.1 *Densidade* $\varphi_T(t; \lambda, \beta, \alpha)$, *para* $\beta = 1.0$ *e a)* $\alpha = 1.75$ *e* $\lambda = 0.75$ *(linha de pontos e tracejada), 1.0 (linha de pontos), 1.5 (linha tracejada) e 1.75 (linha contínua). b)* $\lambda = 1.0$ *e* $\alpha = 0.75$ *(linha de pontos e tracejada), 1.0 (linha de pontos), 2.0 (linha tracejada) e 3.5 (linha contínua).*

Seja $T \sim BSPN(\lambda, \beta, \alpha)$. Então, a função de distribuição cumulativa de T é dada por

$$\mathbb{F}_T(t; \lambda, \beta, \alpha) = \{\Phi(a_t)\}^\alpha, \tag{4.8}$$

pois $\varphi_T(t; \lambda, \beta, \alpha) = \frac{d}{dt}\{\Phi(a_t)\}^\alpha = \frac{d(\mathbb{F}_T(t;\lambda,\beta,\alpha))}{dt}$.

De (4.8) e segundo a metodologia de Chang e Tang (1994) tem-se que o p-ésimo percentil da distribuição, $t_p = \mathbb{F}_T^{-1}(p; \lambda, \beta, \alpha)$, é dado por

$$t_p = \beta \left[\frac{\lambda}{2} z_p + \sqrt{\left(\frac{\lambda}{2} z_p\right)^2 + 1} \right]^2, \tag{4.9}$$

onde z_p é o p-ésimo percentil da distribuição PN padrão dado por

$$z_p = \mathbb{F}_Z^{-1}(p;\alpha) = \Phi^{-1}(p^{1/\alpha}). \tag{4.10}$$

O modelo BS é comumente utilizado para explicar o comportamento de dados de sobrevivência. Como definidas no capítulo (2), as funções de sobrevivência, de taxa de risco acumulado, risco e risco invertido do modelo BSPN são, respectivamente:

$$S(t) = 1 - \{\Phi(a_t(\lambda,\beta))\}^{\alpha}, \quad H(t) = -\log[S(t)],$$

$$r(t) = r_{BS}(t)\frac{\alpha\{\Phi(a_t)\}^{\alpha-1} - \{\Phi(a_t)\}^{\alpha}}{1-\{\Phi(a_t)\}^{\alpha}} \quad \text{e} \quad R(t) = \alpha R_{BS}(t),$$

onde $r_{BS}(t)$ e $R_{BS}(t)$ são os índices de risco e risco invertido do modelo BS sob normalidade, ou seja que a taxa de risco inversa de T é proporcional à taxa de risco da distribuição BS. Porém, é importante observar o risco de aplicar o modelo BS tradicional na presença de assimetria ou curtose fora dos limites admitidos pelo modelo BS classica. Assim, os intervalos onde $R(t)$ é decrescente ou não decrescente, são os mesmos intervalos onde $R_{BS}(t)$ é decrescente ou não decrescente.

Propriedades:

- $\lim_{t\to\infty} r(t) = (2\lambda^2\beta)^{-1}$
- Quando $\lambda \to 0$, r(t) tende a ser uma função não decrescente.

Na Figura 4.2 se pode observar que a função de risco é uma função não decrescente de T, mais é uma função não crescente do parâmetro α. Também podemos concluir que $r(t)$ é função não decrescente do parâmetro λ.

Teorema 4.2.2. *Seja $T \sim BSPN(\lambda,\beta,\alpha)$. Então,*

1. $bT \sim BSPN(\lambda, b\beta, \alpha)$ $b > 0$ e

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

(a) (b)

Figura 4.2 *A função de risco $r(t)$, para $\beta = 1.0$ e α =0.75 (linha de pontos e tracejada), 1 (linha de pontos), 2 (linha tracejada) e 5 (linha contínua). (a) $\lambda = 0.25$ e (b) $\lambda = 0.75$.*

Demostração:

1. Se $T \sim BSPN(\lambda, \beta, \alpha)$, então $T = \beta \left[\frac{\lambda}{2}Z + \sqrt{\left(\frac{\lambda}{2}Z\right)^2 + 1}\right]^2$ com $Z \sim PN(\alpha)$ de onde segue que $bT = (b\beta)\left[\frac{\lambda}{2}Z + \sqrt{\left(\frac{\lambda}{2}Z\right)^2 + 1}\right]^2$ com $Z \sim PN(\alpha)$. Portanto, $bT \sim BSPN(\lambda, b\beta, \alpha)$.

2. Se $Z \sim PN(\alpha)$, então $-Z \sim PN(\alpha)$. Agora, para $Z = \frac{1}{\lambda}\left(\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}}\right) \sim PN(\alpha)$, tem-se $-Z = \frac{1}{\lambda}\left(\sqrt{\frac{T^{-1}}{1/\beta}} - \sqrt{\frac{1/\beta}{T^{-1}}}\right) \sim PN(\alpha)$. Logo, $T^{-1} = \frac{1}{\beta}\left[\frac{\lambda}{2}(-Z) + \sqrt{\left(\frac{\lambda}{2}(-Z)\right)^2 + 1}\right]^2$ com $-Z \sim PN(\alpha)$, de onde conclui-se que $T^{-1} \sim BSPN(\lambda, \beta^{-1}, \alpha)$.

Teorema 4.2.3. *Sejam $T \sim BSPN(\lambda, \beta, \alpha)$, $\mathbb{F}_T$ sua função distribuição cumulativa e $\mathbb{F}_Z$ a função distribuição cumulativa de $Z \sim PN(\alpha)$. Então,*
$\mathbb{F}_T(t; \lambda, \beta, \alpha) = \mathbb{F}_Z(a_t(\lambda, \beta); \alpha) = \{\Phi(a_t(\lambda, \beta); \alpha)\}^\alpha$.

Demostração:

Seja $a_x(\lambda, \beta)$, como definida em (4.3). Assim,

$$\begin{aligned}\mathbb{F}_T(t; \lambda, \beta, \alpha) &= \int_0^t \alpha\phi(a_x(\lambda, \beta))\left\{\Phi(a_x(\lambda, \beta))\right\}^{\alpha-1}\frac{d}{dx}a_x(\lambda, \beta)dx \\ &= \int_{-\infty}^{a_t(\lambda,\beta)} \alpha\phi(x)\left\{\Phi(x)\right\}^{\alpha-1}dx = \left\{\Phi(a_t(\lambda, \beta))\right\}^\alpha = \mathbb{F}_Z(a_t(\lambda, \beta); \alpha).\end{aligned}$$

4.3 Momentos da Distribuição $BSPN(\lambda, \beta, \alpha)$

Teorema 4.3.1. *Sejam $T \sim BSPN(\lambda, \beta, \alpha)$ e $Z \sim PN(\alpha)$. Então, $\mathbb{E}(T^n)$ existe se, e somente se,*

$$\mathbb{E}\left[\left(\frac{\lambda Z}{2}\right)^{k+l}\left(\left(\frac{\lambda Z}{2}\right)^2+1\right)^{\frac{k-l}{2}}\right] \tag{4.11}$$

existe para $k = 1, 2, \ldots, n$ com $l = 0, 1, \ldots, k$.

A prova segue de maneira similar à prova apresentada em Vilca-Labra e Leiva-Sánchez, (2006).

Teorema 4.3.2. *Sejam $T \sim BSPN(\lambda, \beta, \alpha)$ e $Z \sim PN(\alpha)$. Se $\mathbb{E}[Z^r]$ existe para $r = 1, 2, \ldots,$ então*

$$\begin{aligned}
\acute{\mu}_r = \mathbb{E}(T^r) &= \beta^r \sum_{k=0}^{r}\binom{r}{k}\left(\frac{1}{2}\right)^k \mathbb{E}\left[\lambda^2 Z^2 + \lambda Z\sqrt{\lambda^2 Z^2+4}\right]^k \\
&= \beta^r \sum_{[0\le k\le r/2]}\binom{r}{2k}\left(\frac{1}{2}\right)^{2k}\sum_{j=0}^{2k}\binom{2k}{j}\mathbb{E}[(\lambda Z)^{4k-j}(\lambda^2 Z^2+4)^{j/2}]+ \\
\beta^r \sum_{[0\le k<r/2]}\binom{r}{2k+1}\left(\frac{1}{2}\right)^{2k+1}&\sum_{j=0}^{2k+1}\binom{2k+1}{j}\mathbb{E}[(\lambda Z)^{4k+2-j}(\lambda^2 Z^2+4)^{j/2}]
\end{aligned} \tag{4.12}$$

onde [.] no índice da soma é a função parte inteira.

Demostração: A prova é imediata do teorema anterior e dado que:

$$\begin{aligned}
\mathbb{E}\left\{\left[\frac{T}{\beta}\right]^r\right\} &= \mathbb{E}\left\{\left[\frac{\lambda}{2}Z+\sqrt{\left(\frac{\lambda}{2}Z\right)^2+1}\right]^2\right\}^r \\
&= \mathbb{E}\left\{\left[1+\left\{\frac{\lambda^2}{2}Z^2+\lambda Z\sqrt{\left(\frac{\lambda}{2}Z\right)^2+1}\right\}\right]^r\right\}
\end{aligned}$$

agora, pelo teorema do binômio $(x+y)^r = \sum_{k=0}^{r}\binom{r}{k}x^{r-k}y^k$ portanto,

$$\begin{aligned}
\mathbb{E}\left\{\left[\frac{T}{\beta}\right]^r\right\} &= \sum_{k=0}^{r}\binom{r}{k}\mathbb{E}\left\{\left[\frac{\lambda^2}{2}Z^2+\lambda Z\sqrt{\left(\frac{\lambda}{2}Z\right)^2+1}\right]^k\right\} \\
&= \sum_{k=0}^{r}\binom{r}{k}\left(\frac{1}{2}\right)^k\mathbb{E}\left[\lambda^2Z^2+\lambda Z\sqrt{\lambda^2Z^2+4}\right]^k, \\
&= \sum_{[0\le k\le r/2]}^{r}\binom{r}{2k}\left(\frac{1}{2}\right)^{2k}\mathbb{E}\left[\lambda^2Z^2+\lambda Z\sqrt{\lambda^2Z^2+4}\right]^{2k} \\
&+ \sum_{[0\le k<r/2]}^{r}\binom{r}{2k+1}\left(\frac{1}{2}\right)^{2k+1}\mathbb{E}\left[\lambda^2Z^2+\lambda Z\sqrt{\lambda^2Z^2+4}\right]^{2k+1}
\end{aligned}$$

e desenvolvendo novamente o binomio dentro da esperança conclui-se a prova do Teorema. Para o calculo dos momentos de ordem $r=1,2,3,4$ da variável aleatória T, assim como de outras estatísticas de interesse, nós vamos usar a notação $\nu_k = \mathbb{E}[Z^k]$ e $\kappa_k = \mathbb{E}\left[Z^k\left(\lambda^2Z^2+4\right)^{1/2}\right]$.

Corolário 4.3.3. *Sejam $T \sim BSPN(\lambda,\beta,\alpha)$ e $Z \sim PN(\alpha)$. Se $\mathbb{E}[Z^r]$ existe para $r = 1,2,3,4$, então, $\acute{\mu}_1 = \mathbb{E}(T) = \frac{\beta}{2}[2+\lambda^2\nu_2+\lambda\kappa_1]$; $\acute{\mu}_2 = \mathbb{E}(T^2) = \frac{\beta^2}{2}[2+4\lambda^2\nu_2+\lambda^4\nu_4+2\lambda\kappa_1+\lambda^3\kappa_3]$;*
$\acute{\mu}_3 = \mathbb{E}(T^3) = \frac{\beta^3}{2}[2+9\lambda^2\nu_2+6\lambda^4\nu_4+\lambda^6\nu_6+3\lambda\kappa_1+4\lambda^3\kappa_3+\lambda^5\kappa_5]$ e
$\acute{\mu}_4 = \mathbb{E}(T^4) = \frac{\beta^4}{2}[2+16\lambda^2\nu_2+20\lambda^4\nu_4+8\lambda^6\nu_6+\lambda^8\nu_8+4\lambda\kappa_1+10\lambda^3\kappa_3+6\lambda^5\kappa_5+\lambda^7\kappa_7]$.

Demostração: O resultado segue desenvolvendo (4.12) para $r=1,2,3,4$.

De acordo com os resultados do corolário (4.3.3) os primeiros quatro momentos centrais, $\mu_r = \mathbb{E}[T-\mathbb{E}(T)]^r$, ficam dados por: $\mu_1 = 0$,

$$\mu_2 = Var(T) = \mathbb{E}[T^2]-\mathbb{E}^2[T] \quad = \frac{\lambda^2\beta^2}{4}[4\nu_2-\kappa_1^2+2\lambda\kappa_3-2\lambda\nu_2\kappa_1-\lambda^2\nu_2^2+2\lambda^2\nu_4],$$

$$\begin{aligned}
\mu_3 &= \mathbb{E}[T^3]-3\mathbb{E}[T]\mathbb{E}[T^2]+2\mathbb{E}^3[T] \\
&= \frac{\lambda^3\beta^3}{4}\left[2\kappa_3-6\nu_2\kappa_1+\kappa_1^3\right]+\frac{\lambda^4\beta^3}{4}\left[-6\nu_2^2+6\nu_4+3\nu_2\kappa_1^2-3\kappa_1\kappa_3\right] \\
&+ \frac{\lambda^5\beta^3}{4}\left[3\nu_2^2\kappa_1-3\nu_4\kappa_1-3\nu_2\kappa_3+2\kappa_5\right]+\frac{\lambda^6\beta^3}{4}\left[\nu_2^3-3\nu_2\nu_4+2\nu_6\right] \qquad (4.13)
\end{aligned}$$

e

$$
\begin{aligned}
\mu_4 &= \mathbb{E}[T^4] - 4\mathbb{E}[T]\mathbb{E}[T^3] + 6\mathbb{E}^2[T]\mathbb{E}[T^2] - 3\mathbb{E}^4[T] \\
&= -\frac{\lambda^4\beta^4}{16}[-16\nu_4 + 16\kappa_1\kappa_3 - 24\nu_2\kappa_1^2 + 3\kappa_1^4] \\
&\quad - \frac{\lambda^5\beta^4}{16}[16\nu_2\kappa_3 - 16\kappa_5 - 48\nu_2^2\kappa_1 + 48\nu_4\kappa_1 - 12\kappa_1^2\kappa_3 + 12\nu_2\kappa_1^3] - \\
&\quad - \frac{\lambda^6\beta^4}{16}[-24\nu_2^3 + 48\nu_2\nu_4 - 32\nu_6 - 24\nu_2\kappa_1\kappa_3 + 16\kappa_1\kappa_3 + 18\nu_2^2\kappa_1^2 - 12\nu_4\kappa_1^2] - \\
&\quad - \frac{\lambda^7\beta^4}{16}[-12\nu_2^2\kappa_3 + 16\nu_2\kappa_5 - 8\kappa_7 + 12\nu_2^3\kappa_3 - 24\nu_2\nu_4\kappa_1 + 16\nu_6\kappa_1] - \\
&\quad - \frac{\lambda^8\beta^4}{16}[3\nu_2^4 - 12\nu_2^2\nu_4 + 16\nu_2\nu_6 - 8\nu_8]. \qquad (4.14)
\end{aligned}
$$

Corolário 4.3.4. *Sejam $T \sim BSPN(\lambda,\beta,\alpha)$ e $Z \sim PN(\alpha)$. Se $\mathbb{E}[Z^r]$ existe para $r = 1,2,3,4$ então, a variância e os coeficientes de assimetria e curtose são, respectivamente:*

$$
Var(T) = \frac{\lambda^2\beta^2}{4}[4\nu_2 - \kappa_1^2 + 2\lambda\kappa_3 - 2\lambda\nu_2\kappa_1 - \lambda^2\nu_2^2 + 2\lambda^2\nu_4], \qquad (4.15)
$$

$$
\begin{aligned}
\sqrt{\beta_1(T)} &= \frac{2\left[2\kappa_3 - 6\nu_2\kappa_1 + \kappa_1^3\right] + 2\lambda\left[-6\nu_2^2 + 6\nu_4 + 3\nu_2\kappa_1^2 - 3\kappa_1\kappa_3\right]}{[4\nu_2 - \kappa_1^2 + 2\lambda\kappa_3 - 2\lambda\nu_2\kappa_1 - \lambda^2\nu_2^2 + 2\lambda^2\nu_4]^{3/2}} \\
&\quad + \frac{2\lambda^2\left[3\nu_2^2\kappa_1 - 3\nu_4\kappa_1 - 3\nu_2\kappa_3 + 2\kappa_5\right] + 2\lambda^3\left[\nu_2^3 - 3\nu_2\nu_4 + 2\nu_6\right]}{[4\nu_2 - \kappa_1^2 + 2\lambda\kappa_3 - 2\lambda\nu_2\kappa_1 - \lambda^2\nu_2^2 + 2\lambda^2\nu_4]^{3/2}} \qquad (4.16)
\end{aligned}
$$

e

$$
\begin{aligned}
\beta_2(T) &= -\frac{-16\nu_4 + 16\kappa_1\kappa_3 - 24\nu_2\kappa_1^2 + 3\kappa_1^4}{\left[4\nu_2 - \kappa_1^2 + 2\lambda\kappa_3 - 2\lambda\nu_2\kappa_1 - \lambda^2\nu_2^2 + 2\lambda^2\nu_4\right]^2} \\
&\quad - \lambda\frac{16\nu_2\kappa_3 - 16\kappa_5 - 48\nu_2^2\kappa_1 + 48\nu_4\kappa_1 - 12\kappa_1^2\kappa_3 + 12\nu_2\kappa_1^3}{\left[4\nu_2 - \kappa_1^2 + 2\lambda\kappa_3 - 2\lambda\nu_2\kappa_1 - \lambda^2\nu_2^2 + 2\lambda^2\nu_4\right]^2} \\
&\quad - \lambda^2\frac{-24\nu_2^3 + 48\nu_2\nu_4 - 32\nu_6 - 24\nu_2\kappa_1\kappa_3 + 16\kappa_1\kappa_3 + 18\nu_2^2\kappa_1^2 - 12\nu_4\kappa_1^2}{\left[4\nu_2 - \kappa_1^2 + 2\lambda\kappa_3 - 2\lambda\nu_2\kappa_1 - \lambda^2\nu_2^2 + 2\lambda^2\nu_4\right]^2} \\
&\quad - \lambda^3\frac{-12\nu_2^2\kappa_3 + 16\nu_2\kappa_5 - 8\kappa_7 + 12\nu_2^3\kappa_3 - 24\nu_2\nu_4\kappa_1 + 16\nu_6\kappa_1}{\left[4\nu_2 - \kappa_1^2 + 2\lambda\kappa_3 - 2\lambda\nu_2\kappa_1 - \lambda^2\nu_2^2 + 2\lambda^2\nu_4\right]^2} \\
&\quad - \lambda^4\frac{3\nu_2^4 - 12\nu_2^2\nu_4 + 16\nu_2\nu_6 - 8\nu_8}{\left[4\nu_2 - \kappa_1^2 + 2\lambda\kappa_3 - 2\lambda\nu_2\kappa_1 - \lambda^2\nu_2^2 + 2\lambda^2\nu_4\right]^2}. \qquad (4.17)
\end{aligned}
$$

De fato, dado que $\sqrt{\beta_1(T)} = \mu_3/[\mu_2]^{3/2}$ e $\beta_2(T) = \mu_4/[\mu_2]^2$, então do Corolário (4.3.3) e dos momentos centrais desenvolvidos acima, segue o resultado.

Se $\alpha = 1$ então, $Z \sim N(0,1)$. Assim nós obtemos que $\nu_1 = \nu_3 = \nu_5 = \nu_7 = 0$, $\nu_2 = 1,\ \nu_4 = 3,\ \nu_6 = 15,\ \nu_8 = 105$ e $\kappa_1 = \kappa_3 = \kappa_5 = \kappa_7 = 0$, logo nós obtemos que,

$$\sqrt{\beta_1(T)} = \frac{4\lambda(11\lambda^2 + 6)}{(5\lambda^2 + 4)^{3/2}} \quad \text{e} \quad \beta_2(T) = 3 + \frac{6\lambda^2(93\lambda^2 + 40)}{(5\lambda^2 + 4)^2},$$

as quais coincidem com os resultados obtidos por Ng et al. (2003) e Johnson et al. (1995). Entretanto, se $\alpha = 2$ então $Z \sim SN(0,1,1)$, a normal assimétrica de Azzalini (1985) com parâmetro de assimetria igual a um. Assim, novamente obtemos que $\nu_1 = \nu_3 = \nu_5 = \nu_7 = 0$, $\nu_2 = 1,\ \nu_4 = 3,\ \nu_6 = 15,\ \nu_8 = 105$ enquanto que $\kappa_1,\ \kappa_3,\ \kappa_5$ e κ_7 podem ser obtidas por métodos de integração numérica.

Para $\lambda = 0.5$, $\beta = 1.0$ e α entre 0.01 e 5 se estuda o comportamento de: a média, a variância, o CV, e os coeficientes de assimetria e curtose da variável aleatória $T \sim BSPN(\lambda, \beta, \alpha)$. A média sempre mostrou um comportamento linear, enquanto que a figura 4.3 mostra como a variância e o CV têm um comportamento crescente quando α aumenta. Também pode-se ver que os coeficientes de assimetria e a curtose são funções decrescentes do parâmetros de forma α. Este mesmo comportamento se repete para outros valores de λ (0.25, 0.75, 1.0, 1.5, 2.5 e 5.0).

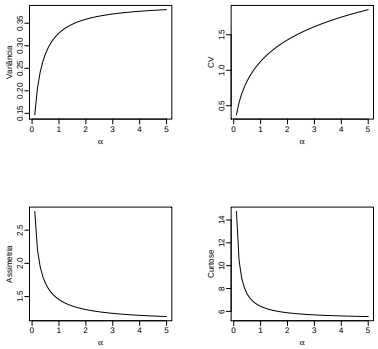


Figura 4.3 *Gráficos da variância, CV e os coeficientes de assimetria e curtose da variável* $T \sim BSPN(\lambda, \beta, \alpha)$ *para* $\lambda = 0.5$, $\beta = 1.0$ *e* α *entre 0.01 e 5.*

4.4 Inferência no Modelo BSPN

4.4.1 Método de Máxima Verossimilhança

Nesta secção apresentamos os estimadores de máxima verossimilhança (EMV), a matriz de informação observada e a matriz de informação esperada do vetor de parâmetros $\boldsymbol{\theta} = (\lambda, \beta, \alpha)'$. Assim, para uma amostra aleatória de tamanho n, $\mathbf{T} = (T_1, \ldots, T_n)'$, da distribuição $BSPN(\lambda, \beta, \alpha)$, a função de log-verossimilhança para $\boldsymbol{\theta} = (\lambda, \beta, \alpha)'$ dado $\mathbf{T}$ pode-se escrever como

$$\begin{aligned}\ell(\boldsymbol{\theta}; \mathbf{T}) &= n\left[\log(\alpha) - \log(\lambda) - \tfrac{1}{2}\log(\beta) - \tfrac{1}{2}\log(2\pi)\right] + \textstyle\sum_{i=1}^n \log(t_i + \beta) - \tfrac{3}{2}\sum_{i=1}^n \log(t_i) \\ &\quad - \tfrac{1}{2\lambda^2}\textstyle\sum_{i=1}^n \left[\tfrac{t_i}{\beta} + \tfrac{\beta}{t_i} - 2\right] + (\alpha - 1)\textstyle\sum_{i=1}^n \log(\Phi(a_{t_i})). \qquad (4.18)\end{aligned}$$

4.4.2 Equações de Verossimilhança

As equações de estimação do método de máxima verossimilhança são obtidas igualando-se a função escore a zero. Para (4.18), temos as seguintes equações

$$U(\lambda) = -n\left[\frac{1}{\lambda} - \frac{1}{n\lambda^3}\sum_{i=1}^n a_{t_i}^2 + \frac{\alpha - 1}{n\lambda}\sum_{i=1}^n a_{t_i} w_i\right] = 0, \qquad U(\alpha) = \frac{n}{\alpha} + \sum_{i=1}^n u_i = 0$$

e

$$\begin{aligned}U(\beta) = -n&\left[\frac{1}{2\beta} - \frac{1}{n}\sum_{i=1}^n \frac{1}{\beta + t_i} + \frac{1}{2n\lambda^2}\sum_{i=1}^n \left[\frac{1}{t_i} - \frac{t_i}{\beta^2}\right] + \frac{\alpha - 1}{2n\beta}\sum_{i=1}^n a_{t_i} w_i\right] \\ &+ \frac{\alpha - 1}{\lambda\beta^{1/2}}\sum_{i=1}^n \frac{w_i}{\sqrt{t_i}} = 0\end{aligned}$$

onde $w_i = \frac{\phi(a_{t_i})}{\Phi(a_{t_i})}$.

As soluções deste sistema de equações não tem forma fechada e devem ser resolvidas por métodos numéricos iterativos.

4.4.3 Matrices de Informação

Matriz de Informação Observada

Os elementos da matriz de informação observada, denotadas por $j_{\lambda\lambda}, j_{\beta\lambda}, j_{\alpha\lambda}, j_{\beta\beta}, \ldots, j_{\alpha\alpha}$, são dados por:

$$j_{\alpha\lambda} = \frac{n}{\lambda}\overline{a_t w}, \qquad j_{\alpha\beta} = n\left[\frac{1}{2\beta}\overline{a_t w} + \frac{1}{\lambda\sqrt{\beta}}\overline{w/\sqrt{t}}\right], \qquad j_{\alpha\alpha} = \frac{n}{\alpha^2},$$

$$j_{\lambda\lambda} = n\left[-\frac{1}{\lambda^2} + \frac{3}{\lambda^4}\left[\bar{t}/\beta + \beta\overline{t^{-1}} - 2\right] + \frac{\alpha-1}{\lambda^2}\left[-2\overline{a_t w} + \overline{a_t^2 w^2} + \overline{a_t^3 w}\right]\right],$$

$$\begin{aligned} j_{\beta\beta} = n\Bigg[&-\frac{1}{2\beta^2} + \overline{(\beta+t)^{-2}} + \frac{1}{\lambda^2\beta^2}\bar{t} + \frac{\alpha-1}{4\beta^2}\left[\overline{a_2^2 w^2} - \overline{a_w} - \left(2\sqrt{\beta}\right)\overline{t^{-1/2}w} + \overline{a_t^3 w}\right]\Bigg] + \\ &+ \frac{n(\alpha-1)}{4\lambda^2\beta^2}\left[4\overline{w^2} - \frac{2\lambda}{\beta}\overline{t^{1/2}w} - 2\beta\overline{t^{-1}w^2} + 4\overline{a_t w}\right] \end{aligned}$$

e

$$\begin{aligned} j_{\beta\lambda} = n\Bigg[&\frac{1}{\lambda^3}\left[\frac{\bar{t}}{\beta^2} - \overline{t^{-1}}\right] - \frac{\alpha-1}{2\lambda\beta}\left[-\overline{a_t w} + \overline{a_t^2 w^2} + \overline{a_t^3 w}\right]\Bigg] + \\ &+ \frac{n(\alpha-1)}{\lambda^3\beta}\left[-\lambda\beta\overline{t^{-1/2}w} + \overline{w^2} - \beta\overline{t^{-1}w^2} + \overline{a_t w} - \beta\overline{t^{-1}w}\right], \end{aligned}$$

onde $\overline{a_t^k w^j} = \sum_{i=1}^n a_{t_i}^k w_i^j/n$, $\overline{(\beta+t)^j} = \sum_{i=1}^n 1/(\beta+t_i)^j/n$, $\overline{t^k w^j} = \sum_{i=1}^n t_i^k w_i^j/n$ e $\overline{w^j} = \sum_{i=1}^n w_i^j/n$.

Matriz de Informação Esperada

Como $T \sim BSPN(\lambda,\beta,\alpha)$ então $\frac{1}{\beta^2}T \sim BSPN(\lambda, 1/\beta, \alpha)$ assim $\frac{1}{\beta^2}T \equiv \frac{1}{T}$ (ou seja, são identicamente distribuidas). Então, $\mathbb{E}[\frac{1}{\beta^2}T - \frac{1}{T}] = 0$. Agora, chamando $a_{kj} = \mathbb{E}[a_T^k W^j]$, $b_0 = \mathbb{E}[\frac{W}{\sqrt{T}}]$, $b_1 = \mathbb{E}[\frac{W}{T}]$, $b_2 = \mathbb{E}[\frac{W^2}{\sqrt{T}}]$, $b_3 = \mathbb{E}[\frac{W^2}{T}]$, $c_0 = \mathbb{E}[\frac{a_T W}{\sqrt{T}}]$, $c_0 = \mathbb{E}[\frac{a_T W}{T}]$, $e_1 = \mathbb{E}[\frac{1}{(\beta+T)^2}]$ e $f_1 = \mathbb{E}[\sqrt{T}W]$, então, os elementos da matriz de informação esperada são dados por:

$$i_{\alpha\alpha} = \frac{1}{\alpha^2}, \qquad i_{\alpha\lambda} = \frac{1}{\lambda}a_{11}, \qquad i_{\alpha\beta} = \frac{1}{2\lambda\beta}a_{11} + \frac{1}{2\lambda\sqrt{\beta}}b_0,$$

$$i_{\lambda\lambda} = -\frac{1}{\lambda^2} + \frac{6}{\lambda^4}\left[\frac{\mu_1}{\beta} - 1\right] + \frac{\alpha-1}{\lambda^2}[a_{22} - 2a_{11} + a_{31}],$$

$$i_{\beta\lambda} = \frac{\alpha-1}{2\lambda\beta}[a_{22} - a_{11} + a_{31}] + \frac{\alpha-1}{\lambda^3\beta^2}[a_{02} - \lambda\beta b_0 - \beta b_3 + a_{11} - \beta a_{21}]$$

e

$$\begin{aligned} i_{\beta\beta} = & -\frac{1}{2\beta^2} + e_1 + \frac{\mu_1}{\lambda^2\beta^2} + \frac{\alpha-1}{4\beta^2}[a_{22} - a_{11} - 2\beta b_0 + a_{31}] + \\ & + \frac{\alpha-1}{4\lambda^2\beta^2}\left[4a_{02} - \frac{2\lambda}{\beta}f_1 - 2\beta b_3 + 4a_{11}\right], \end{aligned}$$

onde as esperanças $a_{kj},\ b_0,\ b_1, ..., f_1$ podem ser calculadas numéricamente. Aqui, para $\alpha = 1$ temos que $a_t \sim N(0,1)$ assim, depois de alguns cálculos numéricos e de acordo com os resultados de Lemonte et al. (2007), temos que

$$I_F(\boldsymbol{\theta}) = \begin{pmatrix} \frac{2}{\lambda^2} & 0 & -\frac{0.5956}{\lambda} \\ 0 & \frac{\sqrt{2\pi}+\lambda p(\lambda)}{\sqrt{2\pi}\lambda^2\beta^2} & -\frac{0.5956}{2\lambda\beta} + \frac{1}{\lambda\sqrt{\beta}}b_0 \\ -\frac{0.5956}{\lambda} & -\frac{0.5956}{2\lambda\beta} + \frac{1}{\lambda\sqrt{\beta}}b_0 & 1 \end{pmatrix},$$

onde $p(\lambda) = \lambda\sqrt{\frac{2}{\pi}} - \frac{\pi\exp(\frac{2}{\lambda^2})}{2}erfc(\frac{2}{\lambda})$, com $erfc(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty \exp(-t^2)dt$.

A matriz superior 2x2 de $I_F(\boldsymbol{\theta})$ é a matriz de informação de Fisher para a distribuição Birnbaum-Saunders sob normalidade, veja Lemonte et al. (2007). As colunas (linhas) da matriz $I_F(\boldsymbol{\theta})$ são linearmente independentes, então a matriz de informação é não singular. Para grandes amostra, temos que o estimador de máxima verossimilhança de $\hat{\boldsymbol{\theta}}$ é assintoticamente consistente e com distribuição normal, isto é,

$$\hat{\boldsymbol{\theta}} \overset{A}{\sim} N_3\left(\boldsymbol{\theta}, I_F^{-1}(\boldsymbol{\theta})\right),$$

resultando que a variância assintótica do EMV de $\hat{\boldsymbol{\theta}}$ é a inversa da $I_F(\hat{\boldsymbol{\theta}})$, a qual denotamos por $\Sigma_{\hat{\boldsymbol{\theta}}} = I_F^{-1}(\boldsymbol{\theta})$.

4.4.4 Estimadores pelo Método dos Momentos Modificados (EMM)

O método dos momentos busca construir uma analogia amostral de uma quantidade que depende do parâmetro a ser estimado e assim a partir das igualdades destas obter uma solução (estimador) para os parâmetros.

Os estimadores obtidos pelo método dos momentos para os parâmetros da BS clássica

nem sempre existem, assim, Ng, Kundu e Balakrishnan 2003, propõem o método dos momentos modificados, o qual consiste em igualar $\mathbb{E}(T)$ e $\mathbb{E}(T^{-1})$ aos seus respectivos momentos amostrais. Barros et al. (2007) seguem essa mesma proposta para a GBS.

Assim, utilizando a expressão do primeiro momento populacional das variáveis T e T^{-1}, $\mathbb{E}(T)$ e $\mathbb{E}(T^{-1})$ respectivamente, e igualando-os aos momentos amostrais $s = \frac{1}{n}\sum_{i=1}^{n} t_i$ e $r^{-1} = \frac{1}{n}\sum_{i=1}^{n} t_i^{-1}$ obtemos os estimadores de momentos modificados para os parâmetros λ e β, dados respectivamente por:

$$\hat{\lambda}_M = \frac{-\kappa_1 + \sqrt{\kappa_1^2 + 8\nu_2\left(\sqrt{\frac{s}{r}} - 1\right)}}{2\nu_2}, \qquad \hat{\beta}_M = \sqrt{sr}, \tag{4.19}$$

enquanto que o EMM para α se pode obter, igualando os segundos momentos amostral e populacional, como a solução da equação:

$$\begin{aligned} s_t^2 = {} & \frac{rs}{32\nu_2^4}(\kappa_1^2 - \kappa_1 D + 4R)\left[8\nu_2(2\nu_2^2(1+R) + \kappa_3(D-\kappa_1) - (3/8)\nu_2\kappa_1^2)\right] \\ & + \frac{rs}{32\nu_2^4}(\kappa_1^2 - \kappa_1 D + 4R)(\kappa_1^2 - 2\kappa_1 D + 8R)(\nu_2 + 2\nu_4), \end{aligned} \tag{4.20}$$

onde s_t^2 é a variância amostral da variável T, $R = \sqrt{s/r} - 1$ e $D = \sqrt{\kappa_1^2 + 8\nu_2 R}$.

Teorema 4.4.1. *Sejam $T \sim BSPN(\lambda, \beta, \alpha)$ e $Z \sim PN(\alpha)$. O estimador de momentos $\hat{\beta}_M$ é consistente para β.*

Demostração: Seja $\tilde{\beta} = \sqrt{SR}$, com S e R variáveis aleatórias, então pela lei forte dos grandes números, com probabilidade um temos que

$$S = \frac{1}{n}\sum_{i=1}^{n} T_i \overset{n\to\infty}{\to} \mathbb{E}(T).$$

Similarmente, temos que

$$\frac{1}{R} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{T_i} \overset{n\to\infty}{\to} \mathbb{E}\left(\frac{1}{T}\right),$$

logo com probabilidade um,

$$(\tilde{\beta}_M)^2 = SR \overset{n\to\infty}{\to} \mathbb{E}(T)\left[\mathbb{E}\left(\frac{1}{T}\right)\right]^{-1} = \beta^2$$

de onde segue o resultado.

Condições sobre as quais os estimadores de momentos do parâmetro θ tem distribuição assintoticamente normal, podem ser estudadas em Sen e Singer (2000).

4.4.5 Outras Propostas de Estimadores

Estimadores Jackknife

Ng et al. (2003) propõem estimadores jackknife para estimar os parâmetros de uma distribuição BS clássica. Esta mesma idéia pode ser usada no caso da BSPN. Como no caso clássico a idéia é remover a observação t_j da amostra aleatória $\mathbf{T} = (T_1, T_2, ..., T_n)'$ estimando os parâmetros baseados nas $n-1$ observações que ficam. Assim, seguindo Ng et al. (2003), temos as seguintes expressões para $\alpha,\ \lambda\ e\ \beta$ no caso da BSPN

$$\alpha_{(j)} = -\frac{1}{\overline{u}_{(j)}}, \quad \lambda_{(j)} = \left\{ \frac{-\overline{u}_{(j)}\left(\frac{s_{(j)}}{\beta_{(j)}} + \frac{\beta_{(j)}}{r_{(j)}} - 2\right)}{\overline{(a_t w_t)}_{(j)} - \overline{u}_{(j)} + \overline{u}_{(j)}\overline{(a_t w_t)}_{(j)}} \right\}^{1/2},$$

$$\beta_{(j)} = \left\{ \frac{s_{(j)}/(2\lambda_{(j)}^2)}{(\alpha_{(j)}-1)\overline{(a_t w_t)}_{(j)}/(2\beta_{(j)}) + (\alpha_{(j)}-1)\overline{(w_t/\sqrt{t})}_{(j)}/(\lambda_{(j)}\sqrt{\beta_{(j)}}) + 1/(2\beta_{(j)}) - 1/k_{(j)}(\beta) + 1/(2r_{(j)}\lambda_{(j)})} \right\}^{1/2}$$

com $\overline{u}_{(j)} = \frac{n\overline{u}-u_j}{n-1}$, $\overline{(a_t w_t)}_{(j)} = \frac{n\overline{a_t w_t} - a_{t_j} w_{t_j}}{n-1}$, $s_{(j)} = \frac{ns-t_j}{n-1}$, $r_{(j)} = \left[\frac{nr^{-1}-t_j^{-1}}{n-1}\right]^{-1}$, $k_{(j)}(\beta) = \left\{\frac{nk^{-1}(\beta)-(\beta+t_j)^{-1}}{n-1}\right\}$, $k(\beta) = \left[\frac{1}{n}\sum(\beta+t_i)^{-1}\right]^{-1}$ e $\overline{(w_t/\sqrt{t})}_{(j)} = \frac{n\overline{wt/\sqrt{t}} - w_{t_j}/t_j}{n-1}$.

Assim os estimadores Jackknife são dados por:

$$\alpha_{JK} = \frac{1}{n}\sum_{i=1}^{n}\alpha_{(j)}, \quad \lambda_{JK} = \frac{1}{n}\sum_{i=1}^{n}\lambda_{(j)} \ \text{ e } \ \beta_{JK} = \frac{1}{n}\sum_{i=1}^{n}\beta_{(j)}.$$

Estimação por Máxima Verossimilhança Perfilada

O objetivo agora é encontrar os estimadores do parâmetro $\boldsymbol{\theta} = (\alpha,\ \lambda,\ \beta)' = (\boldsymbol{\tau}', \phi')'$ por máxima verossimilhança perfilada e máxima verossimilhança perfilada modificada. Então de acordo à teoria estudada no capitulo dois, temos os seguintes resultados para o modelo BSPN. Fixados λ e β, a estimação de máxima verossimilhança para o parâmetro α resulta no estimador

$$\hat{\alpha} = -\frac{n}{\sum_{i=1}^{n}\log\{\Phi(a_{t_i})\}},$$

então, sendo $\boldsymbol{\tau} = (\lambda, \beta)'$ o parâmetro de interesse e $\phi = \alpha$ o parâmetro de incômodo, a função de log verossimilhança perfilada pode ser escrita como (excetuando uma constante),

$$\ell_p(\lambda,\beta) = -n\log\left[-\sum_{i=1}^{n}\log\{\Phi(a_{t_i})\}\right] - \sum_{i=1}^{n}\log\{\Phi(a_{t_i})\} - n\log(\lambda\beta)$$
$$+\sum_{i=1}^{n}\log\left[\left(\frac{\beta}{t_i}\right)^{1/2} + \left(\frac{\beta}{t_i}\right)^{3/2}\right] - \frac{1}{2\lambda^2}\sum_{i=1}^{n}\left(\frac{t_i}{\beta} + \frac{\beta}{t_i} - 2\right).$$

Os elementos da função escore perfilada são dados por

$$u_p(\lambda) = \frac{\partial \ell_p(\tau)}{\partial \lambda} = \frac{n}{\lambda}\frac{\sum_{i=1}^{n} a_{t_i} w_i}{\sum_{i=1}^{n}\log\{\Phi(a_{t_i})\}} + \frac{1}{\lambda}\sum_{i=1}^{n} a_{t_i} w_i - \frac{n}{\lambda} + \frac{1}{\lambda}\sum_{i=1}^{n} a_{t_i}^2$$

e

$$u_p(\beta) = \frac{\partial \ell_p(\tau)}{\partial \beta} = \frac{n}{2\lambda}\frac{\sum_{i=1}^{n}\left(\frac{\sqrt{t_i}}{\beta} + \frac{1}{\sqrt{t_i}}\right) w_i}{\sum_{i=1}^{n}\log\{\Phi(a_{t_i})\}} + \frac{1}{2\lambda\beta^{1/2}}\sum_{i=1}^{n}\left(\frac{\sqrt{t_i}}{\beta} + \frac{1}{\sqrt{t_i}}\right) w_i$$
$$-\frac{n}{2\beta} + \sum_{i=1}^{n}\frac{1}{t_i+\beta} + \frac{1}{2\lambda^2\beta^2}\sum_{i=1}^{n} t_i - \frac{1}{2\lambda^2}\sum_{i=1}^{n}\frac{1}{t_i}$$

e, conseqüentemente, as estimativas de máxima verossimilhança perfilada para λ e β, $\hat{\lambda}_p$ e $\hat{\beta}_p$, são as soluções das equações

$$u_p(\lambda) = 0 \;\text{ e }\; u_p(\beta) = 0.$$

Enquanto que a aproximação para a função de log-verossimilhança perfilada modificada de Barndorff-Nielsen é dada por

$$\breve{\ell}_{BN}(\lambda,\beta) = \ell_p(\lambda,\beta) + \frac{1}{2}\log\left(\frac{n}{\hat{\alpha}}\right) - \log\left|\breve{I}_\alpha(\lambda,\beta,\hat{\alpha}_{\lambda,\beta};\hat{\lambda},\hat{\beta},\hat{\alpha})\right|$$

onde

$$\breve{I}_\alpha(\lambda,\beta,\hat{\alpha}_p;\hat{\lambda},\hat{\beta},\hat{\alpha}) = \sum_{j=1}^{n}\left\{\ell_\alpha^{(j)}(\lambda,\beta,\hat{\alpha}_p)\ell_\alpha^{(j)}(\hat{\lambda},\hat{\beta},\hat{\alpha})^T\right\}$$
$$= \frac{n}{\hat{\alpha}_{\lambda,\beta}\hat{\alpha}} + \frac{1}{\hat{\alpha}_{\lambda,\beta}}\sum_{i=1}^{n}\log\{\Phi(a_{t_i})\} + \frac{1}{\hat{\alpha}}\sum_{i=1}^{n}\hat{a}_{t_i}$$
$$+\sum_{i=1}^{n}\log\{\Phi(\hat{a}_{t_i})\}\log\{\Phi(a_{t_i})\},$$

sendo $a_t = \frac{1}{\lambda}\left\{\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right\}$ e $\hat{a}_t = \frac{1}{\hat{\lambda}}\left\{\sqrt{\frac{t}{\hat{\beta}}} - \sqrt{\frac{\hat{\beta}}{t}}\right\}$. Assim, as estimativas de máxima verosimilhança perfilada modificada de Barndorff-Nielsen são as soluções das equações

$$u_{BN}(\lambda) = 0 \;\; e \;\; u_{BN}(\beta) = 0.$$

4.5 Estudo do Viés

Em relação à estimação pontual dos parâmetros λ, β e α apresenta-se agora as técnicas utilizadas para estudar e propor estimadores de viés corrigidos dos estimadores de máxima verossimilhança de tais parâmetros. Utiliza-se um parâmetros genérico ao invés de se particularizar a exposição aos parâmetros citados. Nós usamos alguns métodos conhecidos na literatura neste tipo de modelo.

4.5.1 Método Jackknife

Para os EMV de λ, β e α os estimadores de viés corrigidos vía Jackknife são dados por (vide Efron, 1982) $\hat{\lambda}_{JKc} = n\hat{\lambda} - (n-1)\hat{\lambda}_{JK}$, $\hat{\beta}_{JKc} = n\hat{\beta} - (n-1)\hat{\beta}_{JK}$ e $\hat{\alpha}_{JKc} = n\hat{\alpha} - (n-1)\hat{\alpha}_{JK}$, onde $\hat{\alpha}_{JK}$, $\hat{\lambda}_{JK}$ e $\hat{\beta}_{JK}$ foram dados acima.

4.5.2 Método de Cox e Snell

Primeiro nós introduzimos alguma notação, $U_\lambda = \frac{\partial l(\theta;T)}{\partial \lambda}$, $U_\beta = \frac{\partial l(\theta;T)}{\partial \beta}$, $U_\alpha = \frac{\partial l(\theta;T)}{\partial \alpha}$, $U_{\alpha\lambda} = \frac{\partial^2 l(\theta;T)}{\partial\alpha\partial\lambda}$, $U_{\alpha\beta} = \frac{\partial^2 l(\theta;T)}{\partial\alpha\partial\beta}$, $U_{\lambda\beta} = \frac{\partial^2 l(\theta;T)}{\partial\lambda\partial\beta}$, *etc.* A notação para os momentos das derivadas anteriores, Lawley (1956): $k_{\alpha\alpha} = \mathbb{E}(U_{\alpha\alpha})$, $k_{\alpha\lambda} = \mathbb{E}(U_{\alpha\lambda})$, $k_{\alpha\beta} = \mathbb{E}(U_{\alpha\beta})$ e assim sucessivamente. $k_{\alpha\alpha,\alpha} = \mathbb{E}(U_{\alpha\alpha}U_\alpha)$, $k_{\alpha\alpha,\lambda} = \mathbb{E}(U_{\alpha\alpha}U_\lambda)$, $k_{\alpha\lambda,\beta} = \mathbb{E}(U_{\alpha\lambda}U_\beta)$, etc. $k_{\alpha\alpha\alpha} = \mathbb{E}(U_{\alpha\alpha\alpha})$, $k_{\alpha\alpha\lambda} = \mathbb{E}(U_{\alpha\alpha\lambda})$, $k_{\alpha\lambda\beta} = \mathbb{E}(U_{\alpha\lambda\beta})$ e assim sucessivamente. Finalmente, $k^{(\alpha)}_{\alpha\alpha} = \frac{\partial k_{\alpha\alpha}}{\partial\alpha}$, $k^{(\alpha)}_{\alpha\lambda} = \frac{\partial k_{\alpha\lambda}}{\partial\alpha}$, $k^{\beta}_{\alpha\lambda} = \frac{\partial k_{\alpha\lambda}}{\partial\beta}$, etc.

Cox e Snell (1968) obtem uma fórmula geral para o viés, de ordem $O(n^{-1})$, do EMV do vetor de parâmetros θ. A expressão geral é

$$B(\hat{\theta}_s) = \sum_{r,s,t} k^{s,r} k^{t,u} \left\{ \frac{1}{2} k_{rtu} + k_{rt,u} \right\} \tag{4.21}$$

onde r, t, u percorrem o espaço dos parâmetros. Depois de algumas manipulações algébricas pode-se escrever (4.21) como (Patriota et al., 2009)

$$B(\hat{\theta}_s) = \sum_{r,s,t} k^{s,r} k^{t,u} \left\{ k^{(u)}_{rt} - \frac{1}{2} k_{rtu} \right\}. \tag{4.22}$$

Aqui $k^{\alpha,\alpha}$ denota o elemento $(1,1)$ da matriz de informação de Fisher, $k^{\alpha,\lambda}$ o elemento

$(1,2)$, $k^{\alpha,\beta}$ o elemento $(1,3)$, $k^{\lambda,\lambda}$ o elemento $(2,2)$ e assim por diante.

Nós nos referimos a $B(\theta)$ como o viés de segunda ordem do EMV $\hat{\theta}_s$, para $s = 1, 2, ..., p$. Então, após algumas manipulações algébricas chegamos às seguintes expressões para os viés dos três parâmetros da BSPN, onde as expressões k_{rtu}, $k_{rt}^{(u)}$ e $k^{s,r}$ são dadas no apêndice B:

$$
\begin{aligned}
B(\hat{\alpha}) = {} & \frac{1}{2}k^{\alpha,\alpha}k^{\alpha,\alpha}k_{\alpha\alpha}^{(\alpha)} + k^{\alpha,\alpha}k^{\lambda,\lambda}\left(k_{\alpha\lambda}^{(\lambda)} - \frac{1}{2}k_{\alpha\lambda\lambda}\right) + k^{\alpha,\alpha}k^{\beta,\beta}\left(k_{\alpha\beta}^{(\beta)} - \frac{1}{2}k_{\alpha\beta\beta}\right) \\
& + k^{\alpha,\lambda}k^{\alpha,\lambda}\left(k_{\lambda\lambda}^{(\alpha)} + k_{\alpha\lambda}^{(\lambda)} - k_{\alpha\lambda\lambda}\right) + k^{\alpha,\alpha}k^{\alpha,\lambda}\left(k_{\alpha\alpha}^{(\lambda)} + 2k_{\alpha\lambda}^{(\alpha)} - \frac{3}{2}k_{\alpha\alpha\lambda}\right) \\
& + k^{\alpha,\beta}k^{\alpha,\beta}\left(k_{\beta\beta}^{(\alpha)} + k_{\alpha\beta}^{(\beta)} - k_{\alpha\beta\beta}\right) + k^{\alpha,\alpha}k^{\alpha,\beta}\left(k_{\alpha\alpha}^{(\beta)} + 2k_{\alpha\beta}^{(\alpha)} - \frac{3}{2}k_{\alpha\alpha\beta}\right) \\
& + k^{\lambda,\lambda}k^{\alpha,\lambda}\left(k_{\lambda\lambda}^{(\alpha)} - \frac{1}{2}k_{\lambda\lambda\lambda}\right) + k^{\lambda,\lambda}k^{\alpha,\beta}\left(k_{\lambda\beta}^{(\lambda)} - \frac{1}{2}k_{\lambda\lambda\beta}\right) \\
& + k^{\alpha,\lambda}k^{\lambda,\beta}\left(k_{\lambda\lambda}^{(\beta)} + k_{\lambda\beta}^{(\lambda)} - \frac{3}{2}k_{\lambda\lambda\beta}\right) + k^{\alpha,\beta}k^{\lambda,\beta}\left(k_{\beta\beta}^{(\lambda)} + k_{\lambda\beta}^{(\beta)} - k_{\lambda\beta\beta}\right) \\
& + k^{\beta,\beta}k^{\alpha,\beta}\left(k_{\beta\beta}^{(\alpha)} - \frac{1}{2}k_{\beta\beta\beta}\right) + k^{\beta,\beta}k^{\alpha,\lambda}\left(k_{\lambda\beta}^{(\beta)} - \frac{1}{2}k_{\lambda\beta\beta}\right) \\
& + k^{\alpha,\alpha}k^{\lambda,\beta}\left(k_{\alpha\lambda}^{(\beta)} + k_{\alpha\beta}^{(\lambda)} - k_{\alpha\lambda\beta}\right) + k^{\alpha,\lambda}k^{\alpha,\beta}\left(k_{\lambda\beta}^{(\alpha)} + k_{\alpha\lambda}^{(\beta)} + k_{\alpha\beta}^{(\lambda)} + k_{\lambda\beta}^{(\alpha)} - 2k_{\alpha\lambda\beta}\right)
\end{aligned}
$$

$$
\begin{aligned}
B(\hat{\lambda}) = {} & k^{\lambda,\lambda}k^{\lambda,\lambda}\left(k_{\lambda\lambda}^{(\lambda)} - \frac{1}{2}k_{\lambda\lambda\lambda}\right) + k^{\alpha,\alpha}k^{\lambda,\lambda}\left(k_{\alpha\lambda}^{(\alpha)} - \frac{1}{2}k_{\alpha\alpha\lambda}\right) + k^{\lambda,\lambda}k^{\beta,\beta}\left(k_{\lambda\beta}^{(\beta)} - \frac{1}{2}k_{\lambda\beta\beta}\right) \\
& + k^{\lambda,\lambda}k^{\alpha,\lambda}\left(k_{\lambda\lambda}^{(\alpha)} + 2k_{\alpha\lambda}^{(\lambda)} - \frac{3}{2}k_{\alpha\lambda\lambda}\right) + k^{\alpha,\lambda}k^{\alpha,\lambda}\left(k_{\alpha\alpha}^{(\lambda)} + k_{\alpha\lambda}^{(\alpha)} - k_{\alpha\alpha\lambda}\right) \\
& + k^{\lambda,\beta}k^{\lambda,\beta}\left(k_{\beta\beta}^{(\lambda)} + k_{\lambda\beta}^{(\beta)} - k_{\lambda\beta\beta}\right) + k^{\lambda,\lambda}k^{\lambda,\beta}\left(k_{\lambda\lambda}^{(\beta)} + 2k_{\lambda\beta}^{(\lambda)} - \frac{3}{2}k_{\lambda\lambda\beta}\right) \\
& + k^{\alpha,\alpha}k^{\alpha,\lambda}\left(k_{\alpha\alpha}^{(\alpha)} - \frac{1}{2}k_{\alpha\alpha\alpha}\right) + k^{\alpha,\alpha}k^{\lambda,\beta}\left(k_{\alpha\beta}^{(\alpha)} - \frac{1}{2}k_{\alpha\alpha\beta}\right) \\
& + k^{\beta,\beta}k^{\lambda,\beta}\left(k_{\beta\beta}^{(\beta)} - \frac{1}{2}k_{\beta\beta\beta}\right) + k^{\beta,\beta}k^{\alpha,\lambda}\left(k_{\alpha\beta}^{(\beta)} - \frac{1}{2}k_{\alpha\beta\beta}\right) \\
& + k^{\alpha,\lambda}k^{\alpha,\beta}\left(k_{\alpha\alpha}^{(\beta)} + k_{\alpha\beta}^{(\alpha)} - k_{\alpha\alpha\beta}\right) + k^{\alpha,\beta}k^{\lambda,\beta}\left(k_{\beta\beta}^{(\alpha)} + k_{\alpha\beta}^{(\beta)} - k_{\alpha\beta\beta}\right) \\
& + k^{\lambda,\lambda}k^{\alpha,\beta}\left(k_{\lambda\beta}^{(\alpha)} + k_{\alpha\lambda}^{(\beta)} - k_{\alpha\lambda\beta}\right) + k^{\alpha,\lambda}k^{\lambda,\beta}\left(k_{\alpha\beta}^{(\lambda)} + k_{\alpha\lambda}^{(\beta)} + k_{\alpha\beta}^{(\lambda)} + k_{\lambda\beta}^{(\alpha)} - 2k_{\alpha\lambda\beta}\right)
\end{aligned}
$$

$$
\begin{aligned}
B(\hat{\beta}) = {} & k^{\beta,\beta}k^{\beta,\beta}\left(k^{(\beta)}_{\beta\beta} - \frac{1}{2}k_{\beta\beta\beta}\right) + k^{\alpha,\alpha}k^{\beta,\beta}\left(k^{(\alpha)}_{\alpha\beta} - \frac{1}{2}k_{\alpha\alpha\beta}\right) + k^{\lambda,\lambda}k^{\beta,\beta}\left(k^{(\lambda)}_{\lambda\beta} - \frac{1}{2}k_{\lambda\beta\beta}\right) \\
& + k^{\beta,\beta}k^{\alpha,\beta}\left(k^{(\alpha)}_{\beta\beta} + 2k^{(\beta)}_{\alpha\beta} - \frac{3}{2}k_{\alpha\beta\beta}\right) + k^{\alpha,\beta}k^{\alpha,\beta}\left(k^{(\beta)}_{\alpha\alpha} + k^{(\alpha)}_{\alpha\beta} - k_{\alpha\alpha\beta}\right) \\
& + k^{\lambda,\beta}k^{\lambda,\beta}\left(k^{(\beta)}_{\lambda\lambda} + k^{(\lambda)}_{\lambda\beta} - k_{\lambda\lambda\beta}\right) + k^{\beta,\beta}k^{\lambda,\beta}\left(k^{(\lambda)}_{\beta\beta} + 2k^{(\beta)}_{\lambda\beta} - \frac{3}{2}k_{\lambda\beta\beta}\right) \\
& + k^{\alpha,\alpha}k^{\alpha,\beta}\left(k^{(\alpha)}_{\alpha\alpha} - \frac{1}{2}k_{\alpha\alpha\alpha}\right) + k^{\alpha,\alpha}k^{\lambda,\beta}\left(k^{(\alpha)}_{\alpha\lambda} - \frac{1}{2}k_{\alpha\alpha\lambda}\right) \\
& + k^{\lambda,\lambda}k^{\lambda,\beta}\left(k^{(\lambda)}_{\lambda\lambda} - \frac{1}{2}k_{\lambda\lambda\lambda}\right) + k^{\lambda,\lambda}k^{\alpha,\beta}\left(k^{(\lambda)}_{\alpha\lambda} - \frac{1}{2}k_{\alpha\lambda\lambda}\right) \\
& + k^{\alpha,\lambda}k^{\alpha,\beta}\left(k^{(\lambda)}_{\alpha\alpha} + k^{(\alpha)}_{\alpha\lambda} - k_{\alpha\alpha\lambda}\right) + k^{\alpha,\lambda}k^{\lambda,\beta}\left(k^{(\alpha)}_{\lambda\lambda} + k^{(\lambda)}_{\alpha\lambda} - k_{\alpha\alpha\lambda}\right) \\
& + k^{\beta,\beta}k^{\alpha,\lambda}\left(k^{(\lambda)}_{\alpha\beta} + k^{(\alpha)}_{\lambda\beta} - k_{\alpha\lambda\beta}\right) + k^{\alpha,\beta}k^{\lambda,\beta}\left(k^{(\lambda)}_{\alpha\beta} + k^{(\beta)}_{\alpha\lambda} + k^{(\beta)}_{\alpha\lambda} + k^{(\alpha)}_{\lambda\beta} - 2k_{\alpha\lambda\beta}\right).
\end{aligned}
$$

4.5.3 Método Bootstrap

Seja $\hat{\theta}$ um estimador de θ derivado da amostra aleatória $\mathbb{X}$, $\hat{\theta} = s(\mathbb{X})$. A partir de um número R de repeticões de $\mathbb{X}$, pode-se obter uma aproximação para a função de distribuição de $\hat{\theta}$ utilizando a função de distribuição empírica construída a partir destas R estimativas.

