Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

A distribuição beta semi-normal generalizada geométrica

Thiago Gentil Ramires

Dissertação apresentada para obtenção do título de
Mestre em Ciências. Área de concentração: Estatística e Experimentação Agronômica

Piracicaba
2013

Thiago Gentil Ramires
Bacharel em Estatística

A distribuição beta semi-normal generalizada geométrica
versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:
Prof. Dr. **EDWIN MOISES MARCOS ORTEGA**

Dissertação apresentada para obtenção do título de Mestre em Ciências. Área de concentração: Estatística e Experimentação Agrônômica

Piracicaba
2013

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - ESALQ/USP**

Ramires, Thiago Gentil

A distribuição beta semi-normal generalizada geométrica / Thiago Gentil
Ramires.- - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - -
Piracicaba, 2013.
114 p: il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2013.

1. Análise de sobrevivência 2. Distribuição beta semi-normal generalizada geométrica
3. Modelos locação-escala I. Título

CDD 519.546
R173d

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

DEDICATÓRIA

Com amor dedico...

*Ao melhor pai e mãe do mundo, **Ademir Ramires** e **Janet Gentil Ramires**, que mesmo com alguma dificuldade sempre estiveram dispostos a me ajudar e lutar para realizar meus sonhos.*

*Aos meus Irmãos **Adriano Gentil Ramires**, **Juliano Gentil Ramires** e **Andreza Gentil Ramires**, os quais sinto muita falta de estar ao meu lado.*

*A minha amiga **Monique Bisconsim Ganasin**, que vem me dando muita força nesses últimos dois anos.*

*Especialmente aos meus avós **José Gentil** e **Zulmira Astolfi Gentil**, os quais não estão mais presentes entre nós e que deixam muitas saudades e boas lembranças. Amo vocês!*

And wherever you've gone and wherever we might go.

It don't seem fair. Today just disappeared.

Your light's reflected now. Reflected from afar.

We were but stones. Your light made us stars.

Pearl Jam - Light Years

AGRADECIMENTOS

Agradeço primeiramente a meus pais, por me apoiar sempre nas minhas escolhas e conquistas dos meus sonhos.

À meu amigo e orientador Edwin Moises Marcos Ortega, o qual me incentivou e colaborou a conquistar mais essa conquista em minha vida. Muito obrigado por tudo.

À todos amigos de Piracicaba que sempre estiveram ao meu lado nos melhores e piores momentos em especial Rodrigo Pescim, Maria Cristina, Gabriel Bruno de Lemos, Natalia da Silva, Alessandra do Santos, Flávio Braga, Rafael Jacomini, André Sanches, Rodrigo Lambert, Florian Schumacher, Erik Augusto, Michel Augusto, Debora Robert, Gabriel Rambarussi, Pedro Cerqueira, Cassio Dessotti, Guilherme Biz, Lucas Santana, Tiago Santana, Ana Julia Righetto, Luiz Ricardo Nakamura, Elisângela Oliveira, Maria Joseane e Edilan Quaresma.

À meus melhores amigos Luiz Fernando Navarro e Pedro Henrique Baggio, Gabriel Polizel Santos, Fábio Casagrande Basseto, Gustavo Gomes Correia, Jeanmichel Cavalaro e Henrique Bersanete, que mesmo distante mostra que uma amizade verdadeira é para sempre.

Às secretárias Luciane Brajão, Solange Sabadin e Mayara Segatto que nesses anos se tornaram partes essenciais de nossas vidas.

À CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, pela bolsa de mestrado concedida.

A todos os professores que convivi durante o curso de mestrado em Estatística e Experimentação Agronômica, me dando a oportunidade de participar desta família ESALQ, fornecendo conhecimentos e possibilidades de alto nível, com as quais tive a oportunidade de trabalhar em minhas pesquisas: Dercio, Roseli A. Leandro, Clarice G. B. Demétrio, Sílvio S. Zocchi, César G. de Lima, Sônia M. D. S. Piedade e Renata Alcarde.

Aos alunos do curso de Pós-Graduação em Estatística e Experimentação Agronômica da ESALQ/USP, os quais fizeram parte desta fase.

Em fim, a todos os amigos que me ajudaram a compor mais um pedaço da minha história.

Muito obrigado!

SUMÁRIO

RESUMO	9
ABSTRACT	11
LISTA DE FIGURAS	13
LISTA DE TABELAS	15
1 INTRODUÇÃO	17
2 REVISÃO BIBLIOGRÁFICA	19
2.1 Análise de sobrevivência	19
2.2 Censura	19
2.3 Função densidade de probabilidade	20
2.4 Função de sobrevivência	21
2.5 Função taxa de falha	22
2.6 Propriedades	23
2.6.1 Momentos	23
2.6.2 Função geradora de momentos	25
2.6.3 Função quantílica	26
2.6.4 Desvios médios	27
2.6.5 Curvas de Bonferroni e Lorenz	28
2.7 Estimadores não paramétricos	28
2.7.1 Risco empírico	29
2.8 Modelos de regressão paramétricos	30
2.8.1 Modelo de locação-escala	30
2.9 Modelos paramétricos	31
2.9.1 Distribuição semi-normal (SN)	31
2.9.2 Distribuição semi-normal generalizada (SNG)	32
2.9.3 Distribuição beta semi-normal generalizada (BSNG)	34
2.9.3.1 Momentos	35
2.9.3.2 Assimetria e curtose	36
2.9.3.3 Função geradora de momentos	38
2.9.3.4 Desvios médios	38
2.9.3.5 Função quantílica	39
2.9.3.6 Curvas de Bonferroni e Lorenz	40

2.9.4 Distribuição log-beta semi-normal generalizada (LBSNG)	41
2.9.4.1 Momentos	42
2.9.4.2 Função geradora de momentos	42
2.9.5 Método de máxima verossimilhança	43
2.9.6 Critérios de seleção	43
3 DISTRIBUIÇÃO BETA SEMI-NORMAL GENERALIZADA GEOMÉTRICA . .	45
3.1 Classe de distribuições geométricas	45
3.2 Distribuição beta semi-normal generalizada geométrica	47
3.2.1 Expansões para momentos	50
3.2.2 Momentos	52
3.2.3 Assimetria e curtose	54
3.2.4 Função geradora de momentos	56
3.2.5 Desvios médios	56
3.2.6 Função quantílica	58
3.2.7 Curvas de Lorenz e Bonferroni	58
3.2.8 Estimação dos parâmetros por máxima verossimilhança	59
3.3 Modelo de regressão log-beta semi-normal generalizada geométrica	62
3.3.1 Expansões para momentos	64
3.3.2 Momentos	65
3.3.3 Assimetria e curtose	66
3.3.4 Função geradora de momentos	67
3.3.5 Estimação dos parâmetros	67
3.4 Aplicações	68
3.4.1 Simulações	68
3.4.2 Dados de leucemia	69
4 CONSIDERAÇÕES FINAIS	75
4.1 Conclusão	75
4.2 Estudos futuros	75
REFERÊNCIAS	77
APÊNDICES	83

RESUMO

A distribuição beta semi-normal generalizada geométrica

Com o avanço tecnológico aprimorado, diferentes comportamentos do tempo de vida vem sendo estudados, e com isso é necessário a criação de novos modelos, muitas vezes mais complexos, para melhor ajuste e inferência sobre a população em estudo. A distribuição beta semi-normal generalizada é útil para modelagem de tempos de vida, e com isso propomos neste trabalho uma distribuição mais ampla chamada distribuição beta semi-normal generalizada geométrica, cuja função de risco pode assumir as formas crescente, decrescente, forma de banheira ou modal. A função densidade da nova distribuição é escrita como uma combinação linear da função densidade da distribuição beta semi-normal generalizada, sendo assim, algumas importantes propriedades da nova distribuição foram obtidas, como: momentos, assimetria, curtose, função geradora de momentos, desvios médios, função quantil e curvas de Lorenz e de Bonferroni. Para a estimação dos parâmetros, é utilizado o método de máxima verossimilhança. Também foi proposto no trabalho, o novo modelo de regressão baseado na distribuição beta semi-normal generalizada geométrica, os quais podem ser muito úteis em análise de dados reais por serem mais flexíveis.

Palavras-chave: Análise de sobrevivência; Distribuição beta semi-normal generalizada geométrica; Modelo locação-escala

ABSTRACT

The beta generalized half-normal géométric distribution

Due to the technological improved advances, different behaviors of the lifetime has been studied and for this reason, it is necessary to create new statistical models, many times more complex, for the better fit and inferences about the population under study. The beta generalized half-normal distribution is useful for modeling lifetime data, and in this sense, we propose, in this work, a wider distribution called the geometric beta generalized half-normal distribution in which the hazard function takes the forms increasing, decreasing, bathtub and unimodal. The density function of the new distribution can be written as a linear combination of the beta generalized half-normal densities, and thereby, some properties of the new distribution can be obtained such as the moments, skewness, kurtosis, moment generating function, mean deviations, quantile function and Lorenz and Bonferroni curves. For the estimation of the parameters, we use the maximum likelihood method considering the presence of censored data. We also propose a new regression model based on the geometric beta generalized half-normal distribution, which can be very useful in the analysis of real data due to their flexibility.

Keywords: Geometric beta generalized half-normal distribution; Location-scale model; Survival analysis

LISTA DE FIGURAS

Figura 1 - Curvas de sobrevivência: (a) Própria e (b) Imprópria	22
Figura 2 - Diferentes comportamentos de função de taxa de falha.	23
Figura 3 - Distribuições com coeficiente de assimetria: (A) Positivo (B) Nulo (C) Negativo	24
Figura 4 - Formas da curva TTT plot.	29
Figura 5 - Formas da função de risco da distribuição SNG para alguns valores dos parâmetros.	33
Figura 6 - Gráficos dos valores dos coeficientes de assimetria da distribuição BSNG como função de a para alguns valores de b e como função de b para alguns valores de a	37
Figura 7 - Gráficos dos valores dos coeficientes de curtose da distribuição BSNG como função de a para alguns valores de b e como função de b para alguns valores de a	37
Figura 8 - Formas para a densidade da distribuição BSNGG variando os valores dos parâmetros: (a) Para diferentes valores de α , a , p com $\theta = 1.5$ e $b = 3$. (b) Para diferentes valores de θ , b , p com $\alpha = 2.5$ e $a = 5$. (c) Para diferentes valores de θ , a , p com $\alpha = b = 1.5$	49
Figura 9 - Formas para função taxa de falha da distribuição BSNGG para alguns valores dos parâmetros: (a) Função taxa de falha considerando $\alpha = 1,5$, $\theta = 55$ e $b = 2,0$ fixos, variando a e p . (b) Função taxa de falha considerando $\alpha = 0,5$, $\theta = 5,5$ e $a = 15$ fixos, variando b e p . (c) Função taxa de falha considerando $\alpha = 1,0$, $a = 1,5$ e $b = 2,0$ fixos, variando θ e p	50
Figura 10 - Gráficos dos valores dos coeficientes de assimetria e curtose da distribuição BSNGG como função de p para alguns valores de a e $b = 2$	55
Figura 11 - Gráficos dos valores dos coeficientes de assimetria e curtose da distribuição BSNGG como função de p para alguns valores de b e $a = 0,5$	55
Figura 12 - Gráficos das curvas de Bonferroni e Lorenz utilizando a distribuição GBGSN variando o parâmetro p com $\alpha = 3,0$, $\theta = 10,0$, $a = 0,3$ e $b = 0,3$	59

Figura 13 - Formas da função densidade da distribuição LBSNGG para alguns valores dos parâmetros. (a) Para alguns valores de b, p com $\mu = 0, \sigma = 1$ e $a = 2.5$. (b) Para alguns valores de σ, a, p com $\mu = 0$ and $b = 0.5$. (c) Para alguns valores de μ, σ, p com $a = b = 1.5$	63
Figura 14 - Assimetria e curtose da distribuição LBSNGG para alguns valores dos parâmetros variando e p fixo. (a) Assimetria para alguns valores de a e p com $\mu = 5, \sigma = 13$ e $b = 0, 3$. (b) Curtose para alguns valores de a e p com $\mu = 5, \sigma = 13$ e $b = 0, 3$	66
Figura 15 - Assimetria e curtose da distribuição LBSNGG para alguns valores dos parâmetros variando e p fixo. (a) Assimetria para alguns valores de b e p com $\mu = 5, \sigma = 13$ e $a = 0, 3$. (b) Curtose para alguns valores de a e p com $\mu = 5, \sigma = 13$ e $b = 0, 3$	66
Figura 16 - Histogramas de simulações de dados com densidades originais para a distribuição BSNGG considerando 100.000,00 observações, $\alpha = 2, \theta = 10, a = 0, 5, b = 0, 5$ e p variando.	69
Figura 17 - (a) Densidade das distribuições BSNGG e BSNG ajustadas para os dados de leucemia. (b) Densidade das distribuições BSNGG e SNG ajustadas para os dados de leucemia	71
Figura 18 - Estimativa da função acumulada da distribuição BSNGG e distribuição acumulada empírica comparadas com: (a) Estimativa da distribuição acumulada da distribuição BSNG. (b) Estimativa da distribuição acumulada da distribuição SNG.	72
Figura 19 - (a) Risco empírico para os dados de leucemia. (b) Função de risco das distribuições BSNGG, BSNG e SNG ajustadas para os dados de leucemia.	72
Figura 20 - Diferença entre as curvas de sobrevivência para os grupos "Auer rods" positivo e negativo com: (a) Contagem de glóbulos brancos = 100. (b) Contagem de glóbulos brancos = 0, 75.	74

LISTA DE TABELAS

Tabela 1 - Casos Particulares da distribuição BSNGG.	49
Tabela 2 - Casos particulares da distribuição LBSNGG.	64
Tabela 3 - Estatísticas descritivas sobre a variável tempo.	70
Tabela 4 - Estimativas dos parâmetros, desvios padrões (em parênteses) e critérios de informação para o modelo BSNGG e seus sub-modelos aplicados aos dados de leucemia.	70
Tabela 5 - Teste da razão de verossimilhança para os modelos BSNGG, BSNG e SNG	71
Tabela 6 - Têste de logrank para as covariáveis X_1 e X_2	73
Tabela 7 - Estimativas dos parâmetros, desvios padrões (em parênteses), p-valor [em chaves] e critérios de informação para o modelo LBSNGG e seus sub-modelos aplicados aos dados de leucemia	73

1 INTRODUÇÃO

Presente em praticamente todas as áreas do conhecimento, a estatística desenvolve um papel de destaque no que diz respeito à análise de dados. Dentre essas áreas, uma das que mais cresceu nos últimos anos foi a de análise de sobrevivência, fato este evidenciado por sua quantidade de aplicações nas mais diversas áreas de pesquisa, tendo um enfoque especial em algumas como medicina, agrônômica, engenharia, biologia, economia e demais áreas relacionadas a saúde e financeiras. Este crescimento se deve em parte ao aprimoramento e desenvolvimento de técnicas e métodos estatísticos, cada vez mais eficientes, aliado ao grande avanço tecnológico e computacional que permite a criação de modelos mais sofisticados para a modelagem de dados com comportamentos diferentes dos casos usuais estudados na literatura.

Os dados em análise de sobrevivência possuem características especiais, em que a variável observada é o tempo, a qual possui suporte positivo, portanto é necessário utilizar distribuições com suporte positivo e, normalmente, assimétricas à direita. Outra característica pertinente em dados de sobrevivência é a presença de censura, que são informações parcialmente observadas, isso ocorre com muita frequência em estudos, em que nem sempre é possível esperar o tempo necessário para que ocorra o evento de interesse.

Com o avanço tecnológico aprimorado, diferentes comportamentos do tempo de falha vem sendo estudados, e com isso é necessário a criação de novos modelos para melhor ajuste aos dados e para melhores inferências sobre a população em estudo. Mistura e transformações entre distribuições geram resultados interessantes quando o interesse é encontrar novas formas para a função densidade de probabilidade e função taxa de falha ou risco.

Com frequência, novas distribuições vem sendo desenvolvidas para a modelagem de dados de sobrevivência, caso da distribuição semi-normal generalizada (SNG) proposta por Cooray e Ananda (2008), que possui uma característica muito importante presente na distribuição gamma generalizada, que é a forma de banheira da função taxa de falha. Por se tratar de uma distribuição mais simples que a gamma generalizada e com características importantes como formas, foi proposto por Pescim *et al.* (2010) a distribuição beta semi-normal generalizada (BSNG), utilizando o método proposto por Eugene *et al.* (2002) para gerar distribuições da classe (Beta-G). Algumas distribuições mais recentes pertencentes a classe de distribuições beta são descritas na Seção 2.9.3.

A distribuição BSNG com dois parâmetros de forma a mais proporcionou uma maior flexibilidade à distribuição semi-normal generalizada, possibilitando a modelagem de dados mais assimétricos.

Com base na distribuição BSNG, foi proposto neste trabalho uma nova distribuição chamada beta semi-normal generalizada geométrica (BSNGG), a qual é resultado da aplicação da transformação geométrica aplicada na distribuição BSNG, que consiste em adicionar um parâmetro de forma $p \in [0, 1)$, o qual permitiu maior flexibilidade sem perder as características importantes herdadas da distribuição SNG, como as diferentes formas de taxa de falha. As distribuições geométricas mais recentes na literatura são referenciadas na seção em 3.1.

Como em muitos casos a variável resposta tempo é influenciada por covariáveis, também foi proposto no trabalho um modelo de regressão baseado na nova distribuição BSNGG, chamado de modelo de locação e escala log-beta semi-normal generalizada geométrica (LBSNGG), que permite verificar a influencia de covariáveis sobre a variável resposta. Assim como a distribuição BSNGG, o modelo de regressão LBSNGG se mostrou mais flexível, resultando em modelagens mais realistas para aplicações em dados reais. Diversos modelos de locação e escala vem surgindo para possibilitar um melhor ajuste sobre o comportamento da variável resposta, alguns desses modelos são apresentados da seção 2.9.4.

Para cada modelo proposto, utilizamos tratamentos matemáticos apresentando algumas propriedades importantes como assimetria, curtose, momentos entre outros. Também foi realizado estudos de simulações e aplicações em dados reais para ilustrar a importância e consistência dos modelos. As novas distribuição também se tornam adequadas para testar a qualidade do ajuste de seus sub-modelos especiais.

2 REVISÃO BIBLIOGRÁFICA

2.1 Análise de sobrevivência

A análise de sobrevivência se caracteriza por ser um conjunto de técnicas estatísticas utilizadas quando se deseja analisar dados correspondentes ao tempo de ocorrência de um determinado evento de interesse, por exemplo, quando a variável resposta é o tempo até a ocorrência desse evento representado por uma variável contínua não negativa. Esse tipo de análise tem se expandido nas ultimas décadas, pelo fato da evolução tecnológica (Colosmo e Giolo; 2006), podendo ser aplicada em diversas áreas, especialmente em engenharia como estudo de confiabilidade, ciências biológicas como estudo do tempo de vida e, recentemente, muito utilizado na área financeira como estudo de inadimplência.

2.2 Censura

Nos dados de sobrevivência é muito comum encontrar informações que foram parcialmente observadas, essa é a principal característica desse tipo de dados. Essas informações parciais podem ser causadas por diversos motivos, como o término do estudo sem a ocorrência do evento de interesse, a desistência de um paciente no estudo entre outras causas. Esse tipo de observação é denominada como censura, que pode ser classificada como censura à direita, à esquerda ou intervalar.

Censura à direita, a mais comum, acontece quando o tempo de ocorrência do evento de interesse está a direita do tempo observado como no caso do término do estudo sem a ocorrência do evento de interesse. Censura a esquerda, que não é muito comum, acontece quando no começo do estudo o evento de interesse já tenha ocorrido, como no caso de estudo sobre alfabetização de crianças, tendo em vista que no começo do estudo algumas já saibam ler e escrever. Por fim, censura intervalar, é dito quando não se tem o tempo exato de ocorrência do evento de interesse, isso ocorre quando as observações são feitas em intervalos de tempo.

As censuras a direita também podem ser classificadas como censura do tipo I, censura do tipo II e censura aleatória.

- Censura do tipo I: Ocorre quando o tempo do estudo é fixo, neste caso, o número de falhas é uma variável aleatória. Considerando x_i para $i = 1, 2, \dots, n$ o tempo observado do i -ésimo indivíduo, então se x_i for menor que o tempo fixado do estudo,

ocorre falha, caso contrário ocorre censura.

- Censura do tipo II: Uma amostra censurada do tipo II é aquela que são observados exatamente um número pré-determinado de falhas. O tempo de duração do experimento será uma variável aleatória.
- Censura do tipo aleatória: Caracterizada pela perda de acompanhamento do indivíduo no decorrer do estudo sem este ter apresentado falha. Este tipo de censura é a que mais aparece na prática, podendo ser representada por meio de duas variáveis aleatórias independentes X_i e C_i , em que a variável aleatória X representa o tempo de falha de um indivíduo e a variável aleatória C representa o tempo de censura associado a este indivíduo. O tempo da i -ésima observação para este caso, é dada por $x_i = \min(X_i, C_i)$.

Para cada indivíduo $i = 1, 2, \dots, n$, os dados de sobrevivência são em geral representados pelo par (x_i, δ_i) , em que x_i representa o tempo observado e δ_i é uma variável indicadora definida por:

$$\delta_i = \begin{cases} 1, & \text{se corresponde ao tempo de falha,} \\ 0, & \text{se corresponde ao tempo de censura.} \end{cases}$$

Sendo a variável aleatória X não negativa e contínua representando o tempo de vida, três funções relacionadas a essa variável são particularmente úteis em análise de sobrevivência.

- Função densidade de probabilidade: $f(x)$.
- Função de sobrevivência: $S(x)$.
- Função taxa de falha: $h(x)$.

2.3 Função densidade de probabilidade

Em estatística, a função densidade de probabilidade (fdp) ou simplesmente função de densidade é uma função não negativa definida por:

$$\begin{aligned}
f(x) &= \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X \leq x + \Delta x)}{\Delta x} \\
&= \lim_{\Delta x \rightarrow 0} \frac{P(X \leq x + \Delta x) - P(X \leq x)}{\Delta x} \\
&= \lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = F'(x) \\
&= \frac{d[F(x)]}{dx}
\end{aligned} \tag{1}$$

2.4 Função de sobrevivência

A função de sobrevivência, também conhecida como função de confiabilidade, está relacionada a qualquer variável aleatória observada em um período de tempo até que ocorra determinado evento de interesse. O nome, função de confiabilidade, é um termo muito comum utilizado em engenharia, enquanto a função de sobrevivência é o termo usado em uma ampla gama de aplicações, principalmente em medicina.

A função de sobrevivência é definida como sendo a probabilidade de um indivíduo sobreviver além de um certo tempo, dado que já sobreviveu até o instante anterior a esse tempo. A função de sobrevivência é representada por

$$\begin{aligned}
S(x) &= P(X \geq x) = 1 - P(X \leq x) = 1 - F(x) \\
&= 1 - \int_0^x f(u)du = \int_x^\infty f(u)du,
\end{aligned} \tag{2}$$

em que $F(x)$ representa a função distribuição acumulada (fda) de X . As principais características da função de sobrevivência são,

$$\lim_{x \rightarrow 0} S(x) = 1, \tag{3}$$

$$\lim_{x \rightarrow \infty} S(x) = 0, \tag{4}$$

e quando a equação (4) acontece, então $S(x)$ será chamado de função de sobrevivência própria. Em alguns casos pode ocorrer de

$$\lim_{x \rightarrow \infty} S(x) = p, \tag{5}$$

no qual $p \in [0, 1]$, e neste caso, a função de sobrevivência é chamada de imprópria.