O método Bootstrap foi introduzido por Efron (1979) e consiste em gerar R amostras $x^* = ({x_1}^*, ..., x^*_N)$ e consequentemente R estimativas $s(x^*)$ a partir de um conjunto de medidas $x = (x_1, ..., x_N)$.

Dependendo da forma como as R amostras são geradas, classifica-se o esquema do Bootstrap em três tipos: paramétrico, não paramétrico e bayesiano. Neste trabalho somente utiliza-se o método não-paramétrico. O Bootstrap não-paramétrico caracteriza-se pela obtenção das R amostras de tamanho N a partir da distribuição $\hat{F}$, ou seja de amostragem com reposição dos elementos da amostra observada. Deste modo, assume-se que a população amostral pode ser descrita pela distribuição $\hat{F}$, onde $\hat{F}$ é a distribuição empírica da amostra original, isto é,

$$\hat{F} = \frac{1}{N}\{x_i \leq t\}, \qquad \forall t \in \mathbb{R}.$$

As R amostras bootstrap proporcionam R replicas de $\hat{\theta}$ utilizadas para a construção

de uma função de distribuição empírica. Assim é possível estimar o erro padrão deste estimador ou corrigir seu viés, como também construir intervalos de confiança para o parâmetro θ. A estimativa bootstrap ideal do viés é definida no limite quando $R \to \infty$. Desta forma, gerando R amostras bootstrap independentes $x^{*1}, ..., x^{*R}$ e calculando as respectivas réplicas bootstrap $\hat{\theta}^{*1}, ..., \hat{\theta}^{*R}$ pode-se aproximar as esperanças bootstrap $\mathbb{E}_{\hat{F}}[s(\mathbb{X}^*)]$ pela média

$$\hat{\theta}^*(\cdot) = \frac{1}{R}\sum_{i=1}^{R}\hat{\theta}^{*i}$$

Assim, as estimativas bootstrap do viés baseadas nas R réplicas do estimador $\hat{\theta}$ é

$$\hat{B}_{\hat{F}}(\theta, \hat{\theta}) = \hat{\theta}^*(\cdot) - s(x).$$

Assim, para o EMV de θ pode-se definir o estimador de viés corrigido de segunda ordem como:

$$\hat{\theta}_{Bc} = s(x) - \hat{B}_{\hat{F}}(\theta, \hat{\theta}) = 2\hat{\theta} - \hat{\theta}^{*\cdot}.$$

No caso da BSPN os estimadores de viés corrigido para os parâmetros do modelo são

$$\hat{\alpha}_{Bc} = 2\hat{\alpha} - \hat{\alpha}^{*\cdot}, \quad \hat{\lambda}_{Bc} = 2\hat{\lambda} - \hat{\lambda}^{*\cdot} \quad e \quad \hat{\beta}_{Bc} = 2\hat{\beta} - \hat{\beta}^{*\cdot},$$

onde

$$\hat{\alpha}^{*\cdot} = \frac{1}{R}\sum_{i=1}^{R}\hat{\alpha}^{*i}, \quad \hat{\lambda}^{*\cdot} = \frac{1}{R}\sum_{i=1}^{R}\hat{\lambda}^{*i} \quad e \quad \hat{\beta}^{*\cdot} = \frac{1}{R}\sum_{i=1}^{R}\hat{\beta}^{*i}.$$

4.6 Resultados Numéricos: estimação

O objetivo é comparar, através de simulação, o desempenho dos estimadores de máxima verossimilhança dos parâmetros que indexam a distribuição BSPN; e as versões corrigidas pelos métodos de verossimilhança perfilada, Cox e Snell, bootstrap e jackknife. Para isso foram calculados medidas de qualidade para a estimação pontual como viés e a raiz do erro quadrático médio ($\sqrt{EQM}$). Além disso, também são calculados os coeficientes de assimetria e curtose dos estimadores de máxima verossimilhança. Para a geração das sequências de números pseudo-aleatórios com distribuição $BSPN(\lambda, \beta, \alpha)$, foram geradas observações

de uma distribuição uniforme $U(0,1)$ e daí se obtêm a variável aleatória $Z = \frac{1}{2}\lambda\Phi^{-1}(U^{1/\alpha})$ com distribuição $Z_i \sim PN(0, \frac{1}{4}\lambda^2, \alpha)$, para $i = 1, 2, \ldots, n$. Agora, a variável aleatória T com distribuição BSPN se obtêm pela transformação $T = \beta\left(1 + 2Z^2 + 2Z\left(1 + Z^2\right)^{1/2}\right)$. Todos os procedimentos de cálculo foram programados usando o pacote estatístico R através do método BFGS do procedimento "optim".

Se desenvolve um experimento de simulação de Monte Carlo baseado em 5.000 réplicas considerando os tamanhos amostrais $n = 70, 90, 120$ e 150. Os valores verdadeiros considerados para os parâmetros foram: $\alpha = 0.85, 1.25, 1.5$; $\lambda = 0.15, 0.5, 0.75$ e para o parâmetro de escala atribui-se, sem perda de generalidade, um único valor $\beta = 1.0$. No caso do estimador de viés corrigido pelo método bootstrap não paramétrico o número de amostras por cada réplica foi R=400. Para todos os estimadores estudados, as estimativas foram obtidas maximizando-se o logaritmo da função de verossimilhança, tomando como valores iniciais de λ e β, para o processo de maximização, os estimadores de momentos modificados para a BS sob normalidade e para o parâmetro α o estimador obtido pelo método do percentil *elemental*, dados conhecidos os estimadores de λ e β obtidos pelo método dos momentos modificados.

Inicialmente avaliamos os resultados que correspondem as estimativas de λ. A Tabela $A.3$ mostra que os viéses e a $\sqrt{EQM}$ do estimador de máxima verossimilhança de λ decrescem quando o tamanho da amostra aumenta, ou seja que o estimador apresenta certa consistência. Entretanto, quando α aumenta, também aumenta o viés do estimador de λ. Os resultados da Tabela $A.5$ mostraram que os estimadores de λ possuem comportamentos diferentes, assim o estimador de viés corrigido de Cox e Snell têm viéses, em modulo, e $\sqrt{EQM}$ menores que os dos estimadores de máxima verossimilhançã, máxima verossimilhança perfilada modificada, bootstrap e jackknife, entretanto, estes últimos apresentam viéses muitos semelhantes para todos os valores de λ e α. Além disso, para todos os estimadores estudados, os viéses subestimam em média. Mesmo assim, a $\sqrt{EQM}$ de todos os estimadores em questão diminuem quando o tamanho da amostra aumenta, o que também acontece para os viéses dos estimadores de máxima verossimilhança perfilada modificada, bootstrap e jackknife.

Para o parâmetro β nós observamos da Tabela $A.3$ que a $\sqrt{EQM}$ diminui quando o tamanho da amostra aumenta. Além disso, nós temos que para valores de $\alpha < 1$, os estimadores subestimam o valor do parâmetro, enquanto que para valores de $\alpha > 1$ estes superestimam o valor do parâmetro. Para $\alpha < 1$, o estimador de viés corrigido de Cox e Snell apresenta um bom desempenho enquanto que para $\alpha > 1$ o resto dos estimadores têm melhor desempenho. Além disso, é claro que quando α e λ aumentam, os viéses do estimador também aumentam. Mesmo assim, a eficiência do estimador aumenta quando o tamanho da amostra aumenta.

Por último, para o parâmetro de forma α, o bom desempenho dos estimadores, viés e $\sqrt{EQM}$, dependem do tamanho da amostra.

Para os coeficientes de assimetria e curtose, nós temos que para o parâmetro β sua curtose sempre esta ao redor de três e sua assimetria ao redor de zero, mais para o parâmetro λ, que apresenta uma assimetria negativa e uma curtose maior de três. Enquanto que para o parâmetro α, sua assimetria esta perto de zero para $\alpha < 1$ e com valores positivos para $\alpha > 1$. Mesmo assim, sua curtose têm valores maiores para valores de $\alpha > 1$.

4.6.1 Ilustração

Nós agora apresentamos a aplicação do modelo BSPN para um conjunto de dados reais previamente analisados por Birnbaum e Saunders (1969b) relacionados com a vida em ciclos x 10^{-3} de peças 6061 - T6 cortadas em um ângulo paralelo ao sentido de rotação, variando a taxa de 18 ciclos por segundo a uma pressão máxima de 21.000 psi. O tamanho amostral total foi de 101 unidades. Os resultados foram obtidos usando a função "nlm" do pacote estatístico R. Os estimadores de máxima verossimilhança foram calculados numericamente maximizando a função de log-verossimilhança que decorre da densidade (4.7). Na Tabela 6.3 podemos ver que os dados apresentam assimetria positiva e curtose menor que a admitida pelo modelo BS.

Assim nós propomos o modelo BSPN como alternativa para analisar o conjunto de dados. Para uma melhor justificação consideramos o teste de hipótese de não diferença

Tabela 4.1 *Estatísticas descritivas para o conjunto de dados*

n	$\bar{t}$	s^2	$\sqrt{b_1}$	b_2
101	1400.91	1529.10	0.142	2.81

do modelo BS sob normalidade com o modelo BSPN, isto é a hipótese

$$H_0: \ \alpha = 1 \ \text{ contra } \ H_1: \ \alpha \neq 1$$

usamos a estatística de razão de verossimilhanças baseada em

$$\Lambda = \frac{L_{BS}\left(\hat{\boldsymbol{\theta}}\right)}{L_{BSPN}\left(\hat{\boldsymbol{\theta}}\right)}.$$

Assim, obtemos que $-2\log(\Lambda) = -2(-751.3322 + 747.5460) = 7.5724$, o qual é um valor maior ao percentil da distribuição qui-quadrado com um grau de liberdade ao nível de 5%, cujo valor é 3.84. Portanto, concluímos que o modelo BSPN ajusta melhor que o modelo BS.

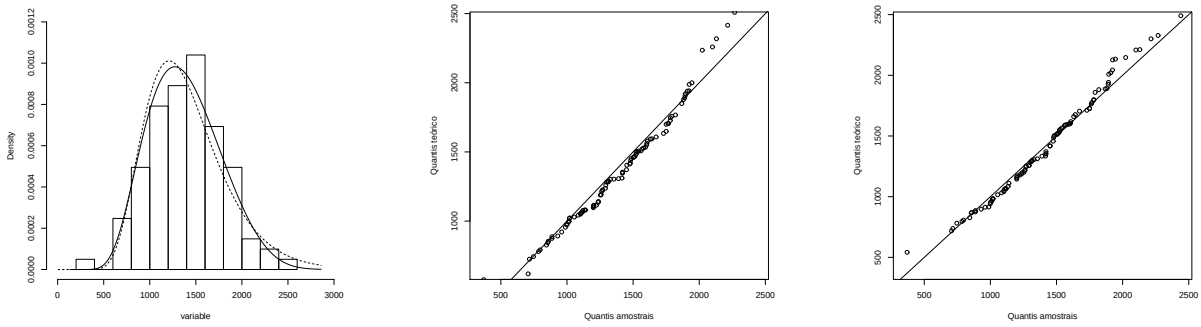
A Tabela 4.2 mostra os valores das estimativas dos parâmetros para o modelo BSPN, em comparação com o modelo BS. Da mesma forma, a Figura 4.4 (a) e os gráficos Q-Q "plot", feitos em forma similar do modelo PSN, mostram que o modelo BSAS é muito mais flexível do que o modelo BS, para modelar a assimetria dos dados. Como uma medida do erro de estimação, nós calculamos intervalos de confiança bootstrap paramétrico (Efron, 1979) para os três parâmetros do modelo. Assim, ao nível do 95% obtemos: $\lambda \in (0.0430, 0.1449)$, $\beta \in (2106.55, 2319.98)$ e $\alpha \in (0.0084, 0.1149)$.

4.7 Extensão da Distribuição BS à Familia de Distribuições α-Potência

Nesta seção nós vamos mudar a suposição de distribuição PN para a variável Z pela suposição de distribuição AP, onde $\alpha \in \mathbb{R}^+$ e F é uma função de distribuição cumulativa absolutamente contínua com função densidade de probabilidade (fdp) $f = dF$. As demostrações para os Teoremas e colorários são da mesma forma que para o caso BSPN.

Tabela 4.2 *Estimativas (erro padrão) dos parâmetros para os modelo BSPN e BS*

Parâmetros	BS	BSPN
λ	0.3101(0.0218)	0.0996(0.0001)
β	1336.5638(40.7572)	2135.99(18.5911)
α	-	0.0527(0.0054)
Log-likelihood	-751.3322	-747.5460



(a) (b) (c)

Figura 4.4 *(a) Gráfico para os modelos BSPN(0.0996,2135.99,0.0527), (linha continua) e BS(0.3101,1336.5638), (linha tracejada). Gráficos Q-Q "plot": (b) BS e (c) BSPN.*

Sem perda de generalidade, assumindo para $Z \sim AP(\alpha)$, a extensão da distribuição BS à família de distribuições α-potência é definida como segue.

Definição 4.7.1. *A variável aleatória T segue a distribuição de Birnbaum-Saunders α-potência, se T é da forma*

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

onde $Z \sim AP(\alpha)$, com $\lambda,\ \alpha > 0$ parâmetros de forma e $\beta > 0$ é um parâmetro de escala.

Se a variável aleatória T segue (4.23), usamos a notação $T \sim BSAP(\lambda, \beta, \alpha)$.

Definição 4.7.2. *Seja $T \sim BSAP(\lambda, \beta, \alpha)$. Então, a função de densidade de probabili-*

dade de T é dada por

$$\varphi_{fT}(t;\lambda,\beta,\alpha) = \alpha f(a_t)\{F(a_t)\}^{\alpha-1}\frac{t^{-3/2}[t+\beta]}{2\lambda\beta^{1/2}}, \tag{4.23}$$

onde a_t é como definida em (4.3).

Teorema 4.7.3. *Seja $T \sim BSAP(\lambda,\beta,\alpha)$. Então,*

1. $bT \sim BSAP(\lambda, b\beta, \alpha)$ $b > 0$ e

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

Teorema 4.7.4. *Sejam $T \sim BSAP(\lambda,\beta,\alpha)$, $\mathbb{F}_{fT}$ sua função distribuição cumulativa e $\mathbb{F}_f$ a função de distribuição cumulativa de $Z \sim AP(\alpha)$. Então,*
$\mathbb{F}_{fT}(t;\alpha) = \mathbb{F}_f(a_t(\lambda,\beta);\alpha) = \{F(a_t(\lambda,\beta);\alpha)\}^{\alpha}$.

4.7.1 Algumas Densidades

Agora apresentamos algumas funções de densidades nas quais é possível fazer análise da influencia dos parâmetros na distribuição BSAP e suas diferenças com a Birnbaum-Saunders generalizada de Díaz-García e Leiva-Sánchez (2005).

Distribuição t-Student (BSPt)

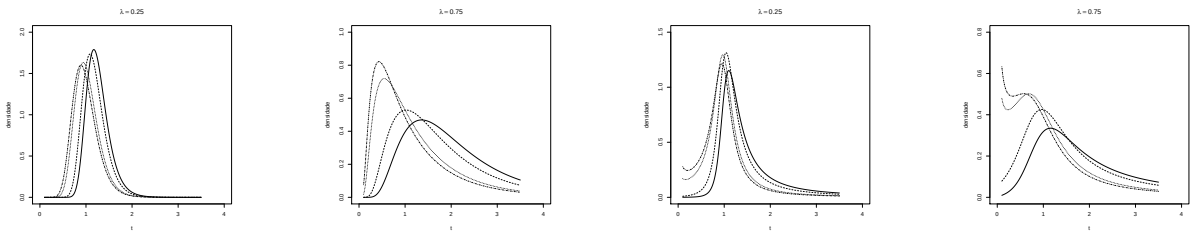
O modelo BSAP para a distribuição t de Student é dado por:

$$\varphi_{fT}(t;\lambda,\beta,\alpha,v) = \alpha\frac{\Gamma(\frac{v+1}{2})}{(v\pi)^{1/2}\Gamma(\frac{v}{2})}\left[1+\frac{1}{v\lambda^2}\left(\frac{t}{\beta}+\frac{\beta}{t}-2\right)\right]^{-(v+1)/2}$$
$$\{F(a_t)\}^{\alpha-1}\frac{t^{-3/2}[t+\beta]}{2\lambda\beta^{1/2}}, \tag{4.24}$$

onde v representa os graus de liberdade e F a função de distribuição cumulativa da distribuição t.

A Figura 4.5 mostra as densidades da BSPt para $\lambda = 0.25$ e 0.75, $\beta = 1.0$ e $\alpha = 0.80, 1, 2$ e 3. Nos gráficos (a) e (b) para $v = 35$, pode-se observar que quando α aumenta, a distribuição é mais leptocúrtica, mesmo quando λ é maior a distribuição é mais assimétrica. Pode-se também comparar como é o comportamento da distribuição GBS (caso $\alpha = 1.0$) para o modelo t-Student, Enquanto que para $v = 1$ temos a distribuição

de Cauchy AP (BSPC). Na Figura 4.5 (c) e (d), observa-se para esta distribuição um comportamento menos estável na cauda esquerda para t ao redor de zero.



(a) (b) (c) (d)

Figura 4.5 *Densidade* $\varphi_T(t;\lambda,\beta,\alpha,v)$, *para valores* $v = 35$ *(a) - (b) e* $v = 1$ *(c) - (d), para* $\lambda = 0.25$ *e* 0.75, $\beta = 1.0$, *e* $\alpha = 0.80$ *(linha de pontos e tracejada), 1 (linha de pontos), 2 (linha tracejada) e 3 (linha contínua).*

Os modelos seguintes também sõ importante em diversas aplicações praticas.

Distribuição Exponencial (BSPE)

A variável aleatória T tem distribução exponencial para a distribuição BSAP se sua função de densidade de probabilidade é dada por

$$\varphi_{fT}(t;\lambda,\beta,\alpha) = \alpha \exp\left\{-\frac{1}{\lambda}\left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right)\right\}\left\{1 - \exp\left\{-\frac{1}{\lambda}\left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right)\right\}\right\}^{\alpha-1} \frac{t^{-3/2}[t+\beta]}{2\lambda\beta^{1/2}}. \tag{4.25}$$

Distribuição de Weibull (BSPW)

A função de densidade de probabilidade de T para a distribuição Weibull padrão (1,c) onde c é um parâmetros de forma, é dada por:

$$\varphi_{fT}(t;\lambda,\beta,\alpha,c) = c\alpha\left\{\frac{1}{\lambda}\left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right)\right\}^{c-1} \exp\left\{-\frac{1}{\lambda^c}\left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right)^c\right\} \left\{1 - \exp\left\{-\frac{1}{\lambda^c}\left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right)^c\right\}\right\}^{\alpha-1} \frac{t^{-3/2}[t+\beta]}{2\lambda\beta^{1/2}}. \tag{4.26}$$

Outros modelos como Tipo Kotz e Tipo Pearson VII também são muito utilizados na estatística moderna.

Distribuição Tipo Kotz (BSPTK)

Para uma variável que segue a distribuição tipo Kotz de parâmetros (q,r,s), a função densidade de probabilidade de T é :

$$\varphi_{fT}(t;\lambda,\beta,\alpha,q,r,s) = \alpha\frac{sr^{\frac{2q-1}{2s}}}{\lambda^{2q-2}\Gamma\left(\frac{2q-1}{2s}\right)}\left(\frac{t}{\beta}+\frac{\beta}{t}-2\right)^{q-1}\exp\left(-\frac{r}{\lambda^{2s}}\left[\frac{t}{\beta}+\frac{\beta}{t}-2\right]^{s}\right)$$
$$\{F(a_t)\}^{\alpha-1}\frac{t^{-3/2}[t+\beta]}{2\lambda\beta^{1/2}}, \quad q>1/2,\ r>0,\ s>0, \tag{4.27}$$

onde F é a função de distribuição cumulativa da Kotz de parâmetros (q,r,s).

Distribuição de Tipo Pearson VII (BSPPT-VII)

Para uma variável que segue a distribuição tipo Pearson VII de parâmetros (q,r), a função densidade de probabilidade de T é:

$$\varphi_{fT}(t;\lambda,\beta,\alpha,q,r) = \alpha\frac{\Gamma(q)}{\sqrt{r\pi}\Gamma(q-1/2)}\left[1+\frac{1}{r\lambda^2}\left(\frac{t}{\beta}+\frac{\beta}{t}-2\right)\right]^{-q}$$
$$\{F(a_t)\}^{\alpha-1}\frac{t^{-3/2}[t+\beta]}{2\lambda\beta^{1/2}}, \quad q>1/2,\ r>0, \tag{4.28}$$

onde F é a função de distribuição cumulativa da Pearson VII de parâmetros (q,r).

4.7.2 Momentos

Os seguentes dois teoremas são uma extensão dos teoremas dados para o caso BSPN.

Teorema 4.7.5. *Sejam $T \sim BSAP(\lambda,\beta,\alpha)$ e $Z \sim AP(\alpha)$. Então, $\mathbb{E}(T^n)$ existe se, e somente se,*

$$\mathbb{E}\left[\left(\frac{\lambda Z}{2}\right)^{k+l}\left(\left(\frac{\lambda Z}{2}\right)+1\right)^{\frac{k-l}{2}}\right] \tag{4.29}$$

existe para $k=1,2,\ldots,n$ com $l=0,1,\ldots,k$.

Teorema 4.7.6. *Sejam $T \sim BSAP(\lambda,\beta,\alpha)$ e $Z \sim AP(\alpha)$. Se $\mathbb{E}[Z^r]$ existe para $r =$*

$1, 2, \ldots,$ *então*

$$\acute{\mu}_r = \mathbb{E}(T^r) = \beta^r \sum_{k=0}^{r} \binom{r}{k} \left(\frac{1}{2}\right)^k \mathbb{E}\left[\lambda^2 Z^2 + \lambda Z\sqrt{\lambda^2 Z^2 + 4}\right]^k$$

$$= \beta^r \sum_{[0\leq k\leq r/2]} \binom{r}{2k} \left(\frac{1}{2}\right)^{2k} \sum_{j=0}^{2k} \binom{2k}{j} \mathbb{E}[(\lambda Z)^{4k-j}(\lambda^2 Z^2 + 4)^{j/2}] +$$

$$\beta^r \sum_{[0\leq k<r/2]} \binom{r}{2k+1} \left(\frac{1}{2}\right)^{2k+1} \sum_{j=0}^{2k+1} \binom{2k+1}{j} \mathbb{E}[(\lambda Z)^{4k+2-j}(\lambda^2 Z^2 + 4)^{j/2}] \tag{4.30}$$

onde [.] *no índice da soma é a função parte inteira.*

Novamente, chamamos de $\nu_k = \mathbb{E}[Z^k]$ e $\kappa_k = \mathbb{E}\left[Z^k\left(\lambda^2 Z^2 + 4\right)^{1/2}\right]$, onde a expressão $h(Z) = Z^n(aZ^2 + 4)^{1/2}$ se pode obter como:

$$E(h(z)) = \alpha \int_0^1 [F^{-1}(z)]^n \left[a\left(F^{-1}(z)\right)^2 + 4\right] z^{\alpha-1} dz.$$

O corolário seguinte é extensão do corolário estudado no caso da distribuição BSPN.

Corolário 4.7.7. *Sejam* $T \sim BSAP(\lambda, \beta, \alpha)$ *e* $Z \sim AP(\alpha)$. *Se* $\mathbb{E}[Z^r]$ *existe para* $r = 1, 2, 3, 4$ *então*

$\acute{\mu}_1 = \mathbb{E}(T) = \frac{\beta}{2}[2 + \lambda^2\nu_2 + \lambda\kappa_1]$; $\quad \acute{\mu}_2 = \mathbb{E}(T^2) = \frac{\beta^2}{2}[2 + 4\lambda^2\nu_2 + \lambda^4\nu_4 + 2\lambda\kappa_1 + \lambda^3\kappa_3]$;

$\acute{\mu}_3 = \mathbb{E}(T^3) = \frac{\beta^3}{2}[2 + 9\lambda^2\nu_2 + 6\lambda^4\nu_4 + \lambda^6\nu_6 + 3\lambda\kappa_1 + 4\lambda^3\kappa_3 + \lambda^5\kappa_5]$ *e*

$\acute{\mu}_4 = \mathbb{E}(T^4) = \frac{\beta^4}{2}[2 + 16\lambda^2\nu_2 + 20\lambda^4\nu_4 + 8\lambda^6\nu_6 + \lambda^8\nu_8 + 4\lambda\kappa_1 + 10\lambda^3\kappa_3 + 6\lambda^5\kappa_5 + \lambda^7\kappa_7]$.

Do corolário (4.7.7), temos que os primeiros quatro momentos centrais, incluindo a variância e os coeficientes de assimetria e curtose ficam dados pelas mesmas expressoẽs do caso BSPN.

4.7.3 Inferência

Discutimos agora a estimação por máxima verossimilhança, a matriz de informação observada e a matriz de informação esperada para o vetor $\boldsymbol{\theta} = (\lambda, \beta, \alpha)'$. Assim, para uma amostra aleatória de tamanho n, $\mathbf{T} = (T_1, \ldots, T_n)'$, da distribuição $BSAP(\lambda, \beta, \alpha)$, a

função de log-verossimilhança para $\boldsymbol{\theta} = (\lambda, \beta, \alpha)'$ dado T é dada por:

$$\ell(\boldsymbol{\theta}; \mathbf{T}) = n[\log(\alpha) - \log(\lambda) - \frac{1}{2}\log(\beta) - \log 2] + \sum_{i=1}^{n} \log(t_i + \beta) - \frac{3}{2}\sum_{i=1}^{n} \log(t_i) + \\ + \sum_{i=1}^{n} \log(f(a_{t_i})) + (\alpha - 1)\sum_{i=1}^{n} \log(F(a_{t_i})) \tag{4.31}$$

Equações de Verossimilhança

Supondo que a derivada de f (f') existe, então os elementos da função escore são dados por:

$$U(\lambda) = -\frac{n}{\lambda}\left[\overline{a_t v} + (\alpha - 1)\overline{a_t w} + 1\right], \qquad U(\alpha) = \frac{n}{\alpha} + \sum_{i=1}^{n} \log\{F(a_{t_i})\} = n\{\overline{u} + 1/\alpha\},$$

$$U(\beta) = -\frac{n}{2\beta}\left[\overline{a_t v} + (\alpha - 1)\overline{a_t w} + 1\right] - \frac{n}{\lambda\beta^{1/2}}\left[\overline{v/\sqrt{t}} + (\alpha - 1)\overline{w/\sqrt{t}}\right] + \overline{(t+\beta)^{-1}},$$

onde $u_i = \log\{F(a_{t_i})\}$, $\overline{u} = \sum_{i=1}^{n} u_i/n$, $v_i = \frac{f'(a_{t_i})}{f(a_{t_i})}$, $w_i = \frac{f(a_{t_i})}{F(a_{t_i})}$, $\overline{a_t v} = \sum_{i=1}^{n} a_{t_i} v_i/n$, $\overline{a_t w} = \sum_{i=1}^{n} a_{t_i} w_i/n$, $\overline{(t+\beta)^{-1}} = \sum_{i=1}^{n} (t_i + \beta)^{-1}/n$, $\overline{v/\sqrt{t}} = \sum_{i=1}^{n} \frac{v_i}{\sqrt{t_i}}$ e $\overline{w/\sqrt{t}} = \sum_{i=1}^{n} \frac{w_i}{\sqrt{t_i}}$.

Nota-se que as soluções dos escores tem que satisfazer as equações:

$$\frac{1}{2\beta}(\overline{a_t v} + (\alpha - 1)\overline{a_t w} + 1) = -\frac{1}{\lambda}\left[\overline{v/\sqrt{t}} + (\alpha - 1)\overline{w/\sqrt{t}}\right] + \overline{(t+\beta)^{-1}}, \quad \alpha = -1/\overline{u}$$

e $1 + \overline{a_t v} = (\alpha - 1)\overline{a_t w}$, as quais geralmente tem soluções por meio de métodos numéricos iterativos.

Matriz de Informação Observada

Supondo que f'' existe e chamando $k_i = f''(a_{t_i})/f(a_{t_i})$ e $s_i = f'(a_{t_i})/F(a_{t_i})$, temos que os elementos da matriz de informação observada, que denotamos por $j_{\lambda\lambda}$, $j_{\beta\lambda}$, $j_{\alpha\lambda}$, $j_{\beta\beta}$, $j_{\alpha\beta}$, $j_{\alpha\alpha}$, são dados por:

$$j_{\lambda\lambda} = \frac{n}{\lambda^2}\left[-2\overline{a_t v} - \overline{a_t^2 k} + \overline{a_t^2 v^2} + (\alpha - 1)\left[\overline{a_t^2 w^2} - 2\overline{a_t w} - \overline{a_t^2 s}\right] - 1\right],$$

$$j_{\alpha\lambda} = \frac{n}{\lambda}\overline{a_t w}, \qquad j_{\alpha\beta} = n\left[\frac{1}{2\beta}\overline{a_t w} + \frac{1}{\lambda\sqrt{\beta}}\overline{w/\sqrt{t}}\right], \qquad j_{\alpha\alpha} = \frac{n}{\alpha^2},$$

$$j_{\beta\lambda} = \frac{n}{2\lambda\beta}\left[\overline{a_t^2 v^2} - \overline{a_t^2 k} - \overline{a_t v} + (\alpha-1)\left[\overline{a_t^2 w^2} - \overline{a_t^2 s} - \overline{a_t w}\right]\right] + \frac{n}{\lambda^2\sqrt{\beta}}\left[\overline{a_t v^2/\sqrt{t}} - \overline{a_t k/\sqrt{t}} - \overline{v/\sqrt{t}} + (\alpha-1)\left[\overline{a_t w^2/\sqrt{t}} - \overline{a_t s/\sqrt{t}} - \overline{w/\sqrt{t}}\right]\right]$$

e

$$j_{\beta\beta} = \frac{n}{4\beta^2}\left[\overline{a_t^2 v^2} - \overline{a_t^2 k} - \overline{a_t v} + 2 + (\alpha-1)\left[\overline{a_t^2 w^2} - \overline{a_t^2 s} - \overline{a_t w}\right]\right] + \frac{n}{2\lambda\beta^{3/2}}\left[2\overline{a_t v^2/\sqrt{t}} - 2\overline{a_t k/\sqrt{t}} - \overline{v/\sqrt{t}}\right] + \frac{n}{\lambda^2\beta}\left[\overline{v^2/\sqrt{t}} - \overline{k/t} + (\alpha-1)\left[\overline{w^2/t} - \overline{s/t}\right]\right] - n\overline{(t+\beta)^{-2}}$$

onde $\overline{a_t^k w^j} = \sum_{i=1}^n a_{t_i}^k w_i^j/n$, $\overline{(\beta+t)^j} = \sum_{i=1}^n 1/(\beta+t_i)^j/n$, $\overline{t^k w^j} = \sum_{i=1}^n t_i^k w_i^j/n$ e $\overline{w^j} = \sum_{i=1}^n w_i^j/n$.

Matriz de Informação Esperada

Os elementos da matriz de informação esperada podem ser expressos como:

$$i_{\lambda\lambda} = \frac{1}{\lambda^2}\left[-2\mathbb{E}(a_T V) - \mathbb{E}(a_T^2 K) + \mathbb{E}(a_T^2 V^2) + (\alpha-1)\left[\mathbb{E}(a_T^2 W^2) - 2\mathbb{E}(a_T W) - \mathbb{E}(a_T^2 S)\right] - 1\right],$$

$$i_{\alpha\lambda} = \frac{1}{\lambda}\mathbb{E}(a_T W), \qquad i_{\alpha\beta} = \left[\frac{1}{2\beta}\mathbb{E}(a_T W) + \frac{1}{\lambda\sqrt{\beta}}\mathbb{E}(W/\sqrt{T})\right], \qquad i_{\alpha\alpha} = \frac{1}{\alpha^2},$$

$$i_{\beta\lambda} = \frac{1}{2\lambda\beta}\left[\mathbb{E}(a_T^2 V^2) - \mathbb{E}(a_T^2 K) - \mathbb{E}(a_T V) + (\alpha-1)\left[\mathbb{E}(a_T^2 W^2) - \mathbb{E}(a_T^2 S) - \mathbb{E}(a_T W)\right]\right] + \frac{1}{\lambda^2\sqrt{\beta}}\left[\mathbb{E}(a_T V^2/\sqrt{T}) - \mathbb{E}(a_T K/\sqrt{T}) - \mathbb{E}(V/\sqrt{T})\right] + \frac{\alpha-1}{\lambda^2\sqrt{\beta}}\left[\left[\mathbb{E}(a_T W^2/\sqrt{T}) - \mathbb{E}(a_T S/\sqrt{T}) - \mathbb{E}(W/\sqrt{T})\right]\right]$$

e

$$i_{\beta\beta} = \frac{1}{4\beta^2}\left[\mathbb{E}(a_T^2 V^2) - \mathbb{E}(a_T^2 K) - \mathbb{E}(a_T V) + 2 + (\alpha-1)\left[\mathbb{E}(a_T^2 W^2) - \mathbb{E}(a_T^2 S) - \mathbb{E}(a_T W)\right]\right] + \frac{1}{2\lambda\beta^{3/2}}\left[2\mathbb{E}(a_T V^2/\sqrt{T}) - 2\mathbb{E}(a_T K/\sqrt{T}) - \mathbb{E}(V/\sqrt{T})\right] + \frac{1}{\lambda^2\beta}\left[\mathbb{E}(V^2/\sqrt{T}) - \mathbb{E}(K/T) + (\alpha-1)\left[\mathbb{E}(W^2/T) - \mathbb{E}(S/T)\right]\right] - \mathbb{E}(T+\beta)^{-2},$$

onde as esperanças acimas são desenvolvidas por meio de métodos numéricos, pois estas não têm forma fechada.

Capítulo 5

MODELOS DE REGRESSÃO α-POTÊNCIA

5.1 Motivação

A análise de regressão é uma técnica estatística para estudar e modelar a relação entre variáveis, sendo uma das mais utilizada na análise de dados. Também é usual supor distribuição normal para os erros do modelo, a qual em muitos casos é uma suposição não realista com o comportamento dos dados. Neste capítulo estudamos o modelo de regressão linear homocedástico com erros PN e introduzimos o modelo log-linear Birnbaum-Saunders PN. Um dos objetivos da análise de regressão é estimar os parâmetros desconhecidos do modelo; existem vários métodos tais como: mínimos quadrados ordinários, máxima verossimilhança, bayesiano e bootstrap. Para cada modelo estudado nós desenvolvemos a estimação dos parâmetros pelo método de máxima verossimilhaça, encontramos a função escore e a matriz de informação observada e fazemos estudo de influência para os modelos. Além disso, estudamos algumas propriedades estatísticas a través de simulações tipo Monte Carlo e por último para cada modelo estudado realizamos aplicações com dados reais.

5.2 Modelo de Regressão Linear Múltiplo Homocedástico com Erros PN

Estudamos agora o modelo de regressão múltiplo, o qual comumente é representado por

$$y_i = x_i'\boldsymbol{\beta} + \epsilon_i, \tag{5.1}$$

onde y_i é a variável resposta para a i-ésima observação, com $i = 1, 2, \ldots, n$, $x_i' = (1, x_{i1}, \ldots, x_{ip})$ representa os valores de $p+1$ $(p+1 < n)$ variáveis explicativas que são assumidas fixas e conhecidas, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$ é um vetor de parâmetros desconhecidos a serem estimados e os ϵ_i são variáveis aleatórias independentes e igualmente distribuídas. Agora, vamos substituir no modelo (5.1) a suposição usual de normalidade ou normalidade assimétrica dos erros do modelo pela suposição $\epsilon_i \sim PN(0, \eta_e, \alpha)$. Portanto, temos que a função de densidade dos ϵ_i é

$$\varphi(\epsilon_i; \boldsymbol{\beta}, \eta_e, \alpha) = \frac{\alpha}{\eta_e}\phi\left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta_e}\right)\left\{\Phi\left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta_e}\right)\right\}^{\alpha-1}, \quad i = 1, 2, ..., n. \tag{5.2}$$

Então de (5.2) conclui-se que y_i dado x_i', $(y_i|x_i')$, também segue a distribuição PN, assim:

$$y_i|x_i' \sim PN(x_i'\boldsymbol{\beta}, \eta_e, \alpha), \quad i = 1, 2, ..., n. \tag{5.3}$$

Neste caso não estamos considerando média zero para os erros, pois

$$E(\epsilon_i) = \alpha\eta_e \int_0^1 \Phi^{-1}(z) z^{\alpha-1} dz \neq 0.$$

Portanto, $E(y_i) \neq x_i'\boldsymbol{\beta}$, para obtemos a reta de regressão como esperança da variável resposta devemos fazer a seguinte correção ao intercepto: $\beta_0^* = \beta_0 + \mu_\epsilon$, onde $\mu_\epsilon = \mathbb{E}(\epsilon_i)$. Assim,

$$E(y_i) = x_i'\boldsymbol{\beta}^* \text{ sendo } \boldsymbol{\beta}^* = (\beta_0^*, \beta_1, ..., \beta_p)'.$$

5.3 Inferência no Modelo Linear Múltiplo PN

5.3.1 Função de Log-Verossimilhança e Função Escore

Agrupando os y_i no vetor Y de dimensão (n×1) e os x_i' na matriz $\mathbf{X}$ de dimensão (n×(p+1)), temos que a função de verossimilhança de $\boldsymbol{\theta} = (\boldsymbol{\beta}', \eta_e, \alpha)'$ para uma amostra

aleatória de n observações independentes $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)'$, pode ser representada por

$$\ell(\boldsymbol{\theta}; \mathbf{Y}) = n[\log(\alpha) - \log \eta_e] - \frac{1}{2\eta_e^2}(Y - \mathbf{X}\boldsymbol{\beta})'(Y - \mathbf{X}\boldsymbol{\beta}) + (\alpha - 1)\sum_{i=1}^{n} \log\left\{\Phi\left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta_e}\right)\right\},$$

com função escore dada por:

$$U(\boldsymbol{\beta}) = \frac{1}{\eta_e^2}\mathbf{X}'(Y - \mathbf{X}\boldsymbol{\beta}) - \frac{\alpha - 1}{\eta_e}\mathbf{X}'\Lambda_1, \qquad U(\alpha) = n\left(\frac{1}{\alpha} + \overline{u}\right), \tag{5.4}$$

$$U(\eta_e) = -\frac{n}{\eta_e} + \frac{1}{\eta_e^3}(Y - \mathbf{X}\boldsymbol{\beta})'(Y - \mathbf{X}\boldsymbol{\beta}) - \frac{\alpha - 1}{\eta_e^2}\left(Y - \mathbf{X}\boldsymbol{\beta}\right)'\Lambda_1 \tag{5.5}$$

sendo

$$\Lambda_1 = (w_1, ..., w_n)' \qquad \text{e} \qquad u_i = \log\left\{\Phi\left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta_e}\right)\right\},$$

com $w_i = \phi\left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta_e}\right)/\Phi\left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta_e}\right)$, para $i = 1, 2, ..., n$. Depois de algumas manipulações algébricas temos as equações de estimação:

$$\boldsymbol{\beta} = \boldsymbol{\beta}_{MQ} - (\alpha - 1)\eta(\mathbf{X}'\mathbf{X})^{-1}X'\Lambda_1, \qquad \alpha = -\frac{1}{\overline{u}}, \tag{5.6}$$

$$\eta = \frac{(1-\alpha)(Y - \mathbf{X}\boldsymbol{\beta})'\Lambda_1}{2n} + \frac{\sqrt{(1-\alpha)^2(Y - \mathbf{X}\boldsymbol{\beta})'\Lambda_1\Lambda_1'(Y - \mathbf{X}\boldsymbol{\beta}) + 4n(Y - \mathbf{X}\boldsymbol{\beta})'(Y - \mathbf{X}\boldsymbol{\beta})}}{2n}, \tag{5.7}$$

onde $\boldsymbol{\beta}_{MQ} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y$. Portanto, o estimador de máxima verossimilhança do vetor de parâmetros β é igual ao estimador de mínimos quadrados de β mais um termo de ajuste pela assimetria. Não existem soluções explicitas para o problema de maximização da função de log-verossimilhança, assim esta pode ser maximizada numericamente.

Para o caso linear simples, p=1, nós temos o seguinte sistema de equações

$$\beta_1 = \eta_e(\alpha - 1)\frac{S_{xw}}{S_x^2} + \frac{S_{xy}}{S_x^2}, \qquad \beta_0 = -\eta_e(\alpha - 1)\overline{w} + \overline{y} - \beta_1\overline{x}, \qquad \alpha = -\frac{1}{\overline{u}},$$

e

$$\eta_e = \frac{(1-\alpha)(\overline{wy} - \beta_0\overline{w} - \beta_1\overline{wx}) + \sqrt{(1-\alpha)^2(\overline{wy} - \beta_0\overline{w} - \beta_1\overline{wx})^2 + 4n\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2}}{2n}$$

com

$$S_x^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2/n, \quad S_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})/n \ \text{ e } \ S_{wx} = \sum_{i=1}^{n}(w_i - \overline{w})(x_i - \overline{x})/n.$$

$\overline{w} = \sum_{i=1}^{n} w_i/n$, $\overline{u} = \sum_{i=1}^{n} u_i/n$, $\overline{x} = \sum_{i=1}^{n} x_i/n$, $\overline{x^2} = \sum_{i=1}^{n} x_i^2/n$, $\overline{y} = \sum_{i=1}^{n} y_i/n$, $\overline{xy} = \sum_{i=1}^{n} x_i y_i/n$, $\overline{wx} = \sum_{i=1}^{n} w_i x_i/n$ e $\overline{wy} = \sum_{i=1}^{n} w_i y_i/n$.

Para $\alpha = 1$ obtemos o modelo de erros com distribuição normal e os EMV já conhecidos $\hat{\beta}_1 = \frac{S_{xy}}{S_x^2}$, $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}$ e $\hat{\eta}_e = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}$.

Para inicializar o processo de optimização nós podemos tomar como valores iniciais para o vetor $\boldsymbol{\beta}$ e para o parâmetro η_e os obtidos pelo método dos mínimos quadrados. Estes podem ser calculados como segue: para $\epsilon_i^* = \epsilon_i - \mu_\epsilon$, temos que $\mathbb{E}(\epsilon^*) = 0$ e $Var(\epsilon^*) = \eta_e^2\Phi_2(\alpha)$. Logo, minimizando a soma dos quadrados dos erros

$$\sum_{i=1}^{n} \epsilon_i^{*2} = \sum_{i=1}^{n} (y_i - x_i'\boldsymbol{\beta}^*)^2$$

obtemos que os estimadores de mínimos quadrados de $\boldsymbol{\beta}^*$ e η_e são:

$$\hat{\boldsymbol{\beta}}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y \ \text{ e } \ \hat{\eta}_e^2 = \frac{\Phi_2^{-1}(\hat{\alpha})}{n-2}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0^* - \hat{\beta}_1 x_i\right)^2.$$

Enquanto que um valor inicial para α pode ser o obtido quando ajustamos o modelo para os erros sob normalidade.

Agora, para obter a matriz de informação de Fisher, temos que os elementos da matriz de informação observada são:

$$j_{\boldsymbol{\beta}'\boldsymbol{\beta}} = \frac{1}{\eta_e^2}\mathbf{X}'\mathbf{X} + \frac{\alpha-1}{\eta_e^2}\mathbf{X}'\Lambda_2\mathbf{X}, \qquad j_{\eta_e\boldsymbol{\beta}} = \frac{2}{\eta_e^3}\mathbf{X}'(Y - \mathbf{X}\boldsymbol{\beta}) + \frac{\alpha-1}{\eta_e^2}\mathbf{X}'\Lambda_3,$$

$$\begin{aligned} j_{\eta_e\eta_e} = &-\frac{n}{\eta_e^2} + \frac{3}{\eta_e^2}\sum_{i=1}^{n}\left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta_e}\right)^2 - \frac{2(\alpha-1)}{\eta_e^2}\sum_{i=1}^{n}\left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta_e}\right)w_i \\ &+ \frac{\alpha-1}{\eta_e^2}\sum_{i=1}^{n}\left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta_e}\right)^3 w_i + \frac{\alpha-1}{\eta_e^2}\sum_{i=1}^{n}\left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta_e}\right)^2 w_i^2 \end{aligned}$$

$$j_{\alpha\boldsymbol{\beta}} = \frac{1}{\eta_e}\mathbf{X}'\Lambda_1, \qquad j_{\alpha\eta_e} = \frac{1}{\eta_e^2}\left(Y - \mathbf{X}\boldsymbol{\beta}\right)'\Lambda_1, \qquad j_{\alpha\alpha} = n/\alpha^2,$$

onde

$$\Lambda_2 = diag\left\{\left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta_e}\right)w_i + w_i^2\right\}_{i=1,2,\ldots,n}$$

e $\Lambda_3 = (a_1, a_2, \ldots, a_n)'$ com

$$a_i = \left\{ \left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta_e} \right)^2 w_i + \left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta_e} \right) w_i^2 - w_i \right\}_{i=1,2,\ldots,n}.$$

Para o caso linear simples, p=1, denotando-os por $i_{\beta_0\beta_0}$, $i_{\beta_1\beta_0}$, $i_{\eta_e\beta_0}$, $i_{\alpha\beta_0}, \ldots, i_{\eta_e\eta_e}$, $i_{\alpha\alpha}$ e fazendo $a_{jk} = \mathbb{E}\left(W^j Y^k\right)$ para $k = 0, 1, 2, 3$ e $j = 0, 1, 2$, se obtêm as informações esperadas:

$$i_{\beta_0\beta_0} = \left\{ 1 + \frac{\alpha - 1}{\eta_e} \left[a_{11} - a_{10}(\beta_0 + \beta_1\overline{x}) \right] + (\alpha - 1)a_{20} \right\} / \eta_e^2,$$

$$i_{\beta_1\beta_0} = \left\{ \overline{x} + \frac{\alpha - 1}{\eta_e} \left[\overline{x}(a_{11} - \beta_0 a_{10}) - \beta_1 a_{10}\overline{x^2} \right] + (\alpha - 1)a_{20}\overline{x} \right\} / \eta_e^2,$$

$$\begin{aligned} i_{\eta_e\beta_0} = {} & \frac{1 - \alpha}{\eta_e^2} a_{10} + \frac{1}{\eta_e^3} \left[2a_{01} + (\alpha - 1)a_{21} - (2 + (\alpha - 1)a_{20})(\beta_0 + \beta_1\overline{x}) \right] + \\ & + \frac{\alpha - 1}{\eta_e^4} \left\{ a_{12} + a_{10}(\beta_0^2 + \beta_1^2\overline{x^2} + 2\beta_0\beta_1\overline{x}) - 2a_{11}(\beta_0 + \beta_1\overline{x}) \right\}, \end{aligned}$$

$$i_{\beta_1\beta_1} = \left\{ \overline{x^2}(1 + (\alpha - 1)a_{20}) + \frac{\alpha - 1}{\eta_e} \left[a_{11}\overline{x^2} - a_{10}(\beta_0\overline{x^2} + \beta_1\overline{x^3}) \right] \right\} / \eta_e^2,$$

$$\begin{aligned} i_{\eta_e\beta_1} = {} & \frac{1 - \alpha}{\eta_e^2} a_{10}\overline{x} + \frac{1}{\eta_e^2} \left[\overline{x}(2(a_{01} - \beta_0) + (\alpha - 1)(a_{21} - \beta_0 a_{20})) - \beta_1(2 + (\alpha - 1)a_{20})\overline{x^2} \right] \\ & + \frac{\alpha - 1}{\eta_e^4} \left[a_{12}\overline{x} + a_{10}(\beta_0^2\overline{x} + 2\beta_0\beta_1\overline{x^2} + \beta_1^2\overline{x^3}) - 2a_{11}(\beta_0\overline{x} + \beta_1\overline{x^2}) \right], \end{aligned}$$

$$\begin{aligned} i_{\eta_e\eta_e} = {} & -\frac{1}{\eta_e^2} + \frac{1}{\eta_e^4} \left[3a_{02} + (\alpha - 1)a_{22} - 2(\beta_0 + \beta_1\overline{x})(3a_{01} + (\alpha - 1)a_{21}) \right] \\ & + \frac{1}{\eta_e^4}(3 + (\alpha - 1)a_{20})(\beta_0^2 + 2\beta_0\beta_1\overline{x} + \beta_1^2\overline{x^2}) - 2\frac{\alpha - 1}{\eta_e^3}(a_{11} - a_{10}(\beta_0 + \beta_1\overline{x})) \\ & + \frac{\alpha - 1}{\eta_e^5} \left[a_{13} - 3a_{12}(\beta_0 + \beta_1\overline{x}) + 3a_{11}(\beta_0^2 + 2\beta_0\beta_1\overline{x} + \beta_1^2\overline{x^2}) \right] \\ & - \frac{\alpha - 1}{\eta_e^5} a_{10}(\beta_0^3 + \beta_1^3\overline{x^3} + 3\beta_0\beta_1\overline{x^2} + 3\beta_0^2\beta_1\overline{x}), \end{aligned}$$

$$i_{\alpha\beta_0} = a_{10}/\eta_e, \quad i_{\alpha\beta_1} = a_{10}\overline{x}/\eta_e, \quad i_{\alpha\eta_e} = [a_{11} - a_{10}(\beta_0 + \beta_1\overline{x})]/\eta_e^2 \ \text{ e } \ i_{\alpha\alpha} = 1/\alpha^2.$$

Estas expressões podem ser calculadas numéricamente.

5.4 Análise de Influência

A análise de diagnóstico é uma técnica para detectar possíveis observações influentes no ajuste da regressão. Esta técnica foi introduzida por Cook (1977) para modelos normais lineares e posteriormente foi adaptada para distintos tipos de modelos. Cook (1986), propõe o método de influência local que é uma importante ferramenta na análise de influência conjunta das observações nos resultados de um ajuste regressão.

Para não mudar a notação original de Cook (1986) na análise de influência local, nós vamos mudar um pouco nesta seção a notação que temos até o capítulo quatro. Assim, nós vamos chamar de

$$\omega = (w_1, w_2, ..., w_n)'$$

o vetor de perturbação e introduzimos a notação

$$\zeta(\cdot) = \hat{\Lambda}_{1i} = \frac{\phi(\cdot)}{\Phi(\cdot)}.$$

5.4.1 Influência Local

O método de influência local tem como principal objetivo avaliar mudanças nos resultados da análise quando pequenas perturbações são incorporadas ao modelo e/ao aos dados. Se essas perturbações causarem efeitos desproporcionais, pode ser indício de que o modelo está mal ajustado ou que possam existir afastamentos sérios das suposições feitas para o mesmo. Nós agorá vamos aplicar esta técnica no modelo de regressão PN linear. Assim nós vamos usar a log-verossimilhança perturbada, como em Cook (1986), para avaliar a influência local.

A influência da perturbação ω sobre o estimador de máxima verossimilhança pode ser avaliada baseando-se na análise do afastamento pela verossimilhança

$$LD(\omega) = 2\{L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}_\omega)\}.$$

Cook (1986) propus estudar o comportamento local de $LD(\omega)$ ao redor de ω_0, usando a curvatura normal C_l no vetor não perturbado em uma direção unitária tal que $\|l\| = 1$ e

então considerar o gráfico de $LD(\omega_0 + al)$ contra a com $a \in \mathbb{R}$. Esse gráfico é chamado de linha projetada. Cada linha projetada pode ser caracterizada pela curvatura normal $C_l(\boldsymbol{\theta})$ em torno de $a = 0$.

Cook mostra que

$$C_l = 2\left|l'\Delta'\ddot{L}^{-1}\Delta l\right|,$$

com $\|l\| = 1$, onde $\ddot{L}$ é a matriz hessiana e Δ é uma matriz $(p+q)\times$n, que depende do esquema de perturbação usado, cujos elementos são $\Delta_{ij} = \frac{\partial^2 \ell(\boldsymbol{\theta}|\omega)}{\partial\theta_j \partial\omega_i}$, $j = 1, 2, ..., p+q$ e $i = 1, 2, ..., n$, com todas as quantidades avaliadas em $\omega = \omega_0$ e $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

Seja l_{max} a direção de máxima curvatura, a qual é a direção que produz a maior mudança em $\hat{\boldsymbol{\theta}}$. O elemento de maior influência dos dados pode ser identificado pela maior componente do vetor l_{max}, correspondente ao maior autovalor de

$$B = -\Delta'\ddot{L}^{-1}\Delta,$$

(veja Galea et al., 2004). Se o interesse é avaliar a influência parcial de um subconjunto θ_1 de $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$, então a curvatura normal na direção do vetor l é dada por

$$C_l(\boldsymbol{\theta}_1) = 2\left|l'\Delta'(\ddot{L}^{-1} - B_1)\Delta l\right|,$$

com

$$B_1 = \begin{pmatrix} \mathbf{0}_{11} & \mathbf{0}_{12} \\ \mathbf{0}_{21} & \ddot{L}_{22}^{-1} \end{pmatrix}$$

onde $\ddot{L}_{22}^{-1} = \frac{\partial^2 \ell(\boldsymbol{\theta}|\omega)}{\partial\boldsymbol{\theta}_2' \partial\boldsymbol{\theta}_2}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$. O gráfico do autovetor associado ao maior autovalor da matriz $-\Delta'(\ddot{L}^{-1} - B_1)\Delta$ contra o índice das observações pode revelar quais observações estão influenciando $\hat{\boldsymbol{\theta}}_1$. Analogamente, se o interesse está em θ_2 então a curvatura normal na direção do vetor l é dada por

$$C_l(\boldsymbol{\theta}_2) = 2\left|l'\Delta'(\ddot{L}^{-1} - B_2)\Delta l\right|,$$

com

$$B_2 = \begin{pmatrix} \ddot{L}_{11}^{-1} & \mathbf{0}_{12} \\ \mathbf{0}_{21} & \mathbf{0}_{22} \end{pmatrix}$$

onde $\ddot{L}_{11}^{-1} = \frac{\partial^2 \ell(\boldsymbol{\theta}|\omega)}{\partial\boldsymbol{\theta}_1' \partial\boldsymbol{\theta}_1}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$. A influência local das observações em θ_2 pode ser avaliada considerando o gráfico l_{max} para a matriz $-\Delta'(\ddot{L}^{-1} - B_2)\Delta$ contra o índice das observações.