A Figura 1 apresenta o comportamento da função de sobrevivência própria dada pela equação (4) e da função de sobrevivência imprópria dada pela equação (5), a qual considera que $p = 0,2$, ou seja, 20% das observações em estudo não estão mais em risco.

Figura 1 - Curvas de sobrevivência: (a) Própria e (b) Imprópria

2.5 Função taxa de falha

Considerando o tempo de ocorrência de um determinado evento de interesse sendo representado por uma variável aleatória X , a função de taxa de falha em um determinado intervalo de tempo $[x, x + \Delta x]$ é definida como a probabilidade de que a falha ocorra nesse determinado intervalo, dado que não ocorreu antes do tempo x , dividida pelo comprimento do intervalo.

Em um intervalo de tempo pequeno, $\Delta x \rightarrow 0$, a função taxa de falha $h(x)$ representa a razão instantânea de falha no tempo x , dado que ele sobreviveu até o instante anterior a x . A função taxa de falha é representada por:

$$h(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X \leq x + \Delta x \mid X \geq x)}{\Delta x} = \frac{f(x)}{S(x)} = \frac{-d\{\ln[S(x)]\}}{dx},$$

em que $f(x)$ representa a função densidade de probabilidade (1) e $S(x)$ a função de sobrevivência (2).

Algumas formas que a função de taxa de falha pode assumir são apresen-

tadas na Figura 2, em que a função h_1 é crescente, o que indica que a taxa de falha da observação aumenta ao decorrer do tempo, h_3 é decrescente, o que indica que a taxa de falha da observação diminui ao decorrer do tempo e a h_2 é constante, o que indica que o risco de falha é igual para todo o tempo. A combinação das três formas de taxa de falha geram a forma de banheira ou forma de “U”, a qual é útil para explicar diversos comportamentos de tempos de vida, como exemplo, a vida de seres humanos, que ao nascer o risco de morrer é alto, ao passar do anos se estabiliza e ao envelhecer o risco novamente aumenta.

Figura 2- Diferentes comportamentos de função de taxa de falha.

2.6 Propriedades

Algumas medidas em estatística são utilizadas para verificar o comportamento das distribuições de probabilidade, medidas essas como desvio padrão, média, moda, variância, assimetria, curtose, desvios médios, assim como propriedades utilizadas em áreas econômicas como curvas de Bonferroni, Lorenz entre outras. Essas medidas são capazes de nos dar informações valiosas sobre o comportamento dos dados a serem estudados e também de indicar possíveis comportamentos desejados de distribuições.

2.6.1 Momentos

Os momentos são muito importantes em estatística para caracterizar distribuições de probabilidade, pois eles dão uma idéia da tendência central, dispersão e assimetria de uma distribuição de probabilidades. A expressão genérica de esperança, o

s -ésimo momento ou momento em torno da média, de ordem s de uma variável aleatória X é dado por $E(X^s)$ e definido na literatura como:

$$\eta'_s = E(X^s) = \int_{-\infty}^{\infty} x^s f(x) dx.$$

Em distribuições que não possuem forma fechada, os momentos são normalmente calculados por métodos de expansões, alguns recentes trabalhos no cálculo de momentos por expansões foram estudados por Paranaíba *et al.* (2011) e Cordeiro *et al.* (2013b) e aplicados às distribuições beta Burr XII e beta exponenciada Weibull, respectivamente.

Outra utilização dos momentos são os cálculos de assimetria e curtose. As medidas de assimetria podem ser positivas ou negativas, ou mesmo indefinidas. Uma medida de assimetria negativa indica que a cauda do lado esquerdo da função densidade de probabilidade é maior do que o lado direito, indicando que o maior volume dos dados, incluindo a mediana, se situam para a direita da média. Por sua vez, uma medida positiva indica que a cauda do lado direito é mais longa do que o lado esquerdo, assim o volume dos valores estão para a esquerda da média. Um valor zero indica que os valores são relativamente uniformemente distribuídas em ambos os lados da média mas não necessariamente implica em uma distribuição simétrica. A Figura 3 apresenta algumas possíveis formas da função de densidade de probabilidade caracterizando os coeficientes de assimetria positivo, negativo e nulo.

Figura 3 - Distribuições com coeficiente de assimetria: (A) Positivo (B) Nulo (C) Negativo

A assimetria de uma variável aleatória X , denotado por γ_1 , pode ser escrita em termos dos momentos como:

$$\gamma_1 = \frac{E(X^3) - 3E(X)E(X^2) + 2[E(X)]^3}{[Var(X)]^{3/2}},$$

em que $Var(X)$ representa a variância de X .

A medida de curtose, que também pode ser obtida por meio dos momentos, em que sua interpretação clássica é o peso da cauda da distribuição, ou seja, quando maior a medida, maior será o tamanho da calda. Sua denotação é dada por γ_2 e definida como

$$\gamma_2 = \frac{E(X^4) - 4E(X)E(X^3) + 6E(X^2)[E(X)]^2 - 3[E(X)]^4}{[Var(X)]^2},$$

2.6.2 Função geradora de momentos

A função geradora de momentos de uma variável aleatória X , é uma especificação alternativa de sua distribuição de probabilidade (no entanto, nota-se que nem todas as distribuições de probabilidade possuem função geradora de momentos). Sua expressão geral é definida como

$$M_X(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx,$$

com $t \in \mathbb{R}$.

Assim, ela fornece uma alternativa diferente para os cálculos dos momentos, em que a s -ésima derivada definida em relação a $t = 0$ fornece o s -ésimo momento

$$E(X^s) = \left. \frac{d^s M_X(t)}{dt^s} \right|_{t=0} = M_X^{(s)}(0).$$

Assim como o cálculo dos momentos, a função geradora de momentos para distribuições que não possuem forma fechada deve ser obtida a partir de expansões. Alguns trabalhos utilizando expansões para o cálculo da função geradora de momentos são citados na Subseção 2.6.1.

2.6.3 Função quantílica

Em probabilidade e estatística, a função quantílica de uma distribuição de probabilidade de uma variável aleatória específica é também chamada de função de distribuição cumulativa inversa. Essa função determina qual o valor máximo que a variável aleatória deve assumir tal que a probabilidade de ser menor ou igual a esse valor seja “ u ”. Desse modo a função quantílica é obtida isolando a variável aleatória X da função distribuição acumulada

$$F(x; \boldsymbol{\lambda}) = u, \quad u \in [0, 1], \quad (6)$$

em que $F(x; \boldsymbol{\lambda})$ representa a função densidade acumulada de X , $\boldsymbol{\lambda}$ representa o vetor de parâmetros da distribuição de X e $u \in [0, 1]$ representa a probabilidade de X ser menor ou igual a um dado valor. Assim a função quantílica é definida na literatura como

$$Q(u, \boldsymbol{\lambda}) = F^{-1}(x; \boldsymbol{\lambda}), \quad (7)$$

em que $F^{-1}(\cdot)$ denota a inversa de F e desempenha um papel importante em estatística, pois permite o cálculo de algumas propriedades de distribuições que não possuem forma fechada, como exemplo, o cálculo de assimetria e curtose demonstrado por Kenney e Keeping (1962) e Moors (1988) respectivamente e recentemente utilizado para os cálculos de assimetria e curtose da distribuição Kumaraswamy generalizada semi-normal (Cordeiro *et al.* (2012)).

Outra propriedade fácil de se demonstrar, é o cálculo dos momentos utilizando a função quantílica dado por

$$\begin{aligned} \eta'_s &= \int_{-\infty}^{\infty} x^s f(x) dx, \quad \text{considerando } u = F(x), \\ &= \int_0^1 Q(u)^s du. \end{aligned}$$

Em estudo de simulações a função quantílica também apresenta um papel importante, pois com ela é possível gerar dados de uma distribuição. Gerando valores aleatórios entre 0 e 1 para o percentíl u , ou seja, $u \sim U(0, 1)$, e substituindo os valores

na equação (6), é possível encontrar os valores simulados de x_i atribuindo os valores dos parâmetros desejados e resolvendo a equação.

Alguns estudos recentes envolvendo a função quantílica em novas distribuições como no caso da beta-Weibull geométrica e Kumaraswamy Weibull modificada são apresentadas por Cordeiro *et al.* (2011) e Cordeiro *et al.* (2013d) respectivamente.

2.6.4 Desvios médios

A quantidade de dispersão em uma população pode ser medida pela totalidade dos valores absolutos dos desvios em relação à média (no caso de uma distribuição simétrica) ou em relação à mediana (no caso de uma distribuição assimétrica). Se X é uma variável aleatória com distribuição qualquer que apresente suporte positivo, com média igual ao primeiro momento $\eta'_1 = E(X)$ e mediana M , então o desvio médio em relação à média é definido na forma

$$\begin{aligned}
 \delta_1(X) &= \int_0^\infty |x - \eta'_1| f(x) dx \\
 &= \int_0^{\eta'_1} (\eta'_1 - x) f(x) dx + \int_{\eta'_1}^\infty (x - \eta'_1) f(x) dx \\
 &= \int_0^{\eta'_1} (\eta'_1 - x) f(x) dx + \int_0^\infty (x - \eta'_1) f(x) dx - \int_0^{\eta'_1} (x - \eta'_1) f(x) dx \\
 &= 2 \int_0^{\eta'_1} (\eta'_1 - x) f(x) dx \\
 &= 2 \eta'_1 F(\eta'_1) - 2 \int_0^{\eta'_1} x f(x) dx \\
 &= 2 \eta'_1 F(\eta'_1) - 2 \eta'_1 + 2 T(\eta'_1),
 \end{aligned} \tag{8}$$

em que $T(q) = \int_q^\infty x f(x) dx$ e $f(x)$ representa a fdp de X .

O desvio médio em relação a mediana é similar a o desvio médio em relação a média, definido como

$$\begin{aligned}
 \delta_2(X) &= \int_0^\infty |x - M| f(x) dx \\
 &= 2T(M) - \eta'_1.
 \end{aligned} \tag{9}$$

Estudos recentes envolvendo desvios médios foram feitos por Cordeiro *et al.* (2013a), Cordeiro *et al.* (2013d), Cordeiro *et al.* (2012) e Paranaíba (2011) para as dis-

tribuições beta generalizada Rayleigh, Kumaraswamy Weibull modificada, Kumaraswamy generalizada semi-normal e beta Burr XII, respectivamente.

2.6.5 Curvas de Bonferroni e Lorenz

As curvas de Bonferroni e Lorenz são amplamente aplicadas não só em economia para estudos de renda e pobreza, mas também em outros campos como confiabilidade, demografia, seguros e medicina. A curva de Bonferroni pode ser expressa por

$$\begin{aligned} B(\rho) &= \frac{1}{\rho\eta'_1} \int_0^q x f(x) dx \\ &= \frac{1}{\rho} - \frac{T(\rho)}{\rho\eta'_1}, \end{aligned} \quad (10)$$

em que $\eta'_1 = E(X)$, q representa a inversa da função acumulada também conhecida como função quantílica expressa na equação (7), tal que $F^{-1}(\rho) = q$, e $T(q) = \int_q^\infty x f(x) dx$.

A curva de Lorenz é uma representação gráfica da função de distribuição cumulativa empírica da distribuição de probabilidade de riqueza. Em tal uso, muitos economistas consideram que é uma medida de desigualdade social para representar a desigualdade da distribuição da riqueza. As curvas de Lorenz são calculadas por

$$\begin{aligned} L(\rho) &= \frac{1}{\eta'_1} \int_0^q x f(x) dx \\ &= 1 - \frac{T(\rho)}{\eta'_1}. \end{aligned} \quad (11)$$

Ao criar uma distribuição nova, é de interesse verificar quais mudanças foram proporcionadas com a adição de novos parâmetros, estudo recém realizado por Cordeiro *et al.* (2013d), verifica os efeitos dos novos parâmetros a e b da distribuição Kumaraswamy Weibull modificada nas curvas de Bonferroni e Lorenz.

2.7 Estimadores não paramétricos

Em dados de análise de sobrevivência, assim como qualquer outro tipo de dados, é importante realizar uma análise descritiva do tempo de sobrevivência como: média, variância, desvio padrão e etc. Considerando a presença de censura, algumas estatísticas descritivas tornam-se um problema, pois os resultados podem estar viesados.

Para isso técnicas estatísticas são necessárias, na qual uma das mais utilizadas em análise de sobrevivência é o estimador não-paramétrico de Kaplan-Meier, utilizado para estimar a função de sobrevivência.

2.7.1 Risco empírico

Outra forma empírica utilizada é o "Total Time on Test"(Aarset, 1987), também conhecido como curva TTT, que é utilizada para estudar o comportamento da função de risco. Esse método gráfico auxilia na seleção do modelo particular para algum fenômeno em estudo.

A curva TTT representa o gráfico de

$$G\left(\frac{r}{n}\right) = \frac{\sum_{i=1}^r T_{i:n} + (n-r)T_{r:n}}{\sum_{i=1}^n T_{i:n}} \quad \text{versus} \quad A = \frac{r}{n},$$

em que $r = 1, \dots, n$ e $T_{i:n}, i = 1, \dots, n$, são estatísticas de ordem da amostra.

Figura 4- Formas da curva TTT plot.

A Figura 4 ilustra as várias formas que a curva TTT pode assumir. Se uma reta é observada (A), a função de risco é constante. Se possuir a forma de uma curva convexa (B) ou côncava (C), a função de risco é monotonamente decrescente ou crescente. No caso em que iniciar como uma curva convexa e depois tornar-se côncava (D), a função de risco assume forma de U, e no caso contrário (E) é unimodal.

2.8 Modelos de regressão paramétricos

É comum, na prática, covariáveis estarem associadas ao tempo até o evento de interesse. Por exemplo, no campo o tempo até a germinação de uma semente pode ser afetado pelo clima, umidade entre outros fatores. Tais covariáveis explicam parte da heterogeneidade do tempo até a ocorrência do evento.

Assim, considerando que $\mathbf{v} = (v_1, v_2, \dots, v_p)^t$ representa o vetor de covariáveis independentes do tempo, que possam ou não estar associadas a ocorrência do evento de interesse, a densidade do tempo até a ocorrência dado as covariáveis é expressos por $f(x|\mathbf{v})$.

O modelo paramétrico de regressão é uma maneira eficiente de estudar o efeito que as covariáveis causam no tempo de resposta. Na literatura esses modelos são conhecidos como modelos de tempo de vida acelerado ou também como locação e escala. Existe uma variedade de estudos sobre diversos tipos de modelos de regressão que pode ser encontrados em Cox (1972), Cox e Oakes (1984), Kalbfleisch e Prentice (1980), Lawless (2003), dentre outros.

2.8.1 Modelo de locação-escala

O modelo de regressão linear é o mais conhecido em análise estatística para verificar a relação entre covariável e variável resposta, no qual leva em conta um componente determinístico ϵ com efeito aditivo seguindo uma distribuição normal padrão.

Em análise de sobrevivência, esse modelo se torna inadequado pelo fato da distribuição se tornar assimétrica na direção dos maiores tempos de sobrevida, tornando viável a transformação na variável resposta para tentar retornar ao modelo linear normal.

A classe de modelos de locação e escala consiste em utilizar a transformação logarítmica nos tempos de falha, $Y = \log(X)$, de tal forma que para um vetor de covariáveis $\mathbf{v}$, o logaritmo do tempo de falha possui uma distribuição com um parâmetro de locação, normalmente denominado com μ , que depende das covariáveis ou variáveis explicativas, um parâmetro de escala usualmente denominado por σ e um componente estocástico Z .

Pode-se então escrever o modelo log-linear da seguinte forma:

$$Y = \mu + \sigma Z,$$

em que Y tenha uma distribuição de probabilidade que pertence à família de distribuições que se caracteriza pelo fato de possuir os dois parâmetro, locação e escala. Neste caso, diferente do modelo linear, as variáveis regressoras possuem um efeito multiplicativo sobre X , ou seja, $X = \exp\{\mu\} \exp\{\sigma Z\}$. O parâmetro de locação μ pode ser escrito em termos de variáveis regressoras da forma $\mu = \mathbf{v}^t \boldsymbol{\beta}$, em que $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ é o vetor de parâmetros desconhecidos. Desta forma, obtém-se

$$Y = \mathbf{v}^t \boldsymbol{\beta} + \sigma Z,$$

tornando-se um modelo log-linear para X com resíduo Z , sendo $\mathbf{V} = (v_1, \dots, v_n)^t$ a matriz do modelo e $\boldsymbol{\beta}$ o vetor de parâmetros associado as covariáveis.

2.9 Modelos paramétricos

Nesta seção são descritas as distribuições de probabilidade utilizadas no contexto do trabalho, são elas as distribuições semi-normal, semi-normal generalizada e beta semi-normal generalizada. A escolha do modelo probabilístico adequado para descrever o tempo de falha deve ser feito com cuidado, pois distribuições inadequadas podem causar conclusões erradas. Por isso é necessário conhecer algumas das características do modelo a ser utilizado, como por exemplo, a flexibilidade da função de risco, caudas flexíveis entre outras.

2.9.1 Distribuição semi-normal (SN)

Dentre as diversas causas de falhas de componentes mecânicos, a mais comum é devida à fadiga do material. Por meio das investigações experimentais, o aumento da ruptura por fadiga aparece como um processo de propriedades aleatórias que variam no tempo. Essas propriedades alteram-se durante o crescimento da ruptura, variando de material para material. Alguns modelos estatísticos permitem estudar a variação aleatória do tempo de falha associados aos materiais expostos à fadiga, como resultado de diferentes padrões cíclicos. Os modelos mais utilizados para descrever o processo do tempo de vida sobre a fadiga dos materiais são as distribuições semi-normal (SN) e Birnbaum-Saunders (BS) proposta por Birnbaum e Saunders (1969).

A distribuição SN, é oriunda de um método chamado semi-normal utilizado para lidar com relações entre erros de medição e magnitude. O método foi proposto por

Bland e Altman (1999) com base em uma sugestão de Altman (1993) para criar gráficos de percentil. O método consiste em que se $Y \sim N(0, \sigma^2)$ e se $X = |Y|$ então $X \sim \text{SN}(\sigma^2)$. A função densidade de probabilidade da distribuição de X é expressa por

$$g_\sigma(x) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad x > 0,$$

em que $\sigma > 0$ representa o parâmetro de escala. A fda pode ser escrita em termos da função erro $\text{erf}(x)$ como

$$G_\sigma(x) = \text{erf}\left(\frac{x}{\sigma\sqrt{2}}\right),$$

em que $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$. A função taxa de falha é expressa por

$$h_\sigma = \frac{\sqrt{2} \exp\left(-\frac{x^2}{2\sigma^2}\right)}{\sigma\sqrt{\pi} \left[1 - \text{erf}\left(\frac{x}{\sigma\sqrt{2}}\right)\right]},$$

a qual pode assumir apenas formas monótonas (crescente e decrescente).

2.9.2 Distribuição semi-normal generalizada (SNG)

Cooray e Ananda (2008) propuseram uma generalização da distribuição semi-normal chamada distribuição semi-normal generalizada (SNG), a qual acomoda funções de risco não-monótonas como, por exemplo, a forma de banheira ou “U”. As possíveis formas da função de risco são apresentadas na Figura 5.

Outra distribuição que possui o comportamento de taxa de falha na forma de banheira é a distribuição gamma generalizada, proposta por Stacy (1962). A desvantagem do uso da distribuição gamma generalizada comparada com a distribuição SNG, se da pelo fato de que as estimativas dos parâmetros são muito mais complexas, Hager e Bain (1970) mostram que mesmo utilizando uma amostra de tamanho 300 não foi possível obter os valores das estimativas dos parâmetros utilizando o método de Newton-Raphson.

A função densidade de probabilidade da distribuição SNG é escrita como

$$g_{\alpha,\theta}(x) = \sqrt{\frac{2}{\pi}} \left(\frac{\alpha}{x}\right) \left(\frac{x}{\theta}\right)^{\alpha} e^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}}, \quad \alpha, \theta > 0, \quad x \geq 0,$$

em que α e θ são os parâmetros de forma e escala respectivamente. A fda pode ser expressa na forma

$$G_{\alpha,\theta}(x) = 2\Phi \left[\left(\frac{x}{\theta}\right)^{\alpha} \right] - 1 = \operatorname{erf} \left[\frac{\left(\frac{x}{\theta}\right)^{\alpha}}{\sqrt{2}} \right],$$

em que $\Phi(\cdot)$ representa a fda da normal padrão. A função taxa de falha pode ser expressa das seguintes maneiras

$$\begin{aligned} h_{\alpha,\theta}(x) &= \frac{\sqrt{\frac{2}{\pi}} \left(\frac{\alpha}{x}\right) \left(\frac{x}{\theta}\right)^{\alpha} e^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}}}{1 - \{2\Phi \left[\left(\frac{x}{\theta}\right)^{\alpha} \right] - 1\}} = \frac{\sqrt{\frac{2}{\pi}} \left(\frac{\alpha}{x}\right) \left(\frac{x}{\theta}\right)^{\alpha} e^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}}}{2 \{1 - \Phi \left[\left(\frac{x}{\theta}\right)^{\alpha} \right]\}} \\ &= \frac{\frac{1}{\sqrt{2\pi}} \left(\frac{\alpha}{x}\right) \left(\frac{x}{\theta}\right)^{\alpha} e^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}}}{1 - \Phi \left[\left(\frac{x}{\theta}\right)^{\alpha} \right]}. \end{aligned} \quad (12)$$

De uma forma mais elegante, a distribuição SNG apresenta uma maior flexibilidade nas caudas e função de risco do que as distribuições usuais na literatura como Weibull, gamma, lognormal, etc. Algumas formas para a função de risco são apresentadas na Figura 5.

Figura 5- Formas da função de risco da distribuição SNG para alguns valores dos parâmetros.

2.9.3 Distribuição beta semi-normal generalizada (BSNG)

A classe de distribuições beta generalizadas vem sendo muito utilizada para análise de dados assimétricos. Proposta por Eugene *et al.* (2002), a função densidade acumulada da nova classe de distribuições é expressa por

$$F(x) = \frac{1}{B(a, b)} \int_0^{G(x)} w^{a-1} (1-w)^{b-1} dw, \quad (13)$$

em que $a > 0$ e $b > 0$ são dois parâmetros de forma, $B(a, b)$ representa a função beta e $G(x)$ a função acumulada de uma variável aleatória contínua.

Desde então diversas distribuições pertencentes a classe beta vem surgindo na literatura como beta normal, beta Gumbel, beta Fréchet e beta exponencial, propostas por Eugene *et al.* (2002), Nadarajah e Kotz (2004), Nadarajah e Gupta (2004) e Nadarajah e Kotz (2005), respectivamente. Recentemente, Carrasco *et al.* (2008) estudaram uma distribuição com quatro parâmetros chamada Weibull modificada generalizada, Barreto-Souza *et al.* (2011) propuseram a distribuição beta exponencial generalizada e Cordeiro *et al.* (2013a) utilizaram a distribuição beta Rayleigh generalizada em aplicações a dados de sobrevivência.

Pescim *et al.* (2010) propuseram a distribuição BSNG, na qual é uma composição baseada na distribuição beta e na distribuição generalizada semi-normal. A utilidade dessa distribuição é que possui uma maior flexibilidade na análise de dados de tempo de vida em relação a outras distribuições existentes na literatura. A função densidade de probabilidade é dada por:

$$\begin{aligned} g_{\alpha, \theta, a, b}(x) &= \frac{\sqrt{\frac{2}{\pi}} \left(\frac{\alpha}{x}\right) \left(\frac{x}{\theta}\right)^{\alpha} e^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}}}{B(a, b)} 2^{b-1} \left\{ 2\Phi \left[\left(\frac{x}{\theta}\right)^{\alpha} \right] - 1 \right\}^{a-1} \\ &\times \left\{ 1 - \Phi \left[\left(\frac{x}{\theta}\right)^{\alpha} \right] \right\}^{b-1} \quad \alpha, \theta, a, b > 0 \quad e \quad x \geq 0. \end{aligned} \quad (14)$$

A distribuição acumulada é descrita pela classe de distribuições beta generalizadas introduzidas por Eugene *et al.* (2002), e pode ser expressa das seguintes formas:

$$\begin{aligned}
G_{\alpha,\theta,a,b}(x) &= \frac{1}{B(a,b)} \int_0^{G_{\alpha,\theta}(x)} w^{a-1}(1-w)^{b-1} dw = I_{G_{\alpha,\theta}(x)}(a,b) \\
&= \frac{1}{B(a,b)} \int_0^{2\Phi[(\frac{x}{\theta})^\alpha]-1} w^{a-1}(1-w)^{b-1} dw = I_{2\Phi[(\frac{x}{\theta})^\alpha]-1}(a,b), \quad (15)
\end{aligned}$$

em que $G_{\alpha,\theta}(x)$ representa a acumulada da distribuição SNG e $I(\cdot)$ a definição da acumulada da distribuição beta.

A função de risco ou de taxa de falha, é expressa por:

$$h_{\alpha,\theta,a,b}(x) = \frac{\sqrt{\frac{2}{\pi}} \left(\frac{\alpha}{x}\right) \left(\frac{x}{\theta}\right)^\alpha e^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}} \{2\Phi[(\frac{x}{\theta})^\alpha] - 1\}^{a-1} \{1 - \Phi[(\frac{x}{\theta})^\alpha]\}^{b-1} 2^{b-1}}{B(a,b) \left\{1 - I_{2\Phi[(\frac{x}{\theta})^\alpha]-1}(a,b)\right\}}.$$

Uma forma alternativa para apresentar a função densidade de probabilidade é apresentada em Pescim *et al.* (2010), e dada por

$$g_{\alpha,\theta,a,b}(x) = \sqrt{\frac{2}{\pi}} \left(\frac{\alpha}{x}\right) \left(\frac{x}{\theta}\right)^\alpha e^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}} \sum_{h,l=0}^{\infty} \sum_{i=0}^l \sum_{j=0}^i u_{h,l,i,j}(a,b) \left\{ \left[\frac{(\frac{x}{\theta})^\alpha}{\sqrt{2}} \right] \right\}^j,$$

em que $u_{h,l,i,j}(a,b)$ é expresso por

$$u_{h,l,i,j}(a,b) = \frac{w_{h,l,i}(a,b)(-1)^j}{B(a,b)2^i} \binom{i}{j} \quad (16)$$

e $w_{h,l,i}(a,b)$ por

$$w_{h,l,i}(a,b) = \frac{2^l(-1)^{h+l+i}\Gamma(b)\Gamma(a+h)}{\Gamma(b-h)\Gamma(a+h-l)h!l!} \binom{l}{i} \quad (17)$$

2.9.3.1 Momentos

O s -ésimo momento da distribuição BSNG é escrito na forma de uma combinação linear com o s -ésimo momento da distribuição SNG, Pescim *et al.* (2010).

$$\eta'_s = \theta^s \sqrt{\frac{2}{\pi}} \sum_{h,l=0}^{\infty} \sum_{i=0}^l \sum_{j=0}^i u_{h,l,i,j}(a,b) I\left(\frac{s}{\alpha}, j\right), \quad (18)$$

em que $u_{h,l,i,j}(a,b)$ é apresentado em (16) e $I\left(\frac{s}{\alpha}, j\right)$ é expresso por

$$I\left(\frac{s}{\alpha}, j\right) = \pi^{-\frac{j}{2}} 2^{j+\frac{s}{2\alpha}-\frac{1}{2}} \sum_{m_1=0}^{\infty} \cdots \sum_{m_j=0}^{\infty} \frac{(-1)^{m_1+\cdots+m_j}}{(m_1+\frac{1}{2}) \cdots (m_j+\frac{1}{2}) m_1! \cdots m_j!} \\ \times \Gamma\left(m_1+\cdots+m_j+\frac{j+\frac{s}{\alpha}+1}{2}\right). \quad (19)$$

Observa-se que quando $j+\frac{s}{\alpha}$ for um número par, os momentos η'_s da distribuição BSNG pode ser escrito em termos da função de Lauricella do tipo A definida por

$$F_A^{(n)}(a; b_1, \dots, b_n; c_1, \dots, c_n; x_1, \dots, x_n) = \sum_{m_1=0}^{\infty} \cdots \sum_{m_n=0}^{\infty} \frac{(a)_{m_1+\cdots+m_n} (b_1)_{m_1} \cdots (b_n)_{m_n}}{(c_1)_{m_1} \cdots (c_n)_{m_n}} \\ \times \frac{x_1^{m_1} \cdots x_n^{m_n}}{m_1! \cdots m_n!}, \quad (20)$$

em que $(a)_k = a(a-1)\cdots(a-k+1)$ representa um fatorial ascendente (o convencional é utilizar $(a)_0 = 1$). Rotinas numéricas para o cálculo da função Lauricella do tipo A são encontradas em Exton (1978) e Mathematica (Trott, 2006). Logo o s -ésimo momento para a distribuição BSNG é dado por

$$\eta'_s = \theta^s \sqrt{\frac{2}{\pi}} \sum_{h,l=0}^{\infty} \sum_{i=0}^l \sum_{j=0}^i u_{h,l,i,j}(a,b) \pi^{-\frac{j}{2}} 2^{j+\frac{s}{2\alpha}-\frac{1}{2}} \Gamma\left(\frac{j+\frac{s}{\alpha}+1}{2}\right) \\ \times F_A^{(j)}\left(\frac{j+\frac{s}{\alpha}+1}{2}; \frac{1}{2}, \dots, \frac{1}{2}; \frac{3}{2}, \dots, \frac{3}{2}; -1, \dots, -1\right).$$

2.9.3.2 Assimetria e curtose

As representações gráficas das medidas de assimetria e curtose quando $\theta = 55$ e $\alpha = 1$, com função do parâmetro a para alguns valores do parâmetro b e como função do parâmetro b para alguns valores escolhidos do parâmetro a , são apresentadas respectivamente pelas Figuras 6 e 7, Pescim *et al.* (2010).

Figura 6 - Gráficos dos valores dos coeficientes de assimetria da distribuição BSNG como função de a para alguns valores de b e como função de b para alguns valores de a

Figura 7 - Gráficos dos valores dos coeficientes de curtose da distribuição BSNG como função de a para alguns valores de b e como função de b para alguns valores de a

As Figuras 6 e 7 revelam imediatamente que a assimetria e a curtose aumentam à medida que b aumenta para o parâmetro a fixo e diminuem à medida que a

aumenta para o parâmetro b fixo.

2.9.3.3 Função geradora de momentos

Utilizando-se a função densidade expressa por (14) e aplicando algumas expansões, Pescim *et al.* (2010), a função geradora de momentos da distribuição BSNG assume a forma

$$M_X(t) = \sqrt{\frac{2}{\pi}} \sum_{h,l=0}^{\infty} \sum_{i=0}^l \sum_{j=0}^i u_{h,l,i,j}(a,b) \pi^{-\frac{j}{2}} \sum_{u=0}^{\infty} 2^{j+\frac{u}{2\alpha}-\frac{1}{2}} \frac{\theta^u t^u}{u!} I\left(\frac{u}{\alpha}, j\right) \quad (21)$$

em que $u_{h,l,i,j}(a,b)$ é definido pela equação (16) e $I\left(\frac{u}{\alpha}, j\right)$ é definido na equação (19).

2.9.3.4 Desvios médios

Os desvios médios em relação a média e mediana são calculados respectivamente pelas equações (8) e (9). Para distribuição BSNG, os desvios médios definidos por Cordeiro *et al.* (2013c), são expressos como:

$$\delta_1(x) = 2\eta'_1 G_{\alpha,\theta,a,b}(\eta'_1) - 2\eta'_1 + 2T_1(\eta'_1),$$

em que $\eta'_1 = E(X)$, $G_{\alpha,\theta,a,b}(\eta'_1)$ pode ser calculado facilmente por (15) e $T_1(q) = \int_q^{\infty} x g_{\alpha,\theta,a,b}(x) dx$. Por meio de expansões, $T_1(q)$ para a distribuição BSNG pode ser expresso por

$$\begin{aligned} T_1(q) &= \theta \sqrt{\frac{2}{\pi}} \sum_{h,l=0}^{\infty} \sum_{i=0}^l \sum_{j=0}^i u_{h,l,i,j}(a,b) \left(\frac{2}{\sqrt{\pi}}\right)^j \\ &\times \sum_{m_1=0}^{\infty} \cdots \sum_{m_j=0}^{\infty} \frac{(-1)^{m_1+\cdots+m_j}}{(2m_1+1) \cdots (2m_j+1) m_1! \cdots m_j!} \\ &\times \Gamma\left[\left(m_1 + \cdots + m_j + \frac{j}{2} + \frac{1}{2\alpha} + \frac{1}{2}\right), \frac{1}{2} \left(\frac{q}{\theta}\right)^{\alpha}\right], \end{aligned} \quad (22)$$

em que $\Gamma(r, x) = \int_x^{\infty} u^{r-1} \exp(-u) du$.

O desvio médio em relação a mediana é definido similarmente como os desvios médios em relação a média definidos como

$$\begin{aligned}\delta_2 &= \int_0^\infty |x - M|f(x)dx \\ &= 2T_1(M) - \eta'_1.\end{aligned}$$

em que a função $T_1(\cdot)$ é definida na equação (22).

2.9.3.5 Função quantílica

A função quantílica da distribuição BSNG, proposta por Cordeiro *et al.* (2013c), é definida como

$$x = \theta \left[Q_N \left(\frac{(1+z)}{2} \right) \right]^{1/\alpha}, \quad (23)$$

em que Q_N representa o quantil da distribuição normal padrão, Steinbrecher (2002), dado por

$$x = Q_N(u) = \Phi^{-1}(u) = \sum_{k=0}^{\infty} c_k (\sqrt{2\pi}(u - 1/2))^{2k+1},$$

sendo que os termos c'_k s são calculados a partir da equação

$$c_{k+1} = \frac{1}{2(2k+3)} \sum_{r=0}^k \frac{(2r+1)(2k-2r+1)c_r c_{k-r}}{(r+1)(2r+1)},$$

assim, $c_0 = 1$, $c_1 = 1/6$, $c_2 = 7/120$, $c_3 = 127/7560, \dots$ e z representa o quantil da distribuição beta, que pode ser encontrado em Wolfram website¹

$$z = Q_B(u) = I_u^{-1}(a, b) = d_1 v + d_2 v^2 + d_3 v^3 + d_4 v^4 + O(v^5),$$

em que $v = [a B(a, b)u]^{1/a}$ para $a > 0$ e $d_1 = 1$, $d_2 = (b-1)/(a+1)$,

$$d_3 = \frac{(b-1)(a^2 + 3ab - a + 5b - 4)}{2(a+1)^2(a+2)},$$

¹<http://functions.wolfram.com/06.23.06.0004.01>

$$d_4 = (b-1)[a^4 + (6b-1)a^3 + (b+2)(8b-5)a^2 + (33b^2 - 30b + 4)a + b(31b-47) + 18]/[3(a+1)^3(a+2)(a+3)], \dots$$

Os coeficientes d'_i s para $i \geq 2$ são derivados da forma

$$d_i = \frac{1}{[i^2 + (\alpha-2)i + (1-\alpha)]} \left\{ (1 - \delta_{i,2}) \sum_{r=2}^{i-1} d_r d_{i+1-r} [r(1-a)(i-r) - r(r-1)] + \sum_{r=1}^{i-1} \sum_{s=1}^{i-r} d_r d_s d_{i+1-r-s} [r(r-a) + s(a+b-2) \times (i+1-r-s)] \right\},$$

em que $\delta_{i,2} = 1$ se $i = 2$ e $\delta_{i,2} = 0$ se $i \neq 2$. Na última equação é possível notar que o termo quadrático só contribuiu quando $i \geq 3$.

2.9.3.6 Curvas de Bonferroni e Lorenz

As curvas de Bonferroni e Lorenz, calculadas a partir das equações (10) e (11), propostas por Cordeiro *et al.* (2013d) para a distribuição BSNG, são representadas respectivamente por

$$B(\rho) = \frac{1}{\rho} - \frac{T_1(\rho)}{\rho \eta'_1}$$

e

$$L(\rho) = 1 - \frac{T_1(\rho)}{\eta'_1},$$

em que $T_1(\rho)$ é calculado a partir da equação definida em (22).

2.9.4 Distribuição log-beta semi-normal generalizada (LBSNG)

Em muitos estudos em medicina, o tempo de vida é afetado por covariáveis como colesterol, pressão sanguínea, contagem de glóbulos brancos entre outras. Modelos paramétricos para estimar funções de sobrevivência são amplamente utilizados e, com isso, diferentes formas para modelos de regressão vem surgindo na literatura. Recentemente, Carrasco *et al.* (2008b) introduziram o modelo de regressão considerando a distribuição Weibull modificada, Ortega *et al.* (2009) propuseram o modelo de regressão log-gama generalizada modificada com fração de cura e Hashimoto *et al.* (2010) introduziram o modelo de regressão log-Weibull exponenciada para dados com censura intervalar.

Pescim *et al.* (2013) introduziram o modelo de regressão log-beta semi-normal generalizada (LBSNG), considerando $Y = \log(X)$, em que $X \sim BSNG(\alpha, \theta, a, b)$, e as reparametrizações $\mu = \log(\theta)$ e $\sigma = \frac{\sqrt{2}}{2\alpha}$, ou seja, $Y \sim \text{LBSNG}(\mu, \sigma, a, b)$. A função densidade de Y , encontrada via método do jacobiano, é expressa por

$$g_{\mu, \sigma, a, b}(y) = \frac{\exp \left\{ \left(\frac{y - \mu}{\sigma} \right) \frac{\sqrt{2}}{2} - \frac{1}{2} \exp \left\{ \left(\frac{y - \mu}{\sigma} \right) \sqrt{2} \right\} \right\} 2^{b-1}}{\sigma \sqrt{\pi} B(a, b)} \left\{ 2\Phi \left[\left(\frac{y - \mu}{\sigma} \right) \frac{\sqrt{2}}{2} \right] - 1 \right\}^{a-1} \\ \times \left\{ 1 - \Phi \left[\left(\frac{y - \mu}{\sigma} \right) \frac{\sqrt{2}}{2} \right] \right\}^{b-1}, \quad -\infty < y < \infty, \quad (24)$$

em que $-\infty < \mu < \infty$, $\sigma > 0$, $a > 0$ e $b > 0$.

Utilizando a expansão

$$(1 - z)^{b-1} = \sum_{j=0}^{\infty} \frac{(-1)^j \Gamma(b)}{\Gamma(b-j) j!} z^j, \quad |z| < 1,$$

na equação (15) e considerando ainda as reparametrizações $\mu = \log(\theta)$ e $\sigma = \frac{\sqrt{2}}{2\alpha}$, obtém-se facilmente a distribuição acumulada da distribuição LBSNG,

$$G_{\mu, \sigma, a, b}(y) = \frac{1}{B(a, b)} \sum_{h=0}^{\infty} \frac{(-1)^h \Gamma(b)}{\Gamma(b-h) h! (a+h)} \left\{ 2\Phi \left[\exp \left\{ \left(\frac{y - \mu}{\sigma} \right) \frac{\sqrt{2}}{2} \right\} \right] - 1 \right\}^{a+h}.$$

2.9.4.1 Momentos

O s -ésimo momento da distribuição LBSNG (Pescim *et al.* (2013)) é dado pela expressão $\eta'_s = \int_{-\infty}^{\infty} y^s g_{\mu,\sigma,a,b}(y) dy$, em que $g_{\mu,\sigma,a,b}(y)$ é dada pela equação (24), e por meio de expansões pode ser escrito como:

$$\begin{aligned} \eta'_s &= \frac{1}{\sigma\sqrt{\pi}} \sum_{h,l=0}^{\infty} \sum_{i=0}^l \sum_{\ell=0}^s e_{h,l,i}(a,b) \mu^{s-o} \left(\sqrt{2}\sigma\right)^{s+1} \binom{s}{\ell} \\ &\times \sum_{m_1=0}^{\infty} \cdots \sum_{m_i=0}^{\infty} \frac{(-1)^{m_1+\dots+m_i}}{(2m_1+1)\dots(2m_i+1)m_1!\dots m_i!} \\ &\times \sum_{q=0}^s \binom{s}{q} [\log(2)]^{s-q} 2^{-1/2-s} \Gamma^q \left(m_1 + \dots + m_i + \frac{i+1}{2}\right), \end{aligned} \quad (25)$$

em que μ representa o parâmetro de locação,

$$e_{h,l,i}(a,b) = \frac{(-1)^{h+l+i} \Gamma(a+b) \Gamma(a+h)}{h! l! \Gamma(a) \Gamma(b-h) \Gamma(a+h-l)} \binom{l}{i} \quad (26)$$

e

$$\Gamma^q \left(m_1 + \dots + m_i + \frac{i+1}{2}\right) = \frac{\partial^q \Gamma(m_1 + \dots + m_i + \frac{i+1}{2})}{\partial(m_1 + \dots + m_i + \frac{i+1}{2})^q}. \quad (27)$$

2.9.4.2 Função geradora de momentos

A função geradora de momentos da distribuição LBSNG (Pescim *et al.* (2013)), $M_Y(t) = \int_{-\infty}^{\infty} e^{yt} g_{\mu,\sigma,a,b}(y) dy$, é expressa na forma

$$\begin{aligned} M_Y(t) &= \frac{\sqrt{2}}{\pi} \sum_{h,l=0}^{\infty} \sum_{i=0}^l e_{h,l,i}(a,b) e^{\mu t} 2^{i+\frac{t\sigma}{\sqrt{2}}-2} \\ &\times \sum_{m_1=0}^{\infty} \cdots \sum_{m_i=0}^{\infty} \frac{(-1)^{m_1+\dots+m_i}}{(m_1+\frac{1}{2})\dots(m_i+\frac{1}{2})m_1!\dots m_i!} \\ &\times \Gamma \left(m_1 + \dots + m_i + \frac{i+\frac{2\sigma t}{\sqrt{2}}+1}{2}\right). \end{aligned} \quad (28)$$

2.9.5 Método de máxima verossimilhança

O estimador de máxima verossimilhança $\hat{\boldsymbol{\lambda}}$ para o vetor de parâmetros $\boldsymbol{\lambda}$ é calculado maximizando a função de verossimilhança ou, equivalentemente, maximizando o logaritmo da função de verossimilhança, $l(\boldsymbol{\lambda})$. Portanto, esses estimadores são obtidos da solução do sistema de equações chamado de vetor escore:

$$U(\boldsymbol{\lambda}) = \frac{\partial l(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}.$$

Como, geralmente, esse sistema é não linear, torna-se necessário usar um algoritmo de otimização. Por exemplo, um algoritmo Newton-Raphson ou quase-Newton (NOCEDAL e WRIGHT, 2006; PRESS *et al.*, 1992).

No caso em que o tamanho da amostra é grande e sob certas condições de regularidade para a função de verossimilhança, intervalos de confiança e testes de hipóteses podem ser obtidos usando o fato de que a distribuição assintótica para os estimadores de máxima verossimilhança é a distribuição normal com média $\boldsymbol{\lambda}$ e matriz de covariância dada pela inversa da matriz de informação de Fisher (SEN & SINGER, 1993). Portanto a matriz de covariância assintótica é dada por $\mathbf{I}^{-1}(\boldsymbol{\lambda})$ com $\mathbf{I}(\boldsymbol{\lambda}) = E[\ddot{\mathbf{L}}(\boldsymbol{\lambda})]$, em que $\ddot{\mathbf{L}}(\boldsymbol{\lambda}) = - \left\{ \frac{\partial^2 l(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}} \right\}$.

Como é complicado calcular a matriz de informação de Fisher $\mathbf{I}(\boldsymbol{\lambda})$ devido às observações censuradas, pode-se usar a matriz Hessiana, $\ddot{\mathbf{L}}(\boldsymbol{\lambda})^{-1}$, avaliada em $\boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}}$, que é um estimador consistente para matriz de covariância assintótica (MUDHOLKAR; SRIVASTAVA; KOLLIA, 1996). Assim, a distribuição assintótica para $\hat{\boldsymbol{\lambda}}$ é dada por $\hat{\boldsymbol{\lambda}} \sim N_{(q)} \left\{ \boldsymbol{\lambda}; \ddot{\mathbf{L}}(\boldsymbol{\lambda})^{-1} \right\}$ em que $\ddot{\mathbf{L}}(\boldsymbol{\lambda})$ é a matriz de informação observada e q é o número de parâmetros do modelo.

2.9.6 Critérios de seleção

Entre os critérios para seleção de modelos, são eles os mais utilizados

- **AIC:** Critério de informação de Akaike (Akaike, 1974), dado por $-2l(\boldsymbol{\lambda}) + 2q$.
- **BIC:** Critério de informação bayesiano (Schwarz, 1978), dado por $-2l(\boldsymbol{\lambda}) + q \log(n)$.
- **CAIC:** Critério de informação de Akaike corrigido (Bozdogan, 1987), dado por $AIC + \frac{2q(q+1)}{n-q-1}$.

em que q representa o número de parâmetros do modelo e n o tamanho da amostra. Os critérios citados levam em conta o valor da verossimilhança e penaliza utilizando o tamanho da amostra e quantidade de parâmetros, e assim como o teste da razão de verossimilhança, são de usos restritos para modelos encaixados. Uma alternativa é a utilizar testes não paramétricos para comparar modelos não encaixados ou utilizar o teste da razão de verossimilhança generalizado proposto por Vuong (1989).

3 DISTRIBUIÇÃO BETA SEMI-NORMAL GENERALIZADA GEOMÉTRICA

Em estudo de confiabilidade, uma ferramenta útil para facilitar a escolha de modelos probabilísticos é analisar o comportamento da função de risco. Ao detectar taxas de risco monótonas (crescentes e decrescentes), a distribuição SN pode ser, em primeiro momento, uma boa escolha por causa das formas de suas funções de risco que são positivamente e negativamente inclinadas. No entanto, essa distribuição não fornece um ajuste razoável para a modelagem de fenômenos com os índices de falha não-monótonas (formas de banheira e unimodal) que são muito comuns nos estudos de análise de sobrevivência. Nesse sentido, Cooray e Ananda (2008) introduziram a distribuição SNG, a qual acomoda funções de risco não monótonas. Com esse intuito, proposta por Pescim *et al.* (2010), a distribuição BSNG se mostrou mais flexível que a distribuição SNG, também assumindo funções crescentes, decrescentes, forma de banheira e unimodal em sua função taxa de falha.