5.4.2 Influência Local Total

A curvatura na direção da i-ésima observação foi sugerida por Lesaffre e Verbeke (1998), ou seja, calcular a curvatura na direção de l_i, em que l_i é um vetor n $\times$ 1 de zeros com um na i-ésima posição. Para Δ_i' denotando a i-ésima linha de Δ, a influência local total do i-ésimo caso é dada por

$$C_i = 2\left|\Delta_i' \ddot{L}^{-1}\Delta_i\right|, \quad i = 1, 2, ..., n.$$

Verbeke e Molenberghs (2000) propoẽm considerar como influentes os casos em que $C_i \geq 2\overline{C}$, sendo $\overline{C} = \frac{1}{n}\sum_{i=1}^{n} C_i$.

5.4.3 A Curvatura Normal Conformada

Poon e Poon (1999), propõem uma segunda alternativa para estudar pontos influentes, eles introduzem a curvatura normal conformada, definida por

$$B_l = -\frac{l'\Delta'(\ddot{L}^{-1})\Delta l}{\sqrt{tr\left[l'\Delta'(\ddot{L}^{-1})\Delta l\right]^2}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\omega=\omega_0},$$

onde $tr(A)$ é o traço da matriz A. Assim, o cálculo de B_l não exige quase nenhum esforço mais do que o cálculo de C_l. Além disso, a curvatura normal conformada goza de varias propriedades interessantes, dentre as quais destacamos:
i)A curvatura normal conformada em qualquer direção a ω_0 é invariante sob reparametrização,
ii) para qualquer direção l, $0 \leq |B_l| \leq 1$, portanto, B_l é uma medida normalizada, o que permitirá a comparação das curvaturas.

Poon e Poon (1999) propuseram que a $i-$ésima observação seja considerada como influente se $M(0)_i = B_{li}$, é maior que o valor

$$\bar{M(0)} + c^* SM(0),$$

onde $\bar{M(0)} = 1/n$ e $SM(0)$ é o erro padrão da amostra $M(0)_s, s = 1, 2, ..., n$ e c^* é uma constante selecionada.

Até o momento, não existe uma regra geral para julgar o quão grande é a influência de

um caso específico dos dados. Além dos pontos de referência estudados aqui, também existem outras propostas comumente consideradas na prática; Poon e Poon (1999) propõem o ponto de referência $2b$, enquanto que Zhu e Lee (2001) propõem para usar $b + 2s$, onde b e s são a média e o erro padrão dos $M(0)_s, s = 1, 2, ..., n$. Por último, Lee e Xu (2004) propõem usar o ponto de referência $b + cs$, onde c é uma constante cujo valor depende da aplicação real. Nós usualmente tomamos $c = 2$ para analisar a influência no vetor $\boldsymbol{\beta}$ e $c = 3$ para o vetor $\theta_2 = (\eta_e, \alpha)'$.

5.4.4 Cálculo das Curvaturas

Seja o modelo de regressão PN onde $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$ com $\boldsymbol{\theta}_1 = \boldsymbol{\beta}$ e $\boldsymbol{\theta}_2 = (\eta_e, \alpha)'$ e

$$\ddot{L}(\theta) = \begin{pmatrix} \ddot{L}_{\boldsymbol{\beta}'\boldsymbol{\beta}} & \ddot{L}_{\boldsymbol{\beta}\boldsymbol{\theta}_2} \\ \ddot{L}_{\boldsymbol{\theta}_2\boldsymbol{\beta}} & \ddot{L}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2} \end{pmatrix}.$$

Nós agora vamos derivar as curvaturas normais para o modelo de regressão PN sob três esquemas de perturbação: 1. Ponderação de casos, 2. perturbação na variável resposta e 3. perturbação na variável explicativa. Em cada caso nós derivamos a matriz Δ sobre a qual se obtêm os autovetores que precisamos conhecer.

Ponderação de Casos

A função de log-verossimilhança perturbada pode ser definida como

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

onde $\omega = (w_1, w_2, ..., w_n)'$ é um vetor de pesos para a contribuição de cada caso à função de verossimilhança e $\omega_0 = (1, 1, ..., 1)'$ é o caso não perturbado, isto é, $\ell(\theta|\omega_0) = \ell(\boldsymbol{\theta})$. Este esquema de perturbação pode ser usado para avaliar quando as observações com distintos pesos pode afetar ou não o estimador de máxima verossimilhança de $\boldsymbol{\theta}$.

Sob este esquema, a função de log-verossimilhança perturbada é

$$\ell(\theta/\omega) \propto (\log(\alpha) - \log(\eta_e)) \sum_{i=1}^{n} \omega_i - \frac{1}{2\eta_e^2} \sum_{i=1}^{n} \omega_i (y_i - x_i'\boldsymbol{\beta})^2 + (\alpha - 1) \sum_{i=1}^{n} \omega_i \log \left\{ \Phi \left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta_e} \right) \right\} \tag{5.8}$$

com $0 \leq \omega_i \leq 1$ e $\omega_0 = (1, ..., 1)'$. A matriz Δ é dada por

$$\Delta = \begin{pmatrix} \Delta_{\boldsymbol{\beta}} \\ \Delta_{\boldsymbol{\theta}_2} \end{pmatrix}$$

em que $\Delta_{\boldsymbol{\beta}}$ é uma matriz de tamanho (p+1)×n e $\Delta_{\boldsymbol{\theta}_2}$ é uma matriz de tamanho 2×n. Neste caso $\Delta_{\boldsymbol{\beta}}$ têm elementos

$$\Delta_{\boldsymbol{\beta}} = \frac{1}{\hat{\eta}_e^2} \mathbf{X}' diag \left\{ (y_i - x_i'\hat{\boldsymbol{\beta}}) - (\hat{\alpha} - 1)\hat{\eta}_e \zeta \left(\frac{\hat{e}_i}{\hat{\eta}_e} \right) \right\}_{i=1,2,...,n},$$

enquanto que

$$\Delta_{\boldsymbol{\theta}_2} = (b_1, b_2, ..., b_n),$$

onde

$$b_i = \left\{ \begin{array}{c} -\frac{1}{\hat{\eta}_e} + \frac{1}{\hat{\eta}_e^3}(y_i - x_i'\hat{\boldsymbol{\beta}})^2 - \frac{\hat{\alpha}-1}{\hat{\eta}_e^2}(y_i - x_i'\hat{\boldsymbol{\beta}})\zeta\left(\frac{\hat{e}_i}{\hat{\eta}_e}\right) \\ \frac{1}{\hat{\alpha}} + \log\left\{\Phi\left(\frac{y_i - x_i'\hat{\boldsymbol{\beta}}}{\hat{\eta}_e}\right)\right\} \end{array} \right\}_{i=1,2,...,n}$$

com $\hat{e}_i = y_i - x_i'\hat{\boldsymbol{\beta}}$ e $\zeta(\cdot) = \hat{\Lambda}_{1i} = \frac{\phi(\cdot)}{\Phi(\cdot)}$, sendo $\hat{\boldsymbol{\beta}}$, $\hat{\eta}_e$ e $\hat{\alpha}$ os estimadores de máxima verossimilhança de $\boldsymbol{\beta}$, η_e e α, respectivamente.

Perturbação na Variável Resposta

Suponha que y_i, apresenta uma perturbação da forma $y_{\omega i} = y_i + \omega_i S_y$ onde S_y é um fator de escala que pode ser estimado como desvio padrão de Y e $\omega_i \in \mathbb{R}$. Assim, o logaritmo da função de verossimilhança perturbada assume a forma,

$$\ell(\boldsymbol{\theta}|\omega) \propto n(\log(\alpha) - \log(\eta_e)) - \frac{1}{2\eta_e^2} \sum_{i=1}^{n} (y_i + \omega_i S_y - x_i'\boldsymbol{\beta})^2 + (\alpha - 1) \sum_{i=1}^{n} \log \left\{ \Phi \left(\frac{y_i + \omega_i S_y - x_i'\boldsymbol{\beta}}{\eta_e} \right) \right\}. \tag{5.9}$$

Os elementos da matriz Δ são

$$\Delta_{\boldsymbol{\beta}} = \frac{S_y}{\hat{\eta}_e^2}\mathbf{X}' diag\left\{1-(\hat{\alpha}-1)\left[\frac{\hat{e}_i}{\hat{\eta}_e}\zeta\left(\frac{\hat{e}_i}{\hat{\eta}_e}\right)+\zeta^2\left(\frac{\hat{e}_i}{\hat{\eta}_e}\right)\right]\right\}_{i=1,2,...,n},$$

enquanto que

$$\Delta_{\boldsymbol{\theta}_2} = (c_1, c_2, ..., c_n),$$

onde

$$c_i = \left\{\begin{array}{c} \frac{S_y}{\hat{\eta}_e^2}\left\{2\frac{\hat{e}_i^2}{\hat{\eta}_e}-(\hat{\alpha}-1)\left[\zeta\left(\frac{\hat{e}_i}{\hat{\eta}_e}\right)-\frac{\hat{e}_i}{\hat{\eta}_e}\left[\frac{\hat{e}_i}{\hat{\eta}_e}\zeta\left(\frac{\hat{e}_i}{\hat{\eta}_e}\right)+\zeta^2\left(\frac{\hat{e}_i}{\hat{\eta}_e}\right)\right]\right]\right\} \\ \frac{S_y}{\hat{\eta}_e}\zeta\left(\frac{\hat{e}_i}{\hat{\eta}_e}\right) \end{array}\right\}_{i=1,2,...,n}.$$

Perturbação na Variável Explicativa

Consideremos agora o caso onde uma variável explicativa, x_q, apresenta uma perturbação aditiva da forma $x_{iqw} = x_{iq} + \omega_i S_q$, onde S_q é um fator de escala que pode ser estimado pelo desvio padrão de x_q e $\omega_i \in \mathbb{R}$, $q \in \{1, 2, ..., p\}$. Assim, o logaritmo da função de verossimilhança perturbada assume a forma

$$\begin{aligned}\ell(\boldsymbol{\theta}|\omega) \propto n(\log(\alpha)-\log(\eta_e)) - \frac{1}{2\eta_e^2}\sum_{i=1}^{n}(y_i - x_i'\boldsymbol{\beta} - \beta_q\omega_i S_q)^2 \\ + (\alpha-1)\sum_{i=1}^{n}\log\left\{\Phi\left(\frac{y_i - x_i'\boldsymbol{\beta} - \beta_q\omega_i S_q}{\eta_e}\right)\right\}. \quad (5.10)\end{aligned}$$

Os elementos da matriz Δ_β são

$$\Delta_{\beta_{ij}} = \begin{cases} -\frac{\hat{\beta}_q S_q}{\hat{\eta}_e^2}x_{ij} - \frac{(\hat{\alpha}-1)\hat{\beta}_q S_q}{\hat{\eta}_e^2}x_{ij}\left[\frac{\hat{e}_i}{\hat{\eta}_e}\zeta\left(\frac{\hat{e}_i}{\hat{\eta}_e}\right)+\zeta^2\left(\frac{\hat{e}_i}{\hat{\eta}_e}\right)\right], & \text{se } j \neq q \\ -\frac{S_q}{\hat{\eta}_e^2}(\hat{\beta}_q x_{iq} - \hat{e}_i) - \frac{(\hat{\alpha}-1)S_q}{\hat{\eta}_e}\zeta\left(\frac{\hat{e}_i}{\hat{\eta}_e}\right) - \frac{(\hat{\alpha}-1)\hat{\beta}_q S_q}{\hat{\eta}_e^2}x_{iq}\left[\frac{\hat{e}_i}{\hat{\eta}_e}\zeta\left(\frac{\hat{e}_i}{\hat{\eta}_e}\right)+\zeta^2\left(\frac{\hat{e}_i}{\hat{\eta}_e}\right)\right], & \text{se } j = q. \end{cases}$$

Por outro lado, nós temos para θ_2

$$\Delta_{\boldsymbol{\theta}_2} = (d_1, d_2, ..., d_n),$$

onde

$$d_i = \left\{\begin{array}{c} 2\frac{\hat{\beta}_q S_q}{\hat{\eta}_e^2}\frac{\hat{e}_i}{\hat{\eta}_e} + \frac{(\hat{\alpha}-1)\hat{\beta}_q S_q}{\hat{\eta}_e^2}\zeta\left(\frac{\hat{e}_i}{\hat{\eta}_e}\right) - \frac{(\hat{\alpha}-1)\hat{\beta}_q S_q}{\hat{\eta}_e}\frac{\hat{e}_i}{\hat{\eta}_e}\left[\frac{\hat{e}_i}{\hat{\eta}_e}\zeta\left(\frac{\hat{e}_i}{\hat{\eta}_e}\right)+\zeta^2\left(\frac{\hat{e}_i}{\hat{\eta}_e}\right)\right] \\ -\frac{\hat{\beta}_q S_q}{\hat{\eta}_e}\zeta\left(\frac{\hat{e}_i}{\hat{\eta}_e}\right) \end{array}\right\}_{i=1,2,...,n.}$$

5.4.5 Análise de Resíduos

Uma ferramenta impostante na validação das suposições feitas para o modelo, é a análise dos resíduos do modelo. Esta técnica também é aplicada na deteção de observações influentes nas estimativas dos parâmetros do modelo. Vamos a considerar dois tipos de resíduos: resíduos componentes do desvio e resíduos tipo martingale (veja Ortega et al., 2003).

Resíduos Componentes do Desvio

O resíduo componente do desvio é definido por (Davison e Gigli, 1989),

$$r_{DC_i} = sinal(y_i - x_i'\hat{\boldsymbol{\beta}})\sqrt{2}[\ell_i(\widetilde{\boldsymbol{\theta}}) - \ell_i(\hat{\boldsymbol{\theta}})]^{\frac{1}{2}},$$

onde $\widetilde{\boldsymbol{\theta}}$ é o estimador de máxima verossimilhança de $\boldsymbol{\theta}$ sob o modelo saturado (com n parâmetros), $\hat{\boldsymbol{\theta}}$ é o estimador de máxima verossimilhança sob o modelo de interesse (com p+3 parâmetros) e $sinal(y_i - x_i'\boldsymbol{\beta})$ significa o sinal de $y_i - x_i'\boldsymbol{\beta}$. Para o modelo de regressão PN o resíduo componente do desvio fica dado por

$$r_{DC_i} = sinal(y_i - x_i'\hat{\boldsymbol{\beta}})\sqrt{2}\left[\left(\frac{y_i - x_i'\hat{\boldsymbol{\beta}}}{\hat{\eta}_e}\right)^2 - (\hat{\alpha} - 1)\log\left\{2\Phi\left(\frac{y_i - x_i'\hat{\boldsymbol{\beta}}}{\hat{\eta}_e}\right)\right\}\right]^{1/2}, \; i = 1, 2, ..., n.$$

Resíduos Tipo Martingale

Therneau, Grambsch e Fleminng (1990) introduzem os resíduos componentes do desvio em processos de contagem usando basicamente resíduos martingales. Estos resíduos são assimétricos, assumem valor máximo em +1 e valor mínimo em $-\infty$. Em modelos de sobrevivência, os resíduos tipo martingale são definidos por $r_{M_i} = 1 + \log(R(\hat{y_i}))$, onde R é a função de sobrevida (veja Leiva et al, 2009). Para o modelo de regressão PN temos que

$$r_{M_i} = 1 + \log\left(1 - \left\{\Phi\left(\frac{y_i - x_i'\hat{\boldsymbol{\beta}}}{\hat{\eta}_e}\right)\right\}^{\hat{\alpha}}\right), \quad i = 1, 2, ..., n.$$

Motivado pelas aplicações dos modelos lineares generalizados, Therneau, Grambsch e Fleminng (1990) propõem os resíduos componentes do desvio como uma transformação

dos resíduos tipo martingale. Para variáveis explicativas não dependendo do tempo, os resíduos componentes do desvio para o modelo de Cook são dados por

$$r_{MD_i} = sinal(r_{M_i}) \left\{-2\left[r_{M_i} + \log(1 - r_{M_i})\right]\right\}^{1/2}, \; i = 1, 2, ..., n.$$

5.4.6 Resíduos Padronizados

Ortega (2001) sugere padronizar os resíduos componentes do desvio r_{DC_i} e os resíduos tipo martingale r_{MD_i} como

$$r^*_{DC_i} = \frac{r_{DC_i}}{\sqrt{1 - GL_{ii}}} \;\text{ e }\; r^*_{MD_i} = \frac{r_{MD_i}}{\sqrt{1 - GL_{ii}}},$$

com GL_{ii} o i-ésimo elemento da diagonal principal da matriz de pontos de alavanca generalizada (Wei, Hu e Fung, 1998) definida por

$$\mathbf{GL}(\boldsymbol{\theta}) = D_{\boldsymbol{\theta}}(-\ddot{L})^{-1}\ddot{L}_{\boldsymbol{\theta}_y},$$

em que $D_{\boldsymbol{\theta}} = \frac{\partial \mu}{\partial \boldsymbol{\theta}'} = (\mathbf{X}, \mathbf{0})$, $\ddot{L}$ é a matriz hessiana e

$$\ddot{L}_{\boldsymbol{\theta}_y} = \begin{pmatrix} \ddot{L}_{\boldsymbol{\beta}_y} \\ \ddot{L}_{\boldsymbol{\theta}_{2y}} \end{pmatrix}.$$

Para o modelo de regressão α-potência temos que

$$\ddot{L}_{\boldsymbol{\beta}_{y_i}} = X' diag\{\nu_i\}$$

com

$$\nu_i = -\frac{1}{\eta_e^2} + \frac{\alpha - 1}{\eta_e^2}\left[\frac{e_i}{\eta_e}\zeta\left(\frac{e_i}{\eta_e}\right) + \zeta^2\left(\frac{e_i}{\eta_e}\right)\right],$$

para $i = 1, 2, ..., n$, enquanto que

$$\ddot{L}_{\boldsymbol{\theta}_{2y_i}} = \begin{pmatrix} \frac{2}{\eta^3}e_i - \frac{\alpha-1}{\eta_e^2}\zeta\left(\frac{e_i}{\eta_e}\right) + \frac{\alpha-1}{\eta_e^2}\frac{e_i}{\eta_e}\left[\frac{e_i}{\eta_e}\zeta\left(\frac{e_i}{\eta_e}\right) + \zeta^2\left(\frac{e_i}{\eta_e}\right)\right] \\ \frac{1}{\eta_e}\zeta\left(\frac{e_i}{\eta_e}\right) \end{pmatrix},$$

para $i = 1, 2, ..., n$, com todas as expressões avaliadas em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e onde os e_i são os resíduos como definidos anteriormente.

5.5 Testes de Hipóteses

Desejamos testar se o conjunto de dados apresenta evidência significativa para o uso da distribuição PN, no lugar da distribuição normal. Então podemos testar $H_0: \ \alpha = 1$ contra $H_1: \ \alpha \neq 1$. Para testar H_0 se considera o teste de razão de verossimilhanças (RV), assim para $\hat{\alpha}$ o estimador de máxima verossimilhança de α, o qual é finito, a estatística do teste de RV fica dada por:

$$\begin{aligned}\xi_{RV} &= 2\{\ell(\hat{\alpha};\hat{e}) - \ell(\alpha_0;\hat{e}_0)\} \\ &= 2\left[n\log(\hat{\alpha}) + n\log\left(\frac{\hat{\eta}_0}{\hat{\eta}_e}\right) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\hat{e}_i^2}{\hat{\eta}_e^2} - \frac{\hat{e}_{i0}^2}{\hat{\eta}_0^2}\right) + (\hat{\alpha}-1)\sum_{i=1}^{n}\log(\Phi(\hat{e}_i))\right], \quad (5.11)\end{aligned}$$

onde $\alpha_0 = 1$ e $\hat{\eta}_0$, $\hat{e}_{i0}$ são estimativas sob H_0. Sob a hipótese nula e condições de regularidades, assintoticamente tem-se que (veja Sen e Senger, 1993),

$$\xi_{RV} \to \chi_1^2.$$

Similarmente, para testar $H_0: \ \alpha = \alpha_0$ contra $H_1: \ \alpha \neq \alpha_0$ a estatística do teste de RV é dada por

$$\begin{aligned}\xi_{RV} &= 2\{\ell(\hat{\alpha};\hat{e}) - \ell(\alpha_0;\hat{e})\} \\ &= 2n\log\left(\frac{\hat{\alpha}}{\alpha_0}\right) + 2n\log\left(\frac{\hat{\eta}_0}{\hat{\eta}_e}\right) - \sum_{i=1}^{n}\left(\frac{\hat{e}_i^2}{\hat{\eta}_e^2} - \frac{\hat{e}_{i0}^2}{\hat{\eta}_0^2}\right) + 2(\hat{\alpha}-1)\sum_{i=1}^{n}\log(\Phi(\hat{e}_i)) \\ &\quad - 2(\alpha_0 - 1)\sum_{i=1}^{n}\log(\Phi(\hat{e}_{i0})). \quad (5.12)\end{aligned}$$

Testes de hipóteses para o parâmetro $\boldsymbol{\beta}$ do tipo

$$H_0: \ \boldsymbol{\beta} = \boldsymbol{\beta}^0 \ \text{ contra } \ H_1: \ \boldsymbol{\beta} \neq \boldsymbol{\beta}^0,$$

se pode definir da mesma forma como aconteceu com o parâmetro α.

Método de Seleção de Modelos AIC

Para a seleção do modelo de regressão existem vários procedimentos, embora nenhum deles seja consistente (veja Paula, 2004). Entre estes temos: forward, backward, stepwise

e o critério de informação de Akaike (AIC) (vide, Neter et al., 1996), além de outros que precisam de métodos computacionais intensivos .

O método AIC foi proposto por Akaike (1974) e é um processo de minimização que não envolve testes estatísticos. A ideia é escolher um modelo que ajuste bem e tenha um número reduzido de parâmetros. A proposta do método é baseada em encontrar o modelo com menor valor para a função $-2\ell(\hat{\boldsymbol{\beta}})+2p$, onde p denota o número de parâmetros livres que são estimados no modelo, pois o máximo do logaritmo da função de verossimilhançã, $\ell(\beta)$, aumenta com o aumento do número de parâmetros do modelo. Para o modelo de regressão com erros PN, esta função é dada por:

$$AIC = \log\left(\frac{\hat{\alpha}}{\hat{\eta}_e}\right)^n - \frac{(Y-\mathbf{X}\hat{\boldsymbol{\beta}})'(Y-\mathbf{X}\hat{\boldsymbol{\beta}})}{2\hat{\eta}_e^2} + \sum_{i=1}^{n}\log\left\{\Phi\left(\frac{y_i - x_i'\hat{\boldsymbol{\beta}}}{\hat{\eta}_e}\right)\right\}^{\hat{\alpha}-1} + 2k. \quad (5.13)$$

5.6 Simulação

Com o objetivo de pesquisar o comportamento dos estimadores de mínimos quadrados $\widetilde{\beta}_0$, $\widetilde{\beta}_1$ e $\widetilde{\eta}$; e de máxima verossimilança $\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\eta}$, dos parâmetros β_0, β_1 e η, se fez um estudo de simulação Monte Carlo com 3000 repeticões para os valores $n = 40,\ 80,\ 120$ e $\alpha = 0.75,\ 1.25,\ 1.75,\ 2.25$. Sem perda de generalidade toma-se $\eta = 1$. Consideramos uma única variável aleatória X com distribuição uniforme U(0,1). Os valores dos parâmetros do modelo são $\beta_0 = 1.5$ e $\beta_1 = -2.5$. Os erros ϵ_i são gerados tal que $\epsilon_i \sim PN(0, \eta, \alpha)$. Foram calculados medidas de qualidades para a estimação pontual como: viés, viés relativo(VR), definido como (viés/valor verdadeiro do parâmetro) e a raiz do erro quadrático médio ($\sqrt{EQM}$). Os resultados foram obtidos mediante a livraría bbmle do pacote estatístico R.

Os resultados mostram, (veja Tabela $A.6$), que o viés e a $\sqrt{EQM}$ dos estimadores de mínimos quadrados e máxima verossimilhança dos parâmetro η_e, β_0 e β_1 decrescem quando o tamanho da amostra é maior, o qual garante a propriedade de consistência assintótica dos estimadores. Também temos que quando o parâmetro α cresce, então o viés dos estimadores de máxima verossimilhança de η_e e $\hat{\beta}_0$ é maior, isto é natural pois o estimador do parâmetro β_0 é uma função de α, como já foi estudado anteriormente.

Assim mesmo, o viés dos estimadores de máxima verossimilhança é maior para tamanhos amostrais pequenos o quando os valores do parâmetro de assimetria são maiores. Nesta situação uma correção dos EMV de η_e, β_0, e β_1, por exemplo Bootstrap ou Jackknife, pode ser feita quando se suspeita que o grau de assimetria é grande (> 2), ou o tamanho da amostra é pequeno.

5.6.1 Ilustração

A seguinte ilustração está baseada nos dados dos atletas australianos, os quais estão incluídos no pacote SN do programa R, disponíveis em http://azzalini.stat.unipd.it/SN/. O modelo considerado é

$$Bfat_i = \beta_0 + \beta_1 Wt_i + \beta_2 sex_i + \epsilon_i, \quad i = 1, 2, ..., 202,$$

onde $Bfat_i$ é a porcentagem de gordura corporal do i-ésimo sujeito, as covariáveis Wt_i e sex_i são o peso e o sexo, respectivamente, do i-ésimo sujeito; a variável sexo esta dicotomizada em 1 para homem e zero para mulher. Uma análise dos resíduos do modelo sob normalidade, forneceu evidência que estes não seguem uma distribuição normal, veja Tabela 5.1 e Figura 5.1, gráficos (a) e (b). O modelo $\epsilon_i \sim PN(0, 4.2583, 2.3336)$ ajusta-se melhor à distribuição dos resíduos. Para uma justificação mais enfática da aplicação do modelo PN nós consideramos o teste paramétrico para normalidade

$$H_0 : \alpha = 1 \text{ contra } H_1 : \alpha \neq 1,$$

mediante a estatística do razão de verossimilhanças

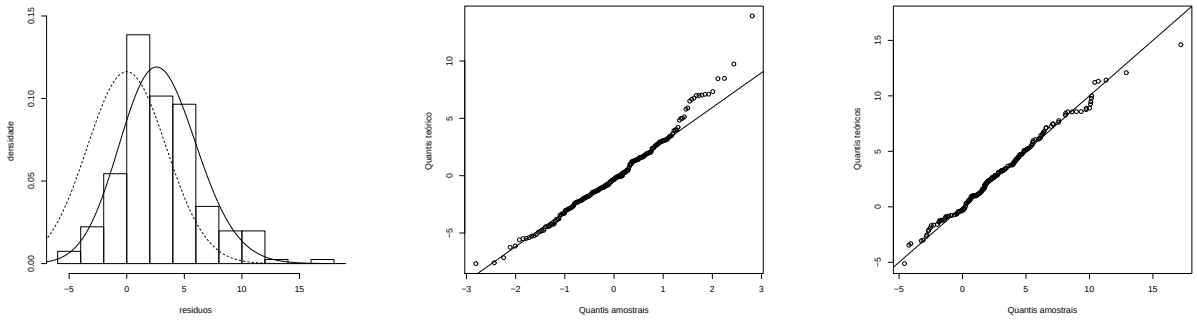
$$\Lambda = \frac{\ell_N(\hat{\boldsymbol{\theta}})}{\ell_{PN}(\hat{\boldsymbol{\theta}})},$$

para o qual o valor observado da estatística de teste é $-2\log(\Lambda) = 5.4566$, assim, $Prob(\chi^2_1 > 5.4566) < 0.05$. Portanto, o modelo de regressão linear proposto com erros PN ajusta melhor os dados que o modelo de regressão com a suposição de distribuição normal para os resíduos, como se pode observar no histograma e os gráficos Q-Q "plot"' da Figura 5.1. O gráfico Q-Q "plot"para o modelo PN foi feito da mesma forma aos

Tabela 5.1 *Estatísticas descritivas para os resíduos sob normalidade*

n	Média	Variância	$\sqrt{b_1}$	b_2
202	0.0050	11.8431	0.6030	3.9321

do modelo PSN. A Tabela 5.2 mostra as estimativas dos parâmetros dos dois modelos. Logo, nós concluímos que, para o modelo de regressão linear PN, a % de gordura corporal depende do peso e do sexo do sujeito.



(a) (b) (c)

Figura 5.1 *(a) Histograma e modelos ajustados dos resíduos,* $N(0, 11.7841)$ *(linha tracejada),* $PN(0, 4.2583, 2.3336)$ *(linha continua). Q-Q "plot"dos resíduos sob: (b) normalidade e (c) PN.*

Nós agora vamos verificar a existência de observações extremas e pontos de alta alavanca, bem como verificar algumas suposições do modelo. Para este objetivo nós vamos usar os resíduos componentes do desvio para o modelo de regressão PN. Nas Figura 5.2 a) e b) respectivamente, se mostram os resíduos do modelo e o gráfico de pontos de alavanca generalizado. O gráfico dos resíduos não mostra nenhuma tendência sistemática mas se observam algumas observações extremas tais como #26, #53, #56, #63, #69, #113 e #160, as primeiras quatro observações correspondem a mulheres com níveis de Wt acima da média das mulheres, enquanto que as observações #113 e #160 correspondem a homens com altos níveis de Wt, também muito acima da média dos homens. O gráfico de pontos de alavanca mostra os pontos #11, #121, #162, #160 e #163 como de alta alavanca; a primeira observação corresponde a uma mulher com o mais alto nível de Wt, enquanto

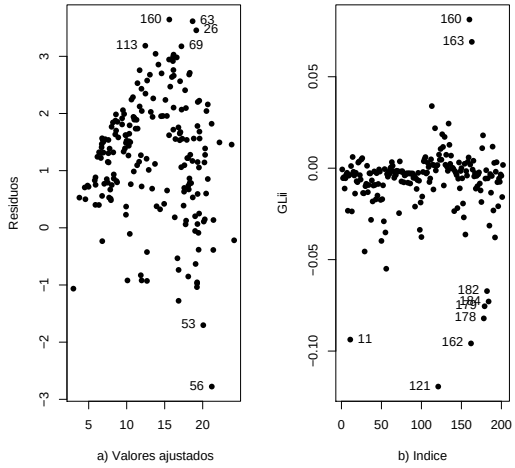


Figura 5.2 *Gráficos a) resíduos r_{DC_i} contra os valores ajustados e b) medida de alavanca generalizada contra os índices das observações.*

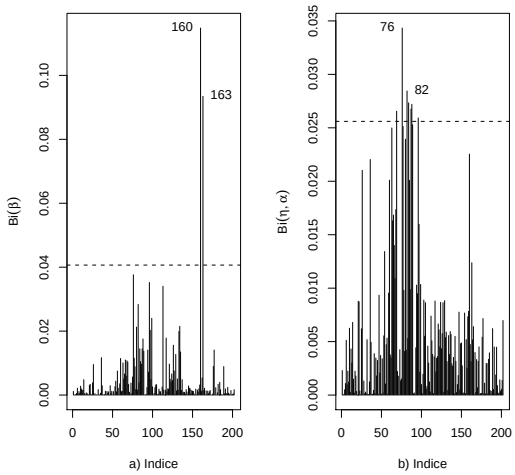


Figura 5.3 *Gráficos a) $B_{\mathbf{e}_i}(\boldsymbol{\beta})$ e b) $B_{\mathbf{e}_i}(\eta, \alpha)$ contra o índice das observações sob o esquema de perturbação de casos.*

Tabela 5.2 *Estimativas (erro padrão) para os parâmetros dos modelos de regressão normal e PN.*

Modelo linear sob normalidade		Modelo linear PN	
Parâmetro	**estimativa**	**Parâmetro**	**estimativa**
$Log-lik$	-535.7707	$Log-lik$	-533.0424
β_0	1.6217(1.4395)	β_0	0.1250(2.0658)
		β_0^*	2.9998(2.1814)
β_1	0.2409(0.0207)	β_1	0.2195(0.0432)
β_2	-12.2563(0.5768)	β_2	-11.8216(1.0682)
σ	3.4328(0.1708)	η	4.2583(0.1244)
		α	2.3336(0.1352)

que as três últimas observações correspondem a homens com níveis de Wt entre os mais altos desse grupo.

Assim, procedendo como em Poon e Poon (1999), considerando ponderação de casos nós vamos identificar as observações influentes do conjunto de dados em estudo, calculando a quantidade $\bar{M}(0)+c^*SM(0)$, obtida a partir da curvatura normal conformada B_l. O valor $SM(0)$ é obtido respectivamente para cada caso na estimação de $\boldsymbol{\beta}$ e $\boldsymbol{\theta}_2$. Para o vetor $\boldsymbol{\beta}$ tomamos $c=2$ enquanto que para o vetor $\boldsymbol{\theta}_2$ o valor $c=3$.

A Figura 5.3 mostra o gráfico da curvatura normal conformada para os dois parâmetros em questão. Dessa forma nós selecionamos os possíveis pontos de alta influência na estimação dos parâmetros do modelo de regressão linear PN. Notamos na Figura 5.3 -a), que novamente as observações #160 e #163 se destacam como as mais influentes; enquanto na Figura 5.3-b) se destaca a observação #76.

Uma análise profunda dos dados levaria analisar todas as possíveis combinações de todas as observações potencialmente influentes, o que seria muito dispendioso. Assim nós vamos ilustrar o impacto dessas observações sobre as estimativas dos parâmetros reajustando o modelo sob algumas situações. Inicialmente nós reajustamos o modelo eliminando cada observação do grupo G_1 com os casos #11, #26, #53, #56, #63, #69, #76, #82,

Tabela 5.3 *Mudança relativa, RC%, para cada parâmetro do modelo de regressão PN.*

Grupo	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\eta}_e$	$\hat{\alpha}$
$\{76\}$	737.25	6.78	2.66	-9.28	-72.58
$\{160\}$	-660.64	-1.32	-2.78	5.69	19.07
$\{163\}$	-712.40	-1.04	-3.56	8.33	17.64
G_0	-446.83	-3.58	-3.96	8.44	17.15
G_1	-1264.50	1.81	-1.87	18.21	18.71
G_2	-909.32	2.80	-3.60	4.09	5.21
G_3	263.32	5.26	0.30	-5.27	-38.95
G_v	-94.88	2.89	3.34	-0.79	-2.53

#96, #113, #121, #160, #162 e #163, depois reajustamos o modelo eliminando todo o grupo de observações G_1; também eliminamos o grupo $G_0 = \{\#76,\ \#160,\ \#163\}$. Posteriormente nós eliminamos o grupo $G_2 = \{\#60,\ \#84,\ \#85\}$, por último eliminamos o grupo $G_3 = \{\#182,\ \#184,\ \#179,\ \#178\}$. Todos os grupos de observações anteriores são considerados potencialmente influentes sobre as estimativas dos parâmetros. Nós também reajustamos o modelo eliminando um conjunto de observações consideradas não influentes as quais foram selecionadas aleatoriamente, $G_v = \{\#4,\ \#30,\ \#71,\ \#108,\ \#202\}$. O impacto foi medido por meio da mudança relativa definida por

$$RC_{\theta_j} = \frac{\hat{\theta}_j - \hat{\theta}_{j(I)}}{\hat{\theta}_j} * 100,$$

onde $\hat{\theta}_j$ denota a estimativa de máxima verossimilhança do parâmetro θ_j com todas as observações e $\hat{\theta}_{j(I)}$ esta denotando a estimativa do mesmo parâmetro obtida depois de remover o "conjunto I"de observações (veja Da silva, 2007). Em cada caso nós não obtemos mudança na significância da estimativa dos parâmetros.

A Tabela 5.3 apresenta a mudança relativa na estimativa dos parâmetros para as observações #76, #160 e #163 mesmo que para os grupos acima mencionados.

Como se pode observar na Tabela 5.3 as observações do grupo G_1 mudam fortemente a estimativa dos parâmetros β_0 e α e em alguns casos também o parâmetro η_e. O grupo

Tabela 5.4 *Estimativas (erro padrão) finais do modelo de regressão PN.*

Log-link	$\hat{\beta}_0$	$\hat{\beta}_0^*$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\eta}_e$	$\hat{\alpha}$
-447.5599	1.9416	4.9237	0.2106	-12.1669	3.4643	2.0355
	(2.5329)	(4.1291)	(0.0363)	(0.8298)	(0.1098)	(0.1050)

G_3 também muda significativamente a estimativa dos parâmetros β_0 e α. O grupo G_2 tem pouca influência na estimativa dos parâmetros do modelo. Ainda os elementos do grupo G_v não são considerados como influentes na estimativa dos parâmetros, eles deixam ver a falta de robustez contra observações extremas na estimativa do parâmetro β_0. Nós estimamos novamente os parâmetros do modelo eliminando em cada grupo as observações consideradas altamente influentes na estimativa dos parâmetros, excetuando a observação #162 e o grupo G_2. Estas estimativas são dadas na Tabela 5.4. Como $\hat{\alpha} \approx 2$, nós concluímos que este modelo se comporta da mesma forma que o modelo de regressão com erros *skew*-normal e parâmetro de assimetria igual a um.

Ainda nós escolhermos este modelo como final, com o exposto acima, notamos que a estimativa de máxima verossimilhança de β_0 apresenta falta de robustez contra observações extremas o qual pode indicar que a suposição de distribuição PN para os erros pode não ser adequada.

5.6.2 Modelo de Regressão Não Linear α-Potência

A partir do modelo de regressão linear, nós podemos definir um modelo mais geral. Assim, o modelo de regressão não linear AP é definido como

$$y_i = f(\boldsymbol{\beta}, x_i) + \epsilon_i, \quad i = 1, 2, ..., n,$$

onde y_i é a variável resposta, f é uma função injetiva contínua e duas vezes diferenciável com respeito o vetor de parâmetro $\boldsymbol{\beta}$, x_i é um vetor de valores de variáveis explicativas e ϵ_i são variáveis aleatórias independentes e identicamente distribuídas $AP(0, \eta, \alpha)$, com $\mu_\epsilon = \alpha\eta \int_0^1 z^{\alpha-1} F^{-1}(z) dz$. Assim, nós temos que $\mathbb{E}(Y_i) = f(\boldsymbol{\beta}, x_i) + \mu_\epsilon$, o qual leva fazer

correções na estimativa do intercepto. Além disso,

$$Y_i \sim AP(f(\boldsymbol{\beta}, x_i), \eta, \alpha), \qquad i = 1, 2, ..., n.$$

No caso da distribuição PN temos a função de densidade

$$f(y_i; \boldsymbol{\beta}, \eta, \alpha) = \frac{\alpha}{\eta}\phi\left(\frac{y_i - f(\boldsymbol{\beta}, x_i)}{\eta}\right)\left\{\Phi\left(\frac{y_i - f(\boldsymbol{\beta}, x_i)}{\eta}\right)\right\}^{\alpha-1} \tag{5.14}$$

a qual denotamos por $Y_i \sim PN(f(\boldsymbol{\beta}, x_i), \eta, \alpha)$. A função de log-verossimilhança, para uma amostra aleatória $\mathbf{Y} = (Y_1, ..., Y_n)'$ dado o vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\beta}', \eta, \alpha)'$ onde Y_i tem distribuição $PN(f(\boldsymbol{\beta}, x_i), \eta, \alpha)$, é dada por

$$\ell(\boldsymbol{\theta}; \mathbf{Y}) = n[\log(\alpha) - \log\eta] - \frac{1}{2\eta^2}\sum_{i=1}^{n}(y_i - f(\boldsymbol{\beta}, x_i))^2 + (\alpha - 1)\sum_{i=1}^{n}\log\left\{\Phi\left(\frac{y_i - f(\boldsymbol{\beta}, x_i)}{\eta}\right)\right\}.$$

A função escore é dada por $U(\boldsymbol{\theta}) = (U(\boldsymbol{\beta}), U(\eta), U(\alpha))'$ onde

$$U(\beta_j) = \frac{1}{\eta^2}\sum_{i=1}^{n}(y_i - f(\boldsymbol{\beta}, x_i))\frac{\partial f(\boldsymbol{\beta}, x_i)}{\partial \beta_j} - \frac{\alpha - 1}{\eta}\sum_{i=1}^{n} w_i \frac{\partial f(\boldsymbol{\beta}, x_i)}{\partial \beta_j}, \qquad U(\alpha) = n\left(\frac{1}{\alpha} + \overline{u}\right),$$

$$U(\eta) = -\frac{n}{\eta} + \frac{1}{\eta^3}\sum_{i=1}^{n}(y_i - f(\boldsymbol{\beta}, x_i))^2 - \frac{\alpha - 1}{\eta^2}\sum_{i=1}^{n}(y_i - f(\boldsymbol{\beta}, x_i))w_i,$$

com $u_i = \log\left\{\Phi\left(\frac{y_i - f(\boldsymbol{\beta}, x_i)}{\eta}\right)\right\}$ e $w_i = \frac{\phi\left(\frac{y_i - f(\boldsymbol{\beta}, x_i)}{\eta}\right)}{\Phi\left(\frac{y_i - f(\boldsymbol{\beta}, x_i)}{\eta}\right)}$. A matriz de informação observada têm elementos

$$\begin{aligned} j_{\beta_k\beta_j} = & \frac{1}{\eta^2}\sum_{i=1}^{n}\left[(\alpha - 1)\left(\frac{y_i - f(\boldsymbol{\beta}, x_i)}{\eta}w_i + w_i^2\right) + 1\right]\frac{\partial f(\boldsymbol{\beta}, x_i)}{\partial \beta_k}\frac{\partial f(\boldsymbol{\beta}, x_i)}{\partial \beta_k} + \\ & \frac{1}{\eta}\sum_{i=1}^{n}\left[-\frac{y_i - f(\boldsymbol{\beta}, x_i)}{\eta} + (\alpha - 1)w_i\right]\frac{\partial^2 f(\boldsymbol{\beta}, x_i)}{\partial \beta_k \partial \beta_j}, \end{aligned}$$

$$\begin{aligned} j_{\eta\beta_j} = & \frac{\alpha - 1}{\eta^2}\sum_{i=1}^{n}\left\{-w_i + \frac{y_i - f(\boldsymbol{\beta}, x_i)}{\eta}\left[\frac{y_i - f(\boldsymbol{\beta}, x_i)}{\eta}w_i + w_i^2\right]\frac{\partial f(\boldsymbol{\beta}, x_i)}{\partial \beta_j}\right\} + \\ & \frac{2}{\eta^2}\sum_{i=1}^{n}\frac{y_i - f(\boldsymbol{\beta}, x_i)}{\eta}\frac{\partial f(\boldsymbol{\beta}, x_i)}{\partial \beta_j}, \end{aligned}$$

$$j_{\alpha\beta_j} = \frac{1}{\eta}\sum_{i=1}^{n} w_i \frac{\partial f(\boldsymbol{\beta}, x_i)}{\partial \beta_j}, \qquad j_{\alpha\eta} = \frac{1}{\eta}\sum_{i=1}^{n}\left(\frac{y_i - f(\boldsymbol{\beta}, x_i)}{\eta}\right) w_i, \qquad j_{\alpha\alpha} = n/\alpha^2,$$

$$\begin{aligned} j_{\eta\eta} = &-\frac{n}{\eta^2} + \frac{3}{\eta^2}\sum_{i=1}^{n}\left(\frac{y_i - f(\boldsymbol{\beta}, x_i)}{\eta}\right)^2 - \frac{2(\alpha-1)}{\eta^2}\sum_{i=1}^{n}\left(\frac{y_i - f(\boldsymbol{\beta}, x_i)}{\eta}\right) w_i \\ &+ \frac{\alpha-1}{\eta^2}\sum_{i=1}^{n}\left(\frac{y_i - f(\boldsymbol{\beta}, x_i)}{\eta}\right)^3 w_i + \frac{\alpha-1}{\eta^2}\sum_{i=1}^{n}\left(\frac{y_i - f(\boldsymbol{\beta}, x_i)}{\eta}\right)^2 w_i^2. \end{aligned}$$

Assim, o estimador de máxima verossimilhança do vetor de parâmetros, $\boldsymbol{\theta}$ pode ser obtido por métodos numéricos iterativos.

5.6.3 Modelo de Regressão Autoregressivo Não Linear PN

Agora consideramos a extensão do modelo de regressão não linear com erros AR(1), para o caso da distribuição PN. Assim, o modelo de regressão PN não linear com erros AR(1) ou modelo auto-regressivo não linear PN, é da forma

$$y_i = f(\boldsymbol{\beta}, x_i) + \epsilon_i, \quad \text{com} \quad \epsilon_i = \rho\epsilon_{i-1} + a_i, \quad i = 1, 2, \cdots, n, \tag{5.15}$$

onde os y_i são as respostas observadas, os x_i são vetores não estocásticos conhecidos, ρ é um coeficiente auto-regressivo desconhecido, com $|\rho| < 1$, β é um vetor de parâmetros desconhecidos p-dimensional, f é uma função contínua e duas vezes diferenciável com respeito a $\boldsymbol{\beta}$, os a_i são variáveis aleatórias independentes com $a_i \sim PN(0, \eta^2, \alpha)$ e $\epsilon_0 = 0$. A esperança da variável resposta para $i = 1, 2, ..., n$, é dada por

$$\mathbb{E}(Y_i) = f(\beta, x_i) + \mathbb{E}(a_i)\sum_{k=0}^{i-1}\rho^k, \tag{5.16}$$

onde $\mathbb{E}(a_i)$ é a esperança de uma variável aleatória com distribuição $PN(0, \eta^2, \alpha)$.

5.6.4 Estimação por Máxima Verossimilhança

Seja o vetor de parâmetros $\boldsymbol{\theta} = (\rho, \boldsymbol{\beta}', \eta^2, \alpha)'$, então a função de log-verossimilhança para o vetor $\boldsymbol{\theta}$ dada a amostra aleatória de tamanho n, $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)'$, é:

$$\ell_n(\boldsymbol{\theta}; \mathbf{Y}) = n\left\{\log(\alpha) - \log(\eta) - \frac{1}{2}\log(2\pi)\right\} - \sum_{i=1}^{n}\frac{(\epsilon_i - \rho\epsilon_{i-1})^2}{2\eta^2} + (\alpha-1)\sum_{i=1}^{n}\log\{\Phi(z_i)\},$$

com $z_i = \frac{\epsilon_i - \rho\epsilon_{i-1}}{\eta}$. Assim, para $w_i = \frac{\phi(z_i)}{\Phi(z_i)}$, $D_i = -\frac{\partial f(\boldsymbol{\beta}, x_i)}{\partial \boldsymbol{\beta}} + \rho \frac{\partial f(\boldsymbol{\beta}, x_{i-1})}{\partial \boldsymbol{\beta}}$ e $Q_i = -w_i(z_i + w_i)$ para $i = 1, 2, \cdots, n$, a função escore $U_\theta = (U_\rho, U'_{\boldsymbol{\beta}}, U_{\eta^2}, U_\alpha)'$ têm elementos:

$$U(\rho) = \sum_{i=1}^{n} \left[\frac{a_i}{\eta^2} - \frac{\alpha - 1}{\eta} w_i \right] \epsilon_{i-1}, \qquad U(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[\frac{a_i}{\eta^2} - \frac{\alpha - 1}{\eta} w_i \right] D_i,$$

$$U(\eta^2) = \sum_{i=1}^{n} \left[-\frac{1}{2\eta^2} + \frac{a_i^4}{2\eta^4} - \frac{\alpha - 1}{2\eta^3} a_i w_i \right], \qquad U(\alpha) = \frac{n}{\alpha} + \sum_{i=1}^{n} \log\{\Phi(z_i)\},$$

onde $a_i = \epsilon_i - \rho\epsilon_{i-1}$ e $\epsilon_i = y_i - \eta(\boldsymbol{\beta}, x_i)$. Agora, tomando $G_i = -\frac{\partial^2 \eta(\boldsymbol{\beta}, x_i)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} + \rho \frac{\partial^2 \eta(\boldsymbol{\beta}, x_{i-1})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}$, obtemos os elementos da matriz hessiana a partir da qual se pode obter a matriz de informação observada:

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \rho^2} = \frac{1}{\eta^2} \sum_{i=1}^{n} \left[-1 + (\alpha - 1) Q_i \right] \epsilon_{i-1}^2,$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}' \partial \rho} = \sum_{i=1}^{n} \left[[1 - (\alpha - 1) Q_i] \frac{\epsilon_{i-1}}{\eta^2} D_i' - \left[\frac{a_i}{\eta^2} - \frac{1}{\eta} w_i \right] \frac{\partial \eta(\boldsymbol{\beta}, x_{i-1})}{\partial \boldsymbol{\beta}'} \right],$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \eta^2 \partial \rho} = \sum_{i=1}^{n} \left[-\frac{a_i}{\eta^4} + \frac{\alpha - 1}{2\eta^2} \left[\frac{a_i Q_i}{\eta^2} + \frac{w_i}{\eta} \right] \right] \epsilon_{i-1}, \qquad \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha \partial \rho} = -\frac{1}{\eta} \sum_{i=1}^{n} w_i \epsilon_{i-1},$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \sum_{i=1}^{n} \left[\frac{1}{\eta^2} [-1 + (\alpha - 1) Q_i] D_i D_i' - \frac{a_i}{\eta^2} G_i + \frac{\alpha - 1}{\eta} w_i G_i \right],$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \eta^2 \partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \left[\frac{a_i}{\eta^4} D_i - \frac{\alpha - 1}{2\eta^2} \left[\frac{a_i}{\eta^2} Q_i D_i + \frac{1}{\eta} w_i D_i \right] \right],$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha \partial \boldsymbol{\beta}} = \frac{1}{\eta} \sum_{i=1}^{n} w_i D_i, \qquad \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha \partial \eta^2} = -\frac{1}{2\eta^3} \sum_{i=1}^{n} a_i w_i, \qquad \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha^2} = -\frac{n}{\alpha^2}.$$

Assim, o estimador de máxima verossimilhança do vetor de parâmetros, $\boldsymbol{\theta}$, pode ser obtido pela solução das equações de estimação, com solução por métodos numéricos iterativos.

5.6.5 Teste de Escore para a Significância do Coeficiente ρ

Se o coeficiente de correlação auto-regressivo $\rho = 0$ então o modelo (5.15) se reduz ao modelo de regressão não linear PN. Assim, nós devemos verificar esta suposição. Portanto, temos o teste de hipótese, sob a suposição que $\boldsymbol{\beta}$, η^2 e α são parâmetros de ruído,

$$H_0 : \rho = 0 \quad \text{contra} \quad H_1 : \rho \neq 0.$$

A estatística do escore para testar H_0 é (Cox and Hinkley, 1974):

$$SC_1 = [U_\rho^2 J^{\rho\rho}(\theta)]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0}, \tag{5.17}$$

onde $J^{\rho\rho}$ é o bloco de J^{-1} correspondente a ρ e $\hat{\boldsymbol{\theta}}_0$ é o estimador de máxima verossimilhança de $\boldsymbol{\theta}$. Sob H_0 a estatística em (5.17) segue assintoticamente, sob condicões de regularidade, uma distribuição χ_1^2.

5.6.6 Ilustração

Uma ilustração com os dados apresentados em Foong (1999) os quais descrevem o comportamento do óleo de palma, e analisado previamente por Xie et al. (2009) com o modelo não linear com erros correlacionados *skew*-normal, é apresentado a seguir. Normalmente, o óleo de palma é colhido após quatro anos de plantio e o rendimento aumenta vigorosamente até o 10° ano de plantação, em seguida continua em uma fase estável até os 25 anos. Esses dados também foram utilizados primeiramente por Azme et al. (2005), onde são obtidas as estimativas para os parâmetros de dois modelos de crescimento não-linear usando o método iterativo de Marquardt. Neste estudo, eles descobriram que o primeiro posto é o modelo com crescimento logístico, (veja, Ratkowsky 1983), ocupando o segundo lugar o modelo de Gompertz e seguido por Morgan-Mercer-Flodin, Chapman-Richard. Cancho et al. (2008) propôs para analisar este conjunto de dados o modelo *skew*-normal não linear com crescimento logístico. Agora nós vamos analisar o conjunto de dados de Foong mediante o modelo de regressão não linear com erros correlacionados PN com crescimento logístico. Assim, o modelo proposto é

$$y_i = \frac{\beta_1}{1 + \beta_2 \exp(-\beta_3 x_i)} + \epsilon_i \tag{5.18}$$

com $\epsilon_i = \rho\epsilon_{i-1} + a_i, \quad i = 1, 2, \cdots, n,\ a_i \sim PN(0, \eta^2, \alpha)$. Nós vamos implementar os modelos não linear com erros correlacionados sob normalidade (NLCN) e não linear com erros correlacionados PN (NLCPN). Os métodos de seleção de modelos de Akaike (1974), $AIC = -2\hat{\ell}(\cdot) + 2k$ e de Akaike modificado pelo critério de informação Bayesiano, $BIC = -2\hat{\ell}(\cdot) + (\log(n))k$, onde k é o número de parâmetros do modelo em questão, sugerem um modelo não linear com erros assimétricos (Tabela 5.5). Os críterios AIC e BIC estão

Tabela 5.5 *Método AIC para seleção de modelo.*

Estatística	Log-lik	AIC	BIC
Normal	-41.2656	92.5312	97.2534
PN	-39.1004	90.2008	95.8674

a favor do modelo PN com erros correlacionados. O seguinte teste paramétrico compara o modelo não linear com erros correlacionado sob normalidade com o modelo não linear com erros correlacionados PN. $H_0 : \alpha = 1$ contra $H_1 : \alpha \neq 1$, leva à estatística de razão de verossimilhanças $\Lambda = \frac{\ell_{NLCN}(\hat{\boldsymbol{\theta}})}{\ell_{NLCPN}(\hat{\boldsymbol{\theta}})}$, de onde $-2\log(\Lambda) = 4.3304$, que é um valor maior que o valor crítico de 5% da distribuição qui-quadrado, com um grau de liberdade, $\chi^2_{1,5\%} = 3.8414$. Portanto, o modelo não linear com erros correlacionados PN, ajusta melhor os dados de óleo de palma, que o modelo sob normalidade.

As estimativas dos parâmetros, com erros padrão em parenteses, para os modelos NLCN e NLCPN são apresentadas na Tabela 5.6.

Na Figura 5.4, (a), se apresenta o gráfico dos modelos estimados e no gráfico (b), os resíduos sob o modelo PN, $\hat{r}_i$ contra $\hat{r}_{i-1} = \hat{r}(1)$, sob a suposição de que $\rho = 0$; no qual não se observa a existência de autocorrelação nos resíduos do modelo. Assim, nós implementamos o modelo não linear com erros $PN(0, \eta, \alpha)$, (NLPN) cujas estimativas dos parâmetros também são dadas na Tabela 5.6.