Neste trabalho, foi utilizada a transformação geométrica na distribuição BSNG, com o intuito de obter maior diversidade entre as funções de probabilidade e taxa de falha. Sabe-se que a transformação geométrica é caracterizada por adicionar mais um parâmetro de forma, e com isso, possibilitando maior flexibilidade nas caudas da nova distribuição.

3.1 Classe de distribuições geométricas

Seja $Y_1, Y_2, \dots, Y_z$ uma amostra aleatória de uma distribuição contínua qualquer com função densidade de probabilidade e função acumulada dadas respectivamente por $g(y)$ e $G(y)$ e Z outra variável aleatória com função de probabilidade dada por

$$p(Z = z) = (1 - p)p^{z-1}, \quad (29)$$

em que $Z \in \mathbb{N}$ e $p \in [0, 1)$, assim dizemos que Z possui uma distribuição geométrica.

Define-se agora uma variável aleatória $X = \min\{Y_i\}_{i=1}^z$. Assumindo independência entre Y e Z , a função distribuição acumulada de X é obtida utilizando a distribuição acumulada condicional de X dado Z pelo método de distribuição do mínimo para variáveis aleatórias contínuas independentes.

$$\begin{aligned}
F(x|z) &= P(X \leq x|z) \\
&= P(\min(Y_1, Y_2, \dots, Y_z) \leq x|z) \\
&= 1 - P(Y_1 > x, Y_2 > x, \dots, Y_z > x) \\
&= 1 - \prod_{i=1}^z P(Y_i > x) \\
&= 1 - [P(Y > x)]^z \\
&= 1 - [1 - P(Y \leq x)]^z \\
&= 1 - [1 - G(x)]^z,
\end{aligned} \tag{30}$$

em que $G(\cdot)$ representa a função acumulada a variável aleatória Y .

Diferenciando a equação (30) em relação a X , obtém-se a função densidade de probabilidade de X condicionada a Z

$$f(x|z) = z [1 - G(x)]^{z-1} g(x). \tag{31}$$

Utilizando a propriedade de independência e as expressões (29) e (31), obtém-se a função densidade de probabilidade conjunta de (X, Z)

$$f(x, z) = f(x|z)P(Z = z) = z g(x) [1 - G(x)]^{z-1} (1 - p)p^{z-1}. \tag{32}$$

Com o intuito de encontrar a função densidade marginal de X , aplica-se o somatório na expressão (32) para Z variando em todo o seu intervalo

$$\begin{aligned}
f(x) &= \sum_{z=1}^{\infty} f(x, z) \\
&= (1 - p)g(x) \sum_{z=1}^{\infty} z \{p[1 - G(x)]\}^{z-1}.
\end{aligned}$$

Considerando a expressão $\sum_{z=1}^{\infty} z \{p[1 - G(x)]\}^{z-1}$ como a derivada de uma série geométrica, a função densidade marginal de X é expressa como

$$f(x) = (1 - p)g(x)\{1 - p[1 - G(x)]\}^{-2}, \quad (33)$$

A expressão apresentada pela equação (33), representa a função distribuição de probabilidade geral para as distribuições da família geométrica, no qual a expressão geral da distribuição acumulada para a família geométrica é dada por

$$F(x) = \int_{-\infty}^x f(x) = G(x)\{1 - p[1 - G(x)]\}^{-1}. \quad (34)$$

em que $g(x)$ e $G(x)$ representam qualquer fdp e função acumulada (fda) contínua, respectivamente.

3.2 Distribuição beta semi-normal generalizada geométrica

Recentemente, diversos modelos pertencendo a família de distribuições G-geométricas vêm surgindo na literatura, Adamidis e Loukas (1998) propuseram a distribuição exponencial geométrica (EG). Após a distribuição EG, Barreto-Souza *et al.* (2008) desenvolveram a distribuição Weibull geométrica (WG), que generaliza a distribuição exponencial geométrica. Recentemente, foi proposto por Cordeiro *et al.*, (2011), a distribuição beta-Weibull geométrica (BWG) e por Mendoza *et al.* (2012) a distribuição log-logística geométrica.

Utilizando a transformação para encontrar as distribuições pertencentes a família de distribuições G-geométricas, apresentada na equação (33), podemos encontrar novas distribuições substituindo em $g(x)$ e $G(x)$, respectivamente, pelas fdp e fda de qualquer distribuição contínua.

Considere agora que a fda da distribuição beta semi-normal generalizada geométrica seja dada por

$$F(x) = \frac{G_{\alpha,\theta,a,b}(x)}{1 - p(1 - G_{\alpha,\theta,a,b}(x))} \quad 0 \leq p < 1, \quad (35)$$

em que $G_{\alpha,\theta,a,b}(x)$ representa a distribuição acumulada da distribuição BSNG. Facilmente encontramos a função densidade de probabilidade, que é definida como a primeira derivada da função densidade de qualquer distribuição contínua. De forma geral, a transformação

geométrica que representa a função densidade de probabilidade é descrita por

$$f(x) = \frac{g_{\alpha, \theta, a, b}(x)(1-p)}{\{1-p[1-G_{\alpha, \theta, a, b}(x)]\}^2}, \quad (36)$$

em que $g_{\alpha, \theta, a, b}(x)$ representa a fdp da distribuição BSNG.

A nova distribuição denominada por beta semi-normal generalizada geométrica (BSNGG), composta por cinco parâmetros, sendo eles α , θ , a , b e p , tem a função densidade de probabilidade obtida pela transformação (36) dada por

$$\begin{aligned} f(x; \alpha, \theta, a, b, p) &= (1-p) \frac{2^{b-\frac{1}{2}} \left(\frac{\alpha}{x}\right) \left(\frac{x}{\theta}\right)^\alpha \exp\left[-\frac{1}{2} \left(\frac{x}{\theta}\right)^{2\alpha}\right]}{\sqrt{\pi} B(a, b)} \left\{2\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right] - 1\right\}^{a-1} \\ &\times \left\{1 - \Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right]\right\}^{b-1} \left\{1 - p \left[1 - I_{2\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right]-1}(a, b)\right]\right\}^{-2}, \quad x > 0, \end{aligned} \quad (37)$$

em que θ , a , $b > 0$ e $0 \leq p < 1$ são parâmetros de forma e $\alpha > 0$ parâmetro de escala.

A função densidade acumulada $F(x, \alpha, \theta, a, b, p)$ obtida pela transformação (35) é expressa por

$$F(x; \alpha, \theta, a, b, p) = \frac{I_{2\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right]-1}(a, b)}{1 - p \left\{1 - I_{2\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right]-1}(a, b)\right\}}, \quad x > 0. \quad (38)$$

A função de sobrevivência definida na literatura como $S(x) = 1 - F(x)$ e a função taxa de falha obtida por $h(x) = \frac{f(x)}{S(x)}$ são expressas, respectivamente, por

$$S(x; \alpha, \theta, a, b, p) = 1 - \frac{I_{2\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right]-1}(a, b)}{1 - p \left\{1 - I_{2\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right]-1}(a, b)\right\}}, \quad x > 0, \quad (39)$$

e

$$\begin{aligned} h(x; \alpha, \theta, a, b, p) &= (1-p) \frac{2^{b-\frac{1}{2}} \left(\frac{\alpha}{x}\right) \left(\frac{x}{\theta}\right)^\alpha \exp\left[-\frac{1}{2} \left(\frac{x}{\theta}\right)^{2\alpha}\right]}{\sqrt{\pi} B(a, b)} \left\{2\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right] - 1\right\}^{a-1} \\ &\times \left\{1 - \Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right]\right\}^{b-1} \left\{1 - p \left[1 - I_{2\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right]-1}(a, b)\right]\right\}^{-2} \\ &\times \left\{1 - \frac{I_{2\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right]-1}(a, b)}{1 - p \left\{1 - I_{2\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right]-1}(a, b)\right\}}\right\}^{-1}, \quad x > 0. \end{aligned} \quad (40)$$

Alguns casos particulares da distribuição são obtidos quando os parâmetros

assumem valores específicos, esses casos são apresentados na Tabela 1.

Tabela 1 - Casos Particulares da distribuição BSNGG.

Distribuição	α	θ	a	b	p
Semi-normal generalizada (SNG)	α	θ	1	1	0
Semi-normal (SN)	1	θ	1	1	0
Beta semi-normal (BSN)	1	θ	a	b	0
Semi-normal geométrica (SNG _{eo})	1	θ	1	1	p
Semi-normal generalizada geométrica (SNGG)	α	θ	1	1	p
Beta semi-normal generalizada (BSNG)	α	θ	a	b	0
Semi-normal generalizada exponenciada (SNGE)	α	θ	a	1	0
Semi-normal generalizada exponenciada geométrica (SNGEG)	α	θ	a	1	p
Beta semi-normal geométrica (BSNG _{eo})	1	θ	a	b	p

Com a adição de mais um parâmetro de forma na distribuição BSNG, a distribuição BSNGG se torna mais flexível. Alguns dos diferentes comportamentos da funções de densidade e função taxa de falha, são apresentados nas Figuras 8 e 9, o qual é possível notar que a função taxa de falha pode assumir as formas crescente, decrescente, unimodal e forma “U”.

Figura 8 - Formas para a densidade da distribuição BSNGG variando os valores dos parâmetros: (a) Para diferentes valores de α , a , p com $\theta = 1.5$ e $b = 3$. (b) Para diferentes valores de θ , b , p com $\alpha = 2.5$ e $a = 5$. (c) Para diferentes valores de θ , a , p com $\alpha = b = 1.5$.

Figura 9 - Formas para função taxa de falha da distribuição BSNGG para alguns valores dos parâmetros: (a) Função taxa de falha considerando $\alpha = 1,5$, $\theta = 55$ e $b = 2,0$ fixos, variando a e p . (b) Função taxa de falha considerando $\alpha = 0,5$, $\theta = 5,5$ e $a = 15$ fixos, variando b e p . (c) Função taxa de falha considerando $\alpha = 1,0$, $a = 1,5$ e $b = 2,0$ fixos, variando θ e p .

3.2.1 Expansões para momentos

Momentos em desempenham um importante papel no teste de controle de qualidade e confiabilidade, em que o pesquisador precisa prever a falha futura baseada no tempo de início das falhas. Para facilitar os cálculos, podemos escrever a função densidade de probabilidade da distribuição BSNGG em forma de uma combinação linear com a distribuição BSNG.

Considere $|\omega| < 1$ e ρ um número real, podemos escrever a seguinte igualdade

$$(1 - \omega)^{-\rho} = \sum_{j=0}^{\infty} (-1)^j \binom{-\rho}{j} \omega^j, \quad (41)$$

aplicando a serie de potências (41) na expressão (36), tem-se

$$f(x; \alpha, \theta, a, b, p) = (1 - p) g_{\alpha, \theta, a, b}(x) \sum_{j=0}^{\infty} (-1)^j \binom{-2}{j} p^j [1 - G_{\alpha, \theta, a, b}(x)]^j. \quad (42)$$

Utilizando o binômio dado por

$$[1 - G(t)]^j = \sum_{k=0}^j (-1)^k \binom{j}{k} G(t)^k, \quad (43)$$

na equação (42), a função densidade é expressa por

$$f(x; \alpha, \theta, a, b, p) = (1 - p) g_{\alpha, \theta, a, b}(x) \sum_{j=0}^{\infty} (-1)^j \binom{-2}{j} p^j \sum_{k=0}^j (-1)^k \binom{j}{k} [G_{\alpha, \theta, a, b}(x)]^k. \quad (44)$$

Invertendo a ordem dos somatórios da forma $\sum_{j=0}^{\infty} \sum_{k=0}^j$ por $\sum_{k=0}^{\infty} \sum_{j=k}^{\infty}$ podemos escrever (44) como

$$f(x; \alpha, \theta, a, b, p) = g_{\alpha, \theta, a, b}(x) \sum_{k=0}^{\infty} t_k [G_{\alpha, \theta, a, b}(x)]^k, \quad (45)$$

em que $t_k = (1 - p) \sum_{j=k}^{\infty} (-1)^{j+k} p^j \binom{-2}{j} \binom{j}{k}$. Note que $\sum_{j=k}^{\infty} (-1)^{j+k} p^j \binom{-2}{j} \binom{j}{k}$ pode ser escrito apenas como $(-1)^{2k} (1 - p)^{-(k+2)} p^k \binom{-2}{k}$, assim t_k é escrito como

$$t_k = (-1)^{2k} (1 - p)^{-(k+1)} p^k \binom{-2}{k}.$$

Utilizando a identidade proposta para a distribuição BSNG (Pescim *et al.*, 2010) dada por $G_{\alpha, \theta, a, b}(x) = \frac{\Gamma(b)}{B(a, b)} \sum_{r=0}^{\infty} a_r [G_{\alpha, \theta}(x)]^{a+r}$ na equação (45), em que $G_{\alpha, \theta}(x)$ representa a função distribuição da distribuição SNG e $a_r = \frac{(-1)^r}{(a+r)r!\Gamma(b-r)}$, tem-se

$$f(x; \alpha, \theta, a, b, p) = g_{\alpha, \theta, a, b}(x) \sum_{k=0}^{\infty} t_k \frac{\Gamma(b)^k}{B(a, b)^k} [G_{\alpha, \theta}(x)]^{ak} \left[\sum_{r=0}^{\infty} a_r [G_{\alpha, \theta}(x)]^r \right]^k. \quad (46)$$

Utilizando a identidade $(\sum_{r=0}^{\infty} a_r x^r)^k = \sum_{r=0}^{\infty} c_{k,r} x^r$ (Gradshteyn e Ryzhik, 2000), em que

$$c_{k,r} = \begin{cases} a_0^k, & \text{para } r = 0; \\ (ra_0)^{-1} \sum_{m=1}^r (mk - r + m) a_m c_{k,r-m}, & \text{para } r \geq 1, \end{cases} \quad (47)$$

a equação (46) é escrita da forma

$$f(x; \alpha, \theta, a, b, p) = g_{\alpha, \theta, a, b}(x) \sum_{k, r=0}^{\infty} t_k \frac{\Gamma(b)^k}{B(a, b)^k} c_{k, r} [G_{\alpha, \theta}(x)]^{ak+r}. \quad (48)$$

Considere que a função densidade de probabilidade da distribuição BSNG pode ser escrita em termos da distribuição SNG na forma

$$g_{\alpha, \theta, a, b}(x) = \frac{g_{\alpha, \theta}(x)}{B(a, b)} G_{\alpha, \theta}(x)^{a-1} [1 - G_{\alpha, \theta}(x)]^{b-1}. \quad (49)$$

Substituindo (49) em (48) e multiplicando por $B[a(k+1)+r, b] B[a(k+1)+r, b]^{-1}$ tem-se

$$\begin{aligned} f(x; \alpha, \theta, a, b, p) &= \sum_{k, r=0}^{\infty} t_k c_{k, r} \frac{\Gamma(b)^k B[a(k+1)+r, b]}{B(a, b)^{k+1}} \frac{g_{\alpha, \theta}(x)}{B[a(k+1)+r, b]} \\ &\times G_{\alpha, \theta}(x)^{a(k+1)+r-1} [1 - G_{\alpha, \theta}(x)]^{b-1}. \end{aligned}$$

Por fim, a distribuição BSNGG pode ser escrita em termos de uma combinação linear da forma

$$f(x; \alpha, \theta, a, b, p) = \sum_{k, r=0}^{\infty} \kappa_{k, r} g_{\alpha, \theta, a(k+1)+r, b}(x), \quad (50)$$

em que

$$\begin{aligned} \kappa_{k, r} &= t_k c_{k, r} \frac{\Gamma(b)^k B[a(k+1)+r, b]}{B(a, b)^{k+1}} \\ &= \frac{(-1)^{2k} p^k (1-p)^{-(k+1)} \binom{-2}{k} \Gamma(b)^k B[a(k+1)+r, b] c_{k, r}}{B(a, b)^{k+1}}. \end{aligned} \quad (51)$$

3.2.2 Momentos

Seja uma variável aleatória com distribuição BSNGG, ou seja, $X \sim \text{BSNGG}(\alpha, \theta, a, b, p)$. O s -ésimo momento de X pode ser obtido a partir da distribuição BSNG na forma de combinação linear (50), ou seja, em termos do s -ésimo momento da distribuição BSNG,

$$\eta'_s = \int_0^{+\infty} x^s \sum_{k, r=0}^{\infty} \kappa_{k, r} g_{\alpha, \theta, a(k+1)+r, b}(x) dx$$

$$\eta'_s = \sum_{k,r=0}^{\infty} \kappa_{k,r}(a, b) \int_0^{+\infty} x^s g_{\alpha, \theta, a(k+1)+r, b}(x) dx \quad (52)$$

O s -ésimo momento da distribuição BSNG (18) reparametrizado na forma $a = a(k+1) + r$ é dado por

$$\eta'_s = \theta^s \sqrt{\frac{2}{\pi}} \sum_{h,l=0}^{\infty} \sum_{i=0}^l \sum_{j=0}^i u_{h,l,i,j}(a(k+1) + r, b) I\left(\frac{s}{\alpha}, j\right). \quad (53)$$

em que

$$u_{h,l,i,j}(a(k+1) + r, b) = \frac{w_{h,l,i}(a(k+1) + r, b)(-1)^j}{B[a(k+1) + r, b]2^i} \binom{i}{j}, \quad (54)$$

$$w_{h,l,i}(a(k+1) + r, b) = \frac{2^l (-1)^{h+l+i} \Gamma(b) \Gamma[a(k+1) + r + h]}{\Gamma(b-h) \Gamma[a(k+1) + r + h - l] h! l!} \binom{l}{i}, \quad (55)$$

e $I\left(\frac{s}{\alpha}, j\right)$ dado pela equação (19).

Substituindo (53) em (52), o s -ésimo momento de X pode ser escrito como

$$\eta'_s = \theta^s \sqrt{\frac{2}{\pi}} \sum_{k,r,h,l=0}^{\infty} \sum_{i=0}^l \sum_{j=0}^i \kappa_{k,r}(a, b) u_{h,l,i,j}(a(k+1) + r, b) I\left(\frac{s}{\alpha}, j\right).$$

Para simplificar os termos, considere que

$$\begin{aligned} \tau_{k,r,h,l,i,j}(a, b) &= \kappa_{k,r}(a, b) u_{h,l,i,j}(a(k+1) + r, b) \\ &= \frac{(-p)^k (k+1) \Gamma(b)^k c_{k,r} B[a(k+1) + r, b]}{B(a, b)^{k+1} (1-p)^{k+1}} \\ &\quad \times \frac{w_{h,l,i}(a(k+1) + r, b)(-1)^j}{B[a(k+1) + r, b]2^i} \binom{i}{j} \end{aligned}$$

$$\begin{aligned}
\tau_{k,r,h,l,i,j}(a,b) &= \frac{(k+1)(-1)^{h+l+i+j}2^{l-i}(-p)^k c_{k,r}}{(1-p)^{k+1}j!h!(l-i)!(i-j)!B(a,b)^{k+1}} \\
&\times \frac{\Gamma[a(k+1)+r+h]\Gamma(b)^{k+1}}{\Gamma(b-h)\Gamma[a(k+1)+r+h-l]}, \tag{56}
\end{aligned}$$

logo o s -ésimo momento de X é escrito como

$$\eta'_s = \theta^s \sqrt{\frac{2}{\pi}} \sum_{k,r,h,l=0}^{\infty} \sum_{i=0}^l \sum_{j=0}^i \tau_{k,r,h,l,i,j}(a,b) I\left(\frac{s}{\alpha}, j\right).$$

Quando $j + \frac{s}{\alpha}$ for um número par, os η'_s poderão ser escritos em termos da função de Lauricella do tipo A dado na equação (20), assim como os momentos da distribuição BSNG,

$$\begin{aligned}
\eta'_s = & \theta^s \sqrt{\frac{2}{\pi}} \sum_{k,r,h,l=0}^{\infty} \sum_{i=0}^l \sum_{j=0}^i \tau_{k,r,h,l,i,j}(a,b) \pi^{\frac{-j}{2}} 2^{2j+\frac{s}{2\alpha}-\frac{1}{2}} \Gamma\left(\frac{j+\frac{s}{\alpha}+1}{2}\right) \\
& \times F_A^{(j)}\left(\frac{j+\frac{s}{\alpha}+1}{2}; \frac{1}{2}, \dots, \frac{1}{2}, \frac{3}{2}, \dots, \frac{3}{2}; -1, \dots, -1\right).
\end{aligned}$$

3.2.3 Assimetria e curtose

Como apresentado nas Figuras 6 e 7, os valores de assimetria e curtose de X aumentam à medida que o valor do parâmetro b aumenta para a fixo e diminuem à medida que o parâmetro a aumenta para b fixo. Com a transformação geométrica, um novo parâmetro foi adicionado, assim algumas representações gráficas dessas medidas considerando alguns valores fixos para a , b , $\theta = 55$, $\alpha = 1$ e variando p , são apresentados nas Figuras 10 e 11.

Figura 10 - Gráficos dos valores dos coeficientes de assimetria e curtose da distribuição BSNGG como função de p para alguns valores de a e $b = 2$.

Figura 11 - Gráficos dos valores dos coeficientes de assimetria e curtose da distribuição BSNGG como função de p para alguns valores de b e $a = 0,5$.

Com as Figuras 10 e 11 observa-se que quanto maior o valor do parâmetro p maior é o valor do coeficiente de assimetria, o que significa que quanto maior p mais assimétrica a direita será a distribuição. Esse resultado nos dá uma indicação de que a distribuição BSNGG se ajusta melhor que a distribuição BSNG à dados assimétricos positivos.

Assim como o coeficiente de assimetria, o coeficiente de curtose aumenta quanto p aumenta, ou seja, quanto maior o valor de p , considerando os outros parâmetros

fixos, maior será o peso da cauda a direita.

3.2.4 Função geradora de momentos

Utilizando a equação (50), é possível obter a função geradora de momentos de X em termos da função geradora de momentos da distribuição BSNG (21), expressa por

$$\begin{aligned} M_X(t) &= \int_0^\infty e^{tx} \sum_{k,r=0}^\infty \kappa_{k,r}(a,b) g_{\alpha,\theta,a(k+1)+r,b}(x) dx \\ &= \sum_{k,r=0}^\infty \kappa_{k,r}(a,b) \int_0^\infty e^{tx} g_{\alpha,\theta,a(k+1)+r,b}(x) dx, \end{aligned} \quad (57)$$

em que $\kappa_{k,r}(a,b)$ é definido em (51) e $\int_0^\infty e^{tx} g_{\alpha,\theta,a(k+1)+r,b}(x) dx$ representa o s -ésimo momento da distribuição BSNG considerando $a = a(k+1) + r$, representada por

$$M_X(t) = \sqrt{\frac{2}{\pi}} \sum_{h,l=0}^\infty \sum_{i=0}^l \sum_{j=0}^i u_{h,l,i,j}(a(k+1) + r, b) \pi^{-\frac{j}{2}} \sum_{u=0}^\infty \frac{\theta^u t^u}{u!} 2^{j+\frac{u}{2\alpha}-\frac{1}{2}} I\left(\frac{u}{\alpha}, j\right), \quad (58)$$

em que $u_{h,l,i,j}(a(k+1) + r, b)$ e $I\left(\frac{u}{\alpha}, j\right)$ são definidos em (54) e (19), respectivamente. Aplicando (58) em (57) tem-se

$$\begin{aligned} M_X(t) &= \sqrt{\frac{2}{\pi}} \sum_{k,r,h,l=0}^\infty \sum_{i=0}^l \sum_{j=0}^i \kappa_{k,r}(a,b) u_{h,l,i,j}(a(k+1) + r, b) \pi^{-\frac{j}{2}} \\ &\quad \times \sum_{u=0}^\infty \frac{\theta^u t^u}{u!} 2^{j+\frac{u}{2\alpha}-\frac{1}{2}} I\left(\frac{u}{\alpha}, j\right). \end{aligned} \quad (59)$$

Utilizando a simplificação (56), a função geradora de momentos da distribuição BSNGG é escrita como

$$M_X(t) = \sqrt{\frac{2}{\pi}} \sum_{k,r,h,l=0}^\infty \sum_{i=0}^l \sum_{j=0}^i \tau_{r,k,h,l,i,j}(a,b) \pi^{-\frac{j}{2}} \sum_{u=0}^\infty \frac{\theta^u t^u}{u!} 2^{j+\frac{u}{2\alpha}-\frac{1}{2}} I\left(\frac{u}{\alpha}, j\right). \quad (60)$$

3.2.5 Desvios médios

Os desvios médios em relação a média e a mediana da distribuição BSNGG são obtidos a partir das equações (8) e (9), respectivamente, assim

$$\delta_1(x) = 2\eta'_1 F(\eta'_1) - 2\eta'_1 + 2T_2(\eta'_1),$$

$$\delta_2(x) = 2T_2(M) - \eta'_1,$$

em que $\eta'_1 = E(x)$, $F(\eta'_1)$ é obtido com a equação (38) e a função $T_2(q) = \int_q^\infty xf(x)dx$, é determinada a partir da fdp escrita na forma de mistura (50) da seguinte forma

$$\begin{aligned} T_2(q) &= \int_q^\infty xf(x; \alpha, \theta, a, b, p)dx \\ &= \int_q^\infty x \sum_{k,r=0}^\infty \kappa_{k,r}(a, b) \times g_{\alpha, \theta, a(k+1)+r, b}(x)dx \\ &= \sum_{k,r=0}^\infty \kappa_{k,r}(a, b) \int_q^\infty x g_{\alpha, \theta, a(k+1)+r, b}(x)dx. \end{aligned} \quad (61)$$

Utilizando a equação (22) e considerando $a = a(k+1) + r$ tem-se

$$\begin{aligned} T_2(q) &= \sqrt{\frac{2}{\pi}}\theta \sum_{k,r,h,l=0}^\infty \sum_{i=0}^l \sum_{j=0}^i \kappa_{k,r}(a, b) u_{h,l,i,j}(a(k+1) + r, b) \left(\frac{2}{\sqrt{\pi}}\right)^j \\ &\times \sum_{m_1=0}^\infty \cdots \sum_{m_j=0}^\infty \frac{(-1)^{m_1+\cdots+m_j}}{(2m_1+1)\cdots(2m_j+1)m_1!\cdots m_j!} \\ &\times \Gamma\left[\left(m_1+\cdots+m_j+\frac{j}{2}+\frac{1}{2\alpha}+\frac{1}{2}\right), \frac{1}{2}\left(\frac{q}{\theta}\right)^\alpha\right]. \end{aligned} \quad (62)$$

Utilizando a simplificação do termo $\kappa_{k,r}(a, b) u_{h,l,i,j}(a(k+1) + r, b)$ já obtida em (56), o termo $T_2(q)$ é escrito na forma

$$\begin{aligned} T_2(q) &= \sqrt{\frac{2}{\pi}}\theta \sum_{k,r,h,l=0}^\infty \sum_{i=0}^l \sum_{j=0}^i \tau_{k,r,h,l,i}(a, b) \left(\frac{2}{\sqrt{\pi}}\right)^j \\ &\times \sum_{m_1=0}^\infty \cdots \sum_{m_j=0}^\infty \frac{(-1)^{m_1+\cdots+m_j}}{(2m_1+1)\cdots(2m_j+1)m_1!\cdots m_j!} \end{aligned}$$

$$\times \Gamma \left[\left(m_1 + \dots + m_j + \frac{j}{2} + \frac{1}{2\alpha} + \frac{1}{2} \right), \frac{1}{2} \left(\frac{q}{\theta} \right)^\alpha \right]. \quad (63)$$

3.2.6 Função quantílica

A função quantílica de uma distribuição de probabilidade é encontrada igualando a função distribuição acumulada à “u” que representa o quantil da distribuição. Utilizando a função densidade descrita na equação (35), é possível encontrar a função quantil de forma geral para a família das distribuições geométricas,

$$\begin{aligned} F(x) &= \frac{G_{\alpha,\theta,a,b}(x)}{[1 - p(1 - G_{\alpha,\theta,a,b}(x))]} \\ \frac{G_{\alpha,\theta,a,b}(x)}{[1 - p(1 - G_{\alpha,\theta,a,b}(x))]} &= u \\ G_{\alpha,\theta,a,b}(x) &= u[1 - p(1 - G_{\alpha,\theta,a,b}(x))] \\ G_{\alpha,\theta,a,b}(x)(1 - pu) &= u - up \\ G_{\alpha,\theta,a,b}(x) &= \frac{u(1 - p)}{(1 - pu)}, \quad u \in (0, 1), \end{aligned}$$

assim a função quantílica de qualquer distribuição gerada a partir da transformação geométrica será descrita em termos da função quantílica da distribuição primaria utilizada para a transformação. Utilizando a função quantílica da distribuição BSNG (23), encontramos a função quantílica da distribuição BSNGG escrita em termos dos quantis da distribuição normal e beta, dada por

$$x = \theta Q_N \left(\frac{1 + Q_B \left(\frac{u(1-p)}{(1-pu)} \right)}{2} \right)^{1/\alpha}. \quad (64)$$

3.2.7 Curvas de Lorenz e Bonferroni

As curvas de Bonferroni e Lorenz, calculadas a partir das equações (10) e (11), respectivamente, são apresentadas nas formas

$$B(\rho) = \frac{1}{\rho} - \frac{T_2(\rho)}{\rho \eta'_1} \quad \text{e} \quad L(\rho) = 1 - \frac{T_2(\rho)}{\eta'_1},$$

em que $\eta'_1 = E(X)$ e $T_2(\rho)$ é dada pela equação (63).

Com a Figura 12 é possível observar o efeito do parâmetro p nas curvas de

Bonferroni e Lorenz, assim pode-se dizer que quanto maior o valor de p , considerando os outros parâmetros fixos, maior a distorção nas curvas, cuja interpretação é de maior desigualdade entre as riquezas em uma população. Esse efeito do parâmetro p sob as curvas era esperado, pois interpretando os coeficiente de assimetria e curtose, observa-se que os mesmos são diretamente proporcionais ao tamanho de n .

Figura 12 - Gráficos das curvas de Bonferroni e Lorenz utilizando a distribuição GBGSN variando o parâmetro p com $\alpha = 3, 0$, $\theta = 10, 0$, $a = 0, 3$ e $b = 0, 3$.

3.2.8 Estimação dos parâmetros por máxima verossimilhança

Para as estimativas do vetor de parâmetros da distribuição BSNGG, iremos utilizar o método de MV. Dessa forma, seja $x_1, \dots, x_n$ uma amostra aleatória e seja $c_1, \dots, c_n$ os tempos de censura, então, os dados observados são definidos por $x_i = \min(x_i, c_i)$. Portanto, assumindo que as censuras são aleatórias e não-informativas, o logaritmo da função de verossimilhança para as distribuições da família geométricas expressa na forma geral pela equação (33) é dada por:

$$\begin{aligned}
l(\boldsymbol{\lambda}) &= \sum_{i \in F} \left[\log \left(\frac{g(x_i)(1-p)}{\{1-p[1-G(x_i)]\}^2} \right) \right] + \sum_{i \in C} \left[\log \left(\frac{(1-p)(1-G(x_i))}{1-p(1-G(x_i))} \right) \right] \\
&= n \log(1-p) + \sum_{i \in F} \log [g(x_i)] - 2 \sum_{i \in F} \log [1-p[1-G(x_i)]] \\
&\quad + \sum_{i \in C} \log [1-G(x_i)] - \sum_{i \in C} \log [1-p[1-G(x_i)]], \tag{65}
\end{aligned}$$

em que F representa o conjunto de observações não censuradas (falha) e C o grupo de observações censuradas.

Considerando que $g(x_i)$ e $G(x_i)$ sejam a fdp e a fda da distribuição BSNGG, dadas respectivamente por (14) e (15), a função log-verossimilhança da equação (65) para o vetor de parâmetros $\boldsymbol{\lambda}=(\alpha, \theta, a, b, p)^T$ da distribuição BSNGG é escrita como

$$\begin{aligned}
l(\boldsymbol{\lambda}) &\propto n \log(1-p) + \sum_{i \in F} \log \left[\frac{\alpha}{x_i} \right] + \alpha \sum_{i \in F} \log \left[\frac{x_i}{\theta} \right] - \frac{1}{2} \sum_{i \in F} \left(\frac{x_i}{\theta} \right)^{2\alpha} - r \log [B(a, b)] \\
&\quad + (b-1)r \log(2) + (a-1) \sum_{i \in F} \log [2u_i - 1] + (b-1) \sum_{i \in F} \log [1 - u_i] \\
&\quad - 2 \sum_{i \in F} \log [1 - p(1 - I_{2u_i-1}(a, b))] + \sum_{i \in C} \log [1 - I_{2u_i-1}(a, b)] \\
&\quad - \sum_{i \in C} \log [1 - p(1 - I_{2u_i-1}(a, b))], \tag{66}
\end{aligned}$$

em que r representa o número de observações não censuradas (falhas) e $u_i = \Phi \left[\left(\frac{x_i}{\theta} \right)^\alpha \right]$.

A diferenciação da função de log-verossimilhança em relação aos parâmetros da distribuição, ou seja, os componentes do vetor escore $\boldsymbol{U} = \left(\frac{\partial l(\boldsymbol{\lambda})}{\partial \alpha}, \frac{\partial l(\boldsymbol{\lambda})}{\partial \theta}, \frac{\partial l(\boldsymbol{\lambda})}{\partial a}, \frac{\partial l(\boldsymbol{\lambda})}{\partial b}, \frac{\partial l(\boldsymbol{\lambda})}{\partial p} \right)^T$ são dados por

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\lambda})}{\partial \alpha} &= \frac{r}{\alpha} + \sum_{i \in F} \log \left[\frac{x_i}{\theta} \right] - \sum_{i \in F} \log \left[\frac{x_i}{\theta} \right] \left(\frac{x_i}{\theta} \right)^{2\alpha} + \frac{2(a-1)}{\sqrt{2\pi}} \sum_{i \in F} \frac{e^{-\frac{1}{2} \left(\frac{x_i}{\theta} \right)^{2\alpha}}}{2u_i - 1} \left(\frac{x_i}{\theta} \right)^\alpha \log \left[\frac{x_i}{\theta} \right] \\
&\quad + \frac{(1-b)}{\sqrt{2\pi}} \sum_{i \in F} \frac{e^{-\frac{1}{2} \left(\frac{x_i}{\theta} \right)^{2\alpha}}}{1 - u_i} \left(\frac{x_i}{\theta} \right)^\alpha \log \left[\frac{x_i}{\theta} \right] - 2 \sum_{i \in F} \frac{p \left[\dot{I}_{2u_i-1}(a, b) \right]_\alpha}{1 - p(1 - I_{2u_i-1}(a, b))} \\
&\quad - \sum_{i \in C} \frac{\left[\dot{I}_{2u_i-1}(a, b) \right]_\alpha}{1 - I_{2u_i-1}(a, b)} - \sum_{i \in C} \frac{p \left[\dot{I}_{2u_i-1}(a, b) \right]_\alpha}{1 - p(1 - I_{2u_i-1}(a, b))},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\lambda})}{\partial \theta} &= \frac{-r\alpha}{\theta} + \sum_{i \in F} \left(\frac{x_i}{\theta}\right)^{2\alpha} \left(\frac{\alpha}{\theta}\right) + \frac{2(1-a)}{\sqrt{2\pi}} \sum_{i \in F} \frac{e^{-\frac{1}{2}\left(\frac{x_i}{\theta}\right)^{2\alpha}}}{2u_i - 1} \left(\frac{x_i}{\theta}\right)^{\alpha} \left(\frac{\alpha}{\theta}\right) \\
&+ \frac{(b-1)}{\sqrt{2\pi}} \sum_{i \in F} \frac{e^{-\frac{1}{2}\left(\frac{x_i}{\theta}\right)^{2\alpha}}}{1 - u_i} \left(\frac{x_i}{\theta}\right)^{\alpha} \left(\frac{\alpha}{\theta}\right) - 2 \sum_{i \in F} \frac{p \left[\dot{I}_{2u_i-1}(a, b) \right]_{\theta}}{1 - p(1 - I_{2u_i-1}(a, b))} \\
&- \sum_{i \in C} \frac{\left[\dot{I}_{2u_i-1}(a, b) \right]_{\theta}}{1 - I_{2u_i-1}(a, b)} - \sum_{i \in C} \frac{p \left[\dot{I}_{2u_i-1}(a, b) \right]_{\alpha}}{1 - p(1 - I_{2u_i-1}(a, b))},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\lambda})}{\partial a} &= r\psi(a+b) - r\psi(a) + \sum_{i \in F} \log[2u_i - 1] - 2 \sum_{i \in F} \frac{p \left[\dot{I}_{2u_i-1}(a, b) \right]_a}{1 - p(1 - I_{2u_i-1}(a, b))} \\
&- \sum_{i \in C} \frac{\left[\dot{I}_{2u_i-1}(a, b) \right]_a}{1 - I_{2u_i-1}(a, b)} - \sum_{i \in C} \frac{p \left[\dot{I}_{2u_i-1}(a, b) \right]_a}{1 - p(1 - I_{2u_i-1}(a, b))},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\lambda})}{\partial b} &= r\psi(a+b) - r\psi(b) + r \log(2) + \sum_{i \in F} \log[1 - u_i] - 2 \sum_{i \in F} \frac{p \left[\dot{I}_{2u_i-1}(a, b) \right]_b}{1 - p(1 - I_{2u_i-1}(a, b))} \\
&- \sum_{i \in C} \frac{\left[\dot{I}_{2u_i-1}(a, b) \right]_b}{1 - I_{2u_i-1}(a, b)} - \sum_{i \in C} \frac{p \left[\dot{I}_{2u_i-1}(a, b) \right]_b}{1 - p(1 - I_{2u_i-1}(a, b))},
\end{aligned}$$

$$\frac{\partial l(\boldsymbol{\lambda})}{\partial p} = \frac{n}{p-1} + 2 \sum_{i \in F} \frac{1 - I_{2u_i-1}(a, b)}{1 - p(1 - I_{2u_i-1}(a, b))} + \sum_{i \in C} \frac{1 - I_{2u_i-1}(a, b)}{1 - p(1 - I_{2u_i-1}(a, b))},$$

em que $\left[\dot{I}_{2u_i-1}(a, b) \right]_{\boldsymbol{\lambda}}$ representa a derivada da função beta regularizada incompleta em relação aos parâmetros indicados, as quais são encontradas no Apêndice A.2.

As estimativas dos parâmetros via MV não possuem forma fechada. Alguns métodos iterativos são utilizados nesses casos, como Newton-Raphson e algoritmo de quase-Newton. Neste caso utilizaremos o procedimento NLMixed do *software* SAS para encontrar as estimativas de máxima verossimilhança, o qual utiliza o método dual quase-Newton.

Considerando amostras grandes, os intervalos de confiança e testes de hipóteses podem ser obtidos utilizando a distribuição assintótica para o vetor de parâmetros, que são apresentadas na Subseção 2.9.5. Assim, a aproximação normal assintótica para

$\hat{\lambda}$ pode ser expressa por $\hat{\lambda} \sim N_{(5)}\{\lambda; \ddot{\mathbf{L}}(\lambda)^{-1}\}$, em que $\ddot{\mathbf{L}}(\lambda)$ é a matriz de informação observada com dimensão 5×5 . Os elementos da matriz de informação, dados pela matriz

$$\ddot{\mathbf{L}}(\lambda) = - \begin{bmatrix} L_{\alpha\alpha} & L_{\alpha\theta} & L_{\alpha a} & L_{\alpha b} & L_{\alpha p} \\ L_{\theta\alpha} & L_{\theta\theta} & L_{\theta a} & L_{\theta b} & L_{\theta p} \\ L_{a\alpha} & L_{a\theta} & L_{aa} & L_{ab} & L_{ap} \\ L_{b\alpha} & L_{b\theta} & L_{ba} & L_{bb} & L_{bp} \\ L_{p\alpha} & L_{p\theta} & L_{pa} & L_{pb} & L_{pp} \end{bmatrix}, \quad (67)$$

são apresentados no Apêndice A.3. Com a matriz de informação observada, é possível construir os intervalos de confiança assintóticos para os respectivos parâmetros da distribuição BSNGG, ou seja, considerando γ o nível de confiança, os intervalos podem ser expressos por

$$IC(\lambda_r, (1 - \gamma)\%) = \left(\hat{\lambda}_r - \frac{z_{\gamma/2}}{\sqrt{\hat{I}(\hat{\lambda}_{rr})}}, \hat{\lambda}_r + \frac{z_{\gamma/2}}{\sqrt{\hat{I}(\hat{\lambda}_{rr})}} \right),$$

em que $z_{\gamma/2}$ representa o quantil da distribuição normal padrão.

3.3 Modelo de regressão log-beta semi-normal generalizada geométrica

Na prática, é comum a ocorrência de situações em que uma ou mais covariáveis estão relacionadas aos tempos de sobrevivência. No contexto dos dados de sobrevivência, uma maneira de incluir covariáveis na análise é utilizando a classe de modelos de locação-escala. Essa classe de modelos consiste em utilizar a transformação logarítmica nos tempo de falha, $Y = \log(X)$, de tal forma que para um vetor de covariáveis, o logaritmo do tempo de falha possui uma distribuição com um parâmetro de locação μ , que depende das variáveis regressoras, e um parâmetro de escala σ .

Seja X uma variável aleatória com distribuição BSNGG com função densidade expressa pela equação (37), assim, $Y = \log(X)$ define a distribuição log-beta semi-normal generalizada geométrica (LBSNGG). Considerando as reparametrizações $\sigma = \frac{\sqrt{2}}{2\alpha}$ e $\mu = \log(\theta)$, a função densidade de $Y \sim \text{LBSNGG}(\mu, \sigma, a, b, p)$ é expressa por

$$f(y; \mu, \sigma, a, b, p) = 2^{b-1}(1-p) \exp \left\{ \frac{\sqrt{2}}{2} \left(\frac{y-\mu}{\sigma} \right) - \frac{1}{2} \exp \left\{ \left(\frac{y-\mu}{\sigma} \right) \sqrt{2} \right\} \right\} \\ \times \frac{\left[2\Phi \left[\exp \left\{ \frac{\sqrt{2}}{2} \left(\frac{y-\mu}{\sigma} \right) \right\} \right] - 1 \right]^{a-1} \left[1 - \Phi \left[\exp \left\{ \frac{\sqrt{2}}{2} \left(\frac{y-\mu}{\sigma} \right) \right\} \right] \right]^{b-1}}{B(a, b) \sigma \sqrt{\pi} \left[1 - p \left(1 - I_{2\Phi \left[\exp \left\{ \frac{\sqrt{2}}{2} \left(\frac{y-\mu}{\sigma} \right) \right\} \right] - 1}(a, b) \right) \right]^2}, \quad (68)$$

em que $a, b > 0$ e $0 \leq p < 1$ são parâmetros de forma, $\sigma > 0$ é parâmetro de escala e $-\infty \leq \mu \leq \infty$ é o parâmetro de locação. Algumas formas que a fdp de Y pode assumir são apresentadas pela Figura 13.

Figura 13 - Formas da função densidade da distribuição LBSNGG para alguns valores dos parâmetros. (a) Para alguns valores de b , p com $\mu = 0$, $\sigma = 1$ e $a = 2.5$. (b) Para alguns valores de σ , a , p com $\mu = 0$ and $b = 0.5$. (c) Para alguns valores de μ , σ , p com $a = b = 1.5$.

Seja a variável aleatória $Z = (Y - \mu)/\sigma$ em que sua função densidade de probabilidade é dada por

$$f(z) = 2^{b-1}(1-p) \exp \left\{ \frac{z\sqrt{2}}{2} - \frac{1}{2} \exp \left\{ z\sqrt{2} \right\} \right\} \times \frac{\left[2\Phi \left[\exp \left\{ \frac{z\sqrt{2}}{2} \right\} \right] - 1 \right]^{a-1} \left[1 - \Phi \left[\exp \left\{ \frac{z\sqrt{2}}{2} \right\} \right] \right]^{b-1}}{B(a, b) \sigma \sqrt{\pi} \left[1 - p \left(1 - I_{2\Phi \left[\exp \left\{ \frac{z\sqrt{2}}{2} \right\} \right] - 1}(a, b) \right) \right]^2}. \quad (69)$$

A função de sobrevivência de Y é expressa como

$$S(y; \mu, \sigma, a, b, p) = 1 - \frac{I_{2\Phi \left[\exp \left\{ \frac{\sqrt{2}}{2} \left(\frac{y-\mu}{\sigma} \right) \right\} \right] - 1}(a, b)}{1 - p \left(1 - I_{2\Phi \left[\exp \left\{ \frac{\sqrt{2}}{2} \left(\frac{y-\mu}{\sigma} \right) \right\} \right] - 1}(a, b) \right)}. \quad (70)$$

Os casos particulares da distribuição LBSNGG, quando alguns parâmetros assumem valores específicos são apresentados na Tabela 2.

Tabela 2- Casos particulares da distribuição LBSNGG.