Tabela 5.6 *Estimativas (erro padrão) dos parâmetros dos modelos NLCN, NLCPN e NLPN.*

	NLCN	NLCPN	NLPN
Parâmetro	**estimativa**	**estimativa**	**estimativa**
ρ	0.3222(0.2757)	0.2574(0.2114)	—
β_1	37.5699(0.3038)	37.9163(0.4041)	38.8798(0.2485)
β_2	11.4310(0.8327)	17.5880(1.2504)	17.5833(1.7888)
β_3	0.5092(0.0227)	0.6140(0.0135)	0.6079(0.0172)
η^2	5.5559(0.7392)	2.6815(0.3658)	1.2010(0.1550)
α	—	0.7010(0.1564)	0.2547(0.0589)

(a) (b)

Figura 5.4 *(a) Gráfico dos modelos ajustado, NLCN(linha tracejada), NLCPN (linha continua) e NLPN(linha de pontos e tracejada); (b) gráfico de $\hat{r}_i$ contra $\hat{r}_{i-1}$*

5.7 Modelo Log-Linear BSPN

5.7.1 A Distribuição senh-Normal

Desenvolvida por Rieck e Nedelman (1991), é definida através de uma transformação da distribuição normal por meio da variável aleatória $Y = (\sigma)arcsenh(\lambda Z/2) + \gamma$, em que $Z \sim N(0,1)$, $\lambda > 0$ é um parâmetro de forma, $\gamma \in \mathbb{R}$ é um parâmetro de localização e

$\sigma > 0$ é um parâmetro de escala. Sua função de densidade de probabilidade é dada por

$$\varphi(y;\lambda,\gamma,\sigma) = \left(\frac{2}{\lambda\sigma\sqrt{2\pi}}\right)\cosh\left(\frac{y-\gamma}{\sigma}\right)\exp\left[-2\lambda^{-2}senh^2\left(\frac{y-\gamma}{\sigma}\right)\right], \quad y \in \mathbb{R}. \quad (5.19)$$

Usamos a notação $Y \sim SHN(\lambda,\gamma,\sigma)$. Se $Y \sim SHN(\lambda,\gamma,\sigma)$, então $Z = 2\lambda^{-1}senh((Y-\gamma)/\sigma) \sim N(0,1)$ (Rieck e Nedelman, 1991). Segue-se então que a função de distribuição cumulativa de Y é

$$F(y) = \Phi\left[\frac{2}{\lambda}senh\left(\frac{y-\gamma}{\sigma}\right)\right]. \quad (5.20)$$

Rieck e Nedelman (1991) provaram que se $T \sim BS(\lambda,\nu)$, então $Y = \log T \sim SHN(\lambda,\gamma,\sigma=2)$, em que $\gamma = \log(\nu)$. Por esta razão, a distribuição senh-normal é também chamada de distribuição log-Birnbaum-Saunders.

5.7.2 Modelo de Regressão Log-Linear Birnbaum-Saunders

O modelo de regressão log-linear com resposta Birnbaum-Saunders é comumente utilizado para modelar a propagação de um dano acumulado até a ocorrência de um processo de falha. Este pode ser escrito como

$$y_i = x_i'\boldsymbol{\beta} + \epsilon_i, \quad i = 1,2,...,n, \quad (5.21)$$

em que y_i é o logaritmo do tempo de sobrevivência para a i-ésima unidade experimental, $\boldsymbol{\beta} = (\beta_1,\beta_2,...,\beta_p)'$ é o vetor de parâmetros desconhecidos a ser estimado, $x_i = (x_{i1},x_{i2},...,x_{ip})'$ são observações de p variáveis explicativas e ϵ_i são variáveis aleatórias independentes e identicamente distribuídas tais que $\epsilon_i \sim SHN(\lambda,0,2)$.

5.7.3 Distribuição senh-PN

A distribuição senh-PN também é definida através de uma transformação da distribuição PN por meio da variável aleatória $Y = (\eta)arcsenh(\lambda Z/2)+\gamma$ em que $Z \sim PN(0,1,\alpha)$, onde $\lambda,\alpha \in \mathbb{R}^+$ são parâmetros de forma, $\gamma \in \mathbb{R}$ é um parâmetro de localização e $\eta > 0$ é um parâmetro de escala. A função de densidade de probabilidade de Y é dada por

$$\varphi(y;\lambda,\gamma,\eta,\alpha) = \alpha\frac{\frac{2}{\lambda}\cosh\left(\frac{y-\gamma}{\eta}\right)}{\eta}\phi\left(\frac{2}{\lambda}senh\left(\frac{y-\gamma}{\eta}\right)\right)\left\{\Phi\left(\frac{2}{\lambda}senh\left(\frac{y-\gamma}{\eta}\right)\right)\right\}^{\alpha-1} \quad (5.22)$$

Usaremos a notação $Y \sim SPN(\lambda, \gamma, \eta, \alpha)$. A função de distribuição cumulativa de Y é

$$\mathcal{F}_{\Phi}(y) = \left\{ \Phi \left[\frac{2}{\lambda} senh \left(\frac{y-\gamma}{\eta} \right) \right] \right\}^{\alpha}. \tag{5.23}$$

A Figura 5.5 mostra o comportamento da densidade (5.22) para valores $\gamma = 0$ e $\eta = 1$, com a) $\alpha = 0.5$ e $\lambda = 0.75$ (linha de pontos), $\alpha = 1.0$ e $\lambda = 1.5$ (linha tracejada) e $\alpha = 5$ e $\lambda = 2$ (linha continua). b) $\alpha = 0.5$ e $\lambda = 3.0$ (linha de pontos), $\alpha = 1.0$ e $\lambda = 4.5$ (linha tracejada), $\alpha = 5$ e $\lambda = 7.5$ (linha continua).

(a) (b)

Figura 5.5 *Densidade* $\varphi(y; \lambda,\ \gamma,\ \eta,\ \alpha)$, *para* $\gamma = 0,\ \eta = 1,\ \alpha = 0.5,\ 1$ *e* 5. *a)*$\lambda \leq 2$, *b)*$\lambda > 2$.

Da figura pode ser visto que:

1. A variável aleatória Y é assimétrica ao redor do parâmetro de localização γ e unimodal para $\lambda < 2$.

2. É bimodal assimétrica para $\lambda = 3.0$, bimodal simétrica ao redor de zero para $\lambda = 4.5$ e unimodal assimétrica negativa para $\lambda = 7.5$.

3. Sempre se tem curtose maior que para o caso $\alpha = 1$, ou senh-normal.

A moda da função de densidade é a raiz da equação não linear

$$\lambda^2 senh \left(\frac{y-\gamma}{\eta} \right) - 4\cosh^2 \left(\frac{y-\gamma}{\eta} \right) + 2\lambda(\alpha - 1) w(\xi_2) \cosh^2 \left(\frac{y-\gamma}{\eta} \right) = 0.$$

A esperança e a variância da variável aleatória Y são dadas por

$$\mathbb{E}(Y) = \eta w_1(\lambda, \alpha) + \gamma \quad \text{e} \quad V(Y) = \eta^2 w_2(\lambda, \alpha)$$

onde w_1 e w_2 são funções de λ e α, dadas por:

$$w_1(\lambda,\alpha) = \alpha \int_{-\infty}^{\infty} arcsenh\left(\frac{\lambda z}{2}\right)\phi(z)\left\{\Phi(z)\right\}^{\alpha-1}dz,$$

e $w_2(\lambda,\alpha)$ é a variância da variável aleatória $arcsenh(\lambda Z/2)$. Em geral não existem expressões fechadas para w_1 e w_2.

Seja $U = \frac{2(Y-\gamma)}{\lambda\sigma}$, onde $Y \sim SPN(\lambda,\gamma,\eta,\alpha)$. Então,

$$f_U(u) = \alpha\cosh\left(\frac{\lambda u}{2}\right)\phi\left(\frac{2}{\lambda}senh\left(\frac{\lambda u}{2}\right)\right)\left\{\Phi\left(\frac{2}{\lambda}senh\left(\frac{\lambda u}{2}\right)\right)\right\}^{\alpha-1}.$$

Agora, quando $\lambda \to 0$, então $\cosh\left(\frac{\lambda u}{2}\right) \to 1$ e $\frac{2senh\left(\frac{\lambda u}{2}\right)}{\lambda} \to u$, portanto, sendo $\phi(x)$ e $\Phi(x)$ funções continuas e diferenciáveis temos que

$$\phi\left(\frac{2senh\left(\frac{\lambda u}{2}\right)}{\lambda}\right) \to \phi(u) \quad \text{e} \quad \left\{\Phi\left(\frac{2senh\left(\frac{\lambda u}{2}\right)}{\lambda}\right)\right\}^{\alpha-1} \to \left\{\Phi(u)\right\}^{\alpha-1},$$

quando $\lambda \to 0$. Logo, conclui-se que $f_U(u) \xrightarrow{D} \alpha\phi(u)\left\{\Phi(u)\right\}^{\alpha-1}$, quando $\lambda \to 0$. Ou seja, U converge em distribuição para uma distribuição PN padrão.

Teorema 5.7.1. *Se $T \sim BSPN(\lambda,\beta,\alpha)$, então $Y = \log(T) \sim SPN(\lambda,\log(\beta),2,\alpha)$.*

5.7.4 Modelo Log-Linear BSPN

Sejam $T_1, T_2, ...T_n$ variáveis aleatórias independentes, onde $T_i \sim BSPN(\lambda_i,\nu_i,\alpha_i)$. Supomos que a distribuição de T_i é independente de um conjunto de p variáveis explicativas denotadas por $x_i = (x_{i1}, x_{i2}, ..., x_{ip})'$, onde

1. $\nu_i = \exp(x_i^T\boldsymbol{\beta})$ para $i = 1,2,...,n$, onde $\boldsymbol{\beta}' = (\beta_1,\beta_2,...,\beta_p)$ é um vetor de parâmetros desconhecidos a serem estimados.

2. Os parâmetros de forma são independentes da variável explicativa x_i; isto é, $\alpha_i = \alpha$ e $\lambda_i = \lambda$ para $i = 1,2,...,n$.

Agora, dado que $T_i \sim BSPN(\lambda_i,\nu_i,\alpha_i)$ e $cT_i \sim BSPN(\lambda_i,c\nu_i,\alpha_i)$ então T_i pode ser expressa como $T_i = \exp(x_i^T\boldsymbol{\beta})\delta_i$ onde $\delta_i \sim BSPN(\lambda,1,\alpha)$. Supomos agora que $Y_i = \log(T_i)$, então

$$y_i = x_i'\boldsymbol{\beta} + \log(\delta_i) = x_i'\boldsymbol{\beta} + \epsilon_i,$$

com $\epsilon_i \sim SPN(\lambda, 0, 2, \alpha)$, onde $cov(\epsilon_i, \epsilon_j) = 0$ pois $Y_1, Y_2, ..., Y_n$, são variáveis aleatórias independentes. Portanto, como acontece no modelo de regressão linear PN concluímos que $Y_i \sim SPN(\lambda, x_i'\boldsymbol{\beta}, 2, \alpha)$.

Agora, fazendo $\boldsymbol{\beta}_0^* = \boldsymbol{\beta}_0 + 2w_1(\lambda, \alpha)$, temos que $\mathbb{E}(y_i) = x_i^T\boldsymbol{\beta}^*$, portanto um estimador linear não viciado de $\boldsymbol{\beta}^* = (\beta_0^*, \boldsymbol{\beta}_1')'$ é desenvolvido pelo método de mínimos quadrados ordinários, com solução e matriz de covariâncias dadas, respectivamente, por

$$\hat{\boldsymbol{\beta}}^* = (X'X)^{-1}X'Y, \quad \text{e} \quad \Sigma_{\hat{\boldsymbol{\beta}}^*} = 4w_2(\lambda, \alpha)(X'X)^{-1}.$$

Também, um estimador de $w_2(\lambda, \alpha)$ é dado por

$$\hat{w}_2(\lambda, \alpha) = \frac{1}{4\Phi_2(\hat{\alpha})(n-p)} \sum_{i=1}^{n} (y_i - x_i'\hat{\boldsymbol{\beta}}^*)^2.$$

De (5.22) temos que a função de log-verossimilhança do vetor $\boldsymbol{\theta} = (\boldsymbol{\beta}', \lambda, \alpha)'$ para uma amostra aleatória de n observações independentes pode ser representada por

$$\ell(\boldsymbol{\theta}; \mathbf{Y}) = n\log(\alpha) + \sum_{i=1}^{n} \log(\xi_{i1}) - \frac{1}{2}\sum_{i=1}^{n} \xi_{i2}^2 + (\alpha - 1)\sum_{i=1}^{n} \xi_{i3}$$

onde $\xi_{i1} = \frac{2}{\lambda}\cosh\left(\frac{y_i - x_i'\boldsymbol{\beta}}{2}\right)$, $\xi_{i2} = \frac{2}{\lambda} senh\left(\frac{y_i - x_i'\boldsymbol{\beta}}{2}\right)$ e $\xi_{i3} = \log\Phi\left\{\frac{2}{\lambda} senh\left(\frac{y_i - x_i'\boldsymbol{\beta}}{2}\right)\right\}$ para $i = 1, 2, ..., n$. Agora, como $Y_i \sim SPN(\lambda, x_i'\boldsymbol{\beta}, 2, \alpha)$ e $\xi_{i2} = \frac{2}{\lambda} senh\left(\frac{y_i - x_i'\boldsymbol{\beta}}{2}\right)$, então segue-se que

$$\xi_{i2} \sim PN(0, 1, \alpha).$$

Os elementos da função escore são dados por:

$$U(\beta_j) = \frac{1}{2}\sum_{i=1}^{n} x_{ij}\left(\xi_{i1}\xi_{i2} - \frac{\xi_{i2}}{\xi_{i1}}\right) - \frac{\alpha - 1}{2}\sum_{i=1}^{n} x_{ij}w_i\xi_{i1}, \quad j = 1, 2, ..., p,$$

$$U(\lambda) = -\frac{n}{\lambda} + \frac{1}{\lambda}\sum_{i=1}^{n} \xi_{i2}^2 - \frac{\alpha - 1}{\lambda}\sum_{i=1}^{n} w_i\xi_{i2}, \qquad U(\alpha) = \frac{n}{\alpha} + \sum_{i=1}^{n} \xi_{i3}.$$

Os elementos da matriz de informação observada são dados por

$$j_{\beta_j\beta_k} = \frac{1}{\lambda}\sum_{i=1}^{n} x_{ij}x_{ik}\left\{2\xi_{i2}^2 + \frac{4}{\lambda^2} - 1 + \frac{\xi_{i2}^2}{\xi_{i2}^2 + 4/\lambda^2}\right\}$$
$$- \frac{\alpha - 1}{4}\sum_{i=1}^{n} x_{ij}x_{ik}\left\{w_i\xi_{i1}(1 - \xi_{i2}^2) - w_i^2\xi_{i1}^2\right\},$$

$$j_{\lambda\beta_j} = \frac{1}{\lambda}\sum_{i=1}^{n} x_{ij}\xi_{i1}\xi_{i2} - \frac{\alpha-1}{2\lambda}\sum_{i=1}^{n} x_{ij}\xi_{i1}\left\{w_i(1-\xi_{i2}^2) - w_i^2\xi_{i2}\right\},$$

$$j_{\lambda\lambda} = -\frac{n}{\lambda^2} + \frac{3}{\lambda^2}\sum_{i=1}^{n}\xi_{i2}^2 + \frac{\alpha-1}{\lambda^2}\sum_{i=1}^{n}\left\{w_i\xi_{i2}^3 + w_i^2\xi_{i2}^2 - 2w_i\xi_{i2}\right\},$$

$$j_{\alpha\beta_j} = \frac{1}{\lambda}\sum_{i=1}^{n} x_{ij}w_i\xi_{i1}, \qquad j_{\alpha\lambda} = \frac{1}{\lambda}\sum_{i=1}^{n} w_i\xi_{i2}, \qquad j_{\alpha\alpha} = \frac{n}{\alpha^2},$$

Porém, a partir da função escore e da matriz de informação observada podemos obter os estimadores de máxima verossimilhança dos coeficientes de regressão e dos parâmetros α e λ através de métodos numéricos iterativos.

5.8 Influência Local

De forma similar ao que foi descrito para o modelo de regressão PN, nós fazemos agora um análise da influência local para o modelo de regressão log-linear BSPN. As matrizes para o estudo da influência local são obtida para os três casos já estudados no modelo de regressão linear. A matriz hessiana para o vetor $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$ com $\boldsymbol{\theta}_1 = \boldsymbol{\beta}$ e $\boldsymbol{\theta}_2 = (\lambda, \alpha)'$, é dada por:

$$\ddot{L}(\boldsymbol{\theta}) = \begin{pmatrix} \ddot{L}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \ddot{L}_{\boldsymbol{\beta}\boldsymbol{\theta}_2} \\ \ddot{L}_{\boldsymbol{\theta}_2\boldsymbol{\beta}} & \ddot{L}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2} \end{pmatrix},$$

cujos elementos podem ser definidos da mesma forma como no modelo de regressão PN. Nós usamos para o análise de diagnóstico no modelo log-linear BSPN a mesma notação usada no caso do modelo de regressão linear PN.

Ponderação de Casos

Baixo este esquema a matriz Δ_θ têm elementos

$$\Delta_{\boldsymbol{\beta}} = \mathbf{X}' diag\left\{\frac{1}{2}\left(\hat{\xi}_{i1}\hat{\xi}_{i2} - \frac{\hat{\xi}_{i2}}{\hat{\xi}_{i1}}\right) - \frac{\alpha-1}{2}\hat{\xi}_{i1}\zeta(\hat{\xi}_{i2}).\right\}_{i=1,2,\ldots,n},$$

enquanto que

$$\Delta_{\boldsymbol{\theta}_2} = (b_1, b_2, ..., b_n),$$

com

$$b_i = \left\{\begin{array}{c} -\frac{1}{\hat{\lambda}} + \frac{1}{\hat{\lambda}}\hat{\xi}_{i2}^2 - \frac{\hat{\alpha}-1}{\hat{\lambda}}\hat{\xi}_{i2}\zeta(\hat{\xi}_{i2}) \\ \frac{1}{\hat{\alpha}} + \hat{\xi}_{i3} \end{array}\right\}_{i=1,2,\ldots,n}$$

onde $\hat{\xi}_{i1}$ e $\hat{\xi}_{i2}$ são os estimadores de máxima verossimilhança de ξ_{i1} e ξ_{i2}.

Perturbação na Variável Resposta

Os elementos da matriz Δ são

$$\Delta_{\boldsymbol{\beta}} = \frac{S_y}{\hat{\lambda}^2}\mathbf{X}'diag\left\{\left[\cosh(\hat{e}_i) - \frac{\hat{\lambda}^2}{4}sech^2\left(\frac{\hat{e}_i}{2}\right)\right]\right\}_{i=1,2,\ldots,n} - \frac{S_y(\hat{\alpha}-1)}{4}\mathbf{X}^T diag\left\{\left[\hat{\xi}_{i2}\zeta(\hat{\xi}_{i2}) - \hat{\xi}_{i1}^2(\hat{\xi}_{i2}\zeta(\hat{\xi}_{i2}) + \zeta^2(\hat{\xi}_{i2}))\right]\right\}_{i=1,2,\ldots,n},$$

enquanto que

$$\Delta_{\boldsymbol{\theta}_2} = (c_1, c_2, ..., c_n),$$

onde

$$c_i = \left\{\begin{array}{c} \frac{S_y}{\hat{\lambda}}\hat{\xi}_{i1}\hat{\xi}_{i2} - \frac{S_y(\hat{\alpha}-1)}{2\hat{\lambda}}\left[\hat{\xi}_{i1}\zeta(\hat{\xi}_{i2}) - \hat{\xi}_{i1}\hat{\xi}_{i2}(\hat{\xi}_{i2}\zeta(\hat{\xi}_{i2}) + \zeta^2(\hat{\xi}_{i2}))\right] \\ \frac{S_y}{2}\hat{\xi}_{i1}\zeta(\hat{\xi}_{i2}) \end{array}\right\}_{i=1,2,\ldots,n}$$

com $\hat{e}_i = y_i - x_i'\hat{\boldsymbol{\beta}}$ e $\zeta(\cdot) = \frac{\phi(\cdot)}{\Phi(\cdot)}$.

Perturbação na variável explicativa

Os elementos da matriz $\Delta_{\boldsymbol{\beta}}$ são:

para $j \neq q$

$$\Delta_{\beta_{ij}} = -\frac{S_q\hat{\theta}_q x_{ij}}{\hat{\lambda}^2}\left[\cosh(\hat{e}_i) - \frac{\hat{\lambda}^2}{4}sech^2\left(\frac{\hat{e}_i}{2}\right)\right] + \frac{S_q\hat{\theta}_q(\hat{\alpha}-1)x_{ij}}{4}\left[\hat{\xi}_{i2}\zeta(\hat{\xi}_{i2})\left(1-\hat{\xi}_{i1}^2\right) - \hat{\xi}_{i1}^2\zeta^2(\hat{\xi}_{i2})\right],$$

enquanto para $j = q$ temos

$$\Delta_{\beta_{iq}} = -\frac{S_q\hat{\theta}_q x_{iq}}{\hat{\lambda}^2}\left[\cosh(\hat{e}_i) - \frac{\hat{\lambda}^2}{4}sech^2\left(\frac{\hat{e}_i}{2}\right)\right] + \frac{S_q}{\hat{\lambda}^2}\left[senh(\hat{e}_i) - \frac{\hat{\lambda}^2}{2}\tanh\left(\frac{\hat{e}_i}{2}\right)\right] + \frac{S_q\hat{\theta}_q(\hat{\alpha}-1)x_{iq}}{4}\left[\hat{\xi}_{i2}\zeta(\hat{\xi}_{i2})\left(1-\hat{\xi}_{i1}^2\right) - \hat{\xi}_{i1}^2\zeta^2(\hat{\xi}_{i2})\right] - \frac{S_q(\hat{\alpha}-1)}{2}\hat{\xi}_{i1}\zeta(\hat{\xi}_{i2}).$$

Similarmente, para $\boldsymbol{\theta}_2$

$$\Delta_{\boldsymbol{\theta}_2} = (d_1, d_2, ..., d_n),$$

com

$$d_i = \left\{\begin{array}{c} -\frac{S_q\hat{\theta}_q}{\hat{\lambda}}\hat{\xi}_{i1}\hat{\xi}_{i2} + \frac{S_q\hat{\theta}_q(\hat{\alpha}-1)}{2\hat{\lambda}}\left[2\hat{\xi}_{i1}\zeta(\hat{\xi}_{i2}) - \hat{\xi}_{i1}\hat{\xi}_{i2}(\hat{\xi}_{i2}\zeta(\hat{\xi}_{i2}) + \zeta^2(\hat{\xi}_{i2}))\right] \\ -\frac{S_q\hat{\theta}_q}{2}\hat{\xi}_{i1}\zeta(\hat{\xi}_{i2}) \end{array}\right\}_{i=1,2,\ldots,n}.$$

5.8.1 Análise de Resíduos

Resíduos Componentes do Desvio

Considerando λ e α fixos ou conhecidos temos, os resíduos componentes do desvio são dados por:

$$r_{DC_i} = sinal(\hat{e}_i)\sqrt{2}\left[-\log\left(\cosh\left(\frac{\hat{e}_i}{2}\right)\right) + \frac{1}{2}\hat{\xi}_{i2}^2 - (\hat{\alpha}-1)\log\left\{2\Phi\left(\hat{\xi}_{i2}\right)\right\}\right]^{1/2} \tag{5.24}$$

para $i = 1, 2, ..., n$, onde $\hat{e}_i = y_i - x_i'\hat{\boldsymbol{\beta}}$.

Resíduos Tipo Martingale

Pela definição do resíduo tipo martingale, temos que

$$r_{MD_i} = sinal(r_{M_i})\left\{-2\left[r_{M_i} + \log(1 - r_{M_i})\right]\right\}^{1/2} \tag{5.25}$$

onde

$$r_{M_i} = 1 + \log\left(1 - \left\{\Phi\left(\hat{\xi}_{i2}\right)\right\}^{\hat{\alpha}}\right), \tag{5.26}$$

para $i = 1, 2, ..., n$.

5.8.2 Resíduos Padronizados

Os resíduos padronizados, componentes do desvio e tipo martingale e a matriz de pontos de alavanca generalizada são definidos da mesma forma que para o caso linear PN, onde

$$v_i = -\frac{1}{2}\tanh\left(\frac{\hat{e}_i}{2}\right) + \frac{1}{4}\hat{\xi}_{i1}\hat{\xi}_{i2} - \frac{\hat{\alpha}-1}{4}\left[\hat{\xi}_{i2}\zeta(\hat{\xi}_{i2}) - \hat{\xi}_{i1}^2\left(\hat{\xi}_{i2}\zeta(\hat{\xi}_{i2}) + \zeta^2(\hat{\xi}_{i2})\right)\right]$$

para $i = 1, 2, ..., n$, enquanto que

$$\ddot{L}_{\boldsymbol{\theta}_2 y_i} = \begin{pmatrix} -\frac{\hat{\alpha}-1}{2\hat{\lambda}}\hat{\xi}_{i1}\zeta\left(\hat{\xi}_{i2}\right) + \frac{1}{2\hat{\lambda}}\hat{\xi}_{i1}\hat{\xi}_{i2}\left[2 + (\hat{\alpha}-1)\left(\hat{\xi}_{i2}\zeta\left(\hat{\xi}_{i2}\right) + \zeta^2\left(\hat{\xi}_{i2}\right)\right)\right] \\ \frac{1}{2}\hat{\xi}_{i1}\zeta\left(\hat{\xi}_{i2}\right) \end{pmatrix},$$

para $i = 1, 2, ..., n$.

5.9 Ilustração

Uma aplicação é feita para o conjunto de dados apresentado em Kalbfleisch e Prentice (2002) e disponíveis em http://lib.stat.cmu.ed/datasets. Este conjunto de dados já foram analisados por meio de modelos de regressão usando para os erros distribuições como: exponencial, gama generalizada, log-logística, log-normal e Weibull (veja Lee e Wang, 2003). Agora nós os vamos analisar supondo uma distribuição senh-PN para os erros do modelo. Os dados correspondem 137 observações de homens com câncer de pulmão avançado. O objetivo do estudo é tentar explicar o tempo de sobrevivência (T, em dias) através de uma estrutura de regressão. As variáveis explicativas que nós vamos considerar nesta aplicação foram as seguintes: uma medida de aleatorização do estado do paciente (Karnofsky): 10-30 completamente aleatorizado, 40-60 parcialmente confinado, 70-90 capacidade do paciente de cuidar-se sozinho (x_1); tempo em meses do diagnóstico até a aleatorização (x_2); idade em anos no instante da aleatorização (x_3); tratamento anterior (x_4), dicotomizada em 10 para sim e 0 para não, tipo histológico do tumor squamous, dicotomizada em 1 para sim e 0 para não (x_5).

O tempo de sobrevivência do câncer é função do dano acumulado provocado por vários fatores, essa degradação leva a um processo de fadiga, onde o tempo de sobrevivência pode ser modelado pela distribuição Birnbaum - Saunders. Então, dado que a média, a variância e os coeficientes de assimetria e curtose dos erros (veja Tabela 5.7), são muito distintos do esperado para o modelo SHN, nós propomos o modelo SPN,

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \epsilon_i, \quad i = 1, 2, ..., 137,$$

para analisar os dados do tempo de sobrevivência de pessoas com câncer de pulmão, onde $\epsilon_i \overset{iid}{\sim} SPN(\lambda, 0, 2, \alpha)$ e a resposta $y_i = \log(t_i)$ é o logaritmo natural do tempo de sobrevivência. A regressão é feita sob $\mathbb{E}(\log(T_i))$, com as respectivas correções para o intercepto nos modelos assimétricos. Uma conclusão mais enfática nessa direção pode ser vista considerando o teste paramétrico para a senh-normalidade, $H_0 : \alpha = 1$ contra $H_1 : \alpha \neq 1$, mediante a estatística de razão de verossimilhanças $\Lambda = \frac{\ell_{SHN}(\hat{\boldsymbol{\theta}})}{\ell_{SPN}(\hat{\boldsymbol{\theta}})}$, para o qual

o valor observado da estatística de teste é $-2\log(\Lambda) = 6.4854$, este é um valor muito maior que o valor crítico de 5% da distribuição qui-quadrado, com um grau de liberdade. Portanto, o modelo log-linear BSPN ajusta melhor os dados de tempo de sobrevivencia de Kalbfleisch que o modelo log-linear BS com a suposição de distribuicao SHN para os erros. A Tabela 5.8 mostra as estimativas dos parâmetros dos modelos.

Tabela 5.7 *Estatísticas descritivas para a variável Z*

n	Média	Variância	$\sqrt{b_1}$	b_2
137	0.0644	1.1716	-0.1875	3.3797

Então, ao nível do 5%, as variáveis explicativas x_2, x_4, x_5 não são estatísticamente significativas para o modelo log-linear BS normal, enquanto que as variáveis x_2, x_3, x_4, x_5 não foram significativas para o modelo log-linear BSPN. Ou seja, que para o modelo log-linear clássico, o tempo de sobrevivência não depende da duração da doença até a entrada no estudo clínico, também não é relevante se houve um tratamento anterior e tipo histólogico do

Tabela 5.8 *Estimativas (erro padrão) dos parâmetros dos modelos log-linear BS e log-linear BSPN.*

Modelo log-linear BS		Modelo log-linear BSPN	
Parâmetro	**estimativa**	**Parâmetro**	**estimativa**
$Log-lik$	10.0879	$Log-lik$	13.3306
β_0	0.6356(0.5539)	β_0	0.2883(0.6246)
		β_0^*	2.0955(0.6314)
β_1	0.0392(0.0049)	β_1	0.0325(0.0053)
β_2	-0.0015(0.0094)	β_2	0.0003(0.0088)
β_3	0.0209(0.0079)	β_3	0.0081(0.0091)
β_4	-0.0028(0.0225)	β_4	0.0044(0.0208)
β_5	-0.2850(0.1984)	β_5	-0.3288(0.1914)
λ	1.2811(0.0779)	λ	2.1262(0.5007)
		α	3.2758(1.0770)

tumor squamous, além disso também não depende da idade no instante da aleatorização para o caso do modelo log-linear BSPN.

Depois de o análise dos resíduos e de influência local pela metodologia de Poon e Poon (1999), nós finalmente chegamos ao modelo ajustado com erros assimétricos SPN, $\hat{\alpha} = 4.7585(0.5850)$ e $\hat{\lambda} = 1.8968(0.1258)$,

$$\begin{array}{rcccc} y_i = & 2.4562 & + & 0.0311x_1 & - & 0.7307x_5 \\ & (0.2264) & & (0.0030) & & (0.1313). \end{array}$$

5.9.1 Caso não Linear da Distribuição BSPN

Vamos substituir no modelo log-linear BSPN a suposição $y_i = \mu_i + \epsilon_i$ pela condição $y_i = f(\boldsymbol{\beta}, x_i) + \epsilon_i$ onde $\epsilon_i \sim SPN(\lambda, 0, 2, \alpha)$ para $i = 1, 2, ..., n$, y_i é o logaritmo da i-ésima observação do tempo de vida, $f(\boldsymbol{\beta}, x_i)$ é uma função não linear, com f assumida conhecida, contínua e duas vezes diferenciável com respeito ao vetor de parâmetros desconhecidos $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)'$ e x_i é um vetor de tamanho m$\times$1 de variáveis explicativas conhecidas e associadas com a i-ésima resposta observada y_i.

Suponha agora que $f(\boldsymbol{\beta}, x_i) = \mu_i$ tem uma estrutura não linear. A função de log-verossimilhança para o vetor $\boldsymbol{\theta} = (\boldsymbol{\beta}', \lambda, \alpha)'$ de uma amostra aleatória $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)'$ obtida do modelo

$$y_i = \mu_i + \epsilon_i, \quad \text{com} \quad \epsilon_i \sim SPN(\lambda, 0, 2, \alpha),$$

é dada por

$$\ell(\boldsymbol{\theta}; \mathbf{Y}) = n \log(\alpha) + \sum_{i=1}^{n} \log(\xi_{i1}) - \frac{1}{2} \sum_{i=1}^{n} \xi_{i2}^2 + (\alpha - 1) \sum_{i=1}^{n} \xi_{i3}$$

onde $\xi_{i1} = \xi_{i1}(\boldsymbol{\theta}) = \frac{2}{\lambda} \cosh\left(\frac{y_i - \mu_i}{2}\right)$, $\xi_{i2} = \xi_{i2}(\boldsymbol{\theta}) = \frac{2}{\lambda} senh\left(\frac{y_i - \mu_i}{2}\right)$ e $\xi_{i3} = \xi_{i3}(\boldsymbol{\theta}) = \log \Phi\left\{\frac{2}{\lambda} senh\left(\frac{y_i - \mu_i}{2}\right)\right\}$, para $i = 1, 2, ..., n$. A função $\ell(\boldsymbol{\theta}; \mathbf{Y})$ é assumida ser regular com respeito a $\boldsymbol{\beta}$ e λ, com derivada até de segunda ordem, Cox e Hinkley, (1974). A matriz, de tamanho n$\times$p, $D = D(\boldsymbol{\beta}) = \frac{\partial \mu}{\partial \boldsymbol{\beta}'}$ de derivadas parciais de $\mu = (\mu_1, \mu_2, ..., \mu_n)'$ com respeito $\boldsymbol{\beta}$, é assumida de posto completo, $posto(D) = p$ para todo $\boldsymbol{\beta}$. Os preditores não lineares $x_1, x_2, ..., x_n$ são elementos de uma sequência infinita de vetores m$\times$1 satisfazendo

certas condições de regularidades para as validações assintóticas. Os elementos da função escore são dados por

$$U(\beta_j) = \frac{1}{2}\sum_{i=1}^{n} d_{ij}\left(\xi_{i1}\xi_{i2} - \frac{\xi_{i2}}{\xi_{i1}}\right) - \frac{\alpha-1}{2}\sum_{i=1}^{n} d_{ij}w_i\xi_{i1}, \quad j = 1, 2, ..., p,$$

$$U(\lambda) = -\frac{n}{\lambda} + \frac{1}{\lambda}\sum_{i=1}^{n}\xi_{i2}^2 - \frac{\alpha-1}{\lambda}\sum_{i=1}^{n} w_i\xi_{i2}, \qquad U(\alpha) = \frac{n}{\alpha} + \sum_{i=1}^{n}\xi_{i3},$$

onde d_{ij} são os elementos da matriz D. Então, $U(\boldsymbol{\beta}) = \frac{1}{2}D'S$, onde S é um vetor de tamanho n com elementos $s_i = \xi_{i1}\xi_{i2} - \frac{\xi_{i2}}{\xi_{i1}} - (\alpha-1)w_i\xi_{i1}$.

Os elementos da matriz de informação observada são dados por

$$\begin{aligned} j_{\beta_j\beta_k} = & \frac{1}{2}\sum_{i=1}^{n} g_{ijk}\left(\xi_{i1}\xi_{i2} - \frac{\xi_{i2}}{\xi_{i1}}\right) - \frac{1}{4}\sum_{i=1}^{n} d_{ij}d_{ik}\left\{2\xi_{i2}^2 + \frac{4}{\lambda^2} - 1 + \frac{\xi_{i2}^2}{\xi_{i2}^2 + 4/\lambda^2}\right\} \\ & - \frac{\alpha-1}{2}\sum_{i=1}^{n} g_{ijk}w_i\xi_{i1} + \frac{\alpha-1}{4}\sum_{i=1}^{n} d_{ij}d_{ik}\left\{w_i\xi_{i1}(1-\xi_{i2}^2) - w_i^2\xi_{i1}^2\right\}, \end{aligned}$$

$$j_{\lambda\beta_j} = -\frac{1}{\lambda}\sum_{i=1}^{n} d_{ij}\xi_{i1}\xi_{i2} + \frac{\alpha-1}{2\lambda}\sum_{i=1}^{n} d_{ij}\xi_{i1}\left\{w_i(1-\xi_{i2}^2) - w_i^2\xi_{i2}\right\},$$

$$j_{\lambda\lambda} = -\frac{n}{\lambda^2} + \frac{3}{\lambda^2}\sum_{i=1}^{n}\xi_{i2}^2 + \frac{\alpha-1}{\lambda^2}\sum_{i=1}^{n}\left\{w_i\xi_{i2}^3 + w_i^2\xi_{i2}^2 - 2w_i\xi_{i2}\right\},$$

$$j_{\alpha\lambda} = \frac{1}{\lambda}\sum_{i=1}^{n} w_i\xi_{i2}, \qquad j_{\alpha\beta_j} = -\frac{1}{2}\sum_{i=1}^{n} d_{ij}w_i\xi_{i1}, \qquad j_{\alpha\alpha} = \frac{n}{\alpha^2},$$

onde $g_{ijk} = \frac{\partial^2\mu_i}{\partial\beta_k\partial\beta_j}$. Assim, os estimadores de máxima verossimilhança dos coeficientes de regressão e dos parâmetros α e λ são soluções das equações $U(\beta_j) = 0$ $(j = 1, 2, ..., p)$, $U(\lambda) = 0$ e $U(\alpha) = 0$, cujas soluções são dadas por métodos numéricos iterativos ponderados.

A variância assintótica de $\hat{\mu}_i$ pode ser expressada explicitamente em termos da covariância de $\hat{\boldsymbol{\beta}}$ por

$$Var(\hat{\mu}_i) = tr\{(d_i d_i')Cov(\hat{\boldsymbol{\beta}})\},$$

onde $Cov(\hat{\boldsymbol{\beta}})\} = J_{\boldsymbol{\beta}}^{-1}$ é a matriz de covariâncias assintóticas ou a matriz de informação de Fisher para $\boldsymbol{\beta}$ e d_i são as filas da matriz D.

5.9.2 Ilustração

Nós consideramos os dados de fadiga biaxial reportados por Rieck e Nedelman (1991) sobre a vida de peças metálicas em ciclos até a falha. A variável resposta N é o número de ciclos para falhas e a variável explicativa W é o trabalho por ciclo (mJ/m^3). Os dados de quarenta e seis observações foram tomados da Tabela 1 de Rieck e Nedelman (1991). O modelo linear

$$y_i = \beta_1 + \beta_2 w_i + \epsilon_i,$$

onde $y_i = \log(N_i)$ e $\epsilon_i \sim SHN(\lambda, 0, 2)$ fornece os estimadores de máxima verossimilhança, e seus erros padrão, $\hat{\beta}_1 = 7.9860$ (0.1544), $\hat{\beta}_2 = -0.0405$ (0.0033) e $\hat{\lambda} = 0.5199$ (0.0541). Uma análise descritiva dos resíduos do modelo, Tabela 5.9, mostra que os coeficientes de viés e a curtose são muito distintos dos valores esperados para a variável aleatória senh-normal de parâmetros $(0.5199, 0.9502, 2.1678)$. Para $\epsilon_i \sim SHN(\lambda, 0, 2)$ temos que a variável $Z_i = \frac{2}{\lambda} senh\left(\frac{\epsilon_i}{2}\right) \sim N(0, 1)$. As estatísticas descritivas da variável Z, dadas na Tabela 5.9 mostram um coeficiente de assimetria muito distinto do coeficiente esperado para a distribuição normal. Entretanto, nós encontramos que o modelo padrão $PN(0.7945)$ pode modelar a assimetria da distribuição.

Tabela 5.9 *Estatísticas descritivas da variável Z.*

n	Média	Variância	$\sqrt{b_1}$	b_2
46	-0.0149	0.2583	0.4793	3.0663

Além disso, o diagrama de dispersão na Figura 5.6 sugere uma relação não linear. Rieck e Nedelman propõem o modelo

$$y_i = \beta_1 + \beta_2 \log(w_i) + \epsilon_i,$$

onde $y_i = \log(N_i)$ e $\epsilon_i \sim SHN(\lambda, 0, 2)$ para $i = 1, 2, ..., 46$. Nós encontramos que os estimadores de máxima verossimilhança dos parâmetros são

$$\hat{\beta}_0 = 12.2795(0.3894) \quad \hat{\beta}_1 = -1.6707(0.1084) \text{ e } \hat{\lambda} = 0.4103(0.0428).$$

Assim, nós propomos o modelo de regressão não linear

$$y_i = \beta_1 + \beta_2 \exp(\beta_3/x_i) + \epsilon_i,$$

com $\epsilon_i \sim SPN(\lambda, 0, 2, \alpha)$ para $i = 1, 2, ..., 46$. Aqui $\mathbb{E}(Y_i) = \beta_1^* + \beta_2 \exp(\beta_3/x_i)$ onde $\beta_1^* = \beta_1 + \mathbb{E}(\epsilon_i)$. Os estimadores de máxima verossimilhança, com erros padrão entre parênteses, fornecido pelo método optim(BFGS) do pacote estatístico R, foram $\hat{\beta}_1 = 9.1095$ (0.0891), $\hat{\beta}_1^* = 8.9557$, $\hat{\beta}_2 = -5.1079$ (0.0206), $\hat{\beta}_3 = -21.1760$ (0.2517), $\hat{\lambda} = 0.3845$ (0.0438) e $\hat{\alpha} = 0.8999$ (0.1078).

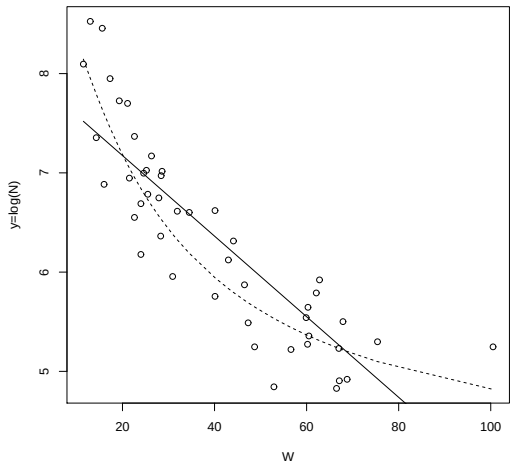


Figura 5.6 *Diagrama de dispersão, modelo linear (linha contínua) com erros SHN e modelo não linear (linha tracejada) com erros SPN para os dados de fadiga biaxial.*

Um teste para comparar o modelo de regressão linear com erros SHN contra o modelo não linear com erros SPN, requer teoria de modelos não encaixados. Portanto, sejam dois modelos não encaixados F_θ com densidade $f(y_i|x_i, \theta)$ e o modelo G_γ com densidade $g(y_i|x_i, \gamma)$.

A estatística, LR, da razão de verossimilhanças para o modelo F_θ contra o modelo G_γ é

$$LR(\hat{\theta}, \hat{\gamma}) \equiv \ell_f(\hat{\theta}) - \ell_g(\hat{\gamma}) = \sum_{i=1}^{n} \log \frac{f(y_i|\mathbf{x}_i, \theta)}{g(y_i|\mathbf{x}_i, \gamma)},$$

a qual não segue uma distribuição χ^2. Vuong (1989) apresentou uma solução para resolver

esse problema. Ele considera o critério de informação de Kullback-Leibler (1951), KLIC, o qual mede a distância entre uma distribuição qualquer e a distribuição verdadeira. Portanto, Vuong faz distinção entre modelos baseando-se nas distâncias entre cada modelo e o verdadeiro processo gerador dos dados, o qual tem densidade $h_0(\mathbf{y}_i, X_i)$, em que a distância é medida usando o critério de informação de Kullback-Leibler. Assim, ele propôs a estatística

$$T_{LR,NN} = \frac{1}{\sqrt{n}} \frac{LR(\hat{\theta}, \hat{\gamma})}{\hat{\omega}^2},$$

em que

$$\hat{\omega}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\log \frac{f(y_i|\mathbf{x}_i, \theta)}{g(y_i|\mathbf{x}_i, \gamma)} \right)^2 - \left(\frac{1}{n} \sum_{i=1}^{n} \left(\log \frac{f(y_i|\mathbf{x}_i, \theta)}{g(y_i|\mathbf{x}_i, \gamma)} \right) \right)^2$$

é uma estimativa da variância de $\frac{1}{\sqrt{n}} LR(\hat{\theta}, \hat{\gamma})$. Para modelos não encaixados

$$T_{LR,NN} \xrightarrow{d} N(0, 1),$$

sob

$$H_0 : \mathbb{E}_h \left[\log \frac{f(y_i|\mathbf{x}_i, \theta)}{g(y_i|\mathbf{x}_i, \gamma)} \right] = 0,$$

onde $\mathbb{E}$ denota o valor esperado com relação ao processo gerador dos dados $h_0(\mathbf{y}_i, X_i)$. Ao nível de 5% a hipótese nula de equivalência dos modelos em favor do modelo F_θ sendo melhor (ou pior) do que o modelo G_γ, não é rejeitada se $T_{LR,NN} > z_{0.025}$, (ou $T_{LR,NN} < -z_{0.025}$).

Assim, tomando F_θ o modelo com erros SPN e G_γ o modelo com erros SHN nós encontramos que $T_{LR,NN} = 2.5164$, portanto, o modelo não linear é melhor, de acordo à estatística generalizada $T_{LR,NN}$, que o modelo linear.

Capítulo 6

Modelos para Dados Censurados AP

6.1 Motivação

Uma importante aplicação do modelo PN para dados contínuos é o caso de dados com uma alta frequência de zeros ou com um valor limite inferior ou superior. Se é o caso de um valor limite inferior, então a variável em estudo pode tomar uma considerável fração de observações amostrais que podem ser obtidas por causa da censura ou por truncamento. Nosso interesse agora é introduzir o estudo de dados censurados e truncados para distribuções assimétricas potência normal, dada a ampla aplicação que têm os modelos para este tipo de situações, onde geralmente o suposto é que os erros do modelo tem distribuição normal, nós mudamos este suposto e estendemos os modelos tobit censorial e de mistura discreta-contínua através de modelos assimétricos.

6.2 Variáveis Aleatorias Censurada e Truncada PN

Uma distribuição truncada é a parte de uma distribuição não truncada antes ou depois de um valor específico. Mesmo assim, a censura é um procedimento no qual os intervalos de uma variável são limitados a priori pelo pesquisador; este procedimento produz uma distorsão estatística similar ao processo de truncamento. Também, uma variável é dita censurada quando os valores do fenômeno medido, acima ou abaixo de um valor limiar,

é fixado como este valor. Vamos supor que y^* tem uma distribuição PN de parâmetros ξ, η e α. Consideramos também uma amostra de tamanho n $(y_1^*, y_2^*, ..., y_n^*)$ e que registramos só aqueles valores y^* maiores que uma constante c. Para aqueles valores $y^* \leq c$ registramos apenas o valor c. Assim, as observações ficam

$$y_i = \begin{cases} y_i^*, & \text{se } y_i^* > c, \\ c, & \text{caso contrario} \end{cases}$$

com $i = 1, 2, ..., n$. A amostra resultante é dita uma amostra censurada. Assim, para as observações $y_i = c$ nós temos que $Pr(y_i = c) = Pr(y_i^* \leq c) = \left\{\Phi\left(\frac{c-\xi}{\eta}\right)\right\}^\alpha$ e para $y_i^* > c$ a distribuição de y_i é mesmo como a distribuição de y_i^* isto é, $y_i \sim PN(\xi, \eta, \alpha)$. Aqui, a função de log-verossimilhança para estimar ξ, η, α é

$$\ell(\theta; \mathbf{Y}) = \alpha \sum_i (1 - I_i) \log\left[\Phi\left(\frac{c-\xi}{\eta}\right)\right] + \\ \sum_i I_i \left\{\log(\alpha) - \log(\eta) + \log\left(\phi\left(\frac{y_i - \xi}{\eta}\right)\right) + (\alpha - 1)\log\left(\Phi\left(\frac{y_i - \xi}{\eta}\right)\right)\right\},$$

onde

$$I_i = \begin{cases} 1, & \text{se } y_i^* > c, \\ 0, & \text{se } y_i^* \leq c. \end{cases}$$

Assim, os elementos da função escore, definidos e denotados da mesma forma que no modelo LPN, são dados por:

$$U(\xi) = -\frac{\alpha}{\eta}\sum_i (1 - I_i) w_c + \frac{1}{\eta}\sum_i I_i \left\{z_i - (\alpha - 1) w_i\right\},$$

$$U(\eta) = -\frac{\alpha}{\eta}\sum_i (1 - I_i) z_c w_c + \frac{1}{\eta}\sum_i I_i \left\{-1 + z_i^2 - (\alpha - 1) z_i w_i\right\},$$

$$U(\alpha) = \sum_i (1 - I_i)\log\left[\Phi\left(\frac{c-\xi}{\eta}\right)\right] + \sum_i I_i \left\{\frac{1}{\alpha} + \log\left(\Phi\left(\frac{y_i - \xi}{\eta}\right)\right)\right\},$$

onde $z_c = \frac{c-\xi}{\eta}$, $z_i = \frac{y_i - \xi}{\eta}$, $w_c = \frac{\phi(z_c)}{\Phi(z_c)}$ e $w_i = \frac{\phi(z_i)}{\Phi(z_i)}$. Definindo e denotando os elementos da matriz de informação observada da mesma forma que no modelo LPN, estes são dados pelas expresões:

$$j_{\xi\xi} = \frac{\alpha}{\eta^2}\sum_i (1 - I_i)\{z_c w_c + w_c^2\} + \frac{1}{\eta^2}\sum_i I_i\{1 + (\alpha - 1)[z_i w_i + w_i^2]\}, \quad j_{\alpha\alpha} = \frac{1}{\alpha^2}\sum_i I_i,$$

$$j_{\eta\xi} = \frac{\alpha}{\eta^2}\sum_i (1-I_i)\{-w_c + z_c^2 w_c + z_c w_c^2\} + \frac{1}{\eta^2}\sum_i I_i\{2z_i + (\alpha-1)[-w_i + z_i^2 w_i + z_i w_i^2]\},$$

$$j_{\eta\eta} = \frac{\alpha}{\eta^2}\sum_i (1-I_i)\{-2z_c w_c + z_c^2 w_c^2 + z_c^3 w_c\} + \frac{1}{\eta^2}\sum_i I_i\{-1 + 3z_i^2 + (\alpha-1)[-2z_i w_i + z_i^2 w_i^2 + z_i^3 w_i]\},$$

$$j_{\alpha\xi} = \frac{1}{\eta}\sum_i (1-I_i) w_c + \frac{1}{\eta}\sum_i I_i w_i, \quad j_{\alpha\eta} = \frac{1}{\eta}\sum_i (1-I_i) z_c w_c + \frac{1}{\eta}\sum_i I_i z_i w_i.$$

Baseado na função escore e na matriz de informação observada os parâmetros são estimados usando algoritmos iterativos.

Suponha agora que antes da amostra ser selecionada, nós truncamos a distribuição de y^* no ponto $y^* = c$ de modo que nenhuma observação para $y^* > c$ é selecionada. Sob estas condições, a função de densidade de probabilidade da distribuição PN truncada é

$$\varphi(y^*|y^* < c) = \frac{\frac{\alpha}{\eta}\phi\left(\frac{y_i-\xi}{\eta}\right)\left\{\Phi\left(\frac{y_i-\xi}{\eta}\right)\right\}^{\alpha-1}}{\left\{\Phi\left(\frac{c-\xi}{\eta}\right)\right\}^{\alpha}}, \quad -\infty < y^* \leq c.$$

Aqui $\frac{1}{\left\{\Phi\left(\frac{c-\xi}{\eta}\right)\right\}^{\alpha}}$ é uma constante de normalização. A função de log-verossimilhança para estimar ξ, η e α é dada por

$$\ell(\theta; \mathbf{Y}) = -N_1\alpha\log\left(\Phi\left(\frac{c-\xi}{\eta}\right)\right) + \sum_{y_i^* \leq c}\left\{\log(\alpha) - \log(\eta) + \log\left(\phi\left(\frac{y_i-\xi}{\eta}\right)\right) + (\alpha-1)\log\left(\Phi\left(\frac{y_i-\xi}{\eta}\right)\right)\right\},$$

onde N_1 é o número de observações tais que $-\infty < y^* \leq c$.

Logo, os elementos da função escore ficam dados por:

$$U(\xi) = \frac{N_1\alpha}{\eta}w_c + \frac{1}{\eta}\sum_{y_i^* \leq c}\{z_i - (\alpha-1)w_i\},$$

$$U(\eta) = \frac{N_1\alpha}{\eta}z_c w_c + \frac{1}{\eta}\sum_{y_i^* \leq c}\left\{-1 + z_i^2 - (\alpha-1)z_i w_i\right\},$$

$$U(\alpha) = -N_1\log\left[\Phi\left(\frac{c-\xi}{\eta}\right)\right] + \sum_{y_i^* \leq c}\left\{\frac{1}{\alpha} + \log\left(\Phi\left(\frac{y_i-\xi}{\eta}\right)\right)\right\}.$$

Enquanto que os elementos da matriz de informação observada são:

$$j_{\xi\xi} = -\frac{N_1\alpha}{\eta^2}\{z_c w_c + w_c^2\} + \frac{1}{\eta^2}\sum_{y_i^* \leq c}\{1 + (\alpha - 1)[z_i w_i + w_i^2]\},$$

$$j_{\eta\xi} = -\frac{N_1\alpha}{\eta^2}\{-w_c + z_c^2 w_c + z_c w_c^2\} + \frac{1}{\eta^2}\sum_{y_i^* \leq c}\{2z_i + (\alpha - 1)[-w_i + z_i^2 w_i + z_i w_i^2]\} \quad \text{e}$$

$$j_{\eta\eta} = -\frac{N_1\alpha}{\eta^2}\{-2z_c w_c + z_c^2 w_c^2 + z_c^3 w_c\} + \frac{1}{\eta^2}\sum_{y_i^* \leq c}\{-1 + 3z_i^2 + (\alpha - 1)[-2z_i w_i + z_i^2 w_i^2 + z_i^3 w_i]\},$$

$$j_{\alpha\xi} = -\frac{N_1}{\eta}w_c + \frac{1}{\eta}\sum_{y_i^* \leq c} w_i, \quad j_{\alpha\eta} = -\frac{N_1}{\eta}z_c w_c + \frac{1}{\eta}\sum_{y_i^* \leq c} z_i w_i, \quad j_{\alpha\alpha} = \frac{N_1}{\alpha^2}.$$

Novamente, os parâmetros podem ser estimados usando algoritmos iterativos, baseado na função escore e a matriz de informação observada.

6.3 Modelo Tobit PN

Agora nós discutimos a possibilidade de estender o modelo Tobit censorial através de modelos assimétricos. O problema de estimação de um modelo de regressão onde a variável dependente é limitada tem sido estudada em diferentes áreas: econometria, ensaios clínicos, fenômenos políticos, entre outros. Em econometria, alguns destes modelos são classificados como: modelos de regressão truncados, modelos de regressão censurados e modelos de regressão discretos.

Quando os dados estão censurados, a distribuição que segue os dados é uma mistura entre uma distribuição discreta e uma continua. Esta ideia foi popularizada por Tobin (1958) e o modelo resultante é conhecido como modelo Tobit, o qual é definido em termos da variável latente

$$y_i = \begin{cases} y_i^*, & \text{se } y_i^* > 0, \\ 0, & \text{se } y_i^* \leq 0, \end{cases}$$

onde y^* segue uma certa distribuição.