Distribuições	μ	σ	a	b	p
Log semi-normal generalizada (LSNG)	$\log(\theta)$	$\sqrt{(2)}/2\alpha$	1	1	0
Log-beta semi-normal generalizada (LBSNG)	$\log(\theta)$	$\sqrt{(2)}/2\alpha$	a	b	0
Log semi-normal generalizada exponenciada (LSNGE)	$\log(\theta)$	$\sqrt{(2)}/2\alpha$	a	1	0
Log-beta semi-normal (LBSN)	$\log(\theta)$	$\sqrt{(2)}/2$	a	b	0
Log semi-normal (LSN)	$\log(\theta)$	$\sqrt{(2)}/2$	1	1	0
Log semi-normal geométrica (LSNG _{eo})	$\log(\theta)$	$\sqrt{(2)}/2$	1	1	p
Log semi-normal generalizada geométrica (LSNGG)	$\log(\theta)$	$\sqrt{(2)}/2\alpha$	1	1	p
Log semi-normal generalizada exponenciada geométrica (LSNGEG)	$\log(\theta)$	$\sqrt{(2)}/2\alpha$	a	1	p
Log-beta semi-normal geométrica (LBSNG _{eo})	$\log(\theta)$	$\sqrt{(2)}/2$	a	b	p

3.3.1 Expansões para momentos

Considerando a distribuição escrita na forma de combinação linear (50), ainda que $Y = \log(X)$ e as transformações $\theta = e^\mu$ e $\alpha = \frac{\sqrt{2}}{2\sigma}$, e novamente pelo método de transformação de variáveis Jacobiano, a distribuição LBSNGG pode ser escrita na forma de uma combinação linear com a distribuição LBSNG (24). Assim

$$\begin{aligned}
f(y) &= \sum_{k,r=0}^{\infty} \kappa_{k,r}(a, b) \frac{\exp \left\{ \left(\frac{y-\mu}{\sigma} \right) \frac{\sqrt{2}}{2} - \frac{1}{2} \exp \left\{ \left(\frac{y-\mu}{\sigma} \right) \sqrt{2} \right\} \right\} 2^{b-1}}{B(a(k+1) + r, b) \sigma \sqrt{\pi}} \\
&\times \left\{ 2\Phi \left[\left(\frac{y-\mu}{\sigma} \right) \frac{\sqrt{2}}{2} \right] - 1 \right\}^{a(k+1)+r-1} \left\{ 1 - \Phi \left[\left(\frac{y-\mu}{\sigma} \right) \frac{\sqrt{2}}{2} \right] \right\}^{b-1} \\
&= \sum_{k,r=0}^{\infty} \kappa_{k,r}(a, b) g_{\mu, \sigma, a(k+1)+r, b}(y),
\end{aligned} \tag{71}$$

em que $\kappa_{k,r}$ é expresso por (51) e $g_{\mu, \sigma, a(k+1)+r, b}(y)$ representa a fdp da distribuição LBSNG repametrizada com $a = a(k+1) + r$.

3.3.2 Momentos

Para calcular os coeficientes de assimetria e curtose é necessário encontrarmos a expressão geral dos s -ésimo momentos da nova distribuição. Esses momentos são obtidos facilmente utilizando a distribuição em forma de mistura (71), pois podem ser escritos em função dos momentos da distribuição LBSNG (25). Logo

$$\begin{aligned}
\eta'_s &= \int_{-\infty}^{\infty} y^s \sum_{k,r=0}^{\infty} \kappa_{k,r}(a, b) g_{\mu, \sigma, a(k+1)+r, b}(y) dy \\
&= \sum_{k,r=0}^{\infty} \kappa_{k,r}(a, b) \int_{-\infty}^{\infty} y^s g_{\mu, \sigma, a(k+1)+r, b}(y) dy \\
&= \sum_{k,r,h,l=0}^{\infty} \sum_{i=0}^l \sum_{\ell=0}^s \kappa_{k,r}(a, b) e_{h,l,i}(a(k+1) + r, b) \frac{1}{\sigma \sqrt{\pi}} \binom{s}{\ell} \mu^{s-\ell} (\sqrt{2}\sigma)^{s+1} \\
&\times \sum_{m_1=0}^{\infty} \dots \sum_{m_i=0}^{\infty} \frac{(-1)^{m_1+\dots+m_i}}{(2m_1+1) \dots (2m_i+1) m_1! \dots m_i!} \\
&\times \sum_{q=0}^s \binom{s}{q} [\log(2)]^{s-q} 2^{-1/2-s} \Gamma^q \left(m_1 + \dots + m_i + \frac{i+1}{2} \right),
\end{aligned}$$

em que $\kappa_{k,r}(a, b)$, $e_{h,l,i}(a(k+1) + r, b)$ e $\Gamma^q \left(m_1 + \dots + m_i + \frac{i+1}{2} \right)$ são definidos anteriormente nas equações (51), (26) e (27), respectivamente.

3.3.3 Assimetria e curtose

Os coeficientes de curtose e assimetria para alguns valores dos parâmetros fixos e outros variando são apresentados graficamente nas Figuras (14) e (15).

Figura 14 - Assimetria e curtose da distribuição LBSNGG para alguns valores dos parâmetros variando e p fixo. (a) Assimetria para alguns valores de a e p com $\mu = 5$, $\sigma = 13$ e $b = 0, 3$. (b) Curtose para alguns valores de a e p com $\mu = 5$, $\sigma = 13$ e $b = 0, 3$.

Figura 15 - Assimetria e curtose da distribuição LBSNGG para alguns valores dos parâmetros variando e p fixo. (a) Assimetria para alguns valores de b e p com $\mu = 5$, $\sigma = 13$ e $a = 0, 3$. (b) Curtose para alguns valores de a e p com $\mu = 5$, $\sigma = 13$ e $b = 0, 3$.

3.3.4 Função geradora de momentos

Assim como os momentos ordinais, a função geradora de momentos de Y também pode ser escrita em termos da função geradora de momentos da distribuição LBSNG (28). Utilizando a expansão (71) temos que

$$\begin{aligned}
 M_Y(t) &= \int_{-\infty}^{\infty} e^{ty} \sum_{k,r=0}^{\infty} \kappa_{k,r}(a,b) g_{\mu,\sigma,a(k+1)+r,b}(y) dy \\
 &= \sum_{k,r=0}^{\infty} \kappa_{k,r} \int_{-\infty}^{\infty} e^{ty} g_{\mu,\sigma,a(k+1)+r,b}(y) dy \\
 &= \frac{\sqrt{2}}{\pi} \sum_{k,r,h,l=0}^{\infty} \sum_{i=0}^l 2^{i+\frac{t\sigma}{\sqrt{2}}-2} e^{\mu t} \kappa_{k,r}(a,b) e_{h,l,i}(a(k+1)+r,b) \\
 &\quad \times \sum_{m_1=0}^{\infty} \cdots \sum_{m_i=0}^{\infty} \frac{(-1)^{m_1+\dots+m_i}}{(m_1+\frac{1}{2}) \dots (m_i+\frac{1}{2}) m_1! \dots m_i!} \\
 &\quad \times \Gamma\left(m_1+\dots+m_i+\frac{i+\frac{2\sigma t}{\sqrt{2}}+1}{2}\right),
 \end{aligned}$$

em que $\kappa_{k,r}(a,b)$ é representado pela equação (51) e $e_{h,l,i}(a,b)$ por (26).

3.3.5 Estimação dos parâmetros

Considere o vetor de covariáveis $\mathbf{v}_i = (v_{i1}, \dots, v_{ip})^T$ associadas a variável resposta y_i . O modelo de regressão linear baseado na distribuição LBSNGG é escrito na forma

$$y_i = \mathbf{v}_i^T \boldsymbol{\beta} + \sigma z_i, \quad i=1, \dots, n,$$

em que a fdp da variável aleatória z_i é definida em (69) e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ é o vetor de parâmetros desconhecidos.

Considere agora uma amostra $(y_1, \mathbf{v}_1), \dots, (y_n, \mathbf{v}_n)$ de observações independentes, em que a variável resposta é definida por $y_i = \min(\log(x_i), \log(c_i))$. Considere os parâmetros $-\infty < \mu < +\infty$, $\sigma > 0$, $a > 0$, $b > 0$ e $p \in [0, 1)$ desconhecidos, sendo μ escrito em termos de variáveis regressoras na forma $\mu = \mathbf{v}_i^T \boldsymbol{\beta}$ e o vetor de parâmetros de locação $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ representados pelo modelo linear $\boldsymbol{\mu} = \mathbf{V}\boldsymbol{\beta}$, em que $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_n)^T$ representa a matriz do modelo. Assumindo censura não informativa e que os tempos de falha e os tempos de censura sejam independentes, o logaritmo da

função de verossimilhança para o vetor de parâmetros $\boldsymbol{\lambda} = (a, b, p, \sigma, \boldsymbol{\beta}^T)^T$ da distribuição LBSNG é dada por

$$\begin{aligned}
l(\boldsymbol{\lambda}) &\propto n \log(1-p) + \frac{\sqrt{2}}{2} \sum_{i \in F} \left(\frac{y_i - \mu_i}{\sigma} \right) - \frac{1}{2} \sum_{i \in F} \exp \left\{ \sqrt{2} \left(\frac{y_i - \mu_i}{\sigma} \right) \right\} - r \log[B(a, b)] \\
&+ (b-1)r \log(2) + (a-1) \sum_{i \in F} \log[2u_i - 1] + (b-1) \sum_{i \in F} \log[1 - u_i] \\
&- 2 \sum_{i \in F} \log[1 - p(1 - I_{2u_i-1}(a, b))] + \sum_{i \in C} \log[1 - I_{2u_i-1}(a, b)] \\
&- \sum_{i \in C} \log[1 - p(1 - I_{2u_i-1}(a, b))], \tag{72}
\end{aligned}$$

em que F representa o conjunto de observações não censuradas (falha), C representa o conjunto de observações censuradas, $u_i = \Phi \left[e^{\frac{\sqrt{2}}{2} \left(\frac{y_i - \mu_i}{\sigma} \right)} \right]$ e r o número de observações não censuradas.

As derivadas parciais de primeira e segunda ordem que compõem o vetor escore $U = \left(\frac{\partial l(\boldsymbol{\lambda})}{\partial \boldsymbol{\beta}_k}, \frac{\partial l(\boldsymbol{\lambda})}{\partial \sigma}, \frac{\partial l(\boldsymbol{\lambda})}{\partial a}, \frac{\partial l(\boldsymbol{\lambda})}{\partial b}, \frac{\partial l(\boldsymbol{\lambda})}{\partial p} \right)$, em que $k = 0, 1, 2, \dots, m$, assim como os elementos da matriz de informação, são apresentados nos apêndices B.3 e B.4 respectivamente. A aproximação normal assintótica para o vetor $\hat{\boldsymbol{\lambda}}$ pode ser expressa por $\hat{\boldsymbol{\lambda}} \sim N_{(4+m)}\{\boldsymbol{\lambda}; \ddot{\mathbf{L}}(\boldsymbol{\lambda})^{-1}\}$, em que $\ddot{\mathbf{L}}(\boldsymbol{\lambda})$ é a matriz de informação observada com dimensão $(4+m) \times (4+m)$.

3.4 Aplicações

3.4.1 Simulações

Para simulações de observações que seguem distribuição BSNGG foi utilizado a função quantílica (64), considerando que $U \sim U(0; 1)$. Os histogramas dos dados gerados juntamente com a densidade da distribuição original que gerou os dados são apresentadas na Figura (16), na qual pode-se observar que o parâmetro p adicionado pela transformação geométrica causa uma certa flexibilidade na distribuição considerando os outros parâmetros fixos.

Figura 16 - Histogramas de simulações de dados com densidades originais para a distribuição BSNGG considerando 100.000,00 observações, $\alpha = 2$, $\theta = 10$, $a = 0,5$, $b = 0,5$ e p variando.

3.4.2 Dados de leucemia

Os dados utilizados nesta aplicação contém os tempos de vida em semanas de pacientes que sofriam de leucemia mieloide aguda. O estudo foi realizado por Feigl e Zelen (1965) e consistiu no acompanhamento de 33 pacientes até a morte, os quais eram classificados em 2 grupos, um com a presença do fator "Auer rods" e o outro a ausência do mesmo. "Auer rods" são grânulos citoplasmáticos existentes nas células leucêmicas da medula óssea, as quais foram constatada durante o diagnóstico, classificando os pacientes para que possa haver a comparação entre os tempos de vida dos dois grupos. Além do fator "Auer rods", a contagem de glóbulos brancos em milhares também foi um fator de interesse observado para comparação. Os dados podem ser encontrados utilizando o pacote SMIR no *software* R.

Inicialmente, iremos trabalhar apenas com o tempo até o óbito de cada indivíduo para o ajuste da distribuição BSNGG e seus sub-modelos. Para o ajuste da

distribuição aos dados, foi utilizado o *software* **SAS** pela rotina NLMixed, a qual utiliza o método de Quasi-Newton para as estimativas dos parâmetros, sendo necessário indicar um chute inicial. Como a distribuição BSNGG não possui formas fechadas para o EMV, as estimativas dos chutes iniciais são obtidas pelo ajuste de distribuições encaixadas mais simples, no caso a distribuição BSNG, que também utiliza da distribuição SNG para estimativas de seus chutes iniciais.

Com o objetivo de conhecer o comportamento do tempo de vida dos pacientes, algumas estatísticas descritivas são apresentadas na Tabela 3, na qual pode-se concluir que a média dos pacientes no estudo foi de 40,88 semanas e com base na mediana, 50% dos pacientes sobreviveram até 22 semanas. Observando a assimetria dos dados igual à 1,16, pode-se esperar que uma distribuição assimétrica a direita se ajuste melhor aos dados, esse é um bom sinal para o uso da distribuição BSNGG, pois o parâmetro p flexibiliza à direita as distribuições como apresentado nos gráficos de assimetria e curtose da Seção 3.2.3.

Tabela 3 - Estatísticas descritivas sobre a variável tempo.

Média	Mediana	Moda	CV	DP	Variância	Assimetria	Curtose	Mínimo	Máximo
40,88	22	4	114,25	46,7	2181,17	1,16	0,12	1	156

Tabela 4 - Estimativas dos parâmetros, desvios padrões (em parênteses) e critérios de informação para o modelo BSNGG e seus sub-modelos aplicados aos dados de leucemia.

Modelo	α	θ	a	b	p	AIC	CAIC	BIC
BSNGG	7,2631 (0,3662)	132,2900 (27,3855)	0,1231 (0,0251)	0,1907 (0,4127)	0,8749 (0,1580)	309,50	311,70	317,00
BSNG	0,3229 (0,2922)	73,8221 (74,3400)	3,2298 (4,5373)	3,3246 (1,1058)	0 -	314,90	316,30	320,90
SNG	0,6209 (0,0921)	46,5178 (10,1120)	1 -	1 -	0 -	311,50	311,90	314,50

Com a Tabela 4 pode observar que a distribuição BSNGG possui menor AIC e CAIC, sendo o critério BIC o segundo menor. Esse fato pode ocorrer pois este critério penaliza a quantidade de parâmetros do modelos, sendo que o modelo BSNGG possui $q = 5$ parâmetros comparado com o SNG que possui $q = 4$ parâmetros. Sendo assim, torna-se viável a aplicação do teste da razão de verossimilhança, tendo em vista que os modelos são encaixados.

Com os resultados dos dois testes da razão de verossimilhanças apresentados na Tabela 5, rejeita-se a hipótese nula do teste da razão de verossimilhança ao nível de 5%, portanto conclui-se que a distribuição que melhor se ajusta aos dados foi a BSNGG.

Tabela 5 - Teste da razão de verossimilhança para os modelos BSNGG, BSNG e SNG

Modelo	Hipótese	Estatística w	p-valor
BSNGG vs BSNG	$H_0: p = 0$	7,40	0,0065
BSNGG vs SNG	$H_0: p = 0 \text{ e } a = b = 1$	8,00	0,0460

Figura 17 - (a) Densidade das distribuições BSNGG e BSNG ajustadas para os dados de leucemia. (b) Densidade das distribuições BSNGG e SNG ajustadas para os dados de leucemia

Com a Figura 17 é possível identificar maior flexibilidade na função densidade de probabilidade do modelo BSNG, o que se confirma pelos critérios de informação e pelo teste da razão de verossimilhança, o qual na comparação entre os modelos BSNG e SNG com BSNGG a hipótese nula é rejeitada indicando o modelo completo o mais adequado. Com a distribuição acumulada empírica comparada com as fda dos modelos ajustados, apresentada na Figura 18, também é possível perceber um melhor ajuste da distribuição BSNGG.

Figura 18 - Estimativa da função acumulada da distribuição BSNGG e distribuição acumulada empírica comparadas com: (a) Estimativa da distribuição acumulada da distribuição BSNG. (b) Estimativa da distribuição acumulada da distribuição SNG.

Figura 19 - (a) Risco empírico para os dados de leucemia. (b) Função de risco das distribuições BSNGG, BSNG e SNG ajustadas para os dados de leucemia.

O gráfico de risco empírico dado pela Figura (19) mostra evidências de que a função de risco possui uma forma de banheira, logo a função de risco ajustada

da distribuição BSNGG se mostrou mais flexível comparada com as funções de risco das distribuições BSNG e SNG.

Considere agora que há o interesse de estudar possíveis covariáveis que possa estar relacionadas ao tempo de sobrevivência dos pacientes no estudo. Estas covariáveis observadas são: contagem de glóbulos brancos no sangue em milhares (X_1) e a presença de "Auer rods" (X_2).

Para a adição de novas covariáveis no modelo, é preciso utilizar métodos de seleção como stepwise, backward, forward entre outros. Neste caso como são apenas duas covariáveis, utilizaremos um teste para identificar se as mesmas estão interferindo no tempo até o evento de interesse. O teste de logrank foi utilizado, apresentado na Tabela 6, e verificou-se ao nível de 5% que as covariáveis X_1 e X_2 estão diretamente relacionadas ao tempo de sobrevivência dos pacientes.

Tabela 6 - Têste de logrank para as covariáveis X_1 e X_2 .

Variável	Estatística do têste	GL	p-valor
X_1	44,3619	27	0,0190
X_2	8,4487	1	0,0037

As novas estimativas dos parâmetros e critérios de seleção para o modelo LBSNGG e seus sub-modelos são apresentadas na Tabela 7.

Tabela 7 - Estimativas dos parâmetros, desvios padrões (em parênteses), p-valor [em chaves] e critérios de informação para o modelo LBSNGG e seus sub-modelos aplicados aos dados de leucemia

Modelo	β_0	β_1	β_2	σ	a	b	p	AIC	CAIC	BIC
LBSNGG	3,9460 (0,1250) [<0.0001]	-0,0045 (0,0014) [0.0035]	0,8199 (0,1111) [<0.0001]	0,1168 (0,0169)	0,1474 (0,0378)	0,0739 (0,0663)	0,8677 (0,1414)	298,4	302,9	308,9
LBSNG	4,0573 (0,0868) [<0.000]	-0,0044 (0,0010) [0.0002]	0,8235 (0,0858) [<0.0001]	0,0948 (0,0036)	0,0703 (0,0147)	0,2755 (0,1648)	- -	300,1	303,3	309,1
LSNG	3,3118 (0,3135) [<0.000]	-0,0056 (0,0044) [0.2143]	1,0569 (0,3516) [0.0050]	0,9456 (0,1437)	-	- -	- -	304,4	305,8	310,4

Observa-se pela Tabela 7 que o modelo LBSNGG apresenta os menores valores dos critérios AIC, CAIC e BIC, indicando o mesmo como sendo o modelo que melhor se ajustou aos dados. Também pela Tabela 7 que os coeficientes de regressão β_1 e β_2 , que representam as covariáveis X_1 contagem de glóbulos brancos e X_2 "Auer rods" respectivamente, foram significativos ao nível de 5% de significância, ou seja, as covariáveis X_1 e X_2 estão diretamente relacionadas ao tempo de vida dos pacientes com leucemia.

Figura 20 - Diferença entre as curvas de sobrevivência para os grupos “Auer rods” positivo e negativo com: (a) Contagem de glóbulos brancos = 100. (b) Contagem de glóbulos brancos = 0,75.

A Figura 20 apresenta as curvas de sobrevivência por extratos, ajustadas pelo modelo LBSNGG considerando os máximos de mínimos das covariáveis contagem de glóbulos brancos X_1 e “Auer rods” X_2 . Com esse gráfico é possível observar a diferença significativa entre os dois grupos de pacientes $X_2 = +$ e $X_2 = -$ considerando a covariável X_1 fixa.

4 CONSIDERAÇÕES FINAIS

4.1 Conclusão

A distribuição beta semi-normal generalizada (BSNG) é utilizada para modelagem do tempo de vida de sistemas. Propomos um novo modelo chamado beta semi-normal generalizada geométrica (BSNGG), cuja função taxa de falha pode estar crescendo, decrescendo, modal e em forma de banheira, que estende as distribuições BSNG, generalizada semi-normal (GSN) e semi-normal (SN), instituídas por Pescim *et al.* (2010) e Cooray & Ananda (2008) respectivamente.

A distribuição BSNGG é bastante flexível para análise de dados positivos, sua função densidade pode ser expressa como uma combinação linear de densidades da distribuição BSNG. Proporcionamos um tratamento matemático da distribuição incluindo expansões para a forma da densidade, momentos, função geradora, assimetria, curtose, desvios médios, função quantíl e curvas de Lorenz e Bonferroni. As estimativas dos parâmetros é abordada pelo método da máxima verossimilhança e para a construção de testes de hipóteses e intervalos de confiança é apresentado os estimadores de variabilidade dos parâmetros via matriz de informação observada. Além disso, definimos o modelo log-beta semi-normal generalizada geométrico (LBSNGG), útil quando se deseja trabalhar com covariáveis que possam interferir na variável resposta.

Algumas utilidades das distribuições BSNGG e LBGSNG são ilustradas com aplicações em conjuntos de dados reais assim como simulações, na qual se mostrou mais flexível para dados cujo o comportamento é assimétrico a direita. Por fim, podemos concluir que propomos um modelo bastante geral e flexível para tempos de vida, para a modelagem de dados do mundo real.

4.2 Estudos futuros

Dando continuidade ao trabalho desenvolvido durante o mestrado, os seguintes temas podem ser abordados:

- Desenvolver análise de diagnóstico para os modelos BSNGG e LBSNGG.
- Desenvolver uma biblioteca com os modelos propostos.
- Propor um modelo de longa-duração, considerando as distribuições apresentadas

neste trabalho.

- Estudar os resíduos para os modelos de regressão.

REFERÊNCIAS

- AARSET, M.V. How to identify bathtub hazard rate. **IEEE Transactions Reliability**, New York, v. 36, p. 106-108, 1987.
- ADAMIDIS K.; LOUKAS S. A lifetime distribution with decreasing failure rate. **Statistics and Probability Letters**, Ioannina, v. 39, p. 35-42, 1998.
- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, Boston, v. 6, p. 716-723, 1974.
- ALTMAN, D.G. Construction of age-related reference centiles using absolute residuals. **Statistical Methods in Laboratory Medicine** London, v. 12, p. 917-924, 1993
- BARRETO-SOUZA, W.; MORAIS, A.L. de; CORDEIRO, G.M. The Weibull-geometric distribution. **Journal of Statistical Computation and Simulation**, Blacksburg, v. 81, p. 645-657, 2011.
- BARRETO-SOUZA, W.; SANTOS, A.H.S.; CORDEIRO, G.M. The beta generalized exponential distribution. **Journal of Statistical Computation and Simulation**, Blacksburg, v. 81, p. 645-657, 2011.
- BIRNBAUM, Z.W.; SAUNDERS, S.C. A new family of life distributions. **Journal of Applied Probability**, Sheffield, v. 6, p. 319-327, 1969
- BLAND, J.M.; ALTMAN, D.G. Measuring agreement in method comparison studies. **Statistical Methods in Laboratory Medicine** London, v. 8, p. 135-160, 1999
- BOZDOGAN, H. Model selection and Akaike's Information Criterion (AIC): the general theory and its analytical extensions. **Psychometrika**, New York, v. 52, p. 345-370, 1987.
- CARRASCO, J.M.F.; ORTEGA, E.M.M.; CORDEIRO, G.M. A generalized modified Weibull distribution for lifetime modeling. **Computational Statistics and Data Analysis**, New York, v. 53, p. 450-462, 2008.
- CARRASCO, J.M.F.; ORTEGA, E.M.M.; PAULA, G.A. Log-modified Weibull regression models with censored data: sensitivity and residual analysis. **Computation Statistics and Data Analysis**, New York, v. 52, p. 4021-1039, 2008.
- COLOSIMO, E.A.; GIOLO, S.R. **Análise de sobrevivência aplicada**. São Paulo: Edgard Blücher, 2006. 369 p.

COLTON, T. **Statistica in Medicine**. Padova: Piccin Editore, v.10, p.132-142, 165-171, 1979.

COORAY, K.; ANANDA, M.M.A. A generalization of the half-normal distribution with applications to lifetime data. **Communication in Statistics - Theory and Methods**, New York, v. 37, p. 1323-1337, 2008.

CORDEIRO, G.M.; PESCI, R.R.; ORTEGA, E.M.M. The Kumaraswamy generalized half-normal distribution for skewed positive data. **Journal of Data Science**, London, v. 10, p. 195-224, 2012

CORDEIRO, G.M.; SILVA, G.O.; ORTEGA, E.M.M. The beta-Weibull geometric distribution **A Journal of Theoretical and Applied Statistics**, DOI:10.1080/02331888.2011.577897. 2011.

CORDEIRO, G.M.; CRISTINO, C.T.; HASHIOMTO, E.M.; ORTEGA, E.M.M. The beta generalized Rayleigh distribution with applications to lifetime data. **Statistical Papers**, London, v. 54, p. 133-161, 2013a.

CORDEIRO, G.M.; GOMES, A.E.; SILVA, C.Q.; ORTEGA, E.M.M. The beta exponentiated Weibull distribution, **Journal of Statistical Computation and Simulation**, New York, v. 83, p. 114-138, 2013b

CORDEIRO, G.M.; PESCI, R.R.; ORTEGA, E.M.M.; DEMÉTRIO, C.G.B. The beta generalized half-normal distribution: New properties. **Journal of Probability and Statistics**, 2013c. In press.

CORDEIRO, G.M.; BRAGA JUNIOR, A.C.R.; DEMÉTRIO, C.G.B.; ORTEGA, E.M.M.; PESCI, R.R. Some new results for the Kumaraswamy modified Weibull distribution. **Journal of Probability and Statistics**, 2013d. In press.

COX, D.R. Regression models and life-tables (with discussion). **Journal of the Royal Statistical Society: Series B, Methodological**, London, v. 34, n. 2, p. 187-220, 1972.

COX, D.R.; OAKES, D. **Analysis of survival data**. London: Chapman and Hall, 1984. 201 p.

EUGENE, N.; LEE, C.; FAMOYE, F. Beta-normal distribution and its applications. **Communication in Statistics - Theory and Methods**, New York, v. 31, p. 497-512, 2002.

EXTON, H. **Handbook of hypergeometric integrals: theory, applications, tables, computer programs**. New York: Halsted Press, 1978. 500 p.

FEIGL, P.; ZELEN, M. Estimation of exponential survival probabilities with concomitant information. **Biometrics**, Washington, v. 21, p. 826-837, 1965.

FISHER, R.A. On an absolute criterion for fitting frequency. **Messenger of Mathematics**, v. 41, p. 155-160, 1912.

GEHAN, E.A. A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. **Biometrika**, London, v. 52, p. 203-223, 1965.

GILCHRIST, W. Statistical modelling with quantile functions. New York: Chapman & Hall, 2000. 346 p.

GRADSHTEYN, I.S.; RYZHIK, I.M. **Table of integrals, series, and products**. San Diego: Academic Press, 2000. 17 p.

HAGER, H.W.; BAIN, L.J. Inferential procedures for the generalized gamma distribution. **Journal of the American Statistical Association**, New York, v. 65, p. 1601-1609, 1970.

HASHIMOTO, E.M.; ORTEGA, E.M.M.; CANCHO, V.G.; CORDEIRO, G.M. The log-exponentiated Weibull regression model for interval-censored data. **Computational Statistics and Data Analysis**, New York, v. 54, p. 1017-1035, 2010.

KALBFLEISCH, J.D.; PRENTICE, R.L. **The statistical analysis of failure time data**. 2nd ed. New York: John Wiley, 2002. 439 p.

KAPLAN, E.L.; MEIER, P. Non parametric estimation from incomplete observation. **Journal of the American Statistics Association**, London, v. 53, p. 457-481, 1959.

KENNEY, J.F.; KEEPING, E.S. **Mathematics of statistics**: 3rd ed. New Jersey: Van Nostrand, 1962. pt. 1, 348 p.

LAWLESS, J. F. **Statistical models and methods for lifetime data**. 2nd ed. New York: Wiley, 2003. 630 p.

MANTEL, N. Evaluation of survival data and two new rank order statistics arising in its consideration. **Cancer Chemotherapy Reports**, Bethesda, v. 50, p. 163-170, 1966.

MENDOZA, N.V.R. **A distribuição log-logística exponenciada geométrica: dupla ativação**. 2012. 100 p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2012.

MOORS, J.J.A. A quantile alternative for kurtosis. **Journal of the Royal Statistical Society**, London, v. 37, p. 25-32, 1975.

MUDHOLKAR, G.S; SRIVASTAVA, D.K.; KOLLIA, G.D. A generalization of the Weibull distribution with application to the analysis of survival data. **Journal of the American Statistical Association**, Boston, v. 91, n. 436, p. 1575-1583, 1996.

NADARAJAH, S.; GUPTA, A.K. The beta Fréchet distribution. **Far East Journal of Theoretical Statistics**, Allahabad, v. 15, p. 15-24, 2004.

NADARAJAH, S.; KOTZ, S. The beta exponential distribution. **Reliability Engineering and System Safety**, Barking, v. 91, p. 689-697, 2005.

NOCEDAL, J.; WRIGHT, S.J. **Numerical optimization**. 2nd ed. New York: Springer, 2006. 664 p.

ORTEGA, E. M. M.; CANCHO, V. G.; PAULA, G. A. Generalized log-gamma regression models with cure fraction. **Lifetime Data Analysis**, London, v. 15, p. 79-106, 2009.

PARANAÍBA, P.F.; ORTEGA, E.M.M; CORDEIRO, G.M; PESCIM, R.R; PASCOA, M.A.R. The Beta Burr XII Distribution with Applications to Lifetime Data. **Computational Statistics & Data Analysis**, London, v. 55, p. 1118-1136, 2011.

PESCIM, R.R.; DEMÉTRIO, C.G.B.; CORDEIRO, G.M.; ORTEGA, E.M.M.; URBANO, M.R. The beta generalized half-normal distribution. **Computational Statistics and Data Analysis**, London, v. 54, p. 945-957, 2010.

PESCIM, R.R.; ORTEGA, E.M.M.; CORDEIRO, G.M.; DEMÉTRIO, C.G.B. The log-beta generalized half-normal regression model. **Journal of Statistical Theory and Applications**, New York 2013. In press.

PRESS, W.H.; TEULOSKY, S.A.; VETERLING, W.T.; FLANNERY, B.P. **Numerical recipes in C: the art of scientific computing**. 2nded. London: Cambridge University Press, 1992. 630 p.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Disponível em: <<http://www.R-project.org>>. Acesso em: 15 mar. 2013.

SAS INSTITUTE. **SAS/STAT** user's guide: version 9. Cary, 2004. 5121 p.

SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics**, Philadelphia, v. 6, p. 461-464, 1978.

SEN, P.K.; SINGER, J.M. **Large sample methods in statistics: an introduction with applications**. New York: Chapman Hall, 1993. 382 p.

STACY, E.W. A generalization of the gamma distribution. **Annals of Mathematical Statistics**, Ann Arbor, v. 33, p. 1187-1192, 1962.

STEINBRECHER, G. **Taylor expansion for inverse error function around origin**: working paper. Craiova: University of Craiova, 2002.

SZKLO, M.; NIETO, F.J. **Epidemiology**: beyond the basics. Annapolis: Aspen Publishers, 2000. 500 p.

VUONG, Q. Likelihood ratio tests for model selection and non-nested hypotheses. **Econometrica**, Chicago, v. 57, p. 307-333, 1989.

WILKS, S.S. The large-sample distribution of the likelihood ratio for testing composite hypotheses. **The Annals of Mathematical Statistics**, New York, v. 9, p. 60-62, 1938.

APÊNDICES

APÊNDICE A

A.1 Derivadas de $u = \Phi \left[\left(\frac{x_i}{\theta} \right)^\alpha \right]$. para a distribuição BSNGG

$$\dot{u}_\alpha = \phi \left[\left(\frac{x_i}{\theta} \right)^\alpha \right] \left(\frac{x_i}{\theta} \right)^\alpha \ln \left(\frac{x_i}{\theta} \right),$$

$$\dot{u}_\theta = -\phi \left[\left(\frac{x_i}{\theta} \right)^\alpha \right] \left(\frac{x_i}{\theta} \right)^\alpha \left(\frac{\alpha}{\theta} \right),$$

$$\ddot{u}_{\alpha\alpha} = \left(\ln \left(\frac{x_i}{\theta} \right) \right)^2 \phi \left[\left(\frac{x_i}{\theta} \right)^\alpha \right] \left(- \left(\frac{x_i}{\theta} \right)^{3\alpha} + \left(\frac{x_i}{\theta} \right)^\alpha \right),$$

$$\ddot{u}_{\alpha\theta} = \left(\frac{1}{\theta} \right) \phi \left[\left(\frac{x_i}{\theta} \right)^\alpha \right] \left[\left(\frac{x_i}{\theta} \right)^{3\alpha} \alpha \ln \left(\frac{x_i}{\theta} \right) - \left(\frac{x_i}{\theta} \right)^\alpha \alpha \ln \left(\frac{x_i}{\theta} \right) - \left(\frac{x_i}{\theta} \right)^\alpha \right],$$

$$\ddot{u}_{\theta\theta} = \phi \left[\left(\frac{x_i}{\theta} \right)^\alpha \right] \left[-\alpha \left(\frac{x_i}{\theta} \right)^{3\alpha} + \alpha \left(\frac{x_i}{\theta} \right)^\alpha + \left(\frac{x_i}{\theta} \right)^\alpha \right] \frac{\alpha}{\theta^2},$$

em que $\phi \left[\left(\frac{x_i}{\theta} \right)^\alpha \right]$ representa a função densidade da distribuição normal padrão.

A.2 Derivadas parciais da função beta incompleta para a distribuição BSNGG

$$I = [I_{2u_i-1}(a, b)] = I_{2\Phi[(\frac{x_i}{\theta})^\alpha]-1}(a, b) = \frac{1}{B(a, b)} \int_0^{2\Phi[(\frac{x_i}{\theta})^\alpha]-1} w^{a-1} (1-w)^{b-1} dw,$$

$$\dot{I}_\alpha = 2(B(a, b))^{-1} \dot{u}_\alpha (2u-1)^{a-1} (2-2u)^{b-1},$$

$$\dot{I}_\theta = 2(B(a, b))^{-1} \dot{u}_\theta (2u-1)^{a-1} (2-2u)^{b-1},$$

$$\dot{I}_a = [\Psi(a+b) - \Psi(a)] I_{2u_i-1}(a, b) + (B(a, b))^{-1} \int_0^{2u-1} w^{a-1} \ln(w) (1-w)^{b-1} dw,$$

$$\dot{I}_b = [\Psi(a+b) - \Psi(b)] I_{2u_i-1}(a, b) + (B(a, b))^{-1} \int_0^{2u-1} w^{a-1} (1-w)^{b-1} \ln(1-w) dw,$$

$$\ddot{I}_{\alpha\alpha} = 2 \left\{ \frac{\ddot{u}_{\alpha\alpha} (2u-1)^{a-1} (2-2u)^{b-1}}{B(a, b)} + \frac{\dot{u}_\alpha}{B(a, b)} \left[\frac{2(2u-1)^{a-1}}{2u-1} (a-1) \dot{u}_\alpha (2-2u)^{b-1} - \frac{2(2u-1)^{a-1}}{2-2u} (2-2u)^{b-1} (b-1) \dot{u}_\alpha \right] \right\},$$

$$\ddot{I}_{\alpha\theta} = 2 \left\{ (B(a, b))^{-1} \ddot{u}_{\alpha\theta} (2u-1)^{a-1} (2-2u)^{b-1} + \dot{u}_\alpha \left[\frac{2(2u-1)^{a-1}}{2u-1} (a-1) \dot{u}_\theta (2-2u)^{b-1} - \frac{2(2u-1)^{a-1}}{2-2u} (2-2u)^{b-1} (b-1) \dot{u}_\theta \right] \right\},$$

$$\begin{aligned}
\ddot{I}_{\alpha\alpha} &= 2 \left[\frac{\Psi(a+b) - \Psi(a)}{B(a,b)} \dot{u}_\alpha (2u-1)^{a-1} (2-2u)^{b-1} + \frac{\dot{u}_\alpha (2u-1)^{a-1} \ln(2u-1) (2-2u)^{b-1}}{B(a,b)} \right], \\
\ddot{I}_{\alpha b} &= 2 \left[\frac{\Psi(a+b) - \Psi(b)}{B(a,b)} \dot{u}_\alpha (2u-1)^{a-1} (2-2u)^{b-1} + \frac{\dot{u}_\alpha (2u-1)^{a-1} (2-2u)^{b-1} \ln(2-2u)}{B(a,b)} \right], \\
\ddot{I}_{\theta\theta} &= 2 \left[\frac{\dot{u}_{\theta\theta} (2u-1)^{a-1} (2-2u)^{b-1}}{B(a,b)} + \frac{2\dot{u}_\theta}{B(a,b)} \left(\frac{(2u-1)^{a-1}}{2u-1} (a-1) \dot{u}_\theta (2-2u)^{b-1} \right. \right. \\
&\quad \left. \left. - \frac{(2u-1)^{a-1} (2-2u)^{b-1}}{2-2u} (b-1) \dot{u}_\theta \right) \right], \\
\ddot{I}_{\theta a} &= 2 \left[\frac{\Psi(a+b) - \Psi(a)}{B(a,b)} \dot{u}_\theta (2u-1)^{a-1} (2-2u)^{b-1} + \frac{\dot{u}_\theta (2u-1)^{a-1} \ln(2u-1) (2-2u)^{b-1}}{B(a,b)} \right], \\
\ddot{I}_{\theta b} &= 2 \left[\frac{\Psi(a+b) - \Psi(b)}{B(a,b)} \dot{u}_\theta (2u-1)^{a-1} (2-2u)^{b-1} + \frac{\dot{u}_\theta (2u-1)^{a-1} (2-2u)^{b-1} \ln(2-2u)}{B(a,b)} \right], \\
\ddot{I}_{aa} &= \left[(\Psi(a) - \Psi(a+b))^2 - \Psi(1,a) - \Psi(1,a+b) \right] I_{2u_i-1}(a,b) + 2 \frac{\Psi(a+b) - \Psi(a)}{B(a,b)} \\
&\quad \times \int_0^{2u-1} w^{a-1} \ln(w) (1-w)^{b-1} dw + \int_0^u \frac{w^{a-1} (\ln(w))^2 (1-w)^{b-1}}{B(a,b)} dw, \\
\ddot{I}_{ab} &= [(\Psi(a) - \Psi(a+b))(\Psi(b) - \Psi(a+b)) + \Psi(1,a+b)] I_{2u_i-1}(a,b) \\
&\quad + \left[\frac{\Psi(a+b) - \Psi(a)}{B(a,b)} \right] \int_0^{2u-1} w^{a-1} (1-w)^{b-1} \ln(1-w) dw \\
&\quad + \left[\frac{\Psi(a+b) - \Psi(b)}{B(a,b)} \right] \int_0^{2u-1} w^{a-1} \ln(w) (1-w)^{b-1} dw \\
&\quad + \int_0^{2u-1} \frac{w^{a-1} \ln(w) (1-w)^{b-1} \ln(1-w)}{\beta(a,b)} dw, \\
\ddot{I}_{bb} &= [(\Psi(b) - \Psi(a+b))^2 - \Psi(1,b) - \Psi(1,a+b)] I_{2u_i-1}(a,b) + 2 [\Psi(a+b) - \Psi(b)] \\
&\quad \times \int_0^{2u-1} w^{a-1} (1-w)^{b-1} \ln(1-w) dw + \int_0^{2u-1} \frac{w^{a-1} (1-w)^{b-1} (\ln(1-w))^2}{B(a,b)} dw.
\end{aligned}$$

A.3 Elementos da matriz informação observada $L(\hat{\lambda})$ da distribuição BSNGG

$$\begin{aligned}
L_{\alpha\alpha} &= \sum_{i=1}^r \left[-\alpha^{-2} - 2 \left(\frac{t_i}{\theta} \right)^{2\alpha} \left(\ln \left(\frac{t_i}{\theta} \right) \right)^2 + 2 \frac{(a-1) \ddot{u}_{\alpha\alpha}}{2u-1} - 4 \frac{(a-1) (\dot{u}_{\alpha})^2}{(2u-1)^2} \right. \\
&\quad \left. - \frac{(b-1) \ddot{u}_{\alpha\alpha}}{1-u} - \frac{(b-1) (\dot{u}_{\alpha})^2}{(1-u)^2} - \frac{2p \ddot{I}_{\alpha\alpha}}{1-p(1-I)} + \frac{2p^2 \dot{I}_{\alpha}^2}{(1-p(1-I))^2} \right] \\
&\quad + \sum_{i=r+1}^n \left[\frac{(-1+I)(1-p+pI) \ddot{I}_{\alpha\alpha}}{(-1+I)^2(1-p+pI)^2} - \frac{2(pI+\frac{1}{2}-p) \dot{I}_{\alpha}^2}{[(-1+I)(1-p+pI)]^2} \right], \\
L_{\alpha\theta} &= \sum_{i=1}^r \left[-\theta^{-1} + 2 \left(\frac{t_i}{\theta} \right)^{2\alpha} \alpha \ln \left(\frac{t_i}{\theta} \right) \theta^{-1} + \left(\frac{t_i}{\theta} \right)^{2\alpha} \theta^{-1} + 2 \frac{(a-1) \ddot{u}_{\alpha\theta}}{2u-1} \right. \\
&\quad \left. - 4 \frac{(a-1) \dot{u}_{\alpha} \dot{u}_{\theta}}{(2u-1)^2} - \frac{(b-1) \ddot{u}_{\alpha\theta}}{1-u} - \frac{(b-1) \dot{u}_{\alpha} \dot{u}_{\theta}}{(1-u)^2} - \frac{2p \ddot{I}_{\alpha\theta}}{1-p(1-i)} \right. \\
&\quad \left. + \frac{2p^2 \dot{I}_{\alpha} \dot{I}_{\theta}}{(1-p(1-i))^2} \right] + \sum_{i=r+1}^n \left[-\frac{\ddot{I}_{\alpha\theta}}{1-i} - \frac{\dot{I}_{\alpha} \dot{I}_{\theta}}{(1-i)^2} - \frac{p \ddot{I}_{\alpha\theta}}{1-p(1-i)} \right. \\
&\quad \left. + \frac{p^2 \dot{I}_{\alpha} \dot{I}_{\theta}}{(1-p(1-i))^2} \right], \\
L_{\alpha a} &= \sum_{i=1}^r \left[\frac{2\dot{u}_{\alpha}}{2u-1} - \frac{2p \ddot{I}_{\alpha a}}{1-p(1-i)} + \frac{2p^2 \dot{I}_{\alpha} \dot{I}_a}{(1-p(1-i))^2} \right] \\
&\quad + \sum_{i=r+1}^n \left[-\frac{\ddot{I}_{\alpha a}}{1-i} - \frac{\dot{I}_{\alpha} \dot{I}_a}{(1-i)^2} - \frac{p \ddot{I}_{\alpha a}}{1-p(1-i)} + \frac{p^2 \dot{I}_{\alpha} \dot{I}_a}{(1-p(1-i))^2} \right], \\
L_{\alpha b} &= \sum_{i=1}^r \left[-\frac{\dot{u}_{\alpha}}{1-u} - \frac{2p \ddot{I}_{\alpha b}}{1-p(1-i)} + \frac{2p^2 \dot{I}_{\alpha} \dot{I}_b}{(1-p(1-i))^2} \right] \\
&\quad + \sum_{i=r+1}^n \left[-\frac{\ddot{I}_{\alpha b}}{1-i} - \frac{\dot{I}_{\alpha} \dot{I}_b}{(1-i)^2} - \frac{p \ddot{I}_{\alpha b}}{1-p(1-i)} + \frac{p^2 \dot{I}_{\alpha} \dot{I}_b}{(1-p(1-i))^2} \right], \\
L_{\alpha p} &= \sum_{i=1}^r \left[-\frac{2\dot{I}_{\alpha}}{1-p(1-i)} + \frac{2p \dot{I}_{\alpha} (-1+i)}{(1-p(1-i))^2} \right] + \sum_{i=r+1}^n \left[\frac{-\dot{I}_{\alpha}}{1-p(1-i)} \right. \\
&\quad \left. + \frac{p \dot{I}_{\alpha} (-1+i)}{(1-p(1-i))^2} \right], \\
L_{\theta\theta} &= \sum_{i=1}^r \left[\frac{\alpha}{\theta^2} - 2 \frac{\alpha^2}{\theta^2} \left(\frac{t_i}{\theta} \right)^{2\alpha} - \left(\frac{t_i}{\theta} \right)^{2\alpha} \frac{\alpha}{\theta^2} + \frac{2(a-1) \ddot{u}_{\theta\theta}}{2u-1} - \frac{4(a-1) (\dot{u}_{\theta})^2}{(2u-1)^2} \right. \\
&\quad \left. - \frac{(b-1) \ddot{u}_{\theta\theta}}{1-u} - \frac{(b-1) (\dot{u}_{\theta})^2}{(1-u)^2} - \frac{2p \ddot{I}_{\theta\theta}}{1-p(1-i)} + \frac{2p^2 (\dot{I}_{\theta})^2}{(1-p(1-i))^2} \right] \\
&\quad + \sum_{i=r+1}^n \left[-\frac{\ddot{I}_{\theta\theta}}{1-i} - \frac{(\dot{I}_{\theta})^2}{(1-i)^2} - \frac{p \ddot{I}_{\theta\theta}}{1-p(1-i)} + \frac{p^2 (\dot{I}_{\theta})^2}{(1-p(1-i))^2} \right],
\end{aligned}$$