O objetivo principal agora é fazer uma extensão do modelo de regressão Tobit normal

(TN), ao caso da distribuição PN, fazendo uma ampla revisão das características estudadas no caso normal, principalmente por Tobin (1958) e Amemiya (1973). A maioria dos resultados sobre o modelo normal de regressão censurado, estão baseados nos desenvolvimentos para o modelo probit, onde a variável de interesse teórica, digamos, "y", não é observada; em vez disso, se observa uma variável dummy, representada por um caso se é observada e zero caso contrario. Em seu exemplo, Tobin refere-se à relação entre os rendimentos das famílias e as despesas em várias categorias de mercadorias. As despesas zero para alguns produtos de luxo informam um baixo nível de renda, comparado com as famílias que fizeram esse tipo de despesas. Também o intervalo de variabilidade das despesas é amplo, o que significa que há um conjunto de observações concentradas em torno de zero. Além disso, este raciocínio implica que não pode haver despesas negativas. Ou seja, para algumas observações a resposta observada não é a resposta concreta, mas sim o valor da censura (zero) e um indicador de que censura ocorreu. A resposta limitada está diretamente relacionada com a distribuição dos erros, os quais geralmente são assumidos ser normal com variância constante. Agora nós vamos estender essa hipótese para o caso em que os termos de erros têm distribuição PN, descrevendo a função de log-verossimilhança para obter os estimadores dos parâmetros do modelo de regressão e a matriz de informação observada que nos conduz aos erros padrão das estimativas.

6.3.1 Modelo Tobit PN

Escrevemos o modelo Tobit PN (TPN) através da variável $y_i = \max\{y_i^*, 0\}$, onde $y_i^* = x_i'\boldsymbol{\beta} + u_i, \quad i = 1, 2, ..., N$; isto é

$$y_i = \begin{cases} x_i'\boldsymbol{\beta} + u_i, & \text{se } x_i'\boldsymbol{\beta} + u_i > 0, \\ 0, & \text{caso contrario} \end{cases}$$

onde $\boldsymbol{\beta}$ é um vetor de parâmetros desconhecidos de tamanho p, x_i é um vetor de constantes não estocásticas, também de tamanho p, e u_i são os residuais; que são variáveis aleatórias $PN(0, \eta, \alpha)$ independentes e identicamente distribuídas. Este modelo é basicamente um modelo de regressão linear PN censurado; como tal, sua estimativa esta relacionada com a

estimativa nas distribuições PN truncada e censurada, estudadas no capítulo dois. Agora nós discutiremos a abordagem da verossimilhança para a estimativa dos parâmetros. Ainda que o valor limite seja zero isto não é muito restritivo pois se o modelo fosse definido por

$$y_i = \begin{cases} x_i'\boldsymbol{\beta} + u_i, & \text{se } x_i'\boldsymbol{\beta} + u_i > c_i, \\ c_i, & \text{caso contrario} \end{cases}$$

podemos fazer $y_i^* = y_i - c_i$, $x_i^* = (x_i, c_i)'$ e $\boldsymbol{\beta}^* = (\boldsymbol{\beta}', -1)'$, assim o modelo anterior pode ser aplicado substituindo $x_i, y_i, \boldsymbol{\beta}$ por $x_i^*, y_i^*, \boldsymbol{\beta}^*$. Similarmente, para o modelo

$$y_i = \begin{cases} x_i'\boldsymbol{\beta} + u_i, & \text{se } x_i'\boldsymbol{\beta} + u_i < 0, \\ 0, & \text{caso contrario} \end{cases}$$

podemos multiplicar y_i, x_i, u_i por -1 e obtemos o primeiro modelo.

6.3.2 Momentos

Para calcular os momentos de y_i vamos a proceder como em Amemiya (1973). Então, seja ψ o subconjunto dos número inteiros $\{1, 2, ..., T\}$ tais que $y_i = 0$ para $i \in \psi$ e seja $\bar{\psi}$ o complemento de ψ no conjunto dos número enteros. Agora, seja u^* uma variável aleatória com função de densidade $h(\cdot)$ dada por

$$h(\lambda) = \begin{cases} \frac{\alpha}{\eta} \frac{1}{1-\Phi^\alpha\left(\frac{-x_i'\boldsymbol{\beta}}{\eta}\right)} \phi\left(\frac{\lambda}{\eta}\right) \Phi^{\alpha-1}\left(\frac{\lambda}{\eta}\right) & \text{se } -x_i'\boldsymbol{\beta} < \lambda < \infty, \\ 0, & \text{caso contrario} \end{cases}$$

Então, temos que

$$y_i = x_i'\boldsymbol{\beta} + u_i^*, \quad \text{para} \quad i \in \bar{\psi}.$$

Assim, os momentos condicionais de y_i, dado $i \in \bar{\psi}$, podem ser calculados a partir dos momentos de u^*. Assim, temos que

$$\mathbb{E}(y_i|y_i > 0) = x_i'\boldsymbol{\beta} + \frac{\alpha\eta}{1 - \Phi^\alpha\left(\frac{-x_i'\boldsymbol{\beta}}{\eta}\right)} \int_{\Phi\left(\frac{-x_i'\boldsymbol{\beta}}{\eta}\right)}^{1} u^{\alpha-1}\Phi^{-1}(u)du. \tag{6.1}$$

Além disso, $Pr(y_i = 0) = Pr(u_i < -x_i'\boldsymbol{\beta}) = \Phi^\alpha\left(-\frac{x_i'\boldsymbol{\beta}}{\eta}\right)$, então,

$$\mathbb{E}(y_i) = Pr(y_i > 0)\mathbb{E}(y_i|y_i > 0) + Pr(y_i = 0)\mathbb{E}(y_i|y_i = 0) = \left(1 - \Phi^\alpha\left(-\frac{x_i'\boldsymbol{\beta}}{\eta}\right)\right)\mathbb{E}(y_i|y_i > 0), \tag{6.2}$$

que estabelece uma relação básica entre o valor esperado de qualquer observação, $\mathbb{E}(y_i)$, é o valor esperado de y_i dado $y_i > 0$ como demostrado por McDonald e Moffitt (1980), para o caso normal.

Dessa forma, a mudança total em y_i pode ser desmembrada em duas partes: (1) a mudança dos y_i acima do limite, ponderada pela probabilidade de estar acima do limite, (2) a mudança na probabilidade de estar acima do limite, ponderada pelo valor esperado de y_i de estar acima do limite.

6.3.3 Predições no Modelo TPN

A predição sobre y_i pode ser obtida da seguinte forma: se nós definimos o modelo em termos da variável latente

$$y_i^* = x_i'\boldsymbol{\beta} + u_i \quad \text{com} \quad u_i \sim PN(0, \eta, \alpha),$$

modelando a variável latente y_i^*, então podemos definir y_i, a variável observada, como

$$y_i = \begin{cases} y_i^*, & \text{se } y_i^* > 0, \\ 0, & \text{caso contrario.} \end{cases}$$

Agora como os erros têm distribuição $PN(0, \eta, \alpha)$ truncada então, como no caso da regressão linear, $E(y_i^*) \neq x_i'\beta$, devemos fazer a seguinte correção no intercepto, $\beta_0^* = \beta_0 + \mu_u$, onde $\mu_u = \mathbb{E}(u_i)$. Assim,

$$E(y_i^*) = x_i'\boldsymbol{\beta}^* \quad \text{sendo} \quad \boldsymbol{\beta}^* = (\beta_0^*, \boldsymbol{\beta}_1)'.$$

Portanto, nós podemos fazer predições, depois de estimar todos os parâmetros do modelo, a partir da variável latente definida. Também, nós temos outras duas opções para fazer predições; a primeira é a partir da esperança condicional dos valores maiores que zero da variável y_i^* ou y_i dada na expressão (6.1), enquanto que a segunda opção é a partir da esperança de y_i, a média de todos os positivos e zero, dada na expressão (6.2), (veja Muñoz, 2009).

6.3.4 Estimação

Inicialmente definimos as seguintes quantidades:

Y_1' é o vetor de observações com $y_i > 0$,

$\mathbf{X_1'}$ é a matriz de valores correspondente a $y_i \neq 0$ e

$\mathbf{X_0'}$ é uma matriz de valores correspondente a $y_i = 0$.

Agora, para observações y_i que são zero, sabemos que

$$Pr(y_i = 0) = \left\{ \Phi\left(-\frac{x_i'\boldsymbol{\beta}}{\eta} \right) \right\}^{\alpha},$$

enquanto que para observações y_i que são maiores que zero, temos

$$\varphi(y_i|y_i > 0)Pr(y_i > 0) = \frac{\alpha}{\eta}\phi\left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta} \right) \left\{ \Phi\left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta} \right) \right\}^{\alpha-1}.$$

Assim, a função de log-verossimilhança tem duas componentes: uma primeira componente que é a soma sobre as observações com $y_i = 0$ e uma segunda componente, que é a soma para $y_i > 0$. Fazendo

$$I_i = \begin{cases} 1, & \text{se } y_i > 0, \\ 0, & \text{se } y_i = 0, \end{cases}$$

temos que para $\boldsymbol{\theta} = (\boldsymbol{\beta}', \eta, \alpha)'$,

$$\begin{aligned} \ell(\boldsymbol{\theta}; \mathbf{Y}) = \alpha \sum_i (1 - I_i) \log\left(\Phi\left(\frac{-x_i'\boldsymbol{\beta}}{\eta} \right) \right) + \sum_i I_i \left\{ \log(\alpha) - \log(\eta) - \frac{1}{2}\left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta} \right)^2 \right\} \\ + (\alpha - 1) \sum_i I_i \log\left(\Phi\left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta} \right) \right) \end{aligned} \tag{6.3}$$

Portanto, os elementos da função escore ficam dados por:

$$U(\beta) = -\frac{\alpha}{\eta} \sum_i (1 - I_i) w_{0i} x_i' + \frac{1}{\eta} \sum_i I_i \left\{ z_i - (\alpha - 1) w_i \right\} x_i', \tag{6.4}$$

$$U(\eta) = -\frac{\alpha}{\eta} \sum_i (1 - I_i) z_{0i} w_{0i} + \frac{1}{\eta} \sum_i I_i \left\{ -1 + z_i^2 - (\alpha - 1) z_i w_i \right\}, \tag{6.5}$$

$$U(\alpha) = \sum_i (1 - I_i) \log \left[\Phi \left(\frac{-x_i'\boldsymbol{\beta}}{\eta} \right) \right] + \sum_i I_i \left\{ \frac{1}{\alpha} + \log \left[\Phi \left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta} \right) \right] \right\}, \tag{6.6}$$

onde $z_{0i} = \frac{-x_i'\beta}{\eta}$, $z_i = \frac{y_i - x_i'\beta}{\eta}$, $w_{0i} = \frac{\phi(z_{0i})}{\Phi(z_{i0})}$ e $w_i = \frac{\phi(z_i)}{\Phi(z_i)}$.

Então, igualando zero as equações (6.6) e (6.4) obtemos as equações de estimação

$$\boldsymbol{\beta} = \boldsymbol{\beta}_{MQO} - \eta(\mathbf{X_1'X_1})^{-1} \left[\alpha \mathbf{X_0'} W_0 + (\alpha - 1)\mathbf{X_1'} W_1 \right]. \tag{6.7}$$

e

$$\alpha = -\frac{n_1}{\sum_i \log \left\{ \left[\Phi \left(\frac{-x_i'\boldsymbol{\beta}}{\eta} \right) \right]^{1-I_i} \left[\Phi \left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta} \right) \right]^{I_i} \right\}} \tag{6.8}$$

onde $\boldsymbol{\beta}_{MQO}$ é o estimador de mínimos quadrados ordinários obtido com as observações $y_i > 0$, n_1 é o número de observações do vetor Y_1 e W_0, W_1 são vetores com elementos w_{0i} e w_i, respectivamente. Consequentemente, multiplicando a equação (6.4) por $\boldsymbol{\beta}$ e somando com a equação (6.5) obtemos que:

$$\eta = \frac{1}{2n_1} \left\{ (1 - \alpha) W_1' Y_1 + \left[(\alpha - 1)^2 W_1' Y_1 Y_1' W_1 + 4 n_1 Y_1' (Y_1 - \mathbf{X_1}\boldsymbol{\beta}) \right]^{1/2} \right\}. \tag{6.9}$$

Logo, a equação (6.7) mostra a relação entre o estimador de máxima verossimilhança de $\boldsymbol{\beta}$ e o estimador de mínimo quadrados ordinários obtido a partir das observações $y_i > 0$. Os elementos da matriz de informação observada são dados por:

$$j_{\boldsymbol{\beta}\boldsymbol{\beta}'} = \frac{1}{\eta^2} \mathbf{X}_1' \mathbf{X}_1 + \frac{\alpha - 1}{\eta^2} \mathbf{X}_1' \Lambda_{21} \mathbf{X}_1 + \frac{\alpha}{\eta^2} \mathbf{X}_0' \Lambda_{20} \mathbf{X}_0,$$

$$j_{\eta\boldsymbol{\beta}} = \frac{2}{\eta^3} \mathbf{X}_1' \left(Y_1 - \mathbf{X}_1 \boldsymbol{\beta} \right) + \frac{\alpha - 1}{\eta^2} \mathbf{X}_1' \Lambda_{31} + \frac{\alpha}{\eta^2} \mathbf{X}_0' \Lambda_{30}$$

$$j_{\alpha\boldsymbol{\beta}} = \frac{1}{\eta} \mathbf{X}_1' \Lambda_{11} + \frac{1}{\eta} \mathbf{X}_0' \Lambda_{10}, \qquad j_{\alpha\alpha} = \frac{1}{\alpha^2} \sum_i Ii = \frac{n_1}{\alpha^2},$$

$$j_{\alpha\eta} = \frac{1}{\eta} \left(Y_1 - \mathbf{X}_1 \boldsymbol{\beta} \right)' \Lambda_{11} + \frac{1}{\eta} \left(Y_0 - \mathbf{X}_0 \boldsymbol{\beta} \right)' \Lambda_{10},$$

$$j_{\eta\eta} = \frac{\alpha}{\eta^2} \sum_i (1 - I_i) \{ -2 w_0 z_0 + w_0^2 z_0^2 + w_0 z_0^3 \} + \frac{1}{\eta^2} \sum_i I_i \{ -1 + 3 z_i^2 + (\alpha - 1)[-2 w_i z_i + w_i^2 z_i^2 + w_i z_i^3] \},$$

onde Y_0 é um vetor para os dados censurados de Y, no nosso caso um vetor de zeros, e Λ_{10}, Λ_{11}, Λ_{20}, Λ_{21}, Λ_{30} e Λ_{31} são definidos como Λ_1, Λ_2 e Λ_3 do modelo de regressão linear, para os casos censurado e não censurado, respectivamente.

Finalmente, os elementos da matriz de informação observada são n^{-1} vezes a esperança dos elementos da matriz de informação observada.

Portanto, o estimador de máxima verossimilhança do vetor $\boldsymbol{\theta} = (\boldsymbol{\beta}', \eta, \alpha)'$ pode ser obtido por métodos numéricos iterativos a partir da função de log-verossimilhança, bem como a função escore e a matriz de informação observada.

6.4 Análise de Diagnóstico

6.4.1 Influência Local

Novamente vamos definir ψ como o conjunto das observações censuradas, $\overline{\psi}$ o conjunto de observações não censuradas e $\boldsymbol{\theta} = (\boldsymbol{\beta}', \eta, \alpha)' = (\boldsymbol{\beta}', \boldsymbol{\theta}_2')'$ com $\boldsymbol{\theta}_2 = (\eta, \alpha)'$. Além disso, nós usamos para o análise de diagnóstico a mesma notação usada no caso do modelo de regressão linear PN.

Ponderação de casos

A função de log-verossimilhança perturbada no modelo de regressão PN censurado é definido como, exceptuando a constante

$$\ell(\boldsymbol{\theta}|\omega) = \alpha \sum_{\psi} \omega_i \log\left(\Phi\left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta}\right)\right) + \sum_{\overline{\psi}} \omega_i \left\{\log(\alpha) - \log(\eta) - \frac{1}{2}\left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta}\right)^2\right\}$$
$$+ (\alpha - 1) \sum_{\overline{\psi}} \omega_i \log\left(\Phi\left(\frac{y_i - x_i'\boldsymbol{\beta}}{\eta}\right)\right) \qquad (6.10)$$

A matriz Δ é dada por

$$\Delta = \begin{pmatrix} \Delta_{\boldsymbol{\beta}} \\ \Delta_{\boldsymbol{\theta}_2} \end{pmatrix}$$

em que $\Delta_{\boldsymbol{\beta}}$ é uma matriz de tamanho (p+1)×n, e $\Delta_{\boldsymbol{\theta}_2}$ é uma matriz de tamanho 2×n. Neste caso $\Delta_{\boldsymbol{\beta}}$ têm elementos

$$\Delta_{\boldsymbol{\beta}} = \frac{1}{\hat{\eta}^2}\mathbf{X}' diag\left\{a_1, a_2, ..., a_n\right\}_{i=1,2,...,n},$$

$$a_i = \begin{cases} \hat{e}_i - (\hat{\alpha}-1)\hat{\eta}\zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right), & \text{se } i \in \overline{\psi}, \\ -\hat{\alpha}\hat{\eta}\zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right), & \text{se } i \in \psi, \end{cases}$$

enquanto que

$$\Delta_{\boldsymbol{\theta}_2} = (b_1, b_2, ..., b_n),$$

onde

$$b_i = \left\{ \begin{array}{c} -\frac{1}{\hat{\eta}} + \frac{1}{\hat{\eta}^3}\hat{e}_i^2 - \frac{\hat{\alpha}-1}{\hat{\eta}^2}\hat{e}_i\zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) \\ \frac{1}{\hat{\alpha}} + \log\left\{\Phi\left(\frac{\hat{e}_i}{\hat{\eta}}\right)\right\} \end{array} \right\}_{i\in\overline{\psi}}$$

e

$$b_i = \left\{ \begin{array}{c} -\frac{\hat{\alpha}}{\hat{\eta}}\frac{\hat{e}_i}{\hat{\eta}}\zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) \\ \log\left\{\Phi\left(\frac{\hat{e}_i}{\hat{\eta}}\right)\right\} \end{array} \right\}_{i\in\psi}$$

onde $\hat{e}_i = y_i - x_i'\hat{\boldsymbol{\beta}}$, $\zeta(\cdot) = \frac{\phi(\cdot)}{\Phi(\cdot)}$, sendo $\hat{\boldsymbol{\beta}}$, $\hat{\eta}$ e $\hat{\alpha}$ os estimadores de máxima verossimilhança de $\boldsymbol{\beta}$, η e α, respectivamente.

Perturbação na Variável Resposta

Os elementos da matriz Δ são

$$\Delta_{\boldsymbol{\beta}} = \frac{S_y}{\hat{\eta}^2}\mathbf{X}' diag\left\{m_i\right\},$$

com

$$m_i = \begin{cases} 1 - (\hat{\alpha}-1)\left[\frac{\hat{e}_i}{\hat{\eta}}\zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) + \zeta^2\left(\frac{\hat{e}_i}{\hat{\eta}}\right)\right], & \text{se } i \in \overline{\psi}, \\ -\hat{\alpha}\left[\frac{\hat{e}_i}{\hat{\eta}}\zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) + \zeta^2\left(\frac{\hat{e}_i}{\hat{\eta}}\right)\right], & \text{se } i \in \psi. \end{cases}$$

enquanto que

$$\Delta_{\boldsymbol{\theta}_2} = (c_1, c_2, ..., c_n),$$

onde

$$c_i = \left\{ \begin{array}{c} \frac{S_y}{\hat{\eta}^2}\left\{2\frac{\hat{e}_i^2}{\hat{\eta}} + (\hat{\alpha}-1)\left[\zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) - \frac{\hat{e}_i}{\hat{\eta}_e}\left[\frac{\hat{e}_i}{\hat{\eta}}\zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) + \zeta^2\left(\frac{\hat{e}_i}{\hat{\eta}}\right)\right]\right]\right\} \\ \frac{S_y}{\hat{\eta}}\zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) \end{array} \right\}_{i\in\overline{\psi}}$$

e

$$c_i = \left\{ \begin{array}{c} \frac{\hat{\alpha}S_y}{\hat{\eta}^2}\left\{\left[\zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) - \frac{\hat{e}_i}{\hat{\eta}_e}\left[\frac{\hat{e}_i}{\hat{\eta}}\zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) + \zeta^2\left(\frac{\hat{e}_i}{\hat{\eta}}\right)\right]\right]\right\} \\ \frac{S_y}{\hat{\eta}}\zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) \end{array} \right\}_{i\in\psi}.$$

Perturbação na Variável Explicativa

1. Para $i \in \overline{\psi}$, os elementos da matriz $\Delta_{\boldsymbol{\beta}}$ são como foram definidos no caso do modelo de regressão PN

$$\Delta_{\beta_{ij}} = \begin{cases} -\frac{\hat{\beta}_q S_q}{\hat{\eta}^2} x_{ij} - \frac{(\hat{\alpha}-1)\hat{\beta}_q S_q}{\hat{\eta}^2} x_{ij} \left[\frac{\hat{e}_i}{\hat{\eta}} \zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) + \zeta^2\left(\frac{\hat{e}_i}{\hat{\eta}}\right) \right], & \text{se } j \neq q \\ -\frac{S_q}{\hat{\eta}^2}(\hat{\beta}_q x_{iq} - \hat{e}_i) - \frac{(\hat{\alpha}-1)S_q}{\hat{\eta}} \zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) - \frac{(\hat{\alpha}-1)\hat{\beta}_q S_q}{\hat{\eta}^2} x_{iq} \left[\frac{\hat{e}_i}{\hat{\eta}} \zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) + \zeta^2\left(\frac{\hat{e}_i}{\hat{\eta}}\right) \right], & \text{se } j = q. \end{cases}$$

2. Para $i \in \psi$ temos

$$\Delta_{\beta_{ij}} = \begin{cases} -\frac{\hat{\alpha}\hat{\beta}_q S_q}{\hat{\eta}^2} x_{ij} \left[\frac{\hat{e}_i}{\hat{\eta}} \zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) + \zeta^2\left(\frac{\hat{e}_i}{\hat{\eta}}\right) \right], & \text{se } j \neq q \\ -\frac{\hat{\alpha} S_q}{\hat{\eta}} \zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) - \frac{\hat{\alpha}\hat{\beta}_q S_q}{\hat{\eta}^2} x_{iq} \left[\frac{\hat{e}_i}{\hat{\eta}} \zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) + \zeta^2\left(\frac{\hat{e}_i}{\hat{\eta}}\right) \right], & \text{se } j = q. \end{cases}$$

Similarmente, nós temos para $\boldsymbol{\theta}_2$

$$\Delta_{\boldsymbol{\theta}_2} = (d_1, d_2, ..., d_n),$$

onde

$$d_i = \left\{ \begin{array}{c} 2\frac{\hat{\beta}_q S_q}{\hat{\eta}^2} \frac{\hat{e}_i}{\hat{\eta}} + \frac{(\hat{\alpha}-1)\hat{\beta}_q S_q}{\hat{\eta}^2} \zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) - \frac{(\hat{\alpha}-1)\hat{\beta}_q S_q}{\hat{\eta}} \frac{\hat{e}_i}{\hat{\eta}} \left[\frac{\hat{e}_i}{\hat{\eta}} \zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) + \zeta^2\left(\frac{\hat{e}_i}{\hat{\eta}}\right) \right] \\ -\frac{\hat{\beta}_q S_q}{\hat{\eta}} \zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) \end{array} \right\}_{i \in \overline{\psi}}$$

e

$$d_i = \left\{ \begin{array}{c} \frac{\hat{\alpha}\hat{\beta}_q S_q}{\hat{\eta}^2} \zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) - \frac{\hat{\alpha}\hat{\beta}_q S_q}{\hat{\eta}} \frac{\hat{e}_i}{\hat{\eta}} \left[\frac{\hat{e}_i}{\hat{\eta}} \zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) + \zeta^2\left(\frac{\hat{e}_i}{\hat{\eta}}\right) \right] \\ -\frac{\hat{\beta}_q S_q}{\hat{\eta}} \zeta\left(\frac{\hat{e}_i}{\hat{\eta}}\right) \end{array} \right\}_{i \in \psi}.$$

6.4.2 Análise de Resíduos

Resíduos componentes do desvio

Para $i \in \overline{\psi}$ temos que o resíduo componente do desvio para o modelo TPN são dados por

$$r_{DC_i} = sinal(\hat{e}_i) \left[-2 \log \left[1 - \left\{ \Phi\left(\frac{\hat{e}_i}{\hat{\eta}}\right) \right\}^{\hat{\alpha}} \right] \right]^{1/2}, \ i \in \psi.$$

Resíduos Tipo Martingale

Os resíduos tipo martingale para o modelo TPN são definidos por

$$r_{M_i} = \delta_i + \log\left(1 - \left\{ \Phi\left(\frac{\hat{e}_i}{\hat{\eta}}\right) \right\}^{\alpha}\right), \quad i = 1, 2, ..., n,$$

onde

$$\delta_i = \begin{cases} 1, & \text{se } i \in \overline{\psi}, \\ 0, & \text{se } i \in \psi. \end{cases}$$

Enquanto que os resíduos componentes do desvio são dados por:

$$r_{MD_i} = sinal(r_{M_i}) \left\{-2\left[r_{M_i} + \delta_i \ln(\delta_i - r_{M_i})\right]\right\}^{1/2}, \ i = 1, 2, ..., n.$$

Nós usaremos os resíduos r_{MD_i} como resíduos tipo martingale, pois estes são distribuídos simetricamente ao redor de zero.

Resíduos Padronizados

Os resíduos padronizados no modelo TPN são definidos como no caso do modelo de regressão PN onde para a matriz de alavanca generalizada temos

$$\nu_i = \begin{cases} -\frac{1}{\eta^2} + \frac{\alpha-1}{\eta^2}\left[\frac{e_i}{\eta}\zeta\left(\frac{e_i}{\eta}\right) + \zeta^2\left(\frac{e_i}{\eta}\right)\right], & \text{se } i \in \overline{\psi}, \\ \frac{\alpha}{\eta^2}\left[\frac{e_i}{\eta}\zeta\left(\frac{e_i}{\eta}\right) + \zeta^2\left(\frac{e_i}{\eta}\right)\right], & \text{se } i \in \psi, \end{cases}$$

enquanto que,

$$\ddot{L}_{\boldsymbol{\theta}_{2y_i}} = \begin{pmatrix} \frac{\alpha-1}{\eta^2}\zeta\left(\frac{e_i}{\eta}\right) + \frac{\alpha-1}{\eta^2}\frac{e_i}{\eta}\left[\frac{e_i}{\eta}\zeta\left(\frac{e_i}{\eta}\right) + \zeta^2\left(\frac{e_i}{\eta}\right)\right] \\ \frac{1}{\eta}\zeta\left(\frac{e_i}{\eta}\right) \end{pmatrix}_{i\in\overline{\psi}},$$

e

$$\ddot{L}_{\boldsymbol{\theta}_{2y_i}} = \begin{pmatrix} \frac{\alpha}{\eta^2}\zeta\left(\frac{e_i}{\eta}\right) + \frac{\alpha}{\eta^2}\frac{e_i}{\eta}\left[\frac{e_i}{\eta}\zeta\left(\frac{e_i}{\eta}\right) + \zeta^2\left(\frac{e_i}{\eta}\right)\right] \\ \frac{1}{\eta}\zeta\left(\frac{e_i}{\eta}\right) \end{pmatrix}_{i\in\psi},$$

com todas as expressões avaliadas em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e onde os e_i são os resíduos do modelo como definidos anteriormente.

6.5 Simulação

Um pequeno experimento de simulação de Monte Carlo com 3000 iterações foi feito com o objetivo de estudar o comportamento dos estimadores de máxima verossimilhança ($\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\eta}$) dos estimadores de β_0, β_1 e η. Os valores do tamanhos de amostras foram $n = 100$, 300 e 500. Os valores dos parâmetros foram: $\alpha = 0.50$, 0.75, 1.25, 1.75, $\beta_0 =$

1.5 e $\beta_1 = 3.5$. Sem perda de generalidade se toma $\eta = 1$. Consideramos uma única variável aleatória X com distribuição normal padrão. Os erros ϵ_i são gerados tal que $\epsilon_i \sim PN(0, \eta, \alpha)$. Os dados y_i são obtido da forma $y_i = max(\beta_0 + \beta_1 x_i - \epsilon_i, 0)$. Foram calculados medidas de qualidades para a estimação pontual como: viés, viés relativo(VR), definido como (viés/valor verdadeiro do parâmetro) e a raiz do erro quadrático médio ($\sqrt{EQM}$). Os resultados foram obtidos mediante o procedimento optim (método "BFGS") do pacote estatístico R.

Os resultados mostram, a partir da Tabela $A.7$, que o viés e a $\sqrt{EQM}$ dos estimadores de máxima verossimilhança dos parâmetros β_0 e β_1 decrescem com o aumento do tamanho amostral, o qual mostra certa consistência dos estimadores, o que é esperado. Não entanto o viés destes estimadores é maior quando o parâmetro α cresce. Portanto, recomenda-se correções para os estimadores dos parâmetros usando, por exemplo, Bootstrap ou Jackknife, quando o valor do parâmetro α é grande. Nós concluímos que o modelo TPN pode ser usado somente quando o tamanho da amostra é grande, por exemplo maior que 100.

6.5.1 Ilustração

Nós vamos ilustrar a utilidade da metodologia proposta aplicando-a a uma parte do conjunto de dados reais analisados em Fair (1978), um artigo interessante e teoricamente inovador. Para uma amostra de 601 homens e mulheres casados pela primeira vez, ele analisou suas respostas a um questionário sobre casos extraconjugais. Algumas das variáveis usadas no estudo foram as seguintes:

Y =número de casos extraconjugais no ano anterior, X_1 =anos de casado,

X_2 =idade, X_3 =religiosidade em uma escala de 1 a 5, sendo 1=ateu, e

X_4 =auto-avaliação do casamento em uma escala de 1 a 5, sendo 1=muito infeliz, 5=muito feliz.

Dos 601 entrevistados, 451 não tiveram nenhum caso extraconjugal e 150 tiveram um ou mais casos; temos portanto uma amostra censurada o que torna adequado o emprego

Tabela 6.1 *Estatísticas descritivas para os resíduos do modelo tobit normal*

n	Média	Variância	$\sqrt{b_1}$	b_2
601	7.4461	17.1100	0.1553	3.7749

do modelo tobit, assim se define

$$y_i = \begin{cases} x_i'\boldsymbol{\beta} + u_i, & \text{se } x_i'\boldsymbol{\beta} + u_i > 0, \\ 0, & \text{caso contrario.} \end{cases}$$

A Tabela 6.1 mostra os coeficientes de assimetria e curtose dos resíduos para os dados. Os valores dos coeficientes de assimetria e curtose justificam um modelo do tipo TPN. Uma conclusão mais enfática nessa direção pode ser vista considerando o teste paramétrico para normalidade, ou seja, para

$$H_0 : \alpha = 1 \quad \text{contra} \quad H_1 : \alpha \neq 1,$$

mediante o uso da estatística de razão de verossimilhanças

$$\Lambda = \frac{\ell_{TN}(\hat{\boldsymbol{\theta}})}{\ell_{TPN}(\hat{\boldsymbol{\theta}})},$$

para o qual o valor observado da estatística de teste é: $-2\log(\Lambda) = 250.3706$, que é um valor maior que o valor crítico da distribuição qui-quadrado, com um grau de liberdade, ao nível de 5%. Portanto, o modelo TPN ajusta melhor os dados de Fair que o modelo TN. A Tabela 6.2 mostra as estimativas dos parâmetros dos modelos.

Na Figura 6.1, a) e b) respectivamente, se mostram os resíduos do modelo e o gráfico de pontos de alavanca generalizado. O gráfico dos resíduos não mostra nenhuma tendência sistemática mas se observam uma observação extrema, a observação #568. Enquanto que o gráfico de pontos de alavanca mostra o ponto #571 como o único ponto suspeito de alta influência nas estimações dos parâmetros. De acordo à mudança relativa, RC_{θ_j}, definida anteriormente, nenhuma destas observações resultou influente nas estimações dos parâmetros.

Tabela 6.2 *Estimativas (erro padrão) dos parâmetros dos modelos TN e TPN.*

Modelo TN		Modelo TPN	
Parâmetro	**estimativa**	**Parâmetro**	**estimativa**
$Log-lik$	-706.4048	$Log-lik$	-581.2195
β_0	9.0830(2.6588)	β_0	8.6170(1.3948)
		β_0^*	16.9027(1.4351)
β_1	-0.1603(0.0777)	β_1	-0.2937(0.0421)
β_2	0.5389(0.1342)	β_2	0.4583(0.0592)
β_3	-1.7233(0.4047)	β_3	-1.5085(0.1864)
β_4	-2.2673(0.4081)	β_4	-2.0937(0.1902)
σ	8.2738(0.5534)	η	5.3403(0.2234)
		α	10.2573(0.5636)

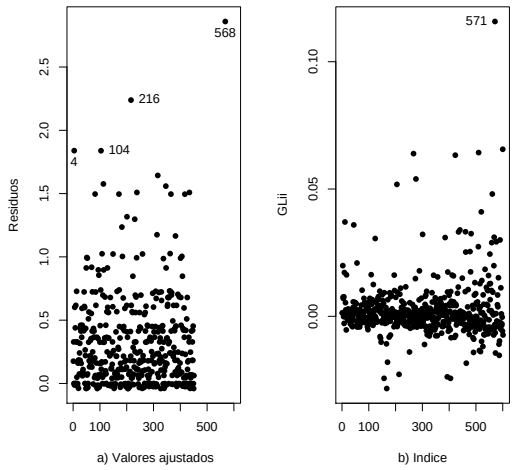


Figura 6.1 *a) Gráficos dos resíduos padronizado $r^*_{DC_i}$ contra os valores ajustados e b) medida de alavanca generalizada contra os índices das observações.*

6.6 Modelo Misto Bernoulli/LPN

Uma aplicação do modelo LPN para dados contínuos é o caso de dados com uma alta frequência de zeros ou com um valor limite. Se é o caso de um valor limite inferior, então a variável em estudo pode tomar uma considerável fração de observações amostrais que podem ser obtidas devido a censura ou truncamento. Um caso de interesse é quando $\log(y_i)$ é função dos parâmetros $\beta_0, ..., \beta_p$ mediante o modelo de regressão linear $\log(y_i) = \beta_0 + \beta_1 x_{1i} + ... + \beta_p x_{pi} + \epsilon_i$, onde $\epsilon_i \sim PN(0, \eta, \alpha)$ e $x_1, ..., x_p$ são constantes fixas e conhecidas. Assim, a função de log-verossimilhança para $\boldsymbol{\theta}$ fica dada por

$$\ell(\boldsymbol{\theta}; \mathbf{Y}) = \alpha \sum_i (1 - I_i) \log \left[\Phi \left(\frac{y_T - x_i' \boldsymbol{\beta}}{\eta} \right) \right] + \\ \sum_i I_i \left\{ \log(\alpha) - \log(\eta y_i) + \log \left(\phi \left(\frac{y_i - x_i' \boldsymbol{\beta}}{\eta} \right) \right) + (\alpha - 1) \log \left(\Phi \left(\frac{y_i - x_i' \boldsymbol{\beta}}{\eta} \right) \right) \right\}.$$

A função escore e a matriz de informação são como encontradas para o modelo censurado estudado anteriormente ou no modelo TPN.

6.6.1 Modelo em Duas Partes

O modelo em duas-partes, "two-part model", foi proposto por Cragg (1971) e é um caminho para analisar situações onde os dados são uma mistura entre uma distribuição discreta e uma continua. Essencialmente, este modelo é uma mistura formada pela combinação linear de uma distribuição continua positiva e uma distribuição pontual com localização em zero. Ainda o modelo tobit pode ser aplicado neste tipo de situações, quando a proporção empírica da distribuição discreta é muito diferente da proporção teórica da distribuição proposta para o modelo, o modelo tobit não é uma boa alternativa para ajustar os dados. Formalmente, a função de densidade de probabilidade de y_i sob o modelo de Cragg pode ser expressado na forma

$$g(y_i) = p_i I_i + (1 - p_i) f^+(y_i)(1 - I_i), \tag{6.11}$$

onde p_i é a probabilidade determinando a contribuição relativa feita pela distribuição pontual, f^+ é uma função de densidade continua com suporte positivo e I_i é uma variável

indicadora definida por

$$I_i = \begin{cases} 0, & \text{se } y_i > 0, \\ 1, & \text{se } y_i \leq 0. \end{cases}$$

Neste modelo, as duas componentes são determinadas por processos estocásticos diferentes, assim, qualquer observação positiva vem necessariamente de f^+, entretanto, um registro zero é gerado pela distribuição pontual.

Moulton e Halsey (1995), generalizam o modelo "two-part", eles admitem a possibilidade que algumas respostas limites sejam o resultado de intervalos sensoriais de f^+. Assim, um zero observado pode ser qualquer realização da distribuição pontual ou uma observação parcial de f^+ com valor atual não precisamente conhecido mas pertencendo a algum ponto do intervalo (0,T), para alguma constante pequena pré-especificada T. Formalmente,

$$g(y_i) = [p_i + (1 - p_i)F^+(T)]I_i + (1 - p_i)f^+(y_i)(1 - I_i), \tag{6.12}$$

onde F^+ é a função de distribuição cumulada de f^+. Assim, uma grande família de distibuições pode ser obtida variando a escolha da distribuição básica f^+ e a função de ligação para modelar a probabilidade p_i. Modelos híbridos: probito/normal truncada, logit/log normal, logit/log gama e probito/log *skew*-normal tem sido usados em pesquisas de diversas áreas como biologia, medicina, agricultura e economia, entre outras.

Agora nós vamos estudar os modelos mistos logit/LPN e probit/LPN, com introdução de covariáveis para cada componente do modelo. Inicialmente suponha que todos os valores observados procedem de uma população LPN com parâmetros de localização e escala ξ e η, respectivamente, e sem covariáveis. A contribuição à verossimilhança para observações não censuradas, isto é para $y > T$, é

$$\frac{\alpha}{\eta y}\phi\left[(\log(y) - \xi)/\eta\right]\left\{\Phi[(\log(y) - \xi)/\eta]\right\}^{\alpha-1},$$

enquanto que para observações censuradas em $y = T$, a contribuição é o fator

$$\left\{\Phi[(\log(y) - \xi)/\eta]\right\}^{\alpha}.$$

A seguir, estendemos este modelo para o caso onde somente a proporção τ de observações

provém da distribuição LPN censurada, enquanto que os outros $(1-\tau)100\%$ das observações provém de outra população de baixa resposta, cuja distribuição está totalmente localizada em ou abaixo do ponto T. Misturamos agora este modelo com a realização de uma variável aleátoria Bernoulli de parâmetro τ onde $y > T$ corresponde a

$$pr(D=1)=\tau,$$

enquanto que se $D=0$, então, $Y \leq T$ com probabilidade um. Assim, condicionando em $D=1$, assumimos que Y segue o modelo LPN. Então, temos que a contribuição à verossimilhança de Y_i pode ser escrita como

$$\left[1+\tau\left\{\left\{\Phi\left(\frac{\log(T)-\xi}{\eta}\right)\right\}^{\alpha}-1\right\}\right]^{I_i}\left[\frac{\tau\alpha}{\eta y}\phi\left(\frac{\log(y)-\xi}{\eta}\right)\left\{\Phi\left(\frac{\log(y)-\xi}{\eta}\right)\right\}^{\alpha-1}\right]^{1-I_i},$$

onde $I_i = 1$ se $y \leq T$ e $I_i = 0$ se $y > T$. Covariáveis são introduzidas para cada variável aleatória, de modo que a função de ligação para o modelo pontual é a função logito, ou seja,

$$logit[pr(D=1|x_{(1)})] = \log\left(\frac{\tau}{1-\tau}\right) = x'_{(1)}\beta_{(1)},$$

onde $x_{(1)}$ é um vetor de covariáveis de dimensão k, com vetor de coeficientes associados a $\beta_{(1)}$. Assim, temos que

$$\tau_i = 1-p_i = \frac{\exp(x'_{(1)i}\boldsymbol{\beta}_{(1)})}{1+\exp(x'_{(1)i}\boldsymbol{\beta}_{(1)})}.$$

Um conjunto de covariáveis $x_{(2)}$ de dimensão q, possivelmente distinto de $x_{(1)}$, pode ser associada à componente LPN, com vetor de coeficientes associado $\beta_{(2)}$, ou seja que

$$log(y_i|y_i>0) \sim PN(x'_{(2)i}\boldsymbol{\beta}_{(2)}, \eta, \alpha),$$

onde os dados observados y_i são variáveis aleatórias independentes e identicamente distribuídas. O conjunto de variáveis explicativas $x_{(1)} = (x_{(1)1}, ..., x_{(1)n})'$ controla a probabilidade de uma verdadeira resposta zero enquanto que a magnitude da resposta não limitada censurada é determinada por $x_{(2)} = (x_{(2)1}, ..., x_{(2)n})'$. Uma vantagem de ter diferentes conjuntos de covariáveis explicativas associadas a p_i e f^+, é que o modelo pode

levar a uma análise de regressão mais exata e informativa (Chai e Bailey, 2008). Assim, o logaritmo da função de verossimilhança de $\boldsymbol{\theta} = (\boldsymbol{\beta}'_{(1)}, \boldsymbol{\beta}'_{(2)}, \eta, \alpha)'$ dados X e Y é dado por,

$$\ell(\boldsymbol{\theta}; \mathbf{Y}) = \sum_i I_i \left\{ \log\left[1 + \exp(x'_{(1)i}\boldsymbol{\beta}_{(1)}) \left\{\Phi(z_{Ti})\right\}^{\alpha}\right] - \log\left[1 + \exp(x'_{(1)i}\boldsymbol{\beta}_{(1)})\right] \right\} +$$
$$\sum_i (1 - I_i) \left\{ \log(\alpha) - \log(\eta y_i) + x'_{(1)i}\boldsymbol{\beta}_{(1)} - \log\left[1 + \exp(x'_{(1)i}\boldsymbol{\beta}_{(1)})\right] - \frac{1}{2} z_i^2 + (\alpha - 1)\log(\Phi(z_i)) \right\} \tag{6.13}$$

onde $z_{Ti} = \frac{\log(T) - x'_{(2)i}\boldsymbol{\beta}_{(2)}}{\eta}$ e $z_i = \frac{\log(y_i) - x'_{(2)i}\boldsymbol{\beta}_{(2)}}{\eta}$. Os elementos da função escore ficam dados por:

$$U(\beta_{(1)j}) = \sum_i I_i \frac{x_{(1)ij} \exp(x'_{(1)i}\boldsymbol{\beta}_{(1)})}{1 + \exp(x'_{(1)i}\boldsymbol{\beta}_{(1)})} \frac{-1 + \left\{\Phi(z_{Ti})\right\}^{\alpha}}{1 + \exp(x'_{(1)i}\boldsymbol{\beta}_{(1)}) \left\{\Phi(z_{Ti})\right\}^{\alpha}} + \sum_i \frac{(1 - I_i) x_{(1)ij}}{1 + \exp(x'_{(1)i}\boldsymbol{\beta}_{(1)})},$$

$$U(\beta_{(2)k}) = -\frac{\alpha}{\eta} \sum_i I_i \frac{x_{(2)ik} \exp(x'_{(1)i}\boldsymbol{\beta}_{(1)}) \left\{\Phi(z_{Ti})\right\}^{\alpha-1} \phi(z_{Ti})}{1 + \exp(x'_{(1)i}\boldsymbol{\beta}_{(1)}) \left\{\Phi(z_{Ti})\right\}^{\alpha}}$$
$$- \frac{1}{\eta} \sum_i (1 - I_i) x_{(2)ik} \left[z_i + (\alpha - 1) \frac{\phi(z_i)}{\Phi(z_i)} \right],$$

$$U(\eta) = -\frac{\alpha}{\eta} \sum_i I_i \frac{z_{Ti} \exp(x'_{(1)i}\boldsymbol{\beta}_{(1)}) \left\{\Phi(z_{Ti})\right\}^{\alpha-1} \phi(z_{Ti})}{1 + \exp(x'_{(1)i}\boldsymbol{\beta}_{(1)}) \left\{\Phi(z_{Ti})\right\}^{\alpha}}$$
$$- \frac{1}{\eta} \sum_i (1 - I_i) \left[1 - z_i^2 + (\alpha - 1) z_i \frac{\phi(z_i)}{\Phi(z_i)} \right]$$

e

$$U(\alpha) = \sum_i I_i \frac{\exp(x'_{(1)i}\boldsymbol{\beta}_{(1)}) \log\left(\Phi(z_{Ti})\right) \left\{\Phi(z_{Ti})\right\}^{\alpha}}{1 + \exp(x'_{(1)i}\boldsymbol{\beta}_{(1)}) \left\{\Phi(z_{Ti})\right\}^{\alpha}} + \sum_i (1 - I_i) \left\{ \frac{1}{\alpha} + \log(\Phi(z_i)) \right\}.$$

Supomos agora que o modelo para a variável D é o modelo probit, então o modelo estocástico específico é

$$p_i = pr(y_i = 0) = \Phi(-x'_{(1)i}\boldsymbol{\beta}_{(1)}) = 1 - \Phi(x'_{(1)i}\boldsymbol{\beta}_{(1)})$$

e

$$\log(y_i | y_i > 0) \sim PN(x'_{(2)i}\boldsymbol{\beta}_{(2)}, \eta, \alpha),$$

onde os dados observados y_i são variáveis aleatórias independentes e identicamente distribuídas. A função de log-verossimilhança se pode escrever como

$$\ell(\boldsymbol{\theta};\mathbf{Y}) = \sum_i I_i \left\{\log\left[1+\Phi(x'_{(1)i}\boldsymbol{\beta}_{(1)})\{\{\Phi(z_{Ti})\}^{\alpha}-1\}\right]\right\} + \sum_i (1-I_i)\left\{\log(\alpha)-\log(\eta)+\log\left(\Phi(x'_{(1)i}\boldsymbol{\beta}_{(1)})\right)-\frac{1}{2}z_i^2+(\alpha-1)\log(\Phi(z_i))\right\} \tag{6.14}$$

onde $z_{Ti} = \frac{\log(T)-x'_{(2)i}\boldsymbol{\beta}_{(2)}}{\eta}$ e $z_i = \frac{\log(y_i)-x'_{(2)i}\boldsymbol{\beta}_{(2)}}{\eta}$. Os escores ficam dados por

$$U(\beta_{(1)j}) = \sum_i I_i \frac{x_{(1)ij}\phi(x'_{(1)i}\boldsymbol{\beta}_{(1)})\{\{\Phi(z_{Ti})\}^{\alpha}-1\}}{1+\Phi(x'_{(1)i}\boldsymbol{\beta}_{(1)})\{\{\Phi(z_{Ti})\}^{\alpha}-1\}} + \sum_i \frac{(1-I_i)x_{(1)ij}\phi(x'_{(1)i}\boldsymbol{\beta}_{(1)})}{\Phi(x'_{(1)i}\boldsymbol{\beta}_{(1)})},$$

$$U(\beta_{(2)k}) = -\frac{\alpha}{\eta}\sum_i I_i \frac{x_{(2)ik}\Phi(x'_{(1)i}\boldsymbol{\beta}_{(1)})\left\{\Phi(z_{Ti})\right\}^{\alpha-1}\phi(z_{Ti})}{1+\Phi(x'_{(1)i}\boldsymbol{\beta}_{(1)})\{\{\Phi(z_{Ti})\}^{\alpha}-1\}} - \frac{1}{\eta}\sum_i (1-I_i)x_{(2)ik}\left[z_i+(\alpha-1)\frac{\phi(z_i)}{\Phi(z_i)}\right],$$

$$U(\eta) = -\frac{\alpha}{\eta}\sum_i I_i \frac{z_{Ti}\Phi(x'_{(1)i}\boldsymbol{\beta}_{(1)})\left\{\Phi(z_{Ti})\right\}^{\alpha-1}\phi(z_{Ti})}{1+\Phi(x'_{(1)i}\boldsymbol{\beta}_{(1)})\{\{\Phi(z_{Ti})\}^{\alpha}-1\}} - \frac{1}{\eta}\sum_i (1-I_i)\left[1-z_i^2+(\alpha-1)z_i\frac{\phi(z_i)}{\Phi(z_i)}\right]$$

e

$$U(\alpha) = \sum_i I_i \frac{\Phi(x'_{(1)i}\boldsymbol{\beta}_{(1)})\log\left(\Phi(z_{Ti})\right)\left\{\Phi(z_{Ti})\right\}^{\alpha}}{1+\Phi(x'_{(1)i}\boldsymbol{\beta}_{(1)})\{\{\Phi(z_{Ti})\}^{\alpha}-1\}} + \sum_i (1-I_i)\left\{\frac{1}{\alpha}+\log(\Phi(z_i))\right\}.$$

6.7 Ilustração

O uso do modelo misto Bernoulli/LPN é ilustrado com dados de um grande estudo de segurança e imonugenicidade da vacinação contra o sarampo, implementada no Haiti entre 1987 e 1990. Entre outros fatores, se estuda a habilidade das doses média e alta vencer as interferências de anticorpos maternos em crianças. O objetivo do estudo era demostrar que altas doses da vacina pode imunizar crianças de até seis meses de idade.

A análise de imonugenicidade indica altas respostas de anticorpos em destinatários com altas doses e entre destinatários com aplicação da vacina Edmonston-zagreb, comparado com destinatários com aplicação da vacina Schwarz (Job et al., 1991).

O foco da análise é estudar mediante o modelo misto logit/LPN a natureza da imonugenicidade diferencial entre crianças de sexo feminino e masculino do estudo. Os dados analisados são dados por Moulton e Halser(1995), os quais correspondem a 330 ensaios de neutralização de anticorpos em crianças de até os 12 meses de idade. A variável em estudo foi a concentração de anticorpos em unidades internacionais (IU) e o limite de deteçao foi de 0.1 unidades internacionais, ou seja $\log(0.1) = -2.3026$ na escala do logaritmo natural. A codificação das covariáveis estudadas foram: EZ (tipo de vacina; 0:Schwarz , 1:Edmonston-Zagreb); HI (dose da vacina; 0:média, 1:alta) e FEM (sexo; 0:homem, 1:mulher). Os resultados foram obtido com a função nlm do pacote estatístico R. Uma análise descritiva dos dados mostra que 26.1% das observações estão abaixo do valor limite de deteção e, portanto, elas são consideradas como respostas censuradas. Na tabela 6.4 se estudam os modelos mistos con distribuição Bernoulli/log-normal e ligação logit. Os modelos (1) e (2) correspondem a modelos sem censura, sem e com covariáveis, enquanto que os modelos (3)-(7) correspondem a modelos mistos com diferentes situações. Em resumo, os modelos estudados foram:

Modelo 1: solo a componente continua com dados censurados e sem covariáveis.

Modelo 2: solo a componente continua com dados censurados e com covariáveis.

Modelo 3: com dados censurados e sem covariáveis nas duas componentes.

Modelo 4: com dados censurados sem covariáveis na resposta limitada e com covariáveis na distribuição localizada em zero.

Modelo 5: com dados censurados e com covariáveis na resposta limitada e sem covariáveis na distribuição localizada em zero.

Modelo 6: com dados censurados, com covariáveis na resposta limitada e na distribuição localizada em zero.

Modelo 7: com dados censurados, com covariáveis na resposta limitada e na distribuição

localizada em zero, um caso particular do modelo 6, exceto covariável "sexo".

Mesmo assim, os dados mostram o seguinte comportamento para a variável $\log(y)$ por acima do valor 0.1:

Tabela 6.3 *Estadísticas descritivas para o conjunto de dados da vacina de Haiti*

n	$\overline{\log(y)}$	s^2	$\sqrt{b_1}$	b_2
330	-0.1793	1.1055	0.7521	2.6286

O alto grau de assimetria recomenda uma distribuição do tipo skew para a análise do modelo misto. O modelo estudado foi o misto Bernoulli/LPN com função de ligação logit. Se ajustam os modelo LPN com censura, modelo (1), e censurado com co-variáveis, modelo (2). Também se ajustam modelos mistos Bernoulli/LPN com covariáveis e ligação logit, modelos (3)-(7), de acordo à tabela 6.5.

Agora levamos a cabo o teste de hipóteses de não diferença do modelo Bernoulli/LPN com o modelo tradicional Bernoulli/log-normal, isto é, hipótese

$$H_0: \ \alpha = 1 \ \text{ contra } \ H_1: \ \alpha \neq 1$$

usando a estatística de razão de verossimilhanças, com base na estatística

$$\Lambda = \frac{L_{Bernoulli-log-normal}(\hat{\boldsymbol{\theta}})}{L_{Bernoulli-LPN}(\hat{\boldsymbol{\theta}})}$$

assim, depois de substituir os valores estimados na expressão anterior obtemos que: $-2\log(\Lambda) = -2(-557.9150 + 544.6657) = 13.2493$, o qual é um valor maior que o valor crítico da distribuição qui-quadrado ao nível de 5%, cujo valor é 3.84. Então rechaçamos a hipóteses nula e concluímos que o modelo Bernoulli/LPN ajusta melhor que o modelo Bernoulli/log-normal. Em geral os modelos mistos com distribuição LPN da Tabela 6.5 ajustaram melhor que seus correspondentes na Tabela 6.4.

O Modelo (6) mostra diferenças estatísticas com os modelos (4) e (5) mas nenhuma com o modelo (7), assim, dos modelos mistos ajustados o melhor foi o modelo (7). Neste

modelo os fatores EZ e HI estão associados à componente bernoulli, e o fator FEM é o único associado à componente LPN. O limite de deteção para mulheres, foi de $\exp(0.2584) = 1.2948$ vezes maior que a concentração de anticorpos de sarampo em homens.

Tabela 6.4 *Estimativas (erro padrão) dos parâmetros e ajuste de uma e duas componentes da mistura Bernoulli/log normal*

		Componente Bernoulli				Componente Log-normal			
Modelo	*Loglik*	INT	EZ	HI	FEM	INT	EZ	HI	FEM
(1)	-557.9130					-0.979			
						(0.0963)			
(2)	-555.59					-1.2870	0.3400	0.1820	0.1150
						(0.1870)	(0.1899)	(0.1895)	(0.1917)
(3)	- 493.4712	1.1946				-0.2795			
		(0.1522)				(0.0859)			
(4)	- 488.1924	0.7704	0.8314	0.4298	-0.2353	-0.2796			
		(0.2677)	(0.3105)	(0.2921)	(0.2901)	(0.0859)			
(5)	- 491.7697	1.1719				-0.3265	-0.1108	-0.0337	0.2793
		(0.1474)				(0.1626)	(0.1570)	(0.1552)	(0.1526)
(6)	- 485.6707	0.8028	0.9117	0.4247	-0.3429	-0.3044	-0.1917	-0.0571	0.3177
		(0.2779)	(0.3289)	(0.2971)	(0.3030)	(0.1622)	(0.1601)	(0.1587)	(0.1605)
(7)	- 487.0980	0.6520	0.8031	0.4237		-0.4004			0.2647
		(0.2206)	(0.3047)	(0.2887)		(0.1125)			(0.1556)

Tabela 6.5 *Estimativas (erro padrão) de parâmetros e ajuste de uma e duas componentes da mistura Bernoulli/LPN*

		Componente Bernoulli				Componente LPN					
Modelo	*Loglik*	INT	EZ	HI	FEM	α	INT	EZ	HI	FEM	η
(1)	-544.6657					0.7024	0.7915				1.2027
						(0.0081)	(0.0319)			(0.0259)	
(2)	-508.9940					0.9276	-1.1669	0.3321	0.1824	0.0726	1.6737
						(0.1446)	(0.0954)	(0.1700)	(0.1763)	(0.1779)	(0.0762)
(3)	-472.3104	0.0259				0.7916	-0.0354				1.1423
		$(5.84 * 10^{-5})$				(0.0041)	(0.0004)				(0.0487)
(4)	-466.4560	0.6585	0.7358	0.3914	-0.2056	0.5621	-0.1972				0.8001
		(0.2394)	(0.2676)	(0.2596)	(0.2591)	(0.0534)	(0.0713)				(0.0580)
(5)	-469.3558	1.0489				0.5597	-0.2170	-0.1527	-0.0615	0.2733	0.7886
		(0.1272)				(0.0528)	(0.1398)	(0.1400)	(0.1394)	(0.1401)	(0.0562)
(6)	-463.9141	0.6638	0.7496	0.3925	-0.2237	0.5627	-0.2145	-0.1664	-0.0658	0.2798	0.7931
		(0.2401)	(0.2694)	(0.2602)	(0.2602)	(0.0534)	(0.1398)	(0.1413)	(0.1405)	(0.1415)	(0.0575)
(7)	-465.0838	0.5601	0.7235	0.3916		0.5615	-0.3217			0.2584	0.7938
		(0.2028)	(0.2667)	(0.2592)		(0.0532)	(0.0989)			(0.1402)	(0.0571)

6.8 Estudo de Simulação

O seguinte estudo de simulação Monte Carlo, com 5000 repetições, tem como objetivo pesquisar o comportamento dos estimadores de máxima verossimilhança na estimação dos parâmetros da distribuição Bernoulli/LPN, com função de ligação probit e covariáveis nas componentes do modelo, para os tamanhos amostrais $n = 300, 500$ e valores $\alpha = 0.75, 1.25, 1.75$, $\boldsymbol{\beta}_{(1)} = (1, -0.5)'$, $\boldsymbol{\beta}_{(2)} = (-2, 1)'$ e fixamos $\eta = 1$. Consideramos uma única variável explicativa X a qual segue uma distribuição normal padrão. Os valores da variável de estudo Y, foram definidos como segue:

$$y_i = \begin{cases} 0, & \text{com probabilidade } \Phi(-\beta_{(1)0} - \beta_{(1)1}x_i), \\ \exp(\beta_{(2)0} + \beta_{(2)1}x_i + \eta\epsilon_i), & \text{com probabilidade } \Phi(\beta_{(1)0} + \beta_{(1)1}x_i), \end{cases}$$

onde $\epsilon_i \sim PN(0, 1, \alpha)$.

Nós calculamos medidas de qualidade tais como: média, viés, viés relativo(VR), definido como 100(viés/valor do parâmetro) e a raiz do erro quadrático médio ($\sqrt{EQM}$). Os resultados foram obtidos com a função nlm do pacote estatístico R. Os resultados mostram que em geral (veja Tabela $A.8$), quando o tamanho da amostra aumenta, o viés, o viés relativo e a raiz do erro quadrático médio decrescem, ou seja que os estimadores são consistentes.

Capítulo 7

Extensão Multivariada do Modelo α-Potência

7.1 Motivação

Para o estudo de distribuições assimétricas com mais de uma variável aleatória, nós agora introduzimos o modelo multivariado α-potência como um passo natural ao caso univariado e ao desenvolvimentos de metodologias estatísticas em estudos que envolvam distribuições multivariadas de dados com coeficientes de assimetria e curtose fora dos intervalos permitidos pela distribuição normal multivariada. Neste contexto, Azzalini e Dalla Valle (1986) estudam a distribuição normal multivariada assimétrica, com propriedades favoraveis para certos conjuntos de observações p-variadas. Portanto, dentro da literatura de modelos multivariados é de interesse ter outras alternativas distintas à distribuição normal assimétrica multivariada, para modelar observações p-variadas com altos(baixos) grau de assimetria e curtose.

Métodos para construir distribuições multivariadas foram desenvolvidos principalmente por Joe (1997), Arnold, Castillo e Sarabia (1999), Nelsen (1999), Koth, Balakrishnan e Johnson (2000), entre eles o métodos baseado em copulas para distribuições marginais conhecidas. Arnold, Castillo e sarabia (1999) desenvolvem a teoria de distribuições mul-

tivariadas a partir de distribuições condicionais especificados, diversos escritos tem sido publicados para distribuições condicionais, normal, exponencial, weibull, SN, etc. Baseados nestes resultados nós vamos desenvolver o modelo α-potência assimétrico multivariado.

7.1.1 Coeficientes de Assimetria e Curtose Multivariados

Os coeficientes de assimetria e curtose multivariados para um vetor aleatório $\mathbf{X} \sim (\mu, \Sigma)$ são definidos, respectivamente, por (Mardial et al., 1979)

$$\beta_{1,p} = \mathbb{E}\left\{(\mathbf{X}-\mu)'\Sigma^{-1}(\mathbf{Y}-\mu)\right\}^3, \qquad \beta_{2,p} = \mathbb{E}\left\{(\mathbf{X}-\mu)'\Sigma^{-1}(\mathbf{X}-\mu)\right\}^2, \tag{7.1}$$

onde p é o número de variáveis. No caso da distribuição normal multivariada temos $\beta_{1,p} = 0$ e $\beta_{2,p} = p(p+2)$. Os coeficientes de assimetria e curtose multivariados amostrais são dados pelas expressões

$$b_{1,p} = \frac{1}{n^2}\sum_{r,s=1}^{n} g_{rs}^3 \quad \text{e} \quad b_{2,p} = \frac{1}{n}\sum_{r=1}^{n} g_{rr}^2,$$

com

$$g_{rs} = (x_r - \bar{x})'S^{-1}(x_s - \bar{x}),$$

onde $\bar{x}$ e S são o vetor de medias e a matriz de covariância amostrais, respectivamente. Em geral para estudar o afastamentos do vetor aleatório $\mathbf{X}$ da distribuição normal multivariada temos como estatísticas de prova as seguintes (sob normalidade)

$$\kappa_1 = \frac{1}{6}nb_{1,p} \xrightarrow{D} \chi_v^2 \quad \text{e} \quad \kappa_2 = \frac{b_{2,p} - p(p+2)}{(8p(p+2)/n)^{1/2}} \xrightarrow{D} N(0,1),$$

onde $v = \frac{1}{6}p(p+1)(p+2)$. Para o caso onde estas estatísticas são significativamente afastados dos valores $\chi^2_{5\%,v}$ e $Z_{5\%}$, se pode fazer uso da normal multivariada assimétrica estudada por Azzalini e Dalla Valle (1986) e da distribuição normal multivariada assimétrica a partir de distribuições condicionais normais assimétricas, desenvolvida por Arnold et al. (2002).

7.2 Distribuições Bivariadas Caracterizadas por Distribuições Condicionais

Seja (X,Y) um vetor aleatório bidimensional de alguma medida produto sobre $S(X) \times S(Y)$, onde $S(X)$ é o conjunto de valores da variável aleatória X e $S(Y)$ o conjunto de valores da variável aleatória Y, onde o conjunto de possíveis valores de $S(X)$ e $S(Y)$ pode ser finito, enumerável ou não enumerável. Sejam $f_X(x)$, $f_Y(y)$, $f_{X|Y}(x|y)$, $f_{Y|X}(y|x)$ e $f(x,y)$, as funcões de densidades marginais, condicionais e conjunta, respectivamente, do vetor (X,Y).

O vetor aleatório bidimensional (X,Y) é dito ser condicionalmente especificado se para qualquer Y, a distribuição condicional de X dado $Y = y$ é um membro de alguma família paramétrica de distribuições conhecida. Considere agora que $\{h_1(x;\underline{\omega}) : \underline{\omega} \in \Omega\}$ denota a família exponencial de densidades de l_1-parâmetros em $\mathbb{R}$ com respeito à medida μ_1, onde $\Omega \in \mathbb{R}^{l_1}$. Também considere a família de densidades de l_2-parâmetros $\{h_2(y;\underline{\tau}) : \underline{\tau} \in T\}$, com respeito à medida μ_2, onde $T \in \mathbb{R}^{l_2}$.

Suponha agora que para todo $y \in S(Y)$ se tem

$$f_{X|Y}(x|y) = h_1(x;\underline{\omega}(y))$$

e para todo $x \in S(X)$, temos que

$$f_{Y|X}(y|x) = h_2(y;\underline{\tau}(x))$$

para certas funções $\underline{\omega} : S(Y) \to \Omega$ e $\underline{\tau} : S(X) \to T$. Supondo a existência das marginais f_X e f_Y, nós temos que

$$f(x,y) = f_Y(y)h_1(x;\underline{\omega}(y)) = f_X(x)h_2(y;\underline{\tau}(x)), \quad \forall x \in S(X), y \in S(Y).$$

Portanto para encontrar todas as densidades conjuntas do vetor aleatório (X,Y), com densidades condicionais da forma h_1 e h_2, temos que resolver a equação funcional anterior para $\underline{\omega}(y)$, $\underline{\tau}(x)$, $f_X(x)$ e $f_Y(y)$.

O seguite Teorema, estudado inicialmente por Suto (1914), é importante na solução desta equação funcional.

Teorema 7.2.1. *As soluções da equação funcional*

$$\sum_{i=1}^{r} f_i(x)\phi_i(y) = \sum_{j=1}^{s} g_j(y)\psi_j(x), \qquad \forall x \in S(X), y \in S(Y) \tag{7.2}$$

com $\{\phi_i\}_{i=1}^r$ *e* $\{\psi_j\}_{j=1}^s$ *sistemas de funções linearmente independentes dados, são da forma* $\underline{f}(x) = C\underline{\psi}(x)$ *e* $\underline{g}(y) = D\underline{\phi}(y)$, *onde* $D = C'$.

Demostração: Veja Arnold, Castilla e Sarabia (2001).

Considere agora que $\{h_1(x;\underline{\omega}) : \underline{\omega} \in \Omega\}$ pertence à família exponencial de densidades de l_1-parâmetros da forma

$$h_1(x;\underline{\omega}) = \gamma_1(x)\beta_1(\underline{\omega}) \exp\left\{\sum_{i=1}^{l_1} \omega_i p_i(x)\right\}, \tag{7.3}$$

onde Ω é um espaço de parâmetros naturais contido em $\mathbb{R}^k$ e as densidades são definidas com respeito alguma medida sobre $\mathbb{R}$, usualmente a medida de Lebesgue. Analogamente, suponha que $\{h_2(y;\underline{\tau}) : \underline{\tau} \in T\}$ pertence à família exponencial de densidades de l_2-parâmetros da forma

$$h_2(y;\underline{\tau}) = \gamma_2(y)\beta_2(\underline{\tau}) \exp\left\{\sum_{j=1}^{l_2} \tau_j q_j(y)\right\}, \tag{7.4}$$

onde T é um espaço de parâmetros naturais contido em $\mathbb{R}^{l_2}$. Aqui assumimos que os $\{p_i\}$ são linearmente independentes, similarmente para os $\{q_j\}$.

Supomos agora que para todo $y \in S(Y)$ e para todo $x \in S(X)$, existem funções $\underline{\omega}(y)$ e $\underline{\tau}(x)$ tais que

$$f_{X|Y}(x|y) = h_1(x;\underline{\omega}(y)), \qquad f_{Y|X}(y|x) = h_2(y;\underline{\tau}(x)), \tag{7.5}$$

para certas funções $\underline{\omega} : S(Y) \to \Omega$ e $\underline{\tau} : S(X) \to T$, onde h_1 e h_2 são definidas como em (7.3) e (7.4).

Teorema 7.2.2. *Seja $f(x,y)$ uma densidade bivariada cujas distribuições condicionais são como em (7.5) para funções $\underline{\omega}(y)$ e $\underline{\tau}(x)$. Então, $f(x,y)$ pode ser escrita na forma*

$$f(x,y) = \gamma_1(x)\gamma_2(y)\exp\left\{P(x)'MQ(y) + a'P(x) + b'Q(y) + c\right\}, \tag{7.6}$$

para M,a,b e c, de dimensões apropriadas tais que $\int_{\mathbb{R}}\int_{\mathbb{R}} f(x,y)dxdy = 1$, com $P(x) = (p_1(x), p_2(x), ..., p_k(x))$ e $Q(y) = (q_1(y), q_2(y), ..., q_l(y))$.

Demostração: Veja Arnold e Strauss (1991).

7.3 Modelo α-Potência Bivariado

Suponha agora que para a distribuição conjunta do vetor aleatório bidimensional (X,Y), a distribuição condicional de X dado $Y = y$ e a distribuição condicional de Y dado $X = x$ são membros da família α-potência, com respeito à medida de Lebesgue. Assim, como considerado no capítulo (1) temos:

$$X|Y = y \sim AP_{f_1}(\underline{\omega}(y)), \qquad Y|X = x \sim AP_{f_2}(\underline{\tau}(x)), \tag{7.7}$$

onde $\underline{\omega}, \underline{\tau}$ são funções de dependência, positivas, as quais devem ser determinadas e f_1, f_2 são funções de densidade, com respeito à medida de Lebesgue, continuas e diferenciáveis.

1) Se $\underline{\omega}(y) = \alpha_1$ e $\underline{\tau}(x) = \alpha_2$, com α_1 e α_2 constantes reais positivas, então as variáveis aleatórias X e Y são independentes e tem-se que

$$f(x,y) = \alpha_1\alpha_2 f_1(x)f_2(y)\{F_1(x)\}^{\alpha_1-1}\{F_2(y)\}^{\alpha_2-1}.$$

2) Se $f_X(x)$ e $f_Y(y)$ são as marginais de $f(x,y)$, então temos que

$$f(x,y) = \tau(x)f_X(x)f_2(y)\{F_2(y)\}^{\tau(x)-1} = \omega(y)f_Y(y)f_1(x)\{F_1(x)\}^{\omega(y)-1}. \tag{7.8}$$

Supomos agora que $\underline{\omega}(y)$ e $\underline{\tau}(x)$ são ambas funções não constantes, então de (7.8) temos que

$$F_Y(y)\{F_1(x)\}^{\underline{\omega}(y)} = F_X(x)\{F_2(y)\}^{\underline{\tau}(x)},$$

onde $F_X(x)$ e $F_Y(y)$ são as funções de distribuição cumuladas marginais de X e Y. Portanto, temos a seguinte equação funcional:

$$\log(F_Y(y)) + \underline{\omega}(y)\log(F_1(x)) = \log(F_X(x)) + \underline{\tau}(x)\log(F_2(y)).$$

Resolvendo esta equação funcional, cuja solução passa pelo origem (0,0), mediante a aplicação do Teorema de Suto (1914), nós encontramos como soluções as seguintes funções de dependência:

$$\underline{\omega}(y) = \alpha_1 + \alpha_{12}\log[F_2(y)], \qquad \underline{\tau}(x) = \alpha_2 + \alpha_{12}\log[F_1(x)], \tag{7.9}$$

onde α_1, α_2 são constantes reais positivas e $\alpha_{12} = \alpha_{21}$ é um número real tal que $\alpha_{12} \leq 0$. Logo, de (7.7) segue que

$$f_{X|Y}(x|y) = f_{f_1}(x;\underline{\omega}(y)) = \underline{\omega}(y)f_1(x)\{F_1(x)\}^{\alpha_1-1}\exp\{\alpha_{12}\log[F_1(x)\log[F_2(y)]]\} \tag{7.10}$$

e

$$f_{Y|X}(y|x) = f_{f_2}(y;\underline{\tau}(x)) = \underline{\tau}(x)f_2(y)\{F_2(y)\}^{\alpha_2-1}\exp\{\alpha_{12}\log[F_1(x)\log[F_2(y)]]\}, \tag{7.11}$$

portanto, segue que as densidades em (7.10) e (7.11) pertencem às famílias exponencial de densidades (7.3) e (7.4), respectivamente, para $X, Y \in \mathbb{R}$, f_i e F_i, para $i = 1, 2$, funções de densidade e de distribuições, respectivamente, fixas e conhecidas. Além disso, de (7.9) concluímos que para $\alpha_{12} = 0$ temos independência das variáveis aleatórias X e Y.

Teorema 7.3.1. *Seja $f(x, y)$ uma densidade bivariada cujas densidades condicionais são como em (7.10) e (7.11). Então,*

$$f(x, y) = k(\boldsymbol{\alpha})f_1(x)f_2(y)\{F_1(x)\}^{\alpha_1-1}\{F_2(y)\}^{\alpha_2-1}\exp\{\alpha_{12}\log[F_1(x)]\log[F_2(y)]\}, \tag{7.12}$$

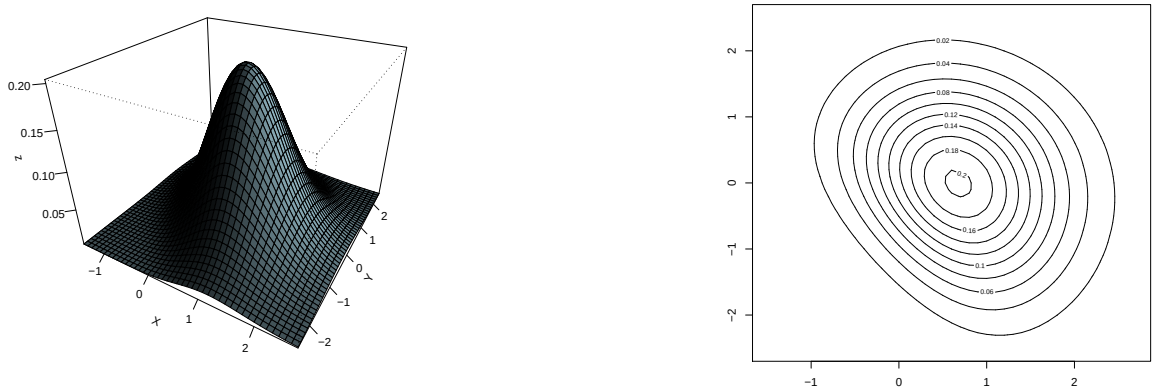
onde $\alpha_1, \alpha_2 > 0$ e $\alpha_{12} \leq 0$, são condições para ter $\int_{\mathbb{R}}\int_{\mathbb{R}} f(x, y)dxdy = 1$ e $k(\boldsymbol{\alpha})$ é uma constante de normalização com $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_{12})'$.

Demostração: Dado que (7.10) e (7.11) são da forma (7.5) o resultado segue do Teorema (7.2.2). Além disso, temos que

$$k(\boldsymbol{\alpha}) = \left[\int_{\mathbb{R}}\int_{\mathbb{R}} f_1(x)f_2(y)\{F_1(x)\}^{\alpha_1-1}\{F_2(y)\}^{\alpha_2-1}\exp\{\alpha_{12}\log[F_1(x)]\log[F_2(y)]\}dxdy\right]^{-1}.$$

Se o vetor aleatório (X,Y) segue o modelo (7.12) utilizaremos a seguinte notação $(X, Y) \sim APB(\alpha_1, \alpha_2, \alpha_{12})$. Para $\alpha_{12} = 0$ temos independência das variáveis aleatórias X e Y, porém, $k(\boldsymbol{\alpha}) = \alpha_1\alpha_2$. A Figura 7.1 mostra o efeito dos parâmetros de assimetria $\alpha_1,\ \alpha_2,\ \alpha_{12}$ quando as distribuições condicionais são normais. As densidades marginais de (7.12) ficam dadas por

$$f_X(x) = k(\boldsymbol{\alpha})\frac{f_1(x)\{F_1(x)\}^{\alpha_1-1}}{\underline{\tau}(x)}, \qquad f_Y(y) = k(\boldsymbol{\alpha})\frac{f_2(y)\{F_2(y)\}^{\alpha_2-1}}{\underline{\omega}(y)}. \tag{7.13}$$



(a) (b)

Figura 7.1 *Gráfico da função de densidade z=f(x,y) e de contorno com distribuições condicionadas normais e parâmetros* $\alpha_1 = 2$, $\alpha_2 = 0.75$ *e* $\alpha_{12} = -0.85$.

As correspondentes funções de regressão são não lineares e tem a forma

$$\mathbb{E}(X|Y = y) = \underline{\omega}(y)\int_0^1 F_1^{-1}(v_1)v_1^{\underline{\omega}(y)-1}dv_1$$

e

$$\mathbb{E}(Y|X = x) = \underline{\tau}(x)\int_0^1 F_2^{-1}(v_2)v_2^{\underline{\tau}(x)-1}dv_2.$$

Para avaliar a correlação entre X e Y temos que a covariância entre X e Y é dada por:

$$cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

onde,

$$\mathbb{E}(XY) = \mathbb{E}(\mathbb{E}[XY|Y]) = k(\boldsymbol{\alpha})\int_0^1\int_0^1 F_1^{-1}(v_1)F_2^{-1}(v_2)v_1^{\underline{\omega}(v_2)-1}v_2^{\alpha_2-1}dv_1dv_2$$

$$\mathbb{E}(X) = k(\boldsymbol{\alpha})\int_0^1 \frac{F_1^{-1}(v_1)v_1^{\alpha_1-1}}{\underline{\tau}(v_1)}dv_1 \quad \text{e} \quad \mathbb{E}(Y) = k(\underline{\boldsymbol{\alpha}})\int_0^1 \frac{F_2^{-1}(v_2)v_2^{\alpha_2-1}}{\underline{\omega}(v_2)}dv_2.$$

Portanto, o coeficiente de correlação entre X e Y, $\rho(X,Y)$, pode ser escrito como

$$\rho_{\boldsymbol{\alpha}}(X,Y) = \frac{m_3(\boldsymbol{\alpha})}{\{m_1(\boldsymbol{\alpha})m_2(\boldsymbol{\alpha})\}^{1/2}} \tag{7.14}$$

onde

$$m_3(\boldsymbol{\alpha}) = \int_0^1\int_0^1 F_1^{-1}(v_1)F_2^{-1}(v_2)v_1^{\alpha_1-1}v_2^{\alpha_2-1} \left[v_1^{\alpha_{12}\log(v_2)} - \frac{1}{\underline{\omega}(v_2)\underline{\tau}(v_1)}\right]dv_1dv_2,$$

$$m_1(\boldsymbol{\alpha}) = \int_0^1 \frac{[F_1^{-1}(v_1)]^2v_1^{\alpha_1-1}}{\underline{\tau}(v_1)}dv_1 - \left\{\int_0^1 \frac{F_1^{-1}(v_1)v_1^{\alpha_1-1}}{\underline{\tau}(v_1)}dv_1\right\}^2$$

e

$$m_2(\boldsymbol{\alpha}) = \int_0^1 \frac{[F_2^{-1}(v_2)]^2v_2^{\alpha_2-1}}{\underline{\omega}(v_2)}dv_2 - \left\{\int_0^1 \frac{F_2^{-1}(v_2)v_2^{\alpha_2-1}}{\underline{\omega}(v_2)}dv_2\right\}^2.$$

Nós obtemos que para $\alpha_{12} = 0$ ou $\alpha_{12} \to -\infty$ então $\rho = 0$. Mesmo assim, se α_{12} é fixo e $\alpha_1 \to \infty$ ou $\alpha_2 \to \infty$, então $\rho = 0$.

Um estimador de $\rho(X,Y)$ fica dado por

$$\hat{\rho}_{\boldsymbol{\alpha}}(X,Y) = \frac{m_3(\hat{\boldsymbol{\alpha}})}{\{m_1(\hat{\boldsymbol{\alpha}})m_2(\hat{\boldsymbol{\alpha}})\}^{1/2}}.$$

Para o caso da distribuição bivariada α-potência normal assimétrica, nós encontramos, depois de um processo de simulação, que para $\alpha_1, \alpha_2 \in (0.4, 100)$, $\alpha_2 \in (0.4, 100)$ e $\alpha_{12} \in (-100, -0.2)$, então $\rho \in (-0.8634, 0.9247)$. Além disso, para $\alpha_1 > 1.4$ e $\alpha_2 > 1.4$ se tem que $\rho > 0$, enquanto que para $\alpha_1 < 1.3$, $\alpha_2 < 1.3$ e $\alpha_{12} > -1$ se tem que $\rho < 0$. Os resultados foram obtidos mediante o método "adapt", o qual calcula integrais múltiplas e o método "integrate"o qual calcula integrais univariadas, do pacote R. Como uma observação ao modelo apresentado em (7.12), temos que as transformações

$V_1 = -\log(F_1(X))$ e $V_2 = -\log(F_2(Y))$ levam ao modelo exponencial bivariado condicional

$$f(v_1, v_2) = k(\boldsymbol{\alpha}) \exp(-\alpha_1 v_1 - \alpha_2 v_2 + \alpha_{12} v_1 v_2) \tag{7.15}$$

amplamente estudado em Arnold, Castillo e Sarabia (1999). Para $\alpha_1 = \alpha_2 = 1$ nós temos que

$$k(\alpha_{12}) = \frac{\alpha_{12} \exp(-1/\alpha_{12})}{-Ei(1/\alpha_{12})}$$

com

$$-Ei(u) = \int_u^\infty \frac{\exp(-w)}{w} dw.$$

Além disso, o coeficiente de correlação entre V_1 e V_2 é dado pela expressão

$$\rho(V_1, V_2) = \frac{\alpha_{12} + k(\alpha_{12}) - k^2(\alpha_{12})}{k(\alpha_{12})(1 + \alpha_{12} - k(\alpha_{12}))}, \tag{7.16}$$

o qual é sempre negativo e limitado inferiormente pelo valor -0.32. Enquanto que para (7.14), $-0.20 < \rho(X, Y) < 0.60$, para $\alpha_{12} \in (-1000, 0)$. Assim, nós concluímos que $\rho(V_1, V_2) = \rho(T_1(X), T_2(Y)) \neq \rho(X, Y)$, para certas funções T_1 e T_2. Portanto as características do modelo (7.12) não podem ser estudadas completamente pelo modelo exponencial condicional (7.15). Ainda algumas vantagens podem ser obtidas desta reparametrização, por exemplo, nos processos de estimação pode-se obter expressões mais simples para processos de cálculo. Assim, nós obtemos estimadores de momentos para os parâmetros do modelo (7.12) a partir do modelo (7.15), dados pelas expressões $\widetilde{\alpha}_1 = \frac{\widetilde{\alpha}_0}{\overline{v}_1(\widetilde{\alpha}_0 + T(\widetilde{\alpha}_0 - 1))}$, $\widetilde{\alpha}_2 = \frac{\widetilde{\alpha}_0}{\overline{v}_2(\widetilde{\alpha}_0 + T(\widetilde{\alpha}_0 - 1))}$ e $\widetilde{\alpha}_{12} = \frac{\widetilde{\alpha}_0(\widetilde{\alpha}_0 - 1)}{\overline{v}_1 \overline{v}_2(\widetilde{\alpha}_0 + T(\widetilde{\alpha}_0 - 1))}$, os quais são consistentes e assintoticamente normal (Arnold, Castillo e Sarabia (1999)) e onde $\widetilde{\alpha}_0 = \frac{T}{1 + R_{v_1 v_2} T}$, com $R_{v_1 v_2} = cor(v_1, v_2)$, e $T = cv(v_1) cv(v_2)$, sendo cor a correlação entre V_1 e V_2 e $cv(u) = \sqrt{S_u^2}/\overline{u}$.

7.4 Caso Localização - Escala Bivariado

Seja $(Z_1, Z_2) \sim APB(\alpha_1, \alpha_2, \alpha_{12})$, então as transformações lineares $X = \xi_1 + \eta_1 Z_1$ e $Y = \xi_2 + \eta_2 Z_2$, onde ξ_1, η_1 são os parâmetros de localização e escala, respectivamente,

da variável X e ξ_2, η_2 são os parâmetros de localização e escala da variável Y, levam ao modelo de localização - escala α-potência assimétrico bivariado.

$$f(x,y) = \frac{k(\boldsymbol{\theta})}{\eta_1\eta_2} f_1\left(\frac{x-\xi_1}{\eta_1}\right) f_2\left(\frac{y-\xi_2}{\eta_2}\right)\left\{F_1\left(\frac{x-\xi_1}{\eta_1}\right)\right\}^{\alpha_1-1}\left\{F_2\left(\frac{y-\xi_2}{\eta_2}\right)\right\}^{\alpha_2-1}$$
$$\exp\left\{\alpha_{12}\log\left[F_1\left(\frac{x-\xi_1}{\eta_1}\right)\right]\log\left[F_2\left(\frac{y-\xi_2}{\eta_2}\right)\right]\right\}, \quad (7.17)$$

onde $\alpha_1, \alpha_2 > 0$ e $\alpha_{12} \leq 0$, são condições para ter $\int_{\mathbb{R}}\int_{\mathbb{R}} f(x,y)dxdy = 1$ e $k(\boldsymbol{\theta})$ é uma constante de normalização. Neste caso utilizaremos a notação $(X,Y) \sim APB(\xi_1, \xi_2, \eta_1, \eta_2, \alpha_1, \alpha_2, \alpha_{12})$.

7.4.1 Estimação no Caso Localização - Escala

Para uma amostra aleatória de tamanho n, $(X_1, Y_1), ..., (X_n, Y_n)$, da densidade $f(x, y; \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$, a função de log-verossimilhança fica dada por

$$\ell(\boldsymbol{\theta}; X, Y) = \sum_{i=1}^{n}\{\log(k(\boldsymbol{\alpha})) - \log(\eta_1\eta_2) + \log[f_1(z_{1i})f_2(z_{2i})] + \alpha_{12}\log[F_1(z_{1i})]\log[F_2(z_{2i})]\}$$
$$+\sum_{i=1}^{n}\{(\alpha_1-1)\log[F_1(z_{1i})] + (\alpha_2-1)\log[F_2(z_{2i})]\},$$

onde $z_{1i} = \frac{x_i-\xi_1}{\eta_1}$ e $z_{2i} = \frac{y_i-\xi_2}{\eta_2}$. Assim, supomos que as derivadas f_1' e f_2' existem, então a função escore fica dada por

$U(\alpha_1) = \sum_{i=1}^{n}\{d_{\alpha_1} + \log[F_1(z_{1i})]\}$, $U(\alpha_2) = \sum_{i=1}^{n}\{d_{\alpha_2} + \log[F_2(z_{2i})]\}$,

$U(\alpha_{12}) = \sum_{i=1}^{n}\{d_{\alpha_{12}} + \log[F_1(z_{1i})]\ln[F_2(z_{2i})]\}$,

$U(\xi_1) = \frac{1}{\eta_1}\sum_{i=1}^{n}\left\{d_{\xi_1} - \frac{f_1'(z_{1i})}{f_1(z_{1i})} - \frac{f_1(z_{1i})}{F_1(z_{1i})}[\alpha_{12}\log[F_2(z_{2i})] + \alpha_1 - 1]\right\}$,

$U(\xi_2) = \frac{1}{\eta_2}\sum_{i=1}^{n}\left\{d_{\xi_2} - \frac{f_2'(z_{2i})}{f_2(z_{2i})} - \frac{f_2(z_{2i})}{F_2(z_{2i})}[\alpha_{12}\log[F_1(z_{1i})] + \alpha_2 - 1]\right\}$,

$U(\eta_1) = \frac{1}{\eta_1}\sum_{i=1}^{n}\left\{d_{\eta_1} - 1 - z_{1i}\frac{f_1'(z_{1i})}{f_1(z_{1i})} - z_{1i}\frac{f_1(z_{1i})}{F_1(z_{1i})}[\alpha_{12}\log[F_2(z_{2i})] + \alpha_1 - 1]\right\}$ e

$U(\eta_2) = \frac{1}{\eta_2}\sum_{i=1}^{n}\left\{d_{\eta_2} - 1 - z_{2i}\frac{f_2'(z_{2i})}{f_2(z_{2i})} - z_{2i}\frac{f_2(z_{2i})}{F_2(z_{2i})}[\alpha_{12}\log[F_1(z_{1i})] + \alpha_2 - 1]\right\}$

onde

$d_{\alpha_1} = \int_{\mathbb{R}}\int_{\mathbb{R}}\log[F_1(z_1)]f(x,y)dxdy$, $d_{\alpha_2} = \int_{\mathbb{R}}\int_{\mathbb{R}}\log[F_2(z_2)]f(x,y)dxdy$,

$d_{\alpha_{12}} = \int_{\mathbb{R}}\int_{\mathbb{R}}\log[F_1(z_1)]\log[F_2(z_2)]f(x,y)dxdy$,

$d_{\xi_1} = -\int_{\mathbb{R}}\int_{\mathbb{R}} f(x,y)\left[\frac{f_1'(z_1)}{f_1(z_1)} + (\alpha_1-1)\{F_1(z_1)\}^{-1} + \alpha_{12}\frac{f_1(z_1)}{F_1(z_1)}\log[F_2(z_2)]\right]dxdy$,

$d_{\xi_2} = -\int_{\mathbb{R}}\int_{\mathbb{R}} f(x,y)\left[\frac{f_2'(z_2)}{f_2(z_2)} + (\alpha_2-1)\{F_2(z_2)\}^{-1} + \alpha_{12}\frac{f_2(z_2)}{F_2(z_2)}\log[F_1(z_1)]\right]dxdy$,

$d_{\eta_1} = 1 - \int_{\mathbb{R}} \int_{\mathbb{R}} z_1 f(x,y) \left[\frac{f_1'(z_1)}{f_1(z_1)} + (\alpha_1 - 1)\{F_1(z_1)\}^{-1} + \alpha_{12} \frac{f_1(z_1)}{F_1(z_1)} \log[F_2(z_2)] \right] dxdy$ e
$d_{\eta_2} = 1 - \int_{\mathbb{R}} \int_{\mathbb{R}} z_2 f(x,y) \left[\frac{f_2'(z_2)}{f_2(z_2)} + (\alpha_2 - 1)\{F_2(z_2)\}^{-1} + \alpha_{12} \frac{f_2(z_2)}{F_2(z_2)} \log[F_1(z_1)] \right] dxdy.$

O sistema de equações acima geralmente tem solução para $\boldsymbol{\theta} = (\xi_1, \xi_2, \eta_1, \eta_2, \alpha_1, \alpha_2, \alpha_{12})'$ por métodos numéricos iterativos.

Estimação por Pseudo-Verossimilhança de Densidades Condicionais

Um caminho para evitar a constante de normalização $k(\boldsymbol{\theta})$ para a estimação dos parâmetros da distribuição APB, esta baseada na analise sobre distribuições condicionais. Então, suponha que $(X_1, Y_1), ..., (X_n, Y_n)$ é uma amostra aleatória de alguma densidade $f(x, y; \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$, a pseudo verossimilhança de $\boldsymbol{\theta}$ dado o vetor (X,Y), é definida a partir das densidades condicionais como

$$PL(\boldsymbol{\theta}; (X,Y)) = \prod_{i=1}^{n} f_{X|Y}(x_i|y_i, \boldsymbol{\theta}) f_{Y|X}(y_i|x_i, \boldsymbol{\theta}). \tag{7.18}$$

Besag (1975), define o estimador de máxima pseudo verossimilhança de $\boldsymbol{\theta}$ como o valor $\boldsymbol{\theta}_0$ de $\boldsymbol{\theta}$ que maximiza a função de pseudo verossimilhança.

Para $X|Y = y \sim AP_{f_1}(\xi_1, \eta_1, \underline{\omega}(y))$ e $Y|X = x \sim AP_{f_2}(\xi_2, \eta_2, \underline{\tau}(x))$, o logaritmo da pseudo-verossimilhança para as distribuições condicionais fica dado por

$$\begin{aligned} P\ell(\boldsymbol{\theta}|X,Y) &= \sum_{i=1}^{n} \left\{ \log[g_{X|Y}(x_i|y_i)] + \log[g_{Y|X}(y_i|x_i)] \right\} \\ &= -n \log(\eta_1 \eta_2) + \sum_{i=1}^{n} \log[(\alpha_2 + \alpha_{12} \log[F_1(z_{1i})])(\alpha_1 + \alpha_{12} \log[F_2(z_{2i})])] \\ &+ \sum_{i=1}^{n} \log[f_1(z_{1i}) f_2(z_{2i})] + \sum_{i=1}^{n} [(\alpha_2 - 1) + \alpha_{12} \log[F_1(z_{1i})]] \log[F_2(z_{2i})] \\ &+ \sum_{i=1}^{n} [(\alpha_1 - 1) + \alpha_{12} \log[F_2(z_{2i})]] \log[F_1(z_{1i})], \end{aligned} \tag{7.19}$$

onde $z_{1i} = \frac{x_i - \xi_1}{\eta_1}$ e $z_{2i} = \frac{y_i - \xi_2}{\eta_2}$. Então a função de pseudo escore é dada por

$$U_P(\alpha_1) = \sum_{i=1}^{n} \left[\frac{1}{\alpha_1 + \alpha_{12} \log[F_2(z_{2i})]} + \log[F_1(z_{1i})] \right],$$

$$U_P(\alpha_2) = \sum_{i=1}^{n} \left[\frac{1}{\alpha_2 + \alpha_{12} \log[F_1(z_{1i})]} + \log[F_2(z_{2i})] \right],$$

$$U_P(\alpha_{12}) = \sum_{i=1}^{n} \left[\frac{\log[F_2(z_{2i})]}{\alpha_1 + \alpha_{12} \log[F_2(z_{2i})]} + \frac{\log[F_1(z_{1i})]}{\alpha_2 + \alpha_{12} \log[F_1(z_{1i})]} + \log[F_1^2(z_{1i})] \log[F_2(z_{2i})] \right],$$

$$U_P(\xi_1) = -\frac{\alpha_{12}}{\eta_1} \sum_{i=1}^{n} \frac{f_1(z_{1i})}{F_1(z_{1i})} \left[\frac{1}{\alpha_2 + \alpha_{12} \log[F_1(z_{1i})]} + 2 \log[F_2(z_{2i})] \right] - \frac{1}{\eta_1} \sum_{i=1}^{n} \left[\frac{f_1'(z_{1i})}{f_1(z_{1i})} + (\alpha_1 - 1) \frac{f_1(z_{1i})}{F_1(z_{1i})} \right],$$

$$U_P(\xi_2) = -\frac{\alpha_{12}}{\eta_2} \sum_{i=1}^{n} \frac{f_2(z_{2i})}{F_2(z_{2i})} \left[\frac{1}{\alpha_1 + \alpha_{12} \log[F_2(z_{2i})]} + 2 \log[F_1(z_{1i})] \right] - \frac{1}{\eta_2} \sum_{i=1}^{n} \left[\frac{f_2'(z_{2i})}{f_2(z_{2i})} + (\alpha_2 - 1) \frac{f_2(z_{2i})}{F_2(z_{2i})} \right],$$

$$U_P(\eta_1) = -\frac{\alpha_{12}}{\eta_1} \sum_{i=1}^{n} z_{1i} \frac{f_1(z_{1i})}{F_1(z_{1i})} \left[\frac{1}{\alpha_2 + \alpha_{12} \log[F_1(z_{1i})]} + 2 \log[F_2(z_{2i})] \right] - \frac{1}{\eta_1} \sum_{i=1}^{n} z_{1i} \left[\frac{f_1'(z_{1i})}{f_1(z_{1i})} + (\alpha_1 - 1) \frac{f_1(z_{1i})}{F_1(z_{1i})} \right],$$

$$U_P(\eta_2) = -\frac{\alpha_{12}}{\eta_2} \sum_{i=1}^{n} z_{2i} \frac{f_2(z_{2i})}{F_2(z_{2i})} \left[\frac{1}{\alpha_1 + \alpha_{12} \log[F_2(z_{2i})]} + 2 \log[F_1(z_{1i})] \right] - \frac{1}{\eta_2} \sum_{i=1}^{n} z_{2i} \left[\frac{f_2'(z_{2i})}{f_2(z_{2i})} + (\alpha_2 - 1) \frac{f_2(z_{2i})}{F_2(z_{2i})} \right].$$

Nosso interesse agora é testar

$$H_0 : (\alpha_1, \alpha_2, \alpha_{12})' = (1, 1, 0)' \ \text{ contra } \ H_1 : (\alpha_1, \alpha_2, \alpha_{12})' \neq (1, 1, 0)',$$

a qual é equivalente a testar

$$H_0 : G\boldsymbol{\theta} = d \ \text{ contra } \ H_1 : G\boldsymbol{\theta} \neq d,$$

onde G e uma matriz e d é um vetor de dimensões apropriadas, respectivamente. Sob H_0

estamos comparando o modelo APB contra um modelo de variáveis aleatórias independentes, X e Y. A estatística de razão de pseudoverossimilhanças

$$G^* = 2\left[lp(\widetilde{\boldsymbol{\theta}}) - lp(\widetilde{\boldsymbol{\theta}}_0)\right],$$

segue uma distribuição que é a soma ponderada $\sum_{i=1}^{3} \lambda_j \chi_{1j}^2$ onde χ_{1j}^2 são variáveis aleatórias independentes qui-quadrado com um grau de liberdade. Para J, a matriz definida como menos a segunda derivada da função de pseudo-verossimilhança e $\Gamma = J^{-1}KJ^{-1}$, onde K é a matriz de produtos cruzados do vetor de primeiras derivadas, $\lambda_1 \geq \lambda_2 \geq \lambda_3$ são os autovalores da matriz $J^{33}\Gamma_{33}$, onde J^{33} e Γ_{33} são submatrizes da inversa J^{-1} e da Γ, respectivamente, correspondente ao vetor $(\alpha_1, \alpha_2, \alpha_{12})'$ (Geys et al., 1999).

Un segundo teste pode ser implementado a partir da estatística tipo Wald como se mostra a continuação. De acordo à normalidade assintótica do pseudo estimador $\widetilde{\boldsymbol{\theta}}$ nós temos que a pseudo estatística tipo Wald para o estimador de máxima pseudo verossimilhança,

$$W_n = \left(G\widetilde{\boldsymbol{\theta}} - d\right)' \left[\hat{\Gamma}_{3x3}(\widetilde{\boldsymbol{\theta}})\right]^{-1} \left(G\widetilde{\boldsymbol{\theta}} - d\right),$$

onde $\hat{\Gamma}_{33}(\widetilde{\boldsymbol{\theta}})$ é a submatriz de $\Gamma(\widetilde{\boldsymbol{\theta}})$ correspondente ao vetor $\boldsymbol{\theta}_3 = (\alpha_1, \alpha_2, \alpha_{12})'$, $G = (\mathbf{I}_3, \mathbf{0}_{3x4})'$ e $d = (1, 1, 0)'$, segue uma distribuição qui-quadrado com 3 graus de liberdade sob a hipóteses nula, (Geys et al., 1999).

Para comparar o modelo normal bivariado contra o modelo α-potência assimétrico bivariado se requer um teste na teoria de modelos não encaixados. Outros critérios como o método de Akaike (1974), AIC, também pode ser usado. Modificações do AIC incluem o critério de informação de Akaike consistente

$$CAIC = -2\hat{\ell}(\cdot) + (1 + \log(n))k,$$

onde k é o número de parâmetros do modelo em questão. O melhor modelo é selecionado visando obter um mínimo AIC.

7.5 Distribuição PN Bivariada

Suponha que

$$X|Y = y \sim PN(\alpha_1), \quad \forall x \in \mathbb{R}, \qquad Y|X = x \sim PN(\alpha_2), \quad \forall y \in \mathbb{R}. \tag{7.20}$$

Então X e Y são variáveis aleatórias independentes e assim,

$$f(x,y) = \alpha_1\alpha_2\phi(x)\phi(y)\{\Phi(x)\}^{\alpha_1-1}\{\Phi(y)\}^{\alpha_2-1}, \tag{7.21}$$

a qual para $\alpha_1 = \alpha_2 = \alpha$ pode-se escrever

$$f(x,y) = \alpha^2 f(x,y)\{F(x,y)\}^{\alpha-1},$$

onde $f(x,y)$ é normal bivariada de parâmetros $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_{xy}$, que nós denotamos por $NB(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_{xy})$. Então segue que para $\alpha_1 = \alpha_2 = 1$ temos a distribuição normal bivariada de variáveis aleatórias independentes, $NBI(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$.

O caso geral de localização - escala da distribuição α-potência normal assimétrica bivariada de parâmetros $\boldsymbol{\theta} = (\xi_1, \xi_2, \eta_1, \eta_2, \alpha_1, \alpha_2, \alpha_{12})'$ fica dado por

$$f(x,y) = \frac{k(\boldsymbol{\theta})}{\eta_1\eta_2}\phi\left(\frac{x-\xi_1}{\eta_1}\right)\phi\left(\frac{y-\xi_2}{\eta_2}\right)\left\{\Phi\left(\frac{x-\xi_1}{\eta_1}\right)\right\}^{\alpha_1-1}\left\{\Phi\left(\frac{y-\xi_2}{\eta_2}\right)\right\}^{\alpha_2-1}$$
$$\exp\left\{\alpha_{12}\log\left[\Phi\left(\frac{x-\xi_1}{\eta_1}\right)\right]\log\left[\Phi\left(\frac{y-\xi_2}{\eta_2}\right)\right]\right\}. \tag{7.22}$$

Usamos a notação $(X,Y) \sim PNB(\xi_1, \xi_2, \eta_1, \eta_2, \alpha_1, \alpha_2, \alpha_{12})'$ para denotar este caso particular da distribuição APB. A função de log-verossimilhança fica dada por

$$\ell(\boldsymbol{\theta}|X,Y) = \sum_{i=1}^{n}\left\{\log(k(\boldsymbol{\theta})) - \log(\eta_1\eta_2) - \log(2\pi) - \frac{1}{2}\left(\frac{x_i-\xi_1}{\eta_1}\right)^2 - \frac{1}{2}\left(\frac{y_i-\xi_2}{\eta_2}\right)\right\}$$
$$+\sum_{i=1}^{n}\left\{\alpha_{12}\log\left[\Phi(z_{1i})\right]\log\left[\Phi(z_{2i})\right] + (\alpha_1-1)\log\left[\Phi(z_{1i})\right] + (\alpha_2-1)\log\left[\Phi(z_{2i})\right]\right\},$$

onde $z_{1i} = \frac{x_i-\xi_1}{\eta_1}$ e $z_{2i} = \frac{y_i-\xi_2}{\eta_2}$. Assim, a função escore fica dada por

$U(\alpha_1) = \sum_{i=1}^{n}\left\{d_{\alpha_1} + \log[\Phi(z_{1i})]\right\}, \quad U(\alpha_2) = \sum_{i=1}^{n}\left\{d_{\alpha_2} + \log[\Phi(z_{2i})]\right\},$

$U(\alpha_{12}) = \sum_{i=1}^{n}\left\{d_{\alpha_{12}} + \log[\Phi(z_{1i})]\log[\Phi(z_{2i})]\right\},$

$U(\xi_1) = \frac{1}{\eta_1}\sum_{i=1}^{n}\left\{d_{\xi_1} + z_{1i} - w_{1i}\left[\alpha_{12}\log[\Phi(z_{2i})] + \alpha_1 - 1\right]\right\},$

$U(\xi_2) = \frac{1}{\eta_2}\sum_{i=1}^{n}\left\{d_{\xi_2} + z_{2i} - w_{2i}\left[\alpha_{12}\log[\Phi(z_{1i})] + \alpha_2 - 1\right]\right\},$
$U(\eta_1) = \frac{1}{\eta_1}\sum_{i=1}^{n}\left\{d_{\eta_1} + z_{1i}^2 - w_{1i}z_{1i}\left[\alpha_{12}\log[\Phi(z_{2i})] + \alpha_1 - 1\right]\right\}$ e
$U(\eta_2) = \frac{1}{\eta_2}\sum_{i=1}^{n}\left\{d_{\eta_2} + z_{2i}^2 - w_{2i}z_{2i}\left[\alpha_{12}\log[\Phi(z_{1i})] + \alpha_2 - 1\right]\right\},$
onde $w_{1i} = \frac{\phi(z_{1i})}{\Phi(z_{1i})}$, $w_{2i} = \frac{\phi(z_{2i})}{\Phi(z_{2i})}$ e $d_{\alpha_1}, d_{\alpha_2}, ..., d_{\eta_2}$ são definidas como acima.

Agora, para $X|Y = y \sim PN(\xi_1, \eta_1, \underline{\omega}(y))$ e $Y|X = x \sim PN(\xi_2, \eta_2, \underline{\tau}(x))$, a função logaritmo da pseudo verossimilhança para as distribuições condicionais fica dada por

$$\begin{aligned}
P\ell(\boldsymbol{\theta}|X,Y) &= \sum_{i=1}^{n}\left\{\log[f_{X|Y}(x_i|y_i)] + \log[f_{Y|X}(y_i|x_i)]\right\} \\
&= -n\log(\eta_1\eta_2) + \sum_{i=1}^{n}\log[(\alpha_2 + \alpha_{12}\log[\Phi(z_{1i})])(\alpha_1 + \alpha_{12}\log[\Phi(z_{2i})])] \\
&+ \sum_{i=1}^{n}\log[\phi(z_{1i})\phi(z_{2i})] + \sum_{i=1}^{n}[(\alpha_2 - 1) + \alpha_{12}\log[\Phi(z_{1i})]]\log[\Phi(z_{2i})] \\
&+ \sum_{i=1}^{n}[(\alpha_1 - 1) + \alpha_{12}\log[\Phi(z_{2i})]]\log[\Phi(z_{1i})], \qquad (7.23)
\end{aligned}$$

onde $z_{1i} = \frac{x_i - \xi_1}{\eta_1}$ e $z_{2i} = \frac{y_i - \xi_2}{\eta_2}$. Então os pseudo escores são dados por

$$U_P(\alpha_1) = \sum_{i=1}^{n}\left[\frac{1}{\alpha_1 + \alpha_{12}\log[\Phi(z_{2i})]} + \log[\Phi(z_{1i})]\right],$$

$$U_P(\alpha_2) = \sum_{i=1}^{n}\left[\frac{1}{\alpha_2 + \alpha_{12}\log[\Phi(z_{1i})]} + \log[\Phi(z_{2i})]\right],$$

$$U_P(\alpha_{12}) = \sum_{i=1}^{n}\left[\frac{\log[\Phi(z_{2i})]}{\alpha_1 + \alpha_{12}\log[\Phi(z_{2i})]} + \frac{\log[\Phi(z_{1i})]}{\alpha_2 + \alpha_{12}\log[\Phi(z_{1i})]} + \log[\Phi^2(z_{1i})]\log[\Phi(z_{2i})]\right],$$

$$\begin{aligned}
U_P(\xi_1) = &-\frac{\alpha_{12}}{\eta_1}\sum_{i=1}^{n}w_{1i}\left[\frac{1}{\alpha_2 + \alpha_{12}\log[\Phi(z_{1i})]} + 2\log[\Phi(z_{2i})]\right] - \\
&\frac{1}{\eta_1}\sum_{i=1}^{n}\left[z_{1i} + (\alpha_1 - 1)w_{1i}\right],
\end{aligned}$$

$$\begin{aligned}
U_P(\xi_2) = &-\frac{\alpha_{12}}{\eta_2}\sum_{i=1}^{n}w_{2i}\left[\frac{1}{\alpha_2 + \alpha_{12}\log[\Phi(z_{2i})]} + 2\log[\Phi(z_{1i})]\right] - \\
&\frac{1}{\eta_2}\sum_{i=1}^{n}\left[z_{2i} + (\alpha_2 - 1)w_{2i}\right],
\end{aligned}$$

$$U_P(\eta_1) = -\frac{n}{\eta_1} - \frac{\alpha_{12}}{\eta_1}\sum_{i=1}^{n} w_{1i}z_{1i}\left[\frac{1}{\alpha_2 + \alpha_{12}\log[\Phi(z_{1i})]} + 2\log[\Phi(z_{2i})]\right] - \frac{1}{\eta_1}\sum_{i=1}^{n}\left[z_{1i}^2 + (\alpha_1 - 1)w_{1i}z_{1i}\right],$$

$$U_P(\eta_2) = -\frac{n}{\eta_2} - \frac{\alpha_{12}}{\eta_2}\sum_{i=1}^{n} w_{2i}z_{2i}\left[\frac{1}{\alpha_1 + \alpha_{12}\log[\Phi(z_{2i})]} + 2\log[\Phi(z_{1i})]\right] - \frac{1}{\eta_2}\sum_{i=1}^{n}\left[z_{2i}^2 + (\alpha_2 - 1)w_{2i}z_{2i}\right].$$

7.6 Extensão Para p Variáveis

Uma extensão natural do Teorema (7.3.1) pode ser feita para o caso no qual $\mathbf{X} = (X_1, X_2, ..., X_p)$ é p-dimensional. A seguinte notação é conveniente para descrever tal extensão. Para cada $j = 1, 2, ..., p$, definimos o vetor $\mathbf{X}_{(j)}$ o qual é o vetor aleatório $(p-1)$-dimensional obtido a partir de $\mathbf{X}$ excluindo X_j. De forma paralela para um vetor real $\mathbf{x} = (x_1, x_2, ..., x_p)$ p-dimensional $\mathbf{x}_{(j)}$ é obtido por exclusão da j-ésima coordenada x_j de $\mathbf{x}$. Suponha agora que para cada $j = 1, 2, ..., p$ a distribuição condicional de $\mathbf{X}$ dado $\mathbf{X}_{(j)} = \mathbf{x}_{(j)}$ é uma distribução α-potência com um parâmetro o qual é função de $\mathbf{x}_{(j)}$. Assim, nós assumimos, para cada $j = 1, 2, ..., p$, que

$$X_j|\mathbf{X}_{(j)} = \mathbf{x}_{(j)} \sim AP_{f_j}(\underline{\omega}_j(\mathbf{x}_{(j)})), \tag{7.24}$$

para alguma densidade continua e diferenciável f_j e alguma função

$$\underline{\omega}_j : \mathbb{R}^{p-1} \to \mathbb{R}_+.$$

Este é um exemplo de uma distribuição p-dimensional com condicionais numa família paramétrica exponencial e, consequentemente, usando argumentos paralelos aos utilizados na seção anterior, para o caso bivariado $p = 2$, podemos identificar a família em geral de todas as densidades $p-$dimensional satisfazendo (7.24), isto é, com condicionais α-potência. A forma geral desta família é dada por

$$f(\mathbf{x};\boldsymbol{\alpha}) = k(\boldsymbol{\alpha})\left[\prod_{j=1}^{p} f_j(x_j)\right] \times exp\left[\sum_{\underline{s}\in\Psi_p} \alpha_{\underline{s}} \prod_{j=1}^{p}[\log F_j(x_j)]^{s_j}\right], \quad \mathbf{x} \in \mathbb{R}^p, \tag{7.25}$$

onde Ψ_p é o conjunto de todos os vetores de 0's e 1's de dimensão p com não todos seus elementos iguais a 0. A dimensão do espaço de parâmetros para esta família é $2^p - 1$ o qual para p grande constituiria um excesso de parâmetros e será razoável considerar submodelos simplificados de (7.25), que ainda justificadamente suas condicionais possam ser chamados distribuições condicionais α-potência.

O modelo que nós recomendamos pode-se dizer que envolve apenas as interações de primeira ordem, ou seja onde $\sum_{j=1}^{p} s_j \leq 2$. Este pode ser escrito usando notação paramétrica simplificada como

$$f(\mathbf{x};\boldsymbol{\alpha}) = k(\boldsymbol{\alpha}) \prod_{j=1}^{p} f_j(x_j)[F_j(x_j)]^{\alpha_j - 1} \times exp\left[\sum_{1\leq j<k\leq p} \alpha_{jk} \log[F_j(x_j)] \log[F_k(x_k)]\right], \quad \mathbf{x} \in \mathbb{R}^p, \tag{7.26}$$

onde $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_p, \alpha_{12}, ..., \alpha_{(p-1)p})'$ no qual $\alpha_j > 0, \;\; j = 1, 2, ..., p$ and $\alpha_{jk} \leq 0, \;\; 1 \leq j < k \leq p$. Se um vetor aleatório $\mathbf{X}$ tem a densidade (7.26), usamos a notação $\mathbf{X} \sim APM(\boldsymbol{\alpha})$.

Para obter mais flexibilidade nós introduzimos parâmetros de localização e escala em (7.26). Assim, nós obtemos a seguinte família de densidades:

$$\begin{aligned} f(\mathbf{x};\boldsymbol{\theta}) = k(\boldsymbol{\theta}) \prod_{j=1}^{p} \frac{1}{\eta_j} f_j\left(\frac{x_j - \xi_j}{\eta_j}\right) \left[F_j\left(\frac{x_j - \xi_j}{\eta_j}\right)\right]^{\alpha_j - 1} \\ \exp\left[\sum_{1\leq j<k\leq p} \alpha_{jk} \log\left[F_j\left(\frac{x_j - \xi_j}{\eta_j}\right)\right] \log\left[F_k\left(\frac{x_k - \xi_k}{\eta_k}\right)\right]\right], \quad \underline{x} \in \mathbb{R}^p, \end{aligned} \tag{7.27}$$

com $\boldsymbol{\theta} = (\boldsymbol{\xi}', \boldsymbol{\eta}', \boldsymbol{\alpha}')'$, onde $\xi_1, ..., \xi_p$ são parâmetros de localização e $\eta_1, \eta_2, ..., \eta_p$ são parâmetros de escala. Utilizaremos a notação $\mathbf{X} \sim APM(\boldsymbol{\xi}', \boldsymbol{\eta}', \boldsymbol{\alpha}')'$.

7.6.1 Estimação

Para uma amostra aleatória de tamanho n, $X_1, ..., X_n$, onde $X_i = (x_{i1}, ..., x_{ip})$, da densidade $f(\mathbf{x};\boldsymbol{\theta})$, a função de log-verossimilhança fica dada por

$$\begin{aligned} \ell(\boldsymbol{\theta};\mathbf{x}) = \sum_{i=1}^{n} \log(k(\boldsymbol{\theta})) - n\sum_{j=1}^{p} \log(\eta_j) + \sum_{i=1}^{n}\sum_{j=1}^{p} \log\left[f_j(z_{ji})\right] + \\ \sum_{i=1}^{n} \sum_{\substack{1\leq j\leq p-1 \\ j<k\leq p}} \alpha_{jk} \log\left[F_j(z_{ji})\right] \log\left[F_k(z_{ki})\right] + \sum_{i=1}^{n}\sum_{j=1}^{p} (\alpha_j - 1) \log\left[F_j(z_{ji})\right], \end{aligned}$$

onde $z_{ji} = \frac{x_j - \xi_j}{\eta_j}$ para $j = 1, 2, ..., p$. Supomos agora que as derivadas f_j' existem para $j = 1, 2, ..., p$, então a função escore fica dada por:

$$U(\alpha_j) = \sum_{i=1}^{n} \left\{ d_{\alpha_j} + \log[F_j(z_{ji})] \right\},$$

$$U(\alpha_{jk}) = \sum_{i=1}^{n} \left\{ d_{\alpha_{jk}} + \log[F_j(z_{ji})] \log[F_k(z_{ki})] \right\},$$

$$U(\xi_j) = \frac{1}{\eta_j} \sum_{i=1}^{n} \left\{ d_{\xi_j} - \frac{f_1'(z_{ji})}{f_j(z_{ji})} - \frac{f_j(z_{ji})}{F_j(z_{ji})} \left[\sum_{\substack{k=1 \\ k \neq j}}^{p} \alpha_{jk} \log[F_k(z_{ki})] + \alpha_j - 1 \right] \right\},$$

$$U(\eta_j) = \frac{1}{\eta_j} \sum_{i=1}^{n} \left\{ d_{\eta_j} - 1 - z_{ji} \frac{f_j'(z_{ji})}{f_j(z_{ji})} - z_{ji} \frac{f_j(z_{ji})}{F_j(z_{ji})} \left[\sum_{\substack{k=1 \\ k \neq j}}^{p} \alpha_{jk} \log[F_k(z_{ki})] + \alpha_j - 1 \right] \right\},$$

para $j, k = 1, 2, ..., p$ e $j \neq k$, onde $d_{\theta_j} = \frac{\partial \log(k(\boldsymbol{\theta}))}{\partial \theta_j}$. O sistema de equações acima geralmente tem soluções por métodos numéricos iterativos.

7.6.2 Estimação por Pseudo-Verossimilhança

Seja

$$X_j | \mathbf{X}_{(j)} = \mathbf{x}_{(j)} \sim AP_{f_j}(\underline{\omega}_j(\mathbf{x}_{(j)})),$$

com $j = 1, 2, ..., p$, a função de pseudo verosimilhança com base nas densidades condicionais α-potência é dada por

$$L_P(\boldsymbol{\theta}; \mathbf{x}) = \prod_{i=1}^{n} \prod_{j=1}^{p} f_{X_j|\mathbf{X}_{(j)}}(x_j | \mathbf{x}_{(j)}, \boldsymbol{\theta}). \tag{7.28}$$

O logaritmo da função de pseudo-verosimilhança é dado por

$$\begin{aligned} \ell_P(\boldsymbol{\theta}|\mathbf{x}) &= \sum_{i=1}^{n} \sum_{j=1}^{p} \left\{ \log[f_{X_j|\mathbf{X}_{(j)}}(x_{ji}|\mathbf{x}_{(j)}, \boldsymbol{\theta})] \right\} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{p} \log(\alpha_j + \sum_{\substack{k=1 \\ k \neq j}}^{p} \alpha_{jk} \log[F_k(z_{ki})]) - n \sum_{j=1}^{p} \log(\eta_j) \\ &+ \sum_{i=1}^{n} \sum_{j=1}^{p} \log[f_j(z_{ji})] + \sum_{i=1}^{n} \sum_{j=1}^{p} (\alpha_j - 1) \log[F_j(z_{ji})] + \\ &\sum_{i=1}^{n} \sum_{j=1}^{p} \sum_{\substack{k=1 \\ k \neq j}}^{p} \alpha_{jk} \log[F_j(z_{ji})] \log[F_k(z_{ki})], \end{aligned} \tag{7.29}$$

onde $z_{ji} = \frac{x_{ji}-\xi_j}{\eta_j}$ para $j = 1, 2, 3, ...p$. Então a função do pseudo escore é dada por

$$U_P(\alpha_j) = \sum_{i=1}^{n} \left[\frac{1}{\alpha_j + \sum_{\substack{l=1 \\ l\neq j}}^{p} \alpha_{jl} \log[F_l(z_{li})]} + \log[F_j(z_{ji})] \right],$$

$$U_P(\alpha_{jk}) = \sum_{i=1}^{n} \left[\frac{\log[F_k(z_{ki})]}{\alpha_j + \sum_{\substack{l=1 \\ l\neq j}}^{p} \alpha_{jl} \log[F_l(z_{li})]} + \frac{\log[F_j(z_{ji})]}{\alpha_k + \sum_{\substack{l=1 \\ l\neq k}}^{p} \alpha_{kl} \log[F_l(z_{li})]} \right]$$
$$+ 2\sum_{i=1}^{n} [\log[F_j(z_{ji})] \log[F_k(z_{ki})]],$$

$$U_P(\xi_j) = -\frac{1}{\eta_j} \sum_{i=1}^{n} \frac{f_j(z_{ji})}{F_j(z_{ji})} \sum_{\substack{q=1 \\ q\neq j}}^{p} \left[\frac{\alpha_{jq}}{\alpha_q + \sum_{\substack{l=1 \\ l\neq q}}^{p} \alpha_{ql} \log[F_l(z_{li})]} + 2\alpha_{jq} \log[F_q(z_{qi})] \right] -$$
$$\frac{1}{\eta_j} \sum_{i=1}^{n} \left[\frac{f_j'(z_{ji})}{f_j(z_{ji})} + (\alpha_j - 1) \frac{f_j(z_{ji})}{F_j(z_{ji})} \right],$$

$$U_P(\eta_j) = -\frac{1}{\eta_j} \sum_{i=1}^{n} z_{ji} \frac{f_j(z_{ji})}{F_j(z_{ji})} \sum_{\substack{q=1 \\ q\neq j}}^{p} \left[\frac{\alpha_{jq}}{\alpha_q + \sum_{\substack{l=1 \\ l\neq q}}^{p} \alpha_{ql} \log[F_l(z_{li})]} + 2\alpha_{jq} \log[F_q(z_{qi})] \right] -$$
$$\frac{1}{\eta_j} \sum_{i=1}^{n} z_{ji} \left[\frac{f_j'(z_{ji})}{f_j(z_{ji})} + (\alpha_j - 1) \frac{f_j(z_{ji})}{F_j(z_{ji})} \right].$$

Resultados Assintóticos

Os seguintes resultados assintóticos para o estimador de máxima pseudo verossimilhança, são apresentado em Arnold e Strauss 1991b.

Teorema 7.6.1. *Sejam $X_1, ..., X_n$ uma amostra aleatória de n vetores p-dimensionais independentes e igualmente distribuídos, com densidade $f(\underline{\mathbf{x}}; \boldsymbol{\theta})$. Então, sobre condições adequadas de regularidade, Lehmann (1983), qualquer sequência consistente $\widetilde{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_n(\boldsymbol{X})$ de raízes da equação de pseudo verossimilhança satisfazem:*

1. O estimado de máxima pseudo verossimilhança $\widetilde{\boldsymbol{\theta}_n}$ é assintoticamente consistente para $\boldsymbol{\theta}$.

2.

$$[\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})] \to N_p(0, J^{-1}(\boldsymbol{\theta})K(\boldsymbol{\theta})J^{-1}(\boldsymbol{\theta})), \tag{7.30}$$

onde $K_{lm}(\boldsymbol{\theta}) = \mathbb{E}\left[\{\frac{\partial}{\partial\theta_l}\log f(x_1, x_2, ..., x_p; \boldsymbol{\theta})\}\{\frac{\partial}{\partial\theta_m}\log f(x_1, x_2, ..., x_p; \boldsymbol{\theta})\}\right]$ *e*
$J_{lm}(\boldsymbol{\theta}) = -\mathbb{E}\left\{\frac{\partial^2}{\partial\theta_l\partial\theta_m}\log f(x_1, x_2, ..., x_p; \boldsymbol{\theta})\right\}$, *para* $l = 1, 2, ..., p$ *e* $m = 1, 2, ..., p$.

A pesar de o pseudo estimador $\widetilde{\boldsymbol{\theta}}$ ser assintoticamente consistente, com distribuição normal e matriz de covariância assintótica

$$\Gamma(\widetilde{\boldsymbol{\theta}}) = \frac{1}{n}J^{-1}(\boldsymbol{\theta})K(\boldsymbol{\theta})J^{-1}(\boldsymbol{\theta}),$$

também temos que $\Gamma(\widetilde{\boldsymbol{\theta}}) - I^{-1}(\boldsymbol{\theta})$, onde $I(\boldsymbol{\theta})$ é a matriz de informação de Fisher, é uma matriz semi definida positiva, portanto, concluímos que o pseudo estimador $\widetilde{\boldsymbol{\theta}}$ é menos eficiente que o estimador de máxima verossimilhança, com perda de eficiência menor em certos cenários (Tibaldi et al., 2004).

Um estimador da matriz de covarianças assintótica $\Gamma(\widetilde{\boldsymbol{\theta}})$ foi encontrado por Cheng e Riu (2006), o qual é conhecido como estimador "sandwich". No caso da distribuição APM esta pode ser desenvolvida como segue:

seja $U_i(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta}P\ell(\boldsymbol{\theta}; X_1, ..., X_p)$, o vetor de funções escores da pseudo log-verossimilhança para a i-ésima observação. Então, define-se $\hat{J}_n(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}'}U_i(\boldsymbol{\theta})|_{\widetilde{\boldsymbol{\theta}}}$, a soma sobre todas as observações da matriz de segundas derivadas de $\ell_P(\boldsymbol{\theta}; X_1, ..., X_p)$ avaliada no pseudo estimador $\widetilde{\boldsymbol{\theta}}$. Define-se então $\hat{K}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}U_i(\boldsymbol{\theta})U_i(\boldsymbol{\theta})'|_{\widetilde{\boldsymbol{\theta}}}$, de modo que um estimador consistente da matriz de covarianças fica

$$\hat{\Gamma}(\widetilde{\boldsymbol{\theta}}) = \frac{1}{n}\hat{J}_n^{-1}(\widetilde{\boldsymbol{\theta}})\hat{K}_n(\widetilde{\boldsymbol{\theta}})\hat{J}_n^{-1'}(\widetilde{\boldsymbol{\theta}}).$$

As expressões para a variância assintótica e seu estimador "sandwich"para a distribuição APM são dadas no apêndice C.

7.6.3 Ilustração

Nós consideramos agora uma aplicação onde estudamos o ajuste do modelo PNB para o conjunto de dados dos atletas australianos. O conjunto de dados consiste de algumas variáveis medidas sobre 202 atletas e nosso foco são RCC(contagem de células vermelhas) e PCS(soma das dobras cutâneas). Os coeficientes de assimetria e curtose multivariados

amostrais, definidos por Mardia et al. (1979), podem ser escritos como

$$b_{1,p} = \frac{1}{n^2}\sum_{r,s=1}^{n} g_{rs}^3 \quad \text{e} \quad b_{2,p} = \frac{1}{n}\sum_{r=1}^{n} g_{rr}^2, \quad \text{com} \quad g_{rs} = (x_r - \bar{x})' S^{-1}(x_s - \bar{x}),$$

onde $\bar{x}$ e S são o vetor de medias e a matriz de covariâncias amostral. Para o conjunto de dados em estudo, segue-se que estes coeficientes são dados por

$$b_{1,2} = 1.605 \quad \text{e} \quad b_{2,2} = 9.792$$

levando a, $\kappa_1 = 54.0552$ e $\kappa_2 = 3.1852$, de modo que, κ_1 é maior que o valor crítico de 5% da distribuição qui-quadrado com $v = 4$ grau de libertades, e, κ_2 é maior que o valor crítico de 5% da distribuição normal. Logo, a distribuição NB não é um modelo susténtavel para os dados em estudo e um modelo alternativo que seja capaz de incorporar algum grau de assimetria, provavelmente se encaixam melhor aos dados. Também é de interese comparar a distribuição PNB com o modelo NB. Para comparar o ajuste dos modelos nós usamos o critério AIC (Akaike, 1974) e o critério de informação de Akaike consistente, CAIC. Para o conjunto de dados em estudo temos,

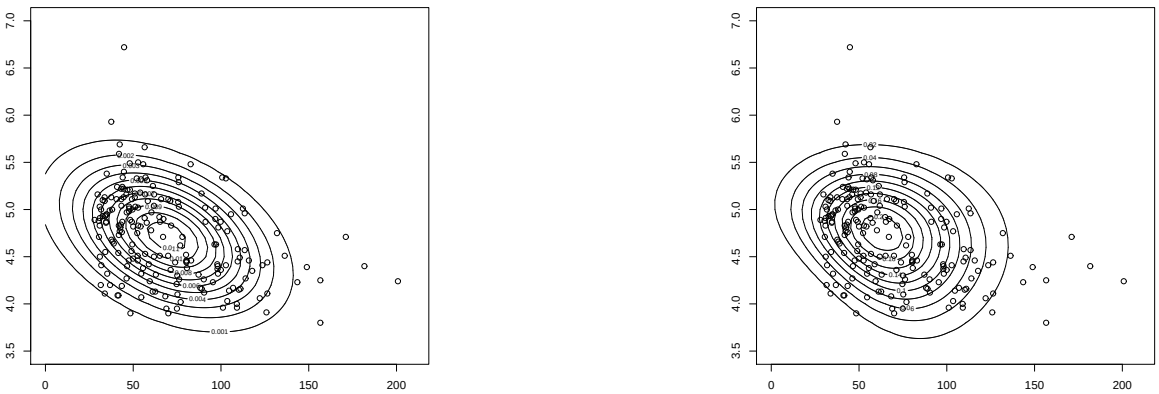
$$AIC_{NB} = 2210.436, \qquad AIC_{PNB} = 2187.584,$$

$$CAIC_{NB} = 2231.977, \qquad CAIC_{PNB} = 2217.742.$$

Para a estatística tipo Wald, pode ser mostrado que $W_n = 763.380$, então, $Prob(\chi_3^2 > W_n) < 5\%$. As estimativas dos parâmetros para os modelos em estudo são apresentados na Tabela 7.1. Contornos para a normal bivariada, $NB(69.021, 4.718, 1060.501, 0.209, -6.010)$ e $PNB(4.197, 0.392, 10.961, 4.930, 45.590, 0.373, -1.349)$, são apresentados na Figura 7.2 a qual indica um melhor ajuste para o modelo PNB.

Tabela 7.1 *Estimativas (erro padrão) e ajuste para as distribuições NB e PNB*

Modelo NB			Modelo PNB	
Parâmetro	**NBI**	**NB**	Parâmetro	PNB
$Log - lik$	-1118.118	-1100.218	$Log - Plik$	-1086.522
μ_x	69.021(2.291)	69.021(2.291)	ξ_1	10.961(0.202)
μ_y	4.718(0.032)	4.718(0.032)	ξ_2	4.930(0.170)
σ_x^2	1060.501(105.523)	1060.501(105.523)	η_1	45.590(0.131)
σ_y^2	0.209(0.020)	0.209(0.020)	η_2	0.373(0.038)
σ_{xy}	–	-6.010(1.131)	α_1	4.197(0.601)
			α_2	0.392(0.061)
			α_{12}	-1.349(0.751)



(a) (b)

Figura 7.2 *Contonos para (a)* $BN(69.021, 4.718, 1060.501, 0.209, -6.010)$ *e (b)* $PNB(4.197, 0.392, 10.961, 4.930, 45.590, 0.373, -1.349)$.

Capítulo 8

Considerações Finais

8.1

Neste trabalho estudamos algumas extensões do modelo α-potência (AP) também conhecido como modelo exponenciado de Gupta e Gupta (2008). Em particular estudamos algumas propriedades da distribuição, também definimos a distribuição log-α-potência e estudamos algumas de suas propriedades tais como momentos, função geradora de momentos, matriz de informação observada e a matriz de informaçõs esperada de Fisher; fazemos uma parte inferêncial encontrando expressões para os estimadores de momentos e de máxima verossimilhança para os parâmetros do modelo. Estimadores por verossimilhança perfilada e verossimilhança perfilada ajustada também são estudados, um experimento de simulação tipo Monte Carlo onde se estudam algumas medidas de qualidades como viés e raiz do erro quadrático médio é apresentado. Uma ilustração com dados reais é feita, onde se mostra a importância do modelo. Extensões para dados censurados e truncados, também são introduzidas. Finalmente nós introduzimos o modelo potência-*skew*-Normal, onde estudamos suas principais características e o processo de estimação de parâmetros. Nós definimos a família de distribuições α-potência flexível e a partir deste modelo derivamos o modelo bimodal PN. Deste modelo estudamos os momentos da variável aleatória e derivamos a função de log-verossimilhança, a função escore e as matrizes de informação observada e esperada, e por último, apresentamos equações a partir das quais podem ser obtidos os estimadores de máxima verossimilhança para os

parâmetros do modelo. Também estudamos os modelos bimodal simétricos e assimétricos e estudamos suas principais propriedades. Fazemos a extensão para o modelo generalizado de Birnbaum-Saunders, assim nós introduzimos o modelo GBS para o caso da família de distribuições AP estudada. Propriedades como momentos, matriz de informação observada, matriz de informaçõs esperada, coeficientes de assimetria, curtose, variância e coeficiente de variação foram desenvolvidos.

Estudamos distintos métodos de estimação para os parâmetros e apresentamos métodos para a correção do viés dos estimadores de máxima verossimilhança. Ilustrações com dados reais também são apresentadas assim como um estudo de simulação tipo Monte Carlo, onde se compara as propriedades estatísticas dos estimadores estudados. Nós também estudamos o modelos de regressão AP para os casos linear, não linear e com erros correlacionados. Introduzimos também o modelo de regressão log-linear Birnbaum-Saunders PN, e o modelo para dados censurados, tipo modelo tobit e mistura de distribuições para erros com distribuição log-α-potência, também são desenvolvidos. Nestes modelos aspectos da análise de diagnóstico e de influência foram derivados amplamente. Nós derivamos curvaturas normais de influência local através do afastamento pela verossimilhança para três esquemas de perturbação e apresentamos algumas estatísticas para conduzir uma análise residual. Em cada modelo estudado nós fazemos aplicações com dados reais e estudamos algumas propriedades particulares de cada modelo. O último aspecto estudado foi a extensão para o caso da distribuição α-potência multivariada a partir de distribuições condicionais AP. Medidas de correlação foram desenvolvidas, estimadores de máxima verossimilhança e pseudo-verossimilhança também foram estudados; e, por último, uma ilustração aos dados dos atletas australianos foi apresentada.

8.1.1 Trabalhos Futuros

Nós propomos, como trabalhos futuros, a extensão bivariada do modelo Birnbaum-Saunders AP através de distribuições condicionais AP. Também se pode estender o modelo de regressão log-linear Birnbaum-Saunders AP para o caso multivariado. Estudo de medidas de dependência tipo copulas podem ser estudadas para variáveis aleatórias com dis-

tribuição AP. Técnicas de diagnóstico para modelos com erros PN e SPN também podem ser estudadas.

Apêndice A

Tabelas-Simulações

A seguir apresentamos as tabelas dos experimentos de simulação feito para os distintos casos estudados na extensão do modelo α-potência. Assim, temos as tabelas para os processos: modelo LPN, modelo de regressão linear com erros PN, modelo Tobit PN e modelo de mistura discreta-contínua Bernoulli/LPN.

Tabela A.1 *Viés para os estimadores* $\hat{\boldsymbol{\theta}}$, $\widetilde{\boldsymbol{\theta}}$ *e* $\widetilde{\widetilde{\boldsymbol{\theta}}}$ *no modelo LPN.*

			$\hat{\boldsymbol{\theta}}$			$\widetilde{\boldsymbol{\theta}}$			$\widetilde{\widetilde{\boldsymbol{\theta}}}$		
α	ξ	n	$\hat{\xi}$	$\hat{\eta}$	$\hat{\alpha}$	$\widetilde{\xi}$	$\widetilde{\eta}$	$\widetilde{\alpha}$	$\widetilde{\widetilde{\xi}}$	$\widetilde{\widetilde{\eta}}$	$\widetilde{\widetilde{\alpha}}$
		20	0.0511	-0.1338	0.4673	0.0140	-0.1705	0.9228	-0.1102	-0.0582	0.4491
	2.0	60	0.0075	-0.0554	0.4104	-0.0638	-0.0612	0.5688	-0.1485	0.0056	0.3359
		100	0.0063	-0.0470	0.3006	-0.0588	-0.0329	0.3810	-0.1670	0.0306	0.2769
		20	-0.0990	-0.1149	0.5337	-0.5996	-0.0191	4.7749	-0.3909	0.0890	1.8052
0.75	4.0	60	-0.0324	-0.0471	0.4305	-0.2884	-0.0034	1.4907	-0.2539	0.0353	0.5494
		100	0.0039	-0.0391	0.3282	-0.1756	-0.0030	0.6476	-0.2290	0.0467	0.3315
		20	0.1844	-0.0266	0.5334	-0.6498	0.0473	4.2885	-0.7230	0.4463	1.6788
	8.0	60	-0.1590	-0.0244	0.3918	-0.3905	0.0335	2.1618	-0.6203	0.1366	0.6097
		100	-0.1295	-0.0009	0.3667	-0.2232	0.0211	0.6920	-0.3499	0.1116	0.3594
		20	0.2770	-0.1875	0.2326	-0.1017	-0.1322	2.1189	0.1123	-0.1244	0.5974
	2.0	60	0.1507	-0.0945	0.2277	-0.0928	-0.0383	1.2432	0.0956	-0.0574	0.2646
		100	0.0855	-0.0617	0.2064	-0.0619	-0.0194	0.8309	0.0785	-0.0544	0.0933
		20	0.1432	-0.1266	0.2699	-0.5195	0.0180	5.4822	-0.2549	0.0177	1.9451
1.25	4.0	60	0.1156	-0.0780	0.2194	-0.2443	0.0033	2.1584	-0.0606	-0.0119	0.5418
		100	0.0754	-0.0515	0.2317	-0.2061	0.0139	1.4047	-0.0415	-0.0087	0.3270
		20	0.1089	-0.1126	0.2392	-0.7010	0.0523	10.9086	-0.3483	0.0284	4.0349
	8.0	60	0.0485	-0.0532	0.2289	-0.3667	0.0301	3.5033	-0.1037	-0.0066	0.9158
		100	0.0215	-0.0299	0.2165	-0.2310	0.0175	1.6734	-0.0024	-0.0215	0.1883
		20	0.4866	-0.2456	-0.2322	-0.0092	-0.1462	3.3459	0.3027	-0.1814	0.5961
	2.0	60	0.2648	-0.1229	-0.0346	-0.1313	-0.0302	2.0918	0.2126	-0.1020	0.3552
		100	0.1770	-0.0826	-0.0319	-0.1437	-0.0044	1.5472	0.2112	-0.0900	0.1296
		20	0.3911	-0.1947	-0.1942	-0.4503	-0.0047	9.3198	0.1505	-0.0472	3.0861
1.75	4.0	60	0.2411	-0.1086	-0.0643	-0.3047	0.0203	3.8247	0.1066	-0.0654	0.7353
		100	0.1878	-0.0820	-0.0180	-0.2261	0.0173	2.4369	0.0604	-0.0694	0.2125
		20	0.2928	-0.1570	-0.1232	-0.7635	0.0626	22.0574	0.1585	-0.0615	7.2367
	8.0	60	0.1774	-0.0852	-0.0347	-0.4111	0.0400	6.3419	0.1136	-0.0608	1.3546
		100	0.1228	-0.0559	-0.0064	-0.2912	0.0337	3.0548	0.0610	-0.0362	0.3182
		20	0.3619	-0.1534	-0.2793	-0.1437	-0.0533	2.2908	0.2928	-0.1309	-0.3182
	2.0	60	0.2601	-0.1046	-0.2181	-0.1057	-0.0159	1.7583	0.2893	-0.1117	-0.1204
		100	0.1770	-0.0826	-0.1811	-0.0677	-0.0044	1.2972	0.2112	-0.0900	-0.0151
		20	0.4346	-0.2017	-0.3581	-0.5197	0.0218	11.7167	0.1960	-0.0848	3.7483
2.0	4.0	60	0.3105	-0.1280	-0.2379	-0.3265	0.0153	5.1731	0.1219	-0.0753	1.3140
		100	0.2466	-0.0976	-0.1879	-0.2133	0.0135	2.7280	0.0131	-0.0615	0.2630
		20	0.3848	-0.1775	-0.3413	-0.7129	0.0524	25.2803	0.1680	-0.0790	7.1471
	8.0	60	0.2467	-0.1034	-0.2028	-0.4044	0.0368	7.9801	0.1609	-0.0886	1.2488
		100	0.1783	-0.0706	-0.1243	-0.3008	0.0344	4.1879	0.0313	-0.0657	0.5314

Tabela A.2 $\sqrt{EQM}$ *para os estimadores* $\hat{\boldsymbol{\theta}}$, $\widetilde{\boldsymbol{\theta}}$ *e* $\widetilde{\widetilde{\boldsymbol{\theta}}}$ *no modelo LPN.*

			$\hat{\boldsymbol{\theta}}$			$\widetilde{\boldsymbol{\theta}}$			$\widetilde{\widetilde{\boldsymbol{\theta}}}$		
α	ξ	n	$\hat{\xi}$	$\hat{\eta}$	$\hat{\alpha}$	$\widetilde{\xi}$	$\widetilde{\eta}$	$\widetilde{\alpha}$	$\widetilde{\widetilde{\xi}}$	$\widetilde{\widetilde{\eta}}$	$\widetilde{\widetilde{\alpha}}$
		20	1.2083	0.4627	1.3903	1.5423	0.5856	2.3690	1.0986	0.4639	1.3666
	2.0	60	1.1068	0.4043	1.2185	1.2497	0.4502	1.6307	0.9178	0.3470	0.9992
		100	1.0027	0.3644	1.0763	1.0567	0.3742	1.2750	0.8007	0.2916	0.7904
		20	1.2594	0.4774	1.4457	2.3303	0.7116	10.2652	1.6135	0.5374	5.5672
0.75	4.0	60	1.0810	0.3869	1.2297	1.6623	0.5301	4.2115	1.1570	0.3888	1.8815
		100	1.0453	0.3421	1.0849	1.2841	0.4131	2.1754	0.9351	0.3078	1.0139
		20	0.9913	0.3510	1.2591	2.1359	0.5961	9.7623	1.4206	0.4104	5.0308
	8.0	60	0.8658	0.2961	1.0481	1.8242	0.5141	7.8037	1.1827	0.3301	1.9915
		100	0.7939	0.2686	0.9495	1.2702	0.3947	2.2298	0.8198	0.2523	0.8615
		20	1.0989	0.4400	1.3685	1.5985	0.5581	4.6961	1.1316	0.4370	2.6851
	2.0	60	0.9145	0.3358	1.2105	1.2813	0.4220	3.1795	0.8750	0.3016	1.7986
		100	0.8039	0.2860	1.1207	1.0800	0.3491	2.4485	0.7667	0.2596	1.3410
		20	0.9141	0.3520	1.3007	1.8860	0.5258	11.1737	1.2008	0.3408	6.2926
1.25	4.0	60	0.8264	0.2931	1.1760	1.4519	0.4253	5.6611	0.9163	0.2676	2.8334
		100	0.7819	0.2629	1.1037	1.2448	0.3684	4.0894	0.8081	0.2338	2.0654
		20	0.8446	0.3331	1.1971	2.2309	0.5742	24.1097	1.5217	0.3995	13.3498
	8.0	60	0.7049	0.2554	1.0430	1.6427	0.4546	9.8359	1.0726	0.2937	4.7055
		100	0.6434	0.2225	0.9624	1.3536	0.3871	5.4148	0.8469	0.2436	1.7464
		20	1.0726	0.4373	1.3806	1.6042	0.5386	7.4929	1.1627	0.4315	3.9814
	2.0	60	0.8328	0.3078	1.1906	1.2943	0.4017	5.0048	0.9690	0.3123	3.0559
		100	0.7227	0.2568	1.1070	1.1187	0.3384	4.0244	0.8784	0.2786	2.4675
		20	0.9072	0.3584	1.2852	1.9012	0.5065	19.5768	1.2728	0.3465	10.8546
1.75	4.0	60	0.7565	0.2749	1.1420	1.4908	0.4136	9.3914	0.9989	0.2839	4.8167
		100	0.6881	0.2398	1.0857	1.2806	0.3582	6.8707	0.8796	0.2505	3.2004
		20	0.7953	0.3244	1.1750	2.3056	0.5615	49.2347	1.5928	0.4092	24.7268
	8.0	60	0.6457	0.2363	1.0417	1.7101	0.4471	17.1824	1.1730	0.3192	7.7183
		100	0.5628	0.1977	0.9718	1.3872	0.3727	9.4500	0.9761	0.2706	3.4761
		20	0.8444	0.3141	1.2231	1.3054	0.4046	5.7816	1.0367	0.3385	3.5652
	2.0	60	0.7227	0.2568	1.1278	1.1340	0.3396	4.7563	0.8905	0.2786	2.7610
		100	0.7167	0.2547	1.1156	1.1187	0.3384	3.9351	0.8784	0.2772	2.46705
		20	0.9059	0.3601	1.3222	1.9337	0.5089	23.9073	1.3120	0.3546	13.0585
2.0	4.0	60	0.7570	0.2735	1.1761	1.5378	0.4143	12.3881	1.1014	0.3080	7.0280
		100	0.6685	0.2350	1.0844	1.2654	0.3510	7.7555	0.9248	0.2630	3.8708
		20	0.8137	0.3282	1.2206	2.2965	0.5554	56.5008	1.5529	0.4028	26.4343
	8.0	60	0.6440	0.2370	1.0773	1.7256	0.4438	21.6409	1.1646	0.3181	8.6144
		100	0.5631	0.1984	1.0087	1.4299	0.3758	13.1584	1.0536	0.2864	5.4628

Tabela A.3 *Medidas de Qualidade dos EMV do vetor de parâmetros no modelo BSPN.*

			$\hat{\lambda}$				$\hat{\beta}$				$\hat{\alpha}$			
λ	α	n	Viés	$\sqrt{EQM}$	Assimetria	Curtoses	Viés	$\sqrt{EQM}$	Assimetria	Curtoses	Viés	$\sqrt{EQM}$	Assimetria	Curtoses
		70	-0.0025	0.0037	-1.7104	6.5833	-0.0218	0.0282	0.0577	2.9470	-0.0201	0.0548	-0.7742	3.8975
		90	-0.0023	0.0033	-1.6796	6.6833	-0.0215	0.0270	-0.0191	2.9432	-0.0043	0.0479	-0.8132	3.9410
	0.85	120	-0.0021	0.0030	-1.5441	5.9837	-0.0212	0.0262	-0.0107	2.8680	0.0107	0.0439	-0.8169	4.0192
		150	-0.0020	0.0028	-1.5151	6.1508	-0.0210	0.0257	0.0539	3.0637	0.0240	0.0403	-0.8164	3.8848
		70	-0.0042	0.0056	-1.3393	5.1115	0.0295	0.0343	0.0451	3.0125	0.0281	0.2065	3.4743	22.8204
		90	-0.0039	0.0050	-1.2311	5.4096	0.0296	0.0332	0.0068	3.0207	-0.0050	0.1801	4.1693	32.4503
0.15	1.25	120	-0.0037	0.0046	-1.1341	4.8697	0.0294	0.0322	0.0211	3.0336	-0.0490	0.1438	3.8979	28.7921
		150	-0.0036	0.0043	-1.0169	4.5388	0.0296	0.0318	0.0437	2.9798	-0.0721	0.1283	3.2473	19.3115
		70	-0.0111	0.0125	-0.6007	3.1300	0.0532	0.0558	-0.0055	2.9636	0.5284	2.4905	12.7123	276.8122
		90	-0.0109	0.0121	-0.6519	3.4781	0.0532	0.0553	0.0931	2.9610	0.5693	2.1994	11.2583	297.5500
	1.50	120	-0.0107	0.0116	-0.5267	3.1533	0.0530	0.0546	0.0802	2.9421	0.5510	2.5831	13.7076	198.5941
		150	-0.0105	0.0102	-0.4494	3.1103	0.0528	0.0540	0.0260	3.0270	0.0750	0.8897	6.7867	103.5370
		70	-0.0090	0.0135	-1.7266	6.8237	-0.0699	0.0888	0.1234	2.8846	-0.0210	0.0555	-0.7094	3.8786
		90	-0.0084	0.0121	-1.6835	6.5696	-0.0688	0.0854	0.2052	3.2492	-0.0029	0.0460	-0.7894	3.9965
	0.85	120	-0.0076	0.0106	-1.4715	5.4325	-0.0674	0.0817	0.1069	2.8485	0.0123	0.0435	-0.8546	4.3224
		150	-0.0071	0.0095	-1.3446	5.1031	-0.0660	0.0793	0.1727	3.0397	0.0234	0.0451	-0.6896	3.5592
		70	-0.0147	0.0199	-1.3693	5.2180	0.1014	0.1182	0.1862	3.1174	0.0207	0.2085	4.7664	42.6167
		90	-0.0138	0.0178	-1.2381	4.9024	0.1012	0.1142	0.1537	3.1342	-0.0091	0.1609	3.8120	28.6467
0.50	1.25	120	-0.0128	0.0160	-1.0507	4.2014	0.0993	0.1091	0.0727	3.0218	-0.0590	0.1271	3.5057	23.7613
		150	-0.0125	0.0151	-0.9806	4.1490	0.0983	0.1073	0.0873	3.0130	-0.0794	0.1234	3.6580	28.2478
		70	-0.0379	0.0428	-0.6330	3.4037	0.1835	0.1932	0.1285	3.0659	0.3464	1.9348	13.3137	274.8027
		90	-0.0377	0.0417	-0.5744	3.1149	0.1824	0.1922	0.1840	2.9701	0.3330	1.9154	13.4555	276.3789
	1.50	120	-0.0370	0.0400	-0.5295	3.2137	0.1866	0.1895	0.1321	2.9725	0.3315	1.8295	11.1673	202.4474
		150	-0.0369	0.0394	-0.4408	3.0783	0.1812	0.1889	0.0868	2.9585	-0.0123	0.6832	6.0682	163.3083
		70	-0.0144	0.0217	-1.9341	8.2431	-0.0968	0.1243	0.1617	3.0019	-0.0206	0.0549	-0.5863	3.5642
		90	-0.0131	0.0190	-1.7436	7.2119	-0.0961	0.1192	0.1704	3.0335	-0.0049	0.0464	-0.6014	3.4750
	0.85	120	-0.0118	0.0163	-1.4455	5.4958	-0.0958	0.1137	0.2067	3.0664	0.0110	0.0429	-0.6527	3.9809
		150	-0.0113	0.0152	-1.3809	5.4569	-0.0956	0.1119	0.1666	3.0079	0.0232	0.0440	-0.5613	3.2260
		70	-0.0225	0.0305	-1.4490	5.9236	0.1487	0.1743	0.2648	3.1031	0.0118	0.1708	3.6120	28.4851
		90	-0.0212	0.0278	-1.3459	5.5556	0.1480	0.1684	0.2590	3.1777	-0.0191	0.1477	3.5267	23.0560
0.75	1.25	120	-0.0201	0.0249	-1.0289	4.1359	0.1477	0.1626	0.1503	3.0150	-0.0600	0.1211	3.1137	19.0723
		150	-0.0198	0.0240	-0.9705	4.2248	0.1465	0.1608	0.1219	3.0485	-0.0829	0.1174	3.1109	21.8769
		70	-0.0596	0.0675	-0.6377	3.3476	0.2799	0.2950	0.2110	3.0035	0.2296	1.2667	8.1808	368.3001
		90	-0.0587	0.0649	-0.5552	3.3833	0.2782	0.2909	0.1256	3.0872	0.2236	1.4117	11.0665	217.5938
	1.50	120	-0.0582	0.0632	-0.4716	2.9983	0.2777	0.2887	0.1891	3.0071	0.1933	1.4049	10.1597	175.0074
		150	-0.0578	0.0616	-0.4084	3.1546	0.2767	0.2863	0.1288	3.0284	-0.0851	0.7498	10.5376	171.6290

Tabela A.4 *Viés para os estimadores corrigidos pelos métodos de MV_{pc}, Cox e Snell, bootstrap e jackknife no modelo BSPN.*

			$\hat{\lambda}$				$\hat{\beta}$				$\hat{\alpha}$			
λ	α	n	EMV_{pc}	CCS	Bootstrap	Jackknife	EMV_{pc}	CCS	Bootstrap	Jackknife	EMV_{pc}	CCS	Bootstrap	Jackknife
		70	-0.0025	-0.0004	-0.0025	-0.0025	-0.0218	-0.0061	-0.0236	-0.0220	-0.0201	-0.0318	-0.0218	-0.0213
		90	-0.0023	-0.0007	-0.0023	-0.0023	-0.0221	-0.0002	-0.0225	-0.0219	-0.0047	-0.0121	-0.0032	-0.0033
	0.85	120	-0.0021	-0.0009	-0.0022	-0.0021	-0.0222	-0.0057	-0.0223	-0.0213	0.0107	0.0029	0.0011	0.0099
		150	-0.0020	-0.0010	-0.0020	-0.0019	-0.0220	-0.0093	-0.0220	-0.0203	0.0242	0.0177	0.0002	0.0229
		70	-0.0042	-0.0021	-0.0042	-0.0041	0.0297	0.0590	0.0398	0.0298	-0.0334	0.0130	0.0314	0.0250
		90	-0.0039	-0.0024	-0.0040	-0.0038	0.0295	0.0530	0.0302	0.0296	-0.0016	-0.0141	0.0045	0.0028
0.15	1.25	120	-0.0038	-0.0025	-0.0038	-0.0037	0.0301	0.0468	0.0290	0.0295	-0.0479	-0.0601	-0.0461	-0.0282
		150	-0.0036	-0.0026	-0.0036	-0.0035	0.0297	0.0433	0.0284	0.0290	-0.0719	-0.0793	-0.0702	-0.0711
		70	-0.0110	-0.0096	-0.0110	-0.0122	0.0529	0.0833	0.0533	0.0534	0.5176	0.4715	0.5202	0.4913
		90	-0.0109	-0.0093	-0.0108	-0.0110	0.0532	0.0530	0.0764	0.0527	0.4682	0.4654	0.5076	0.4688
	1.50	120	-0.0106	-0.0092	-0.0105	-0.0106	0.0529	0.0704	0.0526	0.0531	0.5179	0.2662	0.4645	0.4479
		150	-0.0105	-0.0090	-0.0102	-0.0102	0.0529	0.0669	0.0511	0.0530	0.0708	0.2846	0.0969	0.0784
		70	-0.0091	-0.0021	-0.0090	-0.0090	-0.0685	-0.0422	-0.0784	-0.0682	-0.0196	-0.0326	-0.0191	-0.0197
		90	-0.0081	-0.0027	-0.0085	-0.0084	-0.0692	-0.0481	-0.0750	-0.0676	-0.0031	-0.0132	-0.0023	-0.0016
	0.85	120	-0.0076	-0.0034	-0.0078	-0.0074	-0.0696	-0.0539	-0.0700	-0.0686	0.0116	0.0051	0.0124	0.0127
		150	-0.0073	-0.0039	-0.0072	-0.0069	-0.0706	-0.0574	-0.0696	-0.0661	0.0245	0.0183	0.0241	0.0250
		70	-0.0145	-0.0079	-0.0142	-0.0148	0.1010	0.1335	0.0998	0.1024	0.0217	-0.0009	0.0199	0.0188
		90	-0.0138	-0.0085	-0.0134	-0.0137	0.1012	0.1256	0.0989	0.1003	-0.0127	-0.0244	-0.0140	-0.0077
0.50	1.25	120	-0.0132	-0.0090	-0.0131	-0.0129	0.1011	0.1187	0.1003	0.1002	-0.0572	-0.0650	-0.0576	-0.0383
		150	-0.0125	-0.0094	-0.0124	-0.0127	0.0999	0.1150	0.0992	0.1000	-0.0787	-0.0853	-0.1002	-0.0787
		70	-0.0387	-0.0321	-0.0378	-0.0380	0.1856	0.2201	0.1834	0.1838	0.3570	0.3925	0.3943	0.3470
		90	-0.0373	-0.0318	-0.0371	-0.0374	0.1838	0.2112	0.1824	0.1832	0.3550	0.3450	0.3641	0.3079
	1.50	120	-0.0369	-0.0311	-0.0814	-0.0372	0.1836	0.2037	0.2447	0.1826	0.3644	0.3312	0.3774	0.2784
		150	-0.0366	-0.0305	-0.0369	-0.0370	0.1834	0.1993	0.1812	0.1821	0.3244	0.1454	0.1078	0.1076
		70	-0.0147	-0.0036	-0.0142	-0.0147	-0.0980	-0.0694	-0.0984	-0.0975	-0.0205	-0.0341	-0.0216	-0.0216
		90	-0.0131	-0.0051	-0.0130	-0.0132	-0.0964	-0.0777	-0.0967	-0.0038	-0.0047	-0.0132	-0.0059	-0.0030
	0.85	120	-0.0120	-0.0057	-0.0125	-0.0121	-0.0981	-0.0823	-0.0909	-0.0965	0.0121	0.0044	0.0112	0.0110
		150	-0.0110	-0.0062	-0.0113	-0.0114	-0.0965	-0.0852	-0.0890	-0.0951	0.0246	0.0175	0.0257	0.0240
		70	-0.0225	-0.0136	-0.0225	-0.0226	0.1501	0.1888	0.1506	0.1504	-0.0141	-0.0058	-0.0184	-0.0152
		90	-0.0215	-0.0133	-0.0218	-0.0211	0.1504	0.1750	0.1501	0.1485	-0.0152	-0.0264	-0.0147	-0.0174
0.75	1.25	120	-0.0204	-0.0126	-0.0206	-0.0211	0.1492	0.1696	0.1496	0.1473	-0.0606	-0.0694	-0.0107	-0.0086
		150	-0.0199	-0.0117	-0.0108	-0.0199	0.1488	0.1637	0.1011	0.1450	-0.0822	-0.0884	-0.0104	-0.0051
		70	-0.0598	-0.0521	-0.0605	-0.0607	0.2790	0.3164	0.2810	0.2820	0.2402	0.2235	0.2502	0.2851
		90	-0.0591	-0.0516	-0.0578	-0.0590	0.2800	0.3082	0.2762	0.2789	0.2793	0.2071	0.2433	0.2772
	1.50	120	-0.0577	-0.0512	-0.0575	-0.0585	0.2776	0.2990	0.2632	0.2781	0.1808	0.1514	0.2395	0.2350
		150	-0.0571	-0.0505	-0.0570	-0.0574	0.2797	0.2938	0.2550	0.2779	-0.0840	-0.0884	-0.0743	-0.0894

Tabela A.5 $\sqrt{EQM}$ *para os estimadores corrigidos pelos métodos de* MV_{pc}*, Cox e Snell, bootstrap e jackknife no modelo BSPN.*

			$\hat{\lambda}$				$\hat{\beta}$				$\hat{\alpha}$			
λ	α	n	EMV_{pc}	CCS	Bootstrap	Jackknife	EMV_{pc}	CCS	Bootstrap	Jackknife	EMV_{pc}	CCS	Bootstrap	Jackknife
		70	0.0038	0.0029	0.0037	0.0043	0.0285	0.0198	0.0282	0.0286	0.0560	0.0599	0.0552	0.0566
		90	0.0033	0.0025	0.0034	0.0042	0.0270	0.0164	0.0274	0.0276	0.0473	0.0478	0.0471	0.0458
	0.85	120	0.0030	0.0022	0.0030	0.0046	0.0262	0.0157	0.0265	0.0266	0.0434	0.0424	0.0443	0.0440
		150	0.0027	0.0021	0.0027	0.0051	0.0253	0.0151	0.0256	0.0259	0.0427	0.0420	0.0432	0.0449
		70	0.0056	0.0044	0.0056	0.0059	0.0344	0.0616	0.0344	0.0342	0.2304	0.2101	0.2271	0.2173
		90	0.0051	0.0042	0.0051	0.0057	0.0333	0.0553	0.0337	0.0334	0.1749	0.1901	0.1943	0.1882
0.15	1.25	120	0.0048	0.0037	0.0047	0.0058	0.0329	0.0487	0.0327	0.0326	0.1403	0.1355	0.1402	0.1530
		150	0.0044	0.0030	0.0043	0.0060	0.0319	0.0450	0.0317	0.0323	0.1275	0.1324	0.1292	0.1297
		70	0.0125	0.0108	0.0132	0.0126	0.0555	0.0850	0.0558	0.0559	2.4478	2.8211	2.8066	2.3295
		90	0.0121	0.0106	0.0119	0.0124	0.0552	0.0779	0.0548	0.0549	2.1994	2.5392	2.7351	2.2678
	1.50	120	0.0115	0.0104	0.0113	0.0120	0.0544	0.0616	0.0541	0.0547	2.1831	1.7948	2.3046	2.2494
		150	0.0113	0.0100	0.0110	0.0111	0.0541	0.0680	0.0533	0.0543	0.8774	1.6728	1.0920	0.8942
		70	0.0137	0.0104	0.0135	0.0137	0.0892	0.0718	0.0886	0.0887	0.0540	0.0596	0.0543	0.0545
		90	0.0116	0.0091	0.0122	0.0124	0.0845	0.0701	0.0861	0.0859	0.0463	0.0472	0.0445	0.0452
	0.85	120	0.0106	0.0082	0.0109	0.0110	0.0817	0.0695	0.0827	0.0810	0.0436	0.0410	0.0453	0.0426
		150	0.0099	0.0076	0.0097	0.0103	0.0807	0.0690	0.0798	0.0789	0.0453	0.0413	0.0419	0.0421
		70	0.0195	0.0156	0.0192	0.0199	0.1174	0.1473	0.1165	0.1191	0.1950	0.1863	0.1951	0.1879
		90	0.0179	0.0144	0.0174	0.0179	0.1141	0.1367	0.1120	0.1134	0.1612	0.1601	0.1583	0.1675
0.50	1.25	120	0.0165	0.0133	0.0164	0.0164	0.1113	0.1275	0.1104	0.1099	0.1276	0.1309	0.1255	0.1361
		150	0.0151	0.0128	0.0148	0.0159	0.1077	0.1222	0.1069	0.1085	0.1226	0.1275	0.1272	0.1216
		70	0.0438	0.0379	0.0426	0.0430	0.1956	0.2287	0.1929	0.1937	2.3408	1.6173	1.7460	2.1762
		90	0.0413	0.0375	0.0409	0.0413	0.1914	0.2185	0.1901	0.1910	2.0684	1.5573	1.7074	1.7565
	1.50	120	0.0400	0.0365	0.0399	0.0405	0.1896	0.2092	0.1876	0.1903	1.9698	1.4821	1.0602	1.5064
		150	0.0391	0.0362	0.0395	0.0397	0.1881	0.2037	0.1889	0.1889	0.6732	1.2708	0.7141	0.7358
		70	0.0220	0.0164	0.0211	0.0219	0.1255	0.1060	0.1238	0.1256	0.0545	0.0610	0.0555	0.0556
		90	0.0188	0.0149	0.0187	0.0193	0.1191	0.1052	0.1191	0.1194	0.0468	0.0471	0.0468	0.0459
	0.85	120	0.0165	0.0129	0.0173	0.0171	0.1148	0.1021	0.1169	0.1154	0.0422	0.0412	0.0431	0.0435
		150	0.0149	0.0119	0.0151	0.0161	0.1105	0.1008	0.1119	0.1122	0.0424	0.0407	0.0424	0.0420
		70	0.0304	0.0253	0.0300	0.0304	0.1750	0.2108	0.1744	0.1753	0.1752	0.1674	0.1696	0.1804
		90	0.0275	0.0221	0.0282	0.0277	0.1696	0.1930	0.1708	0.1686	0.1480	0.1534	0.1400	0.1416
0.75	1.25	120	0.0255	0.0213	0.0257	0.0255	0.1641	0.1837	0.1648	0.1631	0.1192	0.1244	0.1220	0.1259
		150	0.0240	0.0200	0.0155	0.0249	0.1610	0.1751	0.1092	0.1622	0.1177	0.1229	0.1210	0.1199
		70	0.0676	0.0594	0.0685	0.0686	0.2952	0.3317	0.2977	0.2984	1.3420	1.3847	1.3917	1.5305
		90	0.0655	0.0588	0.0639	0.0651	0.2930	0.3202	0.2887	0.2910	1.2874	1.3709	1.2121	1.3742
	1.50	120	0.0625	0.0578	0.0624	0.0632	0.2869	0.3089	0.2848	0.2893	1.1715	1.1787	1.0709	1.2769
		150	0.0620	0.0567	0.0619	0.0614	0.2853	0.3010	0.2869	0.2854	0.6272	0.5499	0.6883	0.5479

Tabela A.6 *Medidas de qualidade para os EMV e EMM dos parâmetros no modelo de regressão linear PN.*

		$\alpha = 0.75$			$\alpha = 1.25$			$\alpha = 1.75$			$\alpha = 2.25$		
n	$\hat{\theta}$	Viés	VR(%)	$\sqrt{EQM}$	Viés	VR(%)	$\sqrt{EQM}$	Viés	VR(%)	$\sqrt{EQM}$	Viés	VR(%)	$\sqrt{EQM}$
	$\hat{\alpha}$	0.4892	65.22	1.4696	0.7336	58.69	2.0245	0.4939	28.22	2.0126	0.2235	9.93	1.5209
	$\widetilde{\eta}$	0.1322	13.22	0.4501	0.0380	3.80	0.4125	0.0112	1.12	0.3895	-0.0179	1.79	0.3720
	$\hat{\eta}$	-0.1918	19.18	0.5696	-0.1629	16.29	0.5092	-0.1926	19.26	0.4880	-0.2133	21.33	0.4639
40	$\widetilde{\beta}_0$	-0.5679	37.86	1.5738	-0.3012	20.08	1.4244	-0.1973	13.15	1.3351	-0.0889	5.93	1.2576
	$\hat{\beta}_0$	0.1246	8.30	1.3398	0.1430	9.53	1.2795	0.2827	18.84	1.1921	0.3944	26.29	1.1623
	$\widetilde{\beta}_1$	-0.0062	0.25	0.6083	-0.0047	0.19	0.5235	0.0062	0.25	0.4724	-0.0058	0.23	0.4391
	$\hat{\beta}_1$	-0.0069	0.28	0.6344	-0.0188	0.75	0.5452	-0.0159	0.63	0.4949	0.0323	1.29	0.4526
	$\hat{\alpha}$	0.4339	57.85	1.2526	0.6716	53.73	1.8252	0.4228	24.16	1.7821	0.1800	8.00	1.3486
	$\widetilde{\eta}$	0.0798	7.98	0.3536	0.0312	3.12	0.3456	0.0130	1.30	0.3317	-0.0100	1.00	0.3239
	$\hat{\eta}$	-0.0718	7.18	0.4259	-0.0608	6.08	0.3741	-0.1020	10.20	0.3434	-0.1146	11.46	0.3256
80	$\widetilde{\beta}_0$	-0.3475	23.17	1.1918	-0.2226	14.84	1.1796	-0.1626	10.84	1.1411	-0.0781	5.21	1.1165
	$\hat{\beta}_0$	-0.0092	0.61	1.1046	-0.0035	0.24	1.0453	0.1411	9.41	0.9480	0.2170	14.47	0.8912
	$\widetilde{\beta}_1$	0.0040	0.16	0.4113	-0.0037	0.14	0.3419	-0.0045	0.18	0.3297	0.0056	0.22	0.3167
	$\hat{\beta}_1$	0.0014	0.06	0.4316	-0.0024	0.10	0.3646	-0.0066	0.26	0.3353	0.0138	0.55	0.3292
	$\hat{\alpha}$	0.3073	40.98	1.0623	0.5323	42.59	1.6243	0.3588	20.50	1.6304	0.0722	3.21	1.2669
	$\widetilde{\eta}$	0.0424	4.24	0.3029	0.0222	2.22	0.3005	-0.0135	1.35	0.2765	-0.0040	0.40	0.2843
	$\hat{\eta}$	-0.0532	5.32	0.3457	-0.0384	3.84	0.3078	-0.0627	6.27	0.2690	-0.0671	6.71	0.2504
120	$\widetilde{\beta}_0$	-0.2134	14.23	0.9945	-0.1664	11.10	1.0173	-0.0419	2.79	0.9387	-0.0905	6.04	1.0008
	$\hat{\beta}_0$	0.0049	0.33	0.9452	-0.0006	0.04	0.9034	0.0989	6.59	0.8023	0.1194	7.96	0.7563
	$\widetilde{\beta}_1$	-0.0006	0.03	0.3542	-0.0022	0.09	0.2868	-0.0118	0.47	0.2607	0.0040	0.16	0.2559
	$\hat{\beta}_1$	0.0006	0.02	0.3574	-0.0009	0.03	0.2885	-0.0015	0.06	0.2637	0.0061	0.24	0.2514

Tabela A.7 *Medidas de Qualidade para os EMV dos parâmetros no modelo TPN.*

		$\alpha = 0.50$			$\alpha = 0.75$			$\alpha = 1.25$			$\alpha = 1.75$		
n	$\boldsymbol{\theta}$	Viés	VR(%)	$\sqrt{EQM}$	Viés	VR(%)	$\sqrt{EQM}$	Viés	VR(%)	$\sqrt{EQM}$	Viés	VR(%)	$\sqrt{EQM}$
	$\hat{\alpha}$	-0.0040	0.80	0.1473	-0.0019	0.25	0.1478	-0.0007	0.05	0.1453	-0.0015	0.08	0.1454
100	$\hat{\eta}$	0.1920	19.20	0.2074	0.0374	3.74	0.0992	-0.1142	11.42	0.1399	-0.1904	19.04	0.2045
	$\hat{\beta}_0$	-0.1613	10.75	1.6823	-0.1385	9.23	1.3833	-0.1844	12.29	1.0984	-0.2973	19.82	1.0038
	$\hat{\beta}_1$	-0.0769	2.20	0.2148	0.0585	1.67	0.1990	0.2477	7.08	0.3335	0.4026	11.50	0.4825
	$\hat{\alpha}$	-0.0031	0.62	0.1468	-0.0013	0.18	0.1463	-0.0005	0.04	0.1449	-0.0008	0.05	0.1442
300	$\hat{\eta}$	0.1859	18.59	0.1992	0.0363	3.63	0.0696	-0.1076	10.76	0.1168	-0.1861	18.61	0.1911
	$\hat{\beta}_0$	-0.1502	10.02	1.1971	-0.0599	3.99	0.9515	-0.0430	2.86	0.7307	-0.0822	5.48	0.6018
	$\hat{\beta}_1$	-0.0706	2.01	0.1445	0.0467	1.34	0.1248	0.2235	6.39	0.2564	0.3719	10.63	0.3993
	$\hat{\alpha}$	-0.0014	0.28	0.1450	0.0002	0.03	0.1458	-0.0002	0.02	0.1441	0.0004	0.02	0.1409
500	$\hat{\eta}$	0.1659	16.59	0.1969	0.0228	2.28	0.0616	-0.1047	10.47	0.1103	-0.1833	18.33	0.1864
	$\hat{\beta}_0$	-0.0886	5.91	0.9533	0.0330	2.20	0.7340	0.0372	2.48	0.5610	-0.0310	2.06	0.4663
	$\hat{\beta}_1$	-0.0656	1.88	0.1327	0.0449	1.28	0.1084	0.2189	6.26	0.2411	0.3668	10.48	0.3853

Tabela A.8 *Medidas de qualidade para os EMV dos parâmetros na mistura Discreta-Contínua Bernoulli/LPN.*

		n=300				n=500			
α	$\boldsymbol{\theta}$	Media	Viés	SR	$\sqrt{EQM}$	Media	Viés	SR	$\sqrt{EQM}$
	α	0.9998	0.2498	33.3067	0.7831	0.8819	0.1319	17.5867	0.4950
	$\beta_{(1)0} = 1$	1.0087	0.0087	0.87	0.0892	1.0060	0.0060	0.60	0.0689
0.75	$\beta_{(1)1} = -0.5$	-0.5060	-0.0060	1.2	0.0882	-0.5036	-0.0036	0.72	0.0729
	$\beta_{(2)0} = -2$	-2.1055	-0.1055	5.275	0.7284	-2.0596	-0.0596	2.98	0.5351
	$\beta_{(2)1} = 1$	0.9954	-0.0046	-0.46	0.6730	0.9972	-0.0028	-0.28	0.0550
	η	1.0096	0.0096	0.96	0.2311	1.0043	0.0043	0.43	0.1717
	α	1.6801	0.4301	34.408	1.4003	1.5012	0.2512	20.096	0.8991
	$\beta_{(1)0} = 1$	1.0091	0.0091	0.91	0.0894	1.0047	0.0047	0.47	0.0681
1.25	$\beta_{(1)1} = -0.5$	-0.5093	-0.0093	1.86	0.0887	-0.5047	-0.0047	0.94	0.0709
	$\beta_{(2)0} = -2$	-2.0824	-0.0824	4.12	0.7414	-2.0605	-0.0605	3.025	0.5499
	$\beta_{(2)1} = 1$	0.9990	-0.0010	-0.1	0.0547	0.9994	-0.0006	-0.06	0.0463
	η	1.0018	0.0018	0.18	0.2163	1.0047	0.0047	0.47	0.1637
	α	2.5034	0.7534	43.0514	2.3173	2.1204	0.3704	21.1657	1.2991
	$\beta_{(1)0} = 1$	1.0068	0.0068	0.68	0.0879	1.0050	0.0050	0.50	0.0670
1.75	$\beta_{(1)1} = -0.5$	-0.5054	-0.0054	1.08	0.0907	-0.5051	-0.0051	1.02	0.0738
	$\beta_{(2)0} = -2$	-2.1011	-0.1011	5.055	0.7835	-2.0606	-0.0606	3.03	0.5538
	$\beta_{(2)1} = 1$	1.0003	0.0003	0.03	0.0550	0.9988	-0.0012	-0.12	0.0430
	η	1.0063	0.0063	0.63	0.2165	1.0048	0.0048	0.4800	0.1567

Apêndice B

Correção de Viés de Cox e Snell

Para derivar os estimadores de viés corrigidos de Cox e Snell, nós calculamos as derivadas de terceiro orden da função de log-verossimilhança com respeito aos parâmetros do modelo. Estas derivadas são dadas nas expressões que seguem. Primeiro nós introduzimos a seguinte notação.

Seja $g(a_t) = \log(2\Phi(a_t))$ y $\zeta(k, a_t) = \frac{\partial^k}{\partial a_t} g(a_t)$, para $k = 0, 1, 2, 3, 4$, de onde obtemos que $\zeta(0, a_t) = g(a_t)$, $\zeta(1, at) = \frac{\phi(a_t)}{\Phi(a_t)}$ $\zeta(2, a_t) = -\zeta(1, a_t)(a_t + \zeta(1, a_t))$ e $\zeta(3, a_t) = -(\zeta(1, a_t) + \zeta(2, a_t)[a_t + 2\zeta(1, a_t))]$. Também notamos:
$d_\lambda = -\frac{1}{\lambda^2} a_t$, $d_\beta = -\frac{1}{2\lambda\beta^{1/2}} \left(\frac{\sqrt{t}}{\beta} + \frac{1}{\sqrt{t}}\right)$, $d_{\lambda\lambda} = -\frac{2}{\lambda} d_\lambda$, $d_{\lambda\lambda\lambda} = -\frac{6}{\lambda^2} d_\lambda$, $d_{\beta\beta} = \frac{1}{4\lambda\beta^{3/2}} \left(\frac{3\sqrt{t}}{\beta} + \frac{1}{\sqrt{t}}\right)$, $d_{\beta\beta\beta} = -\frac{3}{8\lambda\beta^{5/2}} \left(\frac{5\sqrt{t}}{\beta} + \frac{1}{\sqrt{t}}\right)$, $d_{\lambda\beta} = -\frac{1}{\lambda} d_\beta$, $d_{\lambda\beta\beta} = -\frac{1}{\lambda} d_{\beta\beta}$, e $d_{\lambda\lambda\beta} = \frac{2}{\lambda^2} d_\beta$, e, Portanto,

$U_{\lambda\lambda} = \frac{n}{\lambda^2} - \frac{3}{\lambda^2} \sum_{i=1}^n a_{t_i}^2 + \frac{\alpha-1}{\lambda^2} \sum_{i=1}^n \left[2a_{t_i}\zeta(1, a_{t_i}) + a_{t_i}^2 \zeta(2, a_{t_i})\right]$

$U_{\lambda\beta} = -\frac{1}{\lambda^3\beta^2} \sum_{i=1}^n t_i + \frac{1}{\lambda^3} \sum_{i=1}^n \frac{1}{t_i} + (\alpha - 1) \sum_{i=1}^n \left[d_{\lambda\beta}\zeta(1, a_{t_i}) + d_\lambda d_\beta \zeta(2, a_{t_i})\right]$

$U_{\beta\beta} = \frac{n}{2\beta^2} - \sum_{i=1}^n \frac{1}{(t_i+\beta)^2} - \frac{1}{\lambda^2\beta^3} \sum_{i=1}^n t_i + (\alpha - 1)[d_{\beta\beta}\zeta(1, a_{t_i}) + d_\beta^2 \zeta(2, a_{t_i})]$

$U_{\beta\alpha} = \sum_{i=1}^n d_\beta \zeta(1, a_{t_i})$, $U_{\lambda\alpha} = -\frac{1}{\lambda} \sum_{i=1}^n a_{t_i}\zeta(1, a_{t_i})$ e $U_{\alpha\alpha} = -\frac{n}{\alpha^2}$.

Adicionalmente,

$$U_{\lambda\lambda\lambda} = -\frac{2n}{\lambda^3} + \frac{4}{\lambda^3} \sum_{i=1}^n a_{t_i}^2 \left[3 - (\alpha - 1)\zeta(2, a_{t_i})\right] + (\alpha - 1) \sum_{i=1}^n \left[d_{\lambda\lambda\lambda}\zeta(1, a_{t_i}) + d_\lambda d_{\lambda\lambda}\zeta(2, a_{t_i}) + d_\lambda^3 \zeta(3, a_{t_i})\right]$$

$$U_{\lambda\lambda\beta} = \frac{3}{\lambda^4\beta^2} \sum_{i=1}^n t_i - \frac{3}{\lambda^4} \sum_{i=1}^n \frac{1}{t_i} + (\alpha - 1) \sum_{i=1}^n \left[d_{\lambda\lambda\beta}\zeta(1, a_{t_i}) + (2d_{\lambda\beta}d_\lambda + d_{\lambda\lambda}d_\beta)\zeta(2, a_{t_i}) + d_\lambda^2 d_\beta \zeta(3, a_{t_i})\right]$$

$U_{\lambda\lambda\alpha} = \sum_{i=1}^{n} [d_{\lambda\lambda}\zeta(1, a_{t_i}) + d_{\lambda}^2\zeta(2, a_{t_i})]$, $U_{\lambda\beta\alpha} = \sum_{i=1}^{n} [d_{\lambda\beta}\zeta(1, a_{t_i}) + d_{\lambda}d_{\beta}\zeta(2, a_{t_i})]$,

$$U_{\lambda\beta\beta} = \frac{2}{\lambda^3\beta^3}\sum_{i=1}^{n} t_i + (\alpha-1)\sum_{i=1}^{n}\left[d_{\lambda\beta\beta}\zeta(1, a_{t_i}) + (2d_{\lambda\beta}d_{\beta} + d_{\beta\beta}d_{\lambda})\zeta(2, a_{t_i}) + d_{\lambda}d_{\beta}^2\zeta(3, a_{t_i})\right]$$

$$U_{\beta\beta\beta} = -\frac{n}{\beta^3} + \sum_{i=1}^{n}\frac{2}{(t_i+\beta)^3} + \frac{3}{\lambda^2\beta^4}\sum_{i=1}^{n} t_i - \frac{\alpha-1}{4\lambda^2\beta^2}\sum_{i=1}^{n}\left[\frac{3t_i}{\beta^2} + \frac{1}{t_i} + \frac{4}{\beta}\right]\zeta(2, a_{t_i}) + (\alpha-1)\sum_{i=1}^{n}\left[d_{\beta\beta\beta}\zeta(1, a_{t_i}) + d_{\beta\beta}d_{\beta}\zeta(2, a_{t_i}) + d_{\beta}^3\zeta(3, a_{t_i})\right]$$

$U_{\beta\beta\alpha} = \sum_{i=1}^{n}\left[d_{\beta\beta}\zeta(1, a_{t_i}) + d_{\beta}^2\zeta(2, a_{t_i})\right]$, $U_{\lambda\alpha\alpha} = 0$, $U_{\beta\alpha\alpha} = 0$, $U_{\alpha\alpha\alpha} = \frac{2n}{\alpha^3}$.

Assim,

$k_{\lambda\lambda} = \int_0^\infty U_{\lambda\lambda}\varphi_T(t;\lambda,\beta,\alpha)dt$, $k_{\lambda\beta} = \int_0^\infty U_{\lambda\beta}\varphi_T(t;\lambda,\beta,\alpha)dt$, $k_{\alpha\alpha} = \int_0^\infty U_{\alpha\alpha}\varphi_T(t;\lambda,\beta,\alpha)dt$,
$k_{\lambda\lambda\lambda} = \int_0^\infty U_{\lambda\lambda\lambda}\varphi_T(t;\lambda,\beta,\alpha)dt$, $k_{\lambda\lambda\beta} = \int_0^\infty U_{\lambda\lambda\beta}\varphi_T(t;\lambda,\beta,\alpha)dt$,..., $k_{\alpha\alpha\alpha} = \int_0^\infty U_{\alpha\alpha\alpha}\varphi_T(t;\lambda,\beta,\alpha)dt$.
Agora, chamando
$h_\lambda(t) = -\frac{1}{\lambda}[1 - a_t^2 + (\alpha-1)a_t\zeta(1, a_t)]$, $h_\alpha(t) = \frac{1}{\alpha}[1 + \alpha\log\{\Phi(a_t)\}]$ e $h_\beta(t) = \frac{1}{2}g(t) + \frac{1}{2\lambda\beta^{1/2}}\left(\frac{\sqrt{t}}{\beta} + \frac{1}{\sqrt{t}}\right)[a_t - (\alpha-1)\zeta(1, a_t)]$, onde $g(t) = \frac{\frac{\beta^{1/2}}{t^{3/2}} - \frac{1}{\beta^{1/2}t^{1/2}}}{\left(\frac{\beta}{t}\right)^{1/2} + \left(\frac{\beta}{t}\right)^{3/2}}$, então obtemos que:

$$k_{\lambda\lambda}^{(\lambda)} = -\frac{2n}{\lambda^3} + \frac{n}{\lambda^3}\int_0^\infty\left[3(4a_t^2 - \lambda a_t^2 h_\lambda(t)) - (\alpha-1)a_t^3\zeta(3, a_t)\right]\varphi_T(t;\lambda,\beta,\alpha)dt - n\frac{\alpha-1}{\lambda^3}\int_0^\infty\left[2a_t\zeta(1, a_t)(3 - \lambda h_\lambda(t)) + a_t^2\zeta(2, a_t)(6 - \lambda h_\lambda(t))\right]\varphi_T(t;\lambda,\beta,\alpha)dt$$

$$k_{\lambda\lambda}^{(\beta)} = -\frac{n}{\lambda^2}\int_0^\infty\left[d_\beta a_t(6 - 2(\alpha-1)a_t\zeta(3, a_t)) + 3(\alpha-1)a_t^2 h_\beta(t)\right]\varphi_T(t;\lambda,\beta,\alpha)dt + n\frac{\alpha-1}{\lambda^2}\int_0^\infty\left[2(d_\beta + a_t h_\beta(t))\zeta(1, a_t) + a_t(4d_\beta + a_t h_\beta(t))\zeta(2, a_t)\right]\varphi_T(t;\lambda,\beta,\alpha)dt$$

$$k_{\lambda\lambda}^{(\alpha)} = -\frac{3n}{\lambda^2}\int_0^\infty a_t^2 h_\alpha(\alpha)\varphi_T(t;\lambda,\beta,\alpha)dt + \frac{n}{\lambda^2}\int_0^\infty\left[2a_t\zeta(1, a_t) + a_t^2\zeta(2, a_t)\right]\left[1 + (\alpha-1)h_\alpha(t)\right]\varphi_T(t;\lambda,\beta,\alpha)dt$$

$$k_{\lambda\beta}^{(\lambda)} = \frac{n}{\lambda^4}\int_0^\infty\left(\frac{1}{t} - \frac{t}{\beta^2}\right)[\lambda h_\lambda(t) + (\alpha-1)\lambda^2 d_\beta a_t^2\zeta(3, a_t) - 3]\varphi_T(t;\lambda,\beta,\alpha)dt + n\frac{\alpha-1}{\lambda^2}\int_0^\infty d_\beta\left[(2 - \lambda h_\lambda(t))\zeta(1, a_t) + (4 - \lambda h_\lambda(t))a_t\zeta(2, a_t)\right]\varphi_T(t;\lambda,\beta,\alpha)dt$$

$$
\begin{aligned}
k_{\lambda\beta}^{(\beta)} = & \frac{n}{\lambda^3}\int_0^\infty \left[\frac{2t}{\beta^3} + \left(\frac{1}{t} - \frac{t}{\beta^2}\right) h_\beta(t)\right] \varphi_T(t;\lambda,\beta,\alpha)dt + \\
& n\frac{\alpha-1}{2\lambda\beta}\int_0^\infty d_\beta\left[\zeta(1,a_t) + a_t\zeta(2,a_t)\right](1-2\lambda h_\beta(t))\varphi_T(t;\lambda,\beta,\alpha)dt \\
& - n\frac{\alpha-1}{\lambda}\int_0^\infty d_\beta^2\left[2\zeta(2,a_t) + a_t\zeta(3,a_t)\right]\varphi_T(t;\lambda,\beta,\alpha)dt \\
& - n\frac{\alpha-1}{2\lambda^2\beta^{5/2}}\int_0^\infty \sqrt{t}\left[\zeta(1,a_t) + a_t\zeta(2,a_t)\right]\varphi_T(t;\lambda,\beta,\alpha)dt
\end{aligned}
$$

$$
\begin{aligned}
k_{\lambda\beta}^{(\alpha)} = & \frac{n}{\lambda^3}\int_0^\infty \left(\frac{1}{t} - \frac{t}{\beta^2}\right) h_\alpha(t)\varphi_T(t;\lambda,\beta,\alpha)dt \\
& - \frac{n}{\lambda}\int_0^\infty d_\beta\left[\zeta(1,a_t) + a_t\zeta(2,a_t)\right](1+(\alpha-1)h_\alpha(t))\varphi_T(t;\lambda,\beta,\alpha)dt
\end{aligned}
$$

$$
\begin{aligned}
k_{\beta\beta}^{(\lambda)} = & -n\int_0^\infty \left[\frac{h_\lambda(t)}{(t+\beta)^2} - \frac{t}{\lambda^3\beta^3}[2-\lambda h_\lambda(t)]\right]\varphi_T(t;\lambda,\beta,\alpha)dt - \\
& 4n\frac{\alpha-1}{\lambda}\int_0^\infty \left[d_{\beta\beta}\zeta(1,a_t) + (d_{\beta\beta}a_t + 2d_\beta^2)\zeta(2,a_t) + d_\beta^2 a_t\zeta(3,a_t)\right]\varphi_T(t;\lambda,\beta,\alpha)dt
\end{aligned}
$$

$$
\begin{aligned}
k_{\beta\beta}^{(\beta)} = & -\frac{n}{\beta^3} + \frac{n}{\lambda^2\beta^3}\int_0^\infty \left[\frac{\lambda^2\beta^3}{(t+\beta)^2}\left[\frac{2}{t+\beta} - h_\beta(t)\right] + t\left[\frac{3}{\beta} - h_\beta(t)\right]\right]\varphi_T(t;\lambda,\beta,\alpha)dt \\
& + n\frac{\alpha-1}{\lambda}\int_0^\infty \left[\lambda\zeta(1,a_t)(d_{\beta\beta\beta} + d_{\beta\beta}h_\beta(t)) + \zeta(2,a_t)(d_{\beta\beta}d_\beta + \lambda d_\beta^2 h_\beta(t))\right]\varphi_T(t;\lambda,\beta,\alpha)dt \\
& + n\frac{\alpha-1}{8\lambda^2\beta^{1/2}}\int_0^\infty \left[\left(\frac{3t}{\beta^2} + \frac{2}{\beta} - \frac{1}{t}\right)\zeta(2,a_t) + 4\lambda\beta^{3/2}d_\beta^2\zeta(3,a_t)\right]\varphi_T(t;\lambda,\beta,\alpha)dt
\end{aligned}
$$

$$
\begin{aligned}
k_{\beta\beta}^{(\alpha)} = & -\frac{n}{\lambda^2\beta^3}\int_0^\infty \left[\frac{\lambda^2\beta^3}{(t+\beta)^2} + t\right] h_\alpha(t)\varphi_T(t;\lambda,\beta,\alpha)dt \\
& + n\int_0^\infty (d_{\beta\beta}\zeta(1,a_t) + d_\beta^2\zeta(2,a_t))[1+(\alpha-1)h_\alpha(t)]\varphi_T(t;\lambda,\beta,\alpha)dt
\end{aligned}
$$

$$
k_{\beta\alpha}^{(\lambda)} = -\frac{2n}{\beta}\int_0^\infty d_\beta\left[\zeta(1,a_t) + a_t\zeta(2,a_t) - \lambda\zeta(1,a_t)h_\lambda(t)\right]\varphi_T(t;\lambda,\beta,\alpha)dt
$$

$$
k_{\beta\alpha}^{(\beta)} = n\int_0^\infty \left[\frac{2\beta}{\lambda}d_{\beta\beta}\zeta(1,a_t) + d_\beta^2\zeta(2,a_t) + d_\beta h_\beta(t)\right]\varphi_T(t;\lambda,\beta,\alpha)dt
$$

$$
k_{\beta\alpha}^{(\alpha)} = n\int_0^\infty d_\beta\zeta(1,a_t)h_\alpha(t)\varphi_T(t;\lambda,\beta,\alpha)dt
$$

$$
k_{\alpha\alpha}^{(\alpha)} = \frac{2n}{\alpha^3}, \quad k_{\alpha\alpha}^{(\lambda)} = 0 \quad \text{e} \quad k_{\alpha\alpha}^{(\beta)} = 0.
$$

Apêndice C

Variância assintótica Do Pseudo Estimador No Caso Multivariado

As expressões para o calculo da variância assintótica do pseudo estimador e seu estimador sandwich no caso multivariado são dadas a continuação. Assumindo que f_j' e f_j'' para $j = 1, 2, ..., p$, existem e notando $\alpha_{jk} = \alpha_{kj}$, então, fazendo $w_j = \frac{f_j(z_j)}{F_j(z_j)}$ nós temos que os elementos da função de pseudo escore da i-ésima observação são:

$U_i(\alpha_j) = \frac{1}{\alpha_j + \sum_{\substack{l=1 \\ l\neq j}}^{p} \alpha_{jl} \log[F_l(z_{li})]} + \log[F_j(z_{ji})]$,

$U_i(\alpha_{jk}) = \frac{\log[F_k(z_{ki})]}{\alpha_j + \sum_{\substack{l=1 \\ l\neq j}}^{p} \alpha_{jl} \log[F_l(z_{li})]} + \frac{\log[F_j(z_{ji})]}{\alpha_k + \sum_{\substack{l=1 \\ l\neq k}}^{p} \alpha_{kl} \log[F_l(z_{li})]} + 2\log[F_j(z_{ji})]\log[F_k(z_{ki})]$,

$U_i(\xi_j) = -\frac{1}{\eta_j} w_{ji} \sum_{\substack{q=1 \\ q\neq j}}^{p} \left[\frac{\alpha_{jq}}{\alpha_q + \sum_{\substack{l=1 \\ l\neq q}}^{p} \alpha_{ql} \log[F_l(z_{li})]} + 2\alpha_{jq} \log[F_q(z_{qi})] \right] - \frac{1}{\eta_j} \left[\frac{f_j'(z_{ji})}{f_j(z_{ji})} + (\alpha_j - 1) w_{ji} \right]$,

$U_i(\eta_j) = -\frac{1}{\eta_j} - \frac{1}{\eta_j} w_{ji} z_{ji} \sum_{\substack{q=1 \\ q\neq j}}^{p} \left[\frac{\alpha_{jq}}{\alpha_q + \sum_{\substack{l=1 \\ l\neq q}}^{p} \alpha_{ql} \log[F_l(z_{li})]} + 2\alpha_{jq} \log[F_q(z_{qi})] \right] - \frac{z_{ji}}{\eta_j} \left[\frac{f_j'(z_{ji})}{f_j(z_{ji})} + (\alpha_j - 1) w_{ji} \right]$,

para $j, k = 1, 2, ..., p$ e $j \neq k$, em que $\alpha_{jk} = \alpha_{kj}$.

Enquanto que fazendo $v_j = \frac{f_j'(z_j)}{f_j(z_j)}$, $t_j = \frac{f_j''(z_j)}{f_j(z_j)}$ e $s_j = \frac{f_j'(z_j)}{F_j(z_j)}$, nós temos que os elementos da matriz hessiana do logaritmo da pseudo verossimilhança para a i-observação são:

$J_i(\alpha_j\alpha_j) = -\frac{1}{(\alpha_j + \sum_{\substack{l=1 \\ l\neq j}}^{p} \alpha_{jl} \ln[F_l(z_{li})])^2}$, $J_i(\alpha_j\xi_j) = -\frac{1}{\eta_j} w_{ji}$, $J_i(\alpha_j\eta_j) = -\frac{z_{ji}}{\eta_j} w_{ji}$,

para $j = 1, 2, ..., p$

$J_i(\alpha_j\alpha_k) = J(\alpha_j\alpha_{lk}) = 0$, para $k \neq l \neq j$ $J(\alpha_j\alpha_{jk}) = -\frac{\log[F_k(z_{ki})]}{(\alpha_j + \sum_{\substack{l=1 \\ l\neq j}}^{p} \alpha_{jl} \log[F_l(z_{li})])^2}$

$J(\alpha_j\alpha_{lk}) = 0$ para $k \neq l \neq j$, $J_i(\alpha_j\xi_j) = -\frac{1}{\eta_j} w_{ji}$ e $J_i(\alpha_j\eta_j) = -\frac{z_{ji}}{\eta_j} w_{ji}$,

$J_i(\alpha_j\xi_k) = \frac{\alpha_{jk}}{\eta_k} \frac{w_{ki}}{(\alpha_j + \sum_{\substack{l=1 \\ l\neq j}}^{p} \alpha_{jl} \log[F_l(z_{li})])^2}$, $J_i(\alpha_j\eta_k) = \frac{\alpha_{jk}}{\eta_k} \frac{z_{ki} w_{ki}}{(\alpha_j + \sum_{\substack{l=1 \\ l\neq j}}^{p} \alpha_{jl} \log[F_l(z_{li})])^2}$,

para $j = 1, 2, ..., p$ e $l, k = 1, 2, ..., p$ $k \neq l$, enquanto que

$J_i(\alpha_{jk}\alpha_{jk}) = -\frac{(\log[F_k(z_{ki})])^2}{(\alpha_j + \sum_{\substack{l=1 \\ l\neq j}}^{p} \alpha_{jl} \log[F_l(z_{li})])^2} - \frac{(\log[F_j(z_{ji})])^2}{(\alpha_k + \sum_{\substack{l=1 \\ l\neq k}}^{p} \alpha_{lk} \log[F_l(z_{li})])^2}$, $J_i(\alpha_{jk}\alpha_{lq}) = 0$

para $j \neq k \neq l \neq q$,

$$J_i(\alpha_{jk}\alpha_{jl}) = -\frac{\log[F_k(z_{ki})]\log[F_l(z_{li})]}{(\alpha_j+\sum_{\substack{l=1\\l\neq j}}^{p}\alpha_{jl}\log[F_l(z_{li})])^2}, \quad J_i(\alpha_{jk}\alpha_{lk}) = -\frac{\log[F_j(z_{ji})]\log[F_l(z_{li})]}{(\alpha_k+\sum_{\substack{l=1\\l\neq k}}^{p}\alpha_{lk}\log[F_l(z_{li})])^2},$$

$$J_i(\alpha_{jk}\xi_j) = -\frac{1}{\eta_j}w_{ji}\left[\frac{\alpha_k+\sum_{\substack{l=1\\l\neq j,k}}^{p}\alpha_{kl}\log[F_l(z_{li})]}{(\alpha_k+\sum_{\substack{l=1\\l\neq k}}^{p}\alpha_{kl}\log[F_l(z_{li})])^2} + 2\log[F_k(z_{ki})]\right],$$

$$J_i(\alpha_{jk}\xi_k) = -\frac{1}{\eta_k}w_{ki}\left[\frac{\alpha_j+\sum_{\substack{l=1\\l\neq j,k}}^{p}\alpha_{jl}\log[F_l(z_{li})]}{(\alpha_j+\sum_{\substack{l=1\\l\neq j}}^{p}\alpha_{jl}\log[F_l(z_{li})])^2} + 2\log[F_j(z_{ji})]\right],$$

$$J_i(\alpha_{jk}\xi_l) = \frac{1}{\eta_l}\frac{\alpha_{jl}w_{li}\log[F_k(z_{ki})]}{(\alpha_j+\sum_{\substack{l=1\\l\neq j}}^{p}\alpha_{jl}\log[F_l(z_{li})])^2} + \frac{1}{\eta_l}\frac{\alpha_{kl}w_{li}\log[F_j(z_{ji})]}{(\alpha_k+\sum_{\substack{l=1\\l\neq k}}^{p}\alpha_{lk}\log[F_l(z_{li})])^2},$$

$$J_i(\alpha_{jk}\eta_j) = -\frac{1}{\eta_j}w_{ji}z_{ji}\left[\frac{\alpha_k+\sum_{\substack{l=1\\l\neq j,k}}^{p}\alpha_{kl}\log[F_l(z_{li})]}{(\alpha_k+\sum_{\substack{l=1\\l\neq k}}^{p}\alpha_{kl}\log[F_l(z_{li})])^2} + 2\log[F_k(z_{ki})]\right],$$

$$J_i(\alpha_{jk}\eta_k) = -\frac{1}{\eta_k}w_{ki}z_{ki}\left[\frac{\alpha_j+\sum_{\substack{l=1\\l\neq j,k}}^{p}\alpha_{jl}\log[F_l(z_{li})]}{(\alpha_j+\sum_{\substack{l=1\\l\neq j}}^{p}\alpha_{jl}\log[F_l(z_{li})])^2} + 2\log[F_j(z_{ji})]\right],$$

$$J_i(\alpha_{jk}\eta_l) = \frac{1}{\eta_l}\frac{\alpha_{jl}w_{li}z_{li}\log[F_k(z_{ki})]}{(\alpha_j+\sum_{\substack{l=1\\l\neq j}}^{p}\alpha_{jl}\log[F_l(z_{li})])^2} + \frac{1}{\eta_l}\frac{\alpha_{kl}w_{li}z_{li}\log[F_j(z_{ji})]}{(\alpha_k+\sum_{\substack{l=1\\l\neq k}}^{p}\alpha_{lk}\log[F_l(z_{li})])^2},$$

$$J_i(\xi_j\xi_j) = \frac{1}{\eta_j^2}(t_{ji}-v_{ji}^2) - \frac{1}{\eta_j^2}w_{ji}^2\sum_{\substack{q=1\\q\neq j}}^{p}\left[\frac{\alpha_{jq}^2}{(\alpha_q+\sum_{\substack{l=1\\l\neq q}}^{p}\alpha_{ql}\log[F_l(z_{li})])^2}\right] +$$

$$\frac{1}{\eta_j^2}(s_{ji}-w_{ji}^2)\left\{\sum_{\substack{q=1\\q\neq j}}^{p}\left[\frac{\alpha_{jq}}{\alpha_q+\sum_{\substack{l=1\\l\neq q}}^{p}\alpha_{ql}\log[F_l(z_{li})]} + 2\alpha_{jq}\log[F_q(z_{qi})]\right] + \alpha_j - 1\right\}$$

$$J_i(\xi_j\xi_k) = -\sum_{\substack{q=1\\q\neq j,k}}^{p}\left[\frac{\alpha_{jq}\alpha_{kq}}{\eta_j\eta_k}\frac{w_{ji}w_{ki}}{(\alpha_q+\sum_{\substack{l=1\\l\neq q}}^{p}\alpha_{ql}\log[F_l(z_{li})])^2} - \frac{2\alpha_{jk}}{\eta_j\eta_k}w_{ji}w_{ki}\right],$$

$$J_i(\eta_j\xi_j) = \frac{z_{ji}}{\eta_j^2}(t_{ji}-v_{ji}^2) - \frac{z_{ji}}{\eta_j^2}w_{ji}^2\sum_{\substack{q=1\\q\neq j}}^{p}\left[\frac{\alpha_{jq}^2}{(\alpha_q+\sum_{\substack{l=1\\l\neq q}}^{p}\alpha_{ql}\log[F_l(z_{li})])^2}\right] +$$

$$\frac{z_{ji}}{\eta_j^2}(s_{ji}-w_{ji}^2)\left\{\sum_{\substack{q=1\\q\neq j}}^{p}\left[\frac{\alpha_{jq}}{\alpha_q+\sum_{\substack{l=1\\l\neq q}}^{p}\alpha_{ql}\log[F_l(z_{li})]} + 2\alpha_{jq}\log[F_q(z_{qi})]\right] + \alpha_j - 1\right\}$$

$$+ \frac{1}{\eta_j^2}w_{ji}\sum_{\substack{q=1\\q\neq j}}^{p}\left[\frac{\alpha_{jq}}{\alpha_q+\sum_{\substack{l=1\\l\neq q}}^{p}\alpha_{ql}\log[F_l(z_{li})]} + 2\alpha_{jq}\log[F_q(z_{qi})]\right] + \frac{1}{\eta_j^2}\left[v_{ji} + (\alpha_j-1)w_{ji}\right],$$

$$J_i(\eta_k\xi_j) = -\sum_{\substack{q=1\\q\neq j,k}}^{p}\left[\frac{\alpha_{jq}\alpha_{kq}}{\eta_j\eta_k}\frac{w_{ji}w_{ki}z_{ki}}{(\alpha_q+\sum_{\substack{l=1\\l\neq q}}^{p}\alpha_{ql}\log[F_l(z_{li})])^2} - \frac{2\alpha_{jk}}{\eta_j\eta_k}w_{ji}w_{ki}z_{ki}\right],$$

$$J_i(\eta_j\eta_j) = \frac{2}{\eta_j^2} + \frac{z_{ji}^2}{\eta_j^2}(t_{ji}-v_{ji}^2) + \frac{2}{\eta_j^2}w_{ji}z_{ji}\sum_{\substack{q=1\\q\neq j}}^{p}\left[\frac{\alpha_{jq}}{\alpha_q+\sum_{\substack{l=1\\l\neq q}}^{p}\alpha_{ql}\log[F_l(z_{li})]} + 2\alpha_{jq}\log[F_q(z_{qi})]\right] +$$

$$\frac{2z_{ji}}{\eta_j^2}\left[v_{ji} + (\alpha_j-1)w_{ji}\right] - \frac{1}{\eta_j^2}w_{ji}^2z_{ji}^2\sum_{\substack{q=1\\q\neq j}}^{p}\left[\frac{\alpha_{jq}^2}{(\alpha_q+\sum_{\substack{l=1\\l\neq q}}^{p}\alpha_{ql}\log[F_l(z_{li})])^2}\right] +$$

$$\frac{z_{ji}^2}{\eta_j^2}(s_{ji}-w_{ji}^2)\left\{\sum_{\substack{q=1\\q\neq j}}^{p}\left[\frac{\alpha_{jq}}{\alpha_q+\sum_{\substack{l=1\\l\neq q}}^{p}\alpha_{ql}\log[F_l(z_{li})]} + 2\alpha_{jq}\log[F_q(z_{qi})]\right] + \alpha_j - 1\right\},$$

$$J_i(\eta_j\eta_k) = -\sum_{\substack{q=1\\q\neq j,k}}^{p}\left[\frac{\alpha_{jq}\alpha_{kq}}{\eta_j\eta_k}\frac{w_{ji}w_{ki}z_{ji}z_{ki}}{(\alpha_q+\sum_{\substack{l=1\\l\neq q}}^{p}\alpha_{ql}\log[F_l(z_{li})])^2} - \frac{2\alpha_{jk}}{\eta_j\eta_k}w_{ji}w_{ki}z_{ji}z_{ki}\right].$$

Referências Bibliográficas

[1] Ahrens, L.H. (1954). The log-normal distribution of the elements (a fundamental law of geochemistry and its subsidiary). Geochemica and Chosmochemica Acta, 5, 49-73.

[2] Aerts, M., Geys, H., Molenberghs, G. and Ryan, L. (2002). Topics in Modelling of Clustered Data. Chapman and Hall: london.

[3] Akaike, H. (1974). A new look at statistical model identification. *IEEE Transaction on Automatic Control*. **AU-19**, 716–722.

[4] Amemiya, T. (1973). Regression Analysis When the Dependent Variable is Truncated Normal. *Econometrica*, **41**, 997–1016.

[5] Amemiya, T. (1984). Tobit models: a survey. *Journal of Econometrics*, **24**, 2–61.

[6] Amemiya, T. (1985). Advanced Econometrics. *Harvard University Press. Cambridge, Massachusetts.*.

[7] Arellano-Valle, R. B., Gómez, H. W. and Quintana, F.A. (2004). A new class of skew-normal distributions. *Communications in Statistics – Theory and Methods*, **33**, 1465–1480.

[8] Arellano-Valle, R. B., Gómez, H. W. and Quintana, F.A. (2005). Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference*, **128**, 427–443.

[9] Arnold, B. C. and Strauss, D. (1988). Bivariate distributions with exponential conditionals. *Journal of the American Statistical Association*, **83**, 522–527.

[10] Arnold, B. C. and Strauss, D. (1991). Bivariate distributions with conditionals in prescribed exponential families. *Journal of the Royal Statistical Society, Ser B*, **53**, 365–375.

[11] Arnold, B. C. and Strauss, D. (1991). Pseudolikelihood Estimation: Some Examples. *Sankhya: The Indian Journal of Statistics, Series B*, **53**, 233–243.

[12] Arnold, B. C., Castillo, E. and Sarabia, J. M. (1992). Conditionally specified distributions, Lecture Notes in Statistics, vol. 73, Springer Verlag, New York.

[13] Arnold, B. C., Beaver, R. J., Groeneveld, R. A. and Meeker, W. Q. (1993). The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrica*, **58**, 471–488.

[14] Arnold, B. C., Castillo, E. and Sarabia, J. M. (1999). Conditionally specification of statistical models, Springer Series in statistics, Springer Verlag, New York.

[15] Arnold, B. C., Castillo, E. and Sarabia, J. M. (2001). Conditionally specified Multivariate Skewed Distributions. *Sankhya: The Journal of Statistics, Series A* , **64**, 206–226.

[16] Arnold, B. C., Castillo, E. and Sarabia, J. M. (2002). Conditionally specified distributions: an introduction. *Statistical Science*, **16**, 249–274.

[17] Arnold, B. C. (2004). discussion of Jones, M.C. Families of distribution arising from distributions of order statistics. *Test* , **13**, 23–25.

[18] Arnold, B. C., Gómez, H.W. and Salinas, H. S. (2009). On multiple contraint skewed models'. *Statistics* , **43:3**, 279–293.

[19] Azme, K., Ismail Z., Haron K. and Ahmad T.M. (2005). Nonlinear growth models for modeling oil palm yield growth. *J Math Stat*, **I(3)**, 225–233.

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

[21] Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **46**, 199–208.

[22] Azzalini, A. and Dalla-Valle, A. (1986). The multivariate skew-normal distribution . *Biometrika*, **83**, 715–726.

[23] Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution . *Journal of the Royal Statistical Society*, **61**, 579–602.

[24] Balakrishnan, N. (2002). Discussion of Arnold, B.C. and Beaver, R.J. Skewed multivariate modelsrelated to hidden truncation and/or selective reporting. *Test*, **11**, 37–39.

[25] Balakrishnan, N., Leiva, V., Sanhueza, A. and Cabrera, E. (2009). Mixture inverse Gaussian distribution and its transformations, moments and applications. *Statistics*, **43**, 91–104.

[26] Barndor-Nielsen, O.E. (1983). On a formula to the distribution of the maximum likelihood estimator. *Biometrika*, **70**, 343–365.

[27] Barndor-Nielsen, O.E. (1994). Adjusted versions of profile likelihood and directed likelihood, and extended likelihood. likelihood estimator. *Journal of the Royal Statistical Society B*, **56**, 125–140.

[28] Barndor-Nielsen, O.E. (1995). Stable and invariant adjusted profile likelihood and directed likelihood for curved exponential models. *Biometrika*, **82**, 489–500.

[29] Barros M. (2007). Modelos de Regressão Birnbaum-Saunders generalizados. Tese de doutorado, Instituto de Matemática e Estatística, Universidade de São Paulo.

[30] Barros, M., Paula, G.A., and Leiva, V. (2008). A new class of survival regression models with heavy-tailed errors: robustness and diagnostics. *Lifetime Data Analysis*, **14**, 316–332.

[31] Besag, J. (1975). Statistical Analysis of Non-Lattice Data. *Journal of the Royal Statistical Society*, **24**, 179–195.

[32] Birnbaum, Z. W. (1950). Effect of linear truncation on a multinormal population. *Ann. Math. Statist.*, **21**, 272–279.

[33] Birnbaum, Z.W., Saunders, S.C. (1969a). A New Family of Life Distributions. *Journal of Applied Probability*, 6, 319–327.

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

[35] Cancho, V.G., Lachos, V.H. and Ortega, E.M. (2008). A nonlinear regression model with skew-normal errors. Statistical papers,51, 547–558.

[36] Castillo, E. and Hadi, A.S. (1995). A method for estimating parameters and quantiles of distributions of continuous random variables. *Computational Statistics and Data Analysis*, 20, 421–439.

[37] Chai, H. and Bailey, K.,(2008). Use of log-normal distribution in analysis of continuous data with a discrete component at zero. *Statistics in Medicine*, **27**, 3643–3655.

[38] Chaibub-Neto, E. and Branco, M.,(2003). Bayesian reference analysis for binomial calibration problem. RT MAE 2003-12: IME-USP.

[39] Chang, D.S. and Tang, L.C.,(1994). Percentile bounds and tolerance limits of the Birnbaum-Saunders distribution. *Comunication in Statistics-Theory and Methods*, **23**, 2853–2863.

[40] Cheng, C. and Riu, J. (2006). On Estimating Linear Relationships When Both Variables Are Subject to Heteroscedastic Measurement Errors. *Technometrics*, **48**, 511–519.

[41] Chiogna, M. (1997). Notes on estimation problems with scalar skew-normal distributions. Technical Report 15, University of Padua, Dept Statistical Sciences.

[42] Cisneiros, A., Cribari-Neto, F. and Araújo, C.(2008). On Birnbaum-Saunders inference. *Computational Statistics and Data Analysis*, **52**, 4939–4950.

[43] Cox, D.R., and Snell, E. (1968). A general definition of residuals. *J. Roy. Statist.Soc.* , B30, 248–275.

[44] Cox, D.R. and Hinkley, D.V. (1974). Theoretical statistics. Chapman and Hall, London.

[45] Cook, R.D. (1977). Detection of influential observations in linear regression . *Tecnometric* , **19**, 15–18.

[46] Cook, R.D. (1986). Assessment of local influence (with discussion) . *J. Roy. statist Soc.* , **B48**, 133–169.

[47] Cragg, J. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods, *ECONOMETRICA*, **39**, 829-844.

[48] Da silva, M.(2005). Estimação e teste de hipótese baseados em verossimilhanças perfiladas. Ph.D. Tese, Universidade de São Paulo, São Paulo, Brasil.

[49] Da silva, M.K.(2007). Modelos de Regressão Birnbaum-Saunders Generalizados. Ph.D. Tese, Universidade de São Paulo, São Paulo, Brasil.

[50] Davison, A. C., and Gigli, A., (1989). Desviance residuals and normal scores plots. *Biometrika*, **76**, 211–221.

[51] Díaz-García, J.A., Leiva-Sánchez, V. (2005). A new family of life distributions based on the elliptically contoured distributions. *J. Statist. Plann. Inference.*, 128, 445–457.

[52] Diaconis, P. and Efron B. (1983). Computer intensive method in statistics. *sci. Amer.* , **248**, 116–130.

[53] Durrans, S. R. (1992). Distributions of fractional order statistics in hydrology. *Water Resources Research*, **28**, 1649–1655.

[54] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist*, **7**, 1–26.

[55] Efron, B.(1981). Nonparametric standard errors and confidence intervals. *Canadian Journal of Statistics*, **9**, 139–172.

[56] Efron, B.(1982). The jackknife, the bootstrap, and other resampling plans. *CBMS 38, SIAM-NSF.*

[57] Efron, B.(1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, **82**, 171–200.

[58] Efron, B. and Tibshirani, R. J. (1993). An Introduction to the Bootstrap. *New York: Chapman and Hall.*

[59] Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics – Theory and Methods*, **31**, 497–512.

[60] Fair, R. (1978). A theory of extramarital Affairs. *Journal of political economy*, **86**, 45–61.

[61] Fernández, C. and Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, **93**, 359–371.

[62] Foong, F. S. (1999). Impact of mixture on potential evapotranspiration, growth and yield of palm oil. Pro 1999, PORIM Interl. Palm Oil Cong (Agric.). 265–287.

[63] Galea, M., Leiva., V. and Paula, G.A. (2004). Influence diagnostic in log-Birnbaum-Saunders regression models. *J Appl Statist*, **31**, 1049–1064.

[64] Geys, H., Molenberghs, G. and Stuart, L. (1998). A Note on the Comparison of Pseudo-Likelihood and Generalized Estimating Equations for Marginally Specified Odds Ratio Models with Exchangeable Association Structure. *Journal of Statistical computation and Simulation*, **62**, 45–72.

[65] Geys, H., Molenberghs, G. and Ryan, L. (1999). Pseudolikelihood Modeling of Multivariate Outcomes in Developmental Toxicology. *Journal of the American Statistical Association*, **94**, 734–745.

[66] Gómez, H. W., Venegas, O. and Bolfarine, H. (2007). Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics*, **18**, 395–407.

[67] Gómez, H. W., Olivero, D.E., Salinas, H.S. and Bolfarine, H. (2009). Bimodal extension based on the skew-normal distribution with application to pollen data. *Environmetrics*, **22**, 50–62.

[68] Gupta, A. K., Kundu, D. (2000). Generalized Exponential Distribution: Different Method Of Estimations . *Journal of Statistical Computation and Simulation*, **00**, 1–22.

[69] Gupta, A. K., Chang, F. C. and Huang, W. J. (2002). Some skew-symmetric models. *Random Operators Stochastic Equations*, **10**, 113–140.

[70] Gupta, A. K. and Nadarajah S. (2004). On the moments of the beta-normal distribution. *Commun Stat theor Methods*, **33**,1–13.

[71] Gupta, D. and Gupta, R. C. (2008). Analyzing skewed data by power normal model. *Test*, **17**, 197–210.

[72] Gupta, R., Kannan, N., and Raychaudhuri, A. (1997). Analysis of Lognormal Survival Data. *Mathematical Biosciences*, **139**,103–115.

[73] Gupta, R.S. and Gupta, R.D. (2004). Generalized skew normal model. *Test*, **13**, 501–524.

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

[75] Job, J. S., Halsey, N. A., Boulos, R., Holt, E., Farrell, D., Albrecht, P., Brutus, J. R., Adrien, M., Andre, J., Chan, E., Kissinger, P., Boulos, C., and the Cite Soleil/JHU Project Team. (1991). Successful immunization of infants at 6 months of age with high dose Edmonston-Zagreb measles vaccine. *Pediatric Infectious Diseases Journal* , **10**, 303–311.

[76] Joe, H. (1997). Multivariate Models and Dependence Concepts. Chapman and Hall, New York

[77] Johnson, S. Kotz, S. and Balakrishnan, N. (1994). Continuous Univariate Distributions. 2nd Ed *Wiley, New York*, Vol. 1.

[78] Johnson, S. Kotz, S. and Balakrishnan, N. (1995). Continuous Univariate Distributions. 2nd Ed *Wiley, New York*, Vol. 2.

[79] Jones, M. C. (2004). Families of distributions arising from the distributions of order statistics. *Test*, **13**, 1–43.

[80] Kalbfleisch, J. D., and Prentice, R. L., (2002). The Statistical Analysis of Faiture Time Data. John Wiley and Sons, New York.

[81] Kamakura, W.A. and Wedel, M. (2001). Explanatory tobit factor analysis for multivariate censored data. *Multivariate Behavioral Research,*, 36 (1), 53-82.

[82] Kim, H. J. (2005).On a class of two-piece skew-normal distribution. *Statistic*, **39:6**, 537–553.

[83] Kotz, S., Balakrishnan, N. and Johnson, N.L. (2000). Continuous Multivariate Distributions, vol. 1: Models and applications, *John Wiley and Sons, New York.*

[84] Kullback, S. and Leibler, R.A. (1951). On Information and Sufficiency. *Annals of Mathematical Statistics*, **22**, 79–86.

[85] Lawley, D. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, 43, 295–303.

[86] Lee, E. T., and Wang, J. W., (2003). Statistical Methods for Survival Data Analysis. John Wiley and Sons, New York.

[87] Lee, S.Y. and Xu, L. (2004). Influence analysis of nonlinear mixed-effects models. *Computational statistics and data Analysis*, **45**, 321–341.

[88] Lehmann, E. L. (1953). The power of rank tests. *Annals of Mathematical Statistics*, **24**, 23–43.

[89] Lehmann, E. L., and Casella, G. (1998). Theory of Point Estimation, 2nd edition. New York: Springer-Verlag.

[90] Lemonte, A., Cribari-Neto, F. and Vasconcellos, K. (2007). Improved statistical inference for the two-parameter Birnbaum-saunders distribution. *Computational Statistics and Data Analysis*, **51**, 4656–4681.

[91] Lemonte, A. and Cordeiro, G.(2009). Birnbaum-Saunders nolinear regression models. *Computational Statistics and Data Analysis.*

[92] Lesaffre, E. and Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, **54**, 570–582.

[93] Ley, C. and Paindaveine, D. (2009). On the Singularity of Multivariate Skew-Symmetric Models. Submitted.

[94] Lin, G.D., Stoyanov, J. (2009). The logarithmic Skew-Normal Distributions are Moment-Indeterminate. *Journal of Applied Probability Trust*, **46**, 909–916.

[95] Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979). Multivariate analysis. Academic Press,London New Yord.

[96] Mateu-Figueras, G. y Pawlosky-Glanh. (2003). Una alternativa a la distribución log-normal. Actas del XXVII Congreso Nacional de Estadística e Investigación Operativa (SEIO) *Sociedade de Estadítica e Investigación Operativa*, 1849–1858.

[97] Mateu-Figueras, G., Pawlosky-Glanh and Barcelo-Vidal, C. (2004). The natural law in geochemistry: lognoraml or logskew-normal?. Proceeding of the 32th International Geological 2004 Congres, Firenze (it), Abs. vol., pt. 2 abs 233-3, p 215. 1849–1858.

[98] McDonald, J.F. and Moffitt, R.A. (1980). The uses of tobit analysis. *The Review of Economics and Statistics.*

[99] Moulton, L. and Halsey N. (1995). A Mixture Model with Detection Limits for Regression Analysses of Antibody Response to Vaccine. *Biometrics*, **51**, 1570–1578.

[100] Mudholkar, G. S. and Hutson, A. D. (2000). The epsilon-skew-normal distribution for analyzing near-normal data. *Journal of Statistical Planning and Inference*, **83**, 291–309.

[101] Muñoz, K.(2009). An Extension of the Normal Censored Regression Model. Estimation and Applications.. Ph.D. Tese, Pontificia Universidad Católica de Chile, Santiago, Chile.

[102] Nadarajah, S. (2008). Some Algebra For Pearson Type VII Random Variables. *Bull.Korean. Math.Soc.* , 45, 339–353.

[103] Nelsen, R.B. (1999). An introduction to copulas. Lecture notes in statistics. 139, Springer-Berlin.

[104] Nelson, W.B. and Hahn, G.J. (1972). Linear estimation of a regression relationship from censored data. Part I-simples methods and their applications. *Technometrics* , 14,247–269.

[105] Ng, H.K.T., Kundu, D., Balakrishnan, N. (2003). Modified moment estimation for the two-parameter Birnbaum-Saunders distribution. *Comput. Statist. Data Anal.*, 43, 283–298.

[106] Neter, J., Kutner, M.H. Nachtsheim, C.J. and Wasserman, W. (1996). Applied linear regression models. 3rd edition. Irwin, Illiois.

[107] O'Hagan, A. and Leonard, T. (1976). Bayes estimation subjectto uncertainty about parameter constraints. *Biometrika*, 63, 201–203.

[108] Olivero, D. E., Gómez, H. W. and Quintana, F.A. (2009). Bayesian modeling using a class of bimodal skew-elliptical distributions. *Journal of Statistical Planning and Inference*, **139**, 1484–1492.

[109] Ortega, E. M. (2001). Influence analysis in generalized log-gamma regression models. Doctor Thesis, Department of Statistics, Universitu of São paulo, Brasil (in portuguese).

[110] Ortega, E. M., Bolfarine, H. and Paula, G.A. (2003). Influence diagnostics in generalized log-gamma regression models. *Comput. Stat Data An.* , **42**,165–186.

[111] Pace, L., and Savan, A. (1997). Pincipled of Statistical Inference from a Neo-Fisherian Perspective. Singapore: World Scientific.

[112] Patriota, A.G., and Lemonte, A.J. (2009). Bias correction in a multivariate normal regression model with general parameterization. *Statistics and probability letters* , **79**,1655–1662.

[113] Paula, G. (2004). Modelos de Regressão com apoio computacional. Instituto de Matemáticas e Estatística, Universidade de São Paulo.

[114] Pewsey, A. (2000). Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics*, **27**, 859–870.

[115] Pewsey, A. (2006). Some observations on a simple means of generating skew distributions. In: N. Balakrishnan, E. Castillo and J.M. Sarabia (Eds.), *Advances in Distribution Theory, Order Statistics, and Inference*, pp. 75–84. Birkhäuser, Boston.

[116] Pewsey, A., Gómez, H. W. and Bolfarine, H. (2010). Likelihood based inference for distributions of fractional order statistics. *submited.*

[117] Poon, W. Y. and Poon, Y.S. (1999). Conformal normal curvature and assessement of local influence. *Journal of the Royal Statistical Society*, **Series B61**, 51–61.

[118] Prudnikov, A.P., A., Brychkov, Y.A., Marichev, O.I. (1990). Integrals and Series. Vols. 1,2 and 3. Amsterdam, Gordon and Breach Science Publishers.

[119] Plungpongpun, A., and Naik, Dayanand (2008). Multivariate Analysis of Variance Using a Kotz type Distribution. *Proceeding of the world Congress on Engineering, London, U.K.*, 2.

[120] Ratkowsky D.A. (1983). Nonlinear regression models. Marcel Dekker, New Yord.

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

[122] Rousseeuw, P.J. (1984). Least median of squares regression . *J. Amer. Statist. Assoc.*, **79**, 871–880.

[123] Roberts, C.D. (1966). A correlation model useful in the study of twins. *Journal of the American Statistical Association*, **61**, 1184–1190.

[124] Ruben, H. (1954). On the moments of order statistics in samples from normal populations. *Biometrika*, **41**,–200.

[125] Sahu, S., Dey, D. and Branco, M. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics*, **31**, 129–150.

[126] Sanhueza, A., Leiva, V. and Balakrishnan, N. (2008). The generalized Birnbaum-Saunders distribution and its theory, methodology and application. *Comunicatios in Statistics-Theory and Methods*, **37**, 645–670.

[127] Sarabia, J. M. and Castillo, E. (2005). About a class of max-stable families with applications to income distributions. *International Journal of Statistics*, **LXIII**, 505–527.

[128] Schneider, H. (1986). Truncated and censored samples from normal populations, vol. 70, Marcel Dekker, New York.

[129] Sen, P.K., Singer, J.(1993). Lange Sample Methods in Statistics: an introdution with applications. *ChapMan & Hall/CRC*

[130] Severini, T.A. (1998). An approximation to the modified profile likelihood function. *Biometrika*, **85**, 403–411.

[131] Severini, T.A. (1999). An empirical adjustment to the likelihhod ratio statistic. *Biometrika*, **86**, 235–247.

[132] Severini, T.A. (2000). Likelihood Methods in Statistics. Oxford: Oxford University press.

[133] Skovgaarrd, I.M. (1996). An explicit large-deviation approximation to one-parameter tests. *Bernoulli*, **2**, 145–165.

[134] Strauss, D. J. and Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, **85**, 204–212.

[135] Suto, O. (1914). Studies on some functional equations, Tohuku Mathematics Journal,6, 1–15.

[136] Therneau, T.M., Grambsch, P.M. and Fleming, T.R. (1990). Martingale-based residuals for survival models. *Biometrika*, **77**, 147–160.

[137] Tibaldi, F., Molenberghs, G., Burzykowski, T. and Geys, H. (2004). Pseudolikelihood estimation for a marginal multivariate survival model. *Statistics in Medicine*, **23**, 947–963.

[138] Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, **26**, 24–36.

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

[140] Vuong, Q. (1989). Likelihood Ratio Tests for Model selection and Non-Nested Hypotheses. *Econometrica*, **57**, 307–333.

[141] Wei, B. C., Hu, Y.Q., Fung, W.K., (1998). Generalized leverage and its applications. *Scand. J. Statist.*, **25**, 25–37.

[142] Whittemore, A. and Altschuler, B. (1976). lung cancer incidence in cigarette smokers: further analysis of doll and hill's data for british physicians. *Biometrics.* , 1, 80–83.

[143] Xie, F. C., Lin, J.G., and Wei, B.C. (2009). Diagnostics for skew nonlinear regression models with AR(1) errors. *Computational Statistics and data analysis*, **53**, 4403–4416.

[144] Zhu, H. and Lee, S. (2001). Local influence for incomplete-data models. *Journal of the Royal Statistical Society*, **Series B63**, 111–126.