$$\begin{aligned}
L_{\theta a} &= \sum_{i=1}^r \left[\frac{2\dot{u}_\theta}{2u-1} - \frac{2p\ddot{I}_{\theta a}}{1-p(1-i)} + \frac{2p^2\dot{I}_\theta\dot{I}_a}{(1-p(1-i))^2} \right] \\
&\quad + \sum_{i=r+1}^n \left[-\frac{\ddot{I}_{\theta a}}{1-i} - \frac{\dot{I}_\theta\dot{I}_a}{(1-i)^2} - \frac{p\ddot{I}_{\theta a}}{1-p(1-i)} + \frac{p^2\dot{I}_\theta\dot{I}_a}{(1-p(1-i))^2} \right], \\
L_{\theta b} &= \sum_{i=1}^r \left[-\frac{\dot{u}_\theta}{1-u} - \frac{2p\ddot{I}_{\theta b}}{1-p(1-i)} + \frac{2p^2\dot{I}_\theta\dot{I}_b}{(1-p(1-i))^2} \right] \\
&\quad + \sum_{i=r+1}^n \left[-\frac{\ddot{I}_{\theta b}}{1-i} - \frac{\dot{I}_\theta\dot{I}_b}{(1-i)^2} - \frac{p\ddot{I}_{\theta b}}{1-p(1-i)} + \frac{p^2\dot{I}_\theta\dot{I}_b}{(1-p(1-i))^2} \right], \\
L_{\theta p} &= \sum_{i=1}^r \left[-\frac{2\dot{I}_\theta}{1-p(1-i)} + \frac{2p\dot{I}_\theta(-1+i)}{(1-p(1-i))^2} \right] \\
&\quad + \sum_{i=r+1}^n \left[-\frac{\dot{I}_\theta}{1-p(1-i)} + \frac{p\dot{I}_\theta(-1+i)}{(1-p(1-i))^2} \right], \\
L_{aa} &= r(\Psi(1, a+b) - \Psi(1, a)) + \left[\sum_{i=1}^r -\frac{2p\ddot{I}_{aa}}{1-p(1-i)} + \frac{2p^2(\dot{I}_a)^2}{(1-p(1-i))^2} \right] \\
&\quad + \sum_{i=r+1}^n \left[-\frac{\ddot{I}_{aa}}{1-i} - \frac{(\dot{I}_a)^2}{(1-i)^2} - \frac{p\ddot{I}_{aa}}{1-p(1-i)} + \frac{p^2(\dot{I}_a)^2}{(1-p(1-i))^2} \right], \\
L_{ab} &= r\Psi(1, a+b) + \sum_{i=1}^r \left[-\frac{2p\ddot{I}_{ab}}{1-p(1-i)} + \frac{2p^2\dot{I}_a\dot{I}_b}{(1-p(1-i))^2} \right] \\
&\quad + \sum_{i=r+1}^n \left[-\frac{\ddot{I}_{ab}}{1-i} - \frac{\dot{I}_a\dot{I}_b}{(1-i)^2} - \frac{p\ddot{I}_{ab}}{1-p(1-i)} + \frac{p^2\dot{I}_a\dot{I}_b}{(1-p(1-i))^2} \right], \\
L_{ap} &= \sum_{i=1}^r \left[-\frac{2\dot{I}_a}{1-p(1-i)} + \frac{2p\dot{I}_a(-1+i)}{(1-p(1-i))^2} \right] \\
&\quad + \sum_{i=r+1}^n \left[-\frac{\dot{I}_a}{1-p(1-i)} + \frac{p\dot{I}_a(-1+i)}{(1-p(1-i))^2} \right], \\
L_{bb} &= r(\Psi(1, a+b) - \Psi(1, b)) + \sum_{i=1}^r \left[-\frac{2p\ddot{I}_{bb}}{1-p(1-i)} + \frac{2p^2(\dot{I}_b)^2}{(1-p(1-i))^2} \right] \\
&\quad + \sum_{i=r+1}^n \left[-\frac{\ddot{I}_{bb}}{1-i} - \frac{(\dot{I}_b)^2}{(1-i)^2} - \frac{p\ddot{I}_{bb}}{1-p(1-i)} + \frac{p^2(\dot{I}_b)^2}{(1-p(1-i))^2} \right],
\end{aligned}$$

$$\begin{aligned}
L_{bp} &= \sum_{i=1}^r \left[-\frac{2\dot{I}_b}{1-p(1-i)} + \frac{2p\dot{I}_b(-1+i)}{(1-p(1-i))^2} \right] \\
&\quad + \sum_{i=r+1}^n \left[-\frac{\dot{I}_b}{1-p(1-i)} + \frac{p\dot{I}_b(-1+i)}{(1-p(1-i))^2} \right], \\
L_{pp} &= \frac{-n}{(1-p)^2} + \sum_{i=1}^r \left[\frac{2(-1+i)^2}{(1-p(1-i))^2} \right] + \sum_{i=r+1}^n \left[\frac{(-1+i)^2}{(1-p(1-i))^2} \right].
\end{aligned}$$

APÊNDICE B

B.1 Derivadas de $u_i = \Phi \left[e^{\frac{\sqrt{2}}{2} \left(\frac{y_i - \mu_i}{\sigma} \right)} \right]$ em que $\mu_i = \mathbf{x}_i^x \boldsymbol{\beta}$ para a distribuição LBSNGG

$$\begin{aligned}
\dot{u}_{\boldsymbol{\beta}_k} &= -\phi \left[e^{\frac{\sqrt{2}}{2} \left(\frac{y_i - \mathbf{x}_i^x \boldsymbol{\beta}}{\sigma} \right)} \right] \mathbf{x}_{ik} e^{\frac{\sqrt{2}(y - \mathbf{x}_i^t \boldsymbol{\beta})}{2\sigma}} \frac{1}{\sqrt{2}\sigma}, \\
\dot{u}_{\sigma} &= -\phi \left[e^{\frac{\sqrt{2}}{2} \left(\frac{y_i - \mathbf{x}_i^x \boldsymbol{\beta}}{\sigma} \right)} \right] (y - \mathbf{x}_i^t \boldsymbol{\beta}) e^{\frac{\sqrt{2}(y - \mathbf{x}_i^t \boldsymbol{\beta})}{2\sigma}} \frac{1}{\sigma^2 \sqrt{2}}, \\
\ddot{u}_{\boldsymbol{\beta}_k \boldsymbol{\beta}_j} &= -\phi \left[e^{\frac{\sqrt{2}}{2} \left(\frac{y_i - \mathbf{x}_i^x \boldsymbol{\beta}}{\sigma} \right)} \right] \frac{2}{4\sigma^2} \left(e^{\frac{3\sqrt{2}(y - \mathbf{x}_i^t \boldsymbol{\beta})}{2\sigma}} \right) \left(\frac{d}{d\boldsymbol{\beta}_k} \mathbf{x}_i^t \boldsymbol{\beta} \right)^2 \\
&\quad - \frac{1}{2} e^{-\frac{1}{2\sigma\sqrt{\pi}}} \left(e^{\frac{\sqrt{2}(y - \mathbf{x}_i^t \boldsymbol{\beta})}{\sigma}} \right) \left(\frac{d^2}{d\boldsymbol{\beta}_k \boldsymbol{\beta}_j} \mathbf{x}_i^t \boldsymbol{\beta} \right) e^{\frac{\sqrt{2}(y - \mathbf{x}_i^t \boldsymbol{\beta})}{2\sigma}} \\
&\quad + \phi \left[e^{\frac{\sqrt{2}}{2} \left(\frac{y_i - \mathbf{x}_i^x \boldsymbol{\beta}}{\sigma} \right)} \right] \frac{2}{4\sigma^2} \left(\frac{d}{d\boldsymbol{\beta}_k} \mathbf{x}_i^t \boldsymbol{\beta} \right)^2 e^{-\frac{1}{2} \frac{\sqrt{2}(-y + \mathbf{x}_i^t \boldsymbol{\beta})}{\sigma}} \\
\ddot{u}_{\sigma\sigma} &= -\phi \left[e^{\frac{\sqrt{2}}{2} \left(\frac{y_i - \mathbf{x}_i^x \boldsymbol{\beta}}{\sigma} \right)} \right] \frac{2}{4\sigma^4} \left(e^{\frac{3\sqrt{2}(y - \mathbf{x}_i^t \boldsymbol{\beta})}{2\sigma}} \right) (-y + \mathbf{x}_i^t \boldsymbol{\beta})^2 \\
&\quad - \phi \left[e^{\frac{\sqrt{2}}{2} \left(\frac{y_i - \mathbf{x}_i^x \boldsymbol{\beta}}{\sigma} \right)} \right] (-y + \mathbf{x}_i^t \boldsymbol{\beta}) e^{\frac{\sqrt{2}(y - \mathbf{x}_i^t \boldsymbol{\beta})}{2\sigma}} \frac{\sqrt{2}}{\sigma^3} \\
&\quad + \phi \left[e^{\frac{\sqrt{2}}{2} \left(\frac{y_i - \mathbf{x}_i^x \boldsymbol{\beta}}{\sigma} \right)} \right] (-y + \mathbf{x}_i^t \boldsymbol{\beta})^2 e^{\frac{\sqrt{2}(y - \mathbf{x}_i^t \boldsymbol{\beta})}{2\sigma}} \frac{2}{4\sigma^4}
\end{aligned}$$

$$\begin{aligned}
\ddot{u}_{\beta_k \sigma} = & \phi \left[e^{\frac{\sqrt{2}}{2} \left(\frac{y_i - \mathbf{x}_i^x \boldsymbol{\beta}}{\sigma} \right)} \right] \frac{2}{4\sigma^3} \left(e^{\frac{3\sqrt{2}(y - \mathbf{x}_i^t \boldsymbol{\beta})}{2\sigma}} \right) (-y + \mathbf{x}_{ik}) \frac{d}{d\beta_k} \mathbf{x}_i^t \boldsymbol{\beta} \\
& + \phi \left[e^{\frac{\sqrt{2}}{2} \left(\frac{y_i - \mathbf{x}_i^x \boldsymbol{\beta}}{\sigma} \right)} \right] \frac{1}{\sigma^2 \sqrt{2}} \left(\frac{d}{d\beta_k} \mathbf{x}_i^t \boldsymbol{\beta} (k) \right) e^{\frac{\sqrt{2}(y - \mathbf{x}_i^t \boldsymbol{\beta})}{2\sigma}} \\
& - \phi \left[e^{\frac{\sqrt{2}}{2} \left(\frac{y_i - \mathbf{x}_i^x \boldsymbol{\beta}}{\sigma} \right)} \right] \frac{2}{4\sigma^3} \left(\frac{d}{d\beta_k} \mathbf{x}_i^t \boldsymbol{\beta} \right) (-y + \mathbf{x}_i^t \boldsymbol{\beta}) e^{\frac{\sqrt{2}(y - \mathbf{x}_{ik})}{2\sigma}}
\end{aligned}$$

B.2 Derivadas parciais da função beta incompleta para a distribuição LBSNGG

$$I = [I_{2u_i-1}(a, b)] = \frac{1}{B(a, b)} \int_0^{2\Phi \left[e^{\frac{\sqrt{2}}{2} \left(\frac{y_i - \mu_i}{\sigma} \right)} \right] - 1} w^{a-1} (1-w)^{b-1} dw,$$

$$\dot{I}_\sigma = 2(B(a, b))^{-1} \dot{u}_\sigma (2u-1)^{a-1} (2-2u)^{b-1},$$

$$\dot{I}_{\beta_k} = 2(B(a, b))^{-1} \dot{u}_{\beta_k} (2u-1)^{a-1} (2-2u)^{b-1},$$

$$\dot{I}_a = [\Psi(a+b) - \Psi(a)] I_{2u_i-1}(a, b) + (B(a, b))^{-1} \int_0^{2u-1} w^{a-1} \ln(w) (1-w)^{b-1} dw,$$

$$\dot{I}_b = [\Psi(a+b) - \Psi(b)] I_{2u_i-1}(a, b) + (B(a, b))^{-1} \int_0^{2u-1} w^{a-1} (1-w)^{b-1} \ln(1-w) dw,$$

$$\begin{aligned}
\ddot{I}_{\sigma\sigma} = & 2 \left\{ \frac{\ddot{u}_{\sigma\sigma} (2u-1)^{a-1} (2-2u)^{b-1}}{B(a, b)} + \frac{\dot{u}_\sigma}{B(a, b)} \left[\frac{2(2u-1)^{a-1}}{2u-1} (a-1) \dot{u}_\sigma (2-2u)^{b-1} \right. \right. \\
& \left. \left. - \frac{2(2u-1)^{a-1}}{2-2u} (2-2u)^{b-1} (b-1) \dot{u}_\sigma \right] \right\},
\end{aligned}$$

$$\begin{aligned}
\ddot{I}_{\sigma\beta_k} = & 2 \left\{ (B(a, b))^{-1} \ddot{u}_{\sigma\beta_k} (2u-1)^{a-1} (2-2u)^{b-1} + \dot{u}_\sigma \left[\frac{2(2u-1)^{a-1}}{2u-1} (a-1) \dot{u}_{\beta_k} (2-2u)^{b-1} \right. \right. \\
& \left. \left. - \frac{2(2u-1)^{a-1}}{2-2u} (2-2u)^{b-1} (b-1) \dot{u}_{\beta_k} \right] \right\},
\end{aligned}$$

$$\begin{aligned}
\ddot{I}_{\sigma a} &= 2 \left[\frac{\Psi(a+b) - \Psi(a)}{B(a,b)} \dot{u}_\sigma (2u-1)^{a-1} (2-2u)^{b-1} + \frac{\dot{u}_\sigma (2u-1)^{a-1} (2-2u)^{b-1}}{\ln(2u-1)^{-1} B(a,b)} \right], \\
\ddot{I}_{\sigma b} &= 2 \left[\frac{\Psi(a+b) - \Psi(b)}{B(a,b)} \dot{u}_\sigma (2u-1)^{a-1} (2-2u)^{b-1} + \frac{\dot{u}_\sigma (2u-1)^{a-1} (2-2u)^{b-1}}{\ln(2-2u)^{-1} B(a,b)} \right], \\
\ddot{I}_{\beta_k \beta_j} &= 2 \left[\frac{\ddot{\beta}_k \beta_j (2u-1)^{a-1} (2-2u)^{b-1}}{B(a,b)} + \frac{2\dot{\beta}_k}{B(a,b)} \left(\frac{(2u-1)^{a-1}}{2u-1} (a-1) \dot{\beta}_k (2-2u)^{b-1} \right. \right. \\
&\quad \left. \left. - \frac{(2u-1)^{a-1} (2-2u)^{b-1}}{2-2u} (b-1) \dot{\beta}_k \right) \right], \\
\ddot{I}_{\beta_k a} &= 2 \left[\frac{\Psi(a+b) - \Psi(a)}{B(a,b)} \dot{\beta}_k (2u-1)^{a-1} (2-2u)^{b-1} + \frac{\dot{\beta}_k (2u-1)^{a-1} (2-2u)^{b-1}}{\ln(2u-1)^{-1} B(a,b)} \right], \\
\ddot{I}_{\beta_k b} &= 2 \left[\frac{\Psi(a+b) - \Psi(b)}{B(a,b)} \dot{\beta}_k (2u-1)^{a-1} (2-2u)^{b-1} + \frac{\dot{\beta}_k (2u-1)^{a-1} (2-2u)^{b-1}}{\ln(2-2u)^{-1} B(a,b)} \right], \\
\ddot{I}_{aa} &= \left[(\Psi(a) - \Psi(a+b))^2 - \Psi(1,a) - \Psi(1,a+b) \right] I_{2u_i-1}(a,b) \\
&\quad - 2 \frac{\Psi(a) - \Psi(a+b)}{B(a,b)} \int_0^{2u-1} w^{a-1} \ln(w) (1-w)^{b-1} dw \\
&\quad + \int_0^u \frac{w^{a-1} (\ln(w))^2 (1-w)^{b-1}}{B(a,b)} dw, \\
\ddot{I}_{ab} &= [(\Psi(a) - \Psi(a+b))(\Psi(b) - \Psi(a+b)) + \Psi(1,a+b)] I_{2u_i-1}(a,b) \\
&\quad + \left[\frac{\Psi(a+b) - \Psi(a)}{B(a,b)} \right] \int_0^{2u-1} w^{a-1} (1-w)^{b-1} \ln(1-w) dw \\
&\quad + \left[\frac{+\Psi(a+b) - \Psi(b)}{B(a,b)} \right] \int_0^{2u-1} w^{a-1} \ln(w) (1-w)^{b-1} dw \\
&\quad + \int_0^{2u-1} \frac{w^{a-1} \ln(w) (1-w)^{b-1} \ln(1-w)}{B(a,b)} dw, \\
\ddot{I}_{bb} &= \left[(\Psi(b) - \Psi(a+b))^2 - \Psi(1,b) - \Psi(1,a+b) \right] I_{2u_i-1}(a,b) \\
&\quad - 2 \frac{\Psi(b) - \Psi(a+b)}{B(a,b)} \int_0^{2u-1} w^{a-1} (1-w)^{b-1} \ln(1-w) dw \\
&\quad + \int_0^{2u-1} \frac{w^{a-1} (1-w)^{b-1} (\ln(1-w))^2}{B(a,b)} dw.
\end{aligned}$$

B.3 Elementos do vetor escore U para a distribuição LBSNGG

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\lambda})}{\partial \beta_k} &= \sum_{i=1}^r \left[-\frac{\sqrt{2} x_i^t \frac{d}{d\beta_k} \boldsymbol{\beta}}{2\sigma} + \frac{\sqrt{2} \left(x_i^t \frac{d}{d\beta_k} \boldsymbol{\beta} \right) e^{\frac{\sqrt{2}(y-x_i^t \boldsymbol{\beta})}{\sigma}}}{2\sigma} + \frac{2(a-1) \dot{u}_{\beta_k}}{2u-1} \right. \\
&\quad \left. - \frac{(b-1) \dot{u}_{\beta_k}}{1-u} - \frac{2p \dot{I}_{\beta_k}}{1-p(1-I)} \right] + \sum_{i=r+1}^n \left[-\frac{\dot{I}_{\beta_k}}{1-I} - \frac{p \dot{I}_{\beta_k}}{1-p(1-I)} \right], \\
\frac{\partial l(\boldsymbol{\lambda})}{\partial \sigma} &= \sum_{i=1}^r \left[-\frac{\sqrt{2} (y-x_i^t \boldsymbol{\beta})}{2\sigma^2} + \frac{\sqrt{2} (y-x_i^t \boldsymbol{\beta}) e^{\frac{\sqrt{2}(y-x_i^t \boldsymbol{\beta})}{\sigma}}}{2\sigma^2} + \frac{2(a-1) \dot{u}_{\sigma}}{2u-1} \right. \\
&\quad \left. - \frac{(b-1) \dot{u}_{\sigma}}{1-u} - \frac{2p \dot{I}_{\sigma}}{1-p(1-I)} \right] + \sum_{i=r+1}^n \left[-\frac{\dot{I}_{\sigma}}{1-I} - \frac{p \dot{I}_{\sigma}}{1-p(1-I)} \right], \\
\frac{\partial l(\boldsymbol{\lambda})}{\partial a} &= r [\Psi(a+b) - \Psi(a)] + \sum_{i=1}^r \left[\ln(2u-1) - \frac{2p \dot{I}_a}{1-p(1-I)} \right] - \sum_{i=r+1}^n \left[\frac{\dot{I}_a}{1-I} + \frac{p \dot{I}_a}{1-p(1-I)} \right], \\
\frac{\partial l(\boldsymbol{\lambda})}{\partial b} &= r [\Psi(a+b) - \Psi(b)] + r \ln(2) + \sum_{i=1}^r \left[\ln(1-u) - \frac{2p \dot{I}_b}{1-p(1-I)} \right] \\
&\quad - \sum_{i=r+1}^n \left[\frac{\dot{I}_b}{1-I} + \frac{p \dot{I}_b}{1-p(1-I)} \right], \\
\frac{\partial l(\boldsymbol{\lambda})}{\partial p} &= \frac{-n}{1-p} - 2 \sum_{i=1}^r \left[\frac{-1+I}{1-p(1-I)} \right] - \sum_{i=r+1}^n \left[\frac{-1+I}{1-p(1-I)} \right].
\end{aligned}$$

B.4 Elementos da matriz informação $L(\hat{\boldsymbol{\lambda}})$ para a distribuição LBSNGG

$$\begin{aligned}
L_{\beta_k \beta_j} &= \sum_{i=1}^r \left[-\frac{\sqrt{2} x_i^t \frac{d^2}{d\beta_k d\beta_j} \boldsymbol{\beta}}{2\sigma} + \frac{\sqrt{2} x_i^t \frac{d^2}{d\beta_k d\beta_j} \boldsymbol{\beta} e^{\frac{\sqrt{2}(y-x_i^t \boldsymbol{\beta})}{\sigma}}}{2\sigma} - \frac{(x_i^t)^2 \left(\frac{d}{d\beta_k} \boldsymbol{\beta} \right)^2 e^{\frac{\sqrt{2}(y-x_i^t \boldsymbol{\beta})}{\sigma}}}{\sigma^2} \right. \\
&\quad + \frac{2(a-1) \ddot{u}_{\beta_k \beta_j}}{2u-1} - \frac{4(a-1) \left(\dot{u}_{\beta_k} \right)^2}{(2u-1)^2} - \frac{(b-1) \ddot{u}_{\beta_k \beta_j}}{1-u} - \frac{(b-1) \left(\dot{u}_{\beta_k} \right)^2}{(1-u)^2} - \frac{2p \ddot{I}_{\beta_k \beta_j}}{1-p(1-i)} \\
&\quad \left. + \frac{2p^2 \left(\dot{I}_{\beta_k} \right)^2}{(1-p(1-i))^2} \right] + \sum_{i=r+1}^n \left[-\frac{\ddot{I}_{\beta_k \beta_j}}{1-i} - \frac{\left(\dot{I}_{\beta_k} \right)^2}{(1-i)^2} - \frac{p \ddot{I}_{\beta_k \beta_j}}{1-p(1-i)} + \frac{p^2 \left(\dot{I}_{\beta_k} \right)^2}{(1-p(1-i))^2} \right],
\end{aligned}$$

$$\begin{aligned}
L_{\beta_k \sigma} &= \sum_{i=1}^r \left[\frac{\sqrt{2} x_i^t \frac{d}{d\beta_k} \beta}{2\sigma^2} - \frac{\sqrt{2} x_i^t \frac{d}{d\beta_k} \beta e^{\frac{\sqrt{2}(y-x_i^t \beta)}{\sigma}}}{2\sigma^2} - \frac{x_i^t \frac{d}{d\beta_k} \beta (y-x_i^t \beta) e^{\frac{\sqrt{2}(y-x_i^t \beta)}{\sigma}}}{\sigma^3} \right. \\
&\quad + \frac{2(a-1) \ddot{\beta}_{k\sigma}}{2u-1} - \frac{4(a-1) \dot{\beta}_k \dot{\beta}_\sigma}{(2u-1)^2} - \frac{(b-1) \ddot{\beta}_{k\sigma}}{1-u} - \frac{(b-1) \dot{\beta}_k \dot{\beta}_\sigma}{(1-u)^2} - \frac{2p \ddot{\beta}_{k\sigma}}{1-p(1-i)} \\
&\quad \left. + \frac{2p^2 \dot{\beta}_k \dot{\beta}_\sigma}{(1-p(1-i))^2} \right] + \sum_{i=r+1}^n \left[-\frac{\ddot{\beta}_{k\sigma}}{1-i} - \frac{\dot{\beta}_k \dot{\beta}_\sigma}{(1-i)^2} - \frac{p \ddot{\beta}_{k\sigma}}{1-p(1-i)} + \frac{p^2 \dot{\beta}_k \dot{\beta}_\sigma}{(1-p(1-i))^2} \right], \\
L_{\beta_k a} &= 2 \sum_{i=1}^r \left[\frac{\dot{\beta}_k}{2u-1} - \frac{p \ddot{\beta}_{ka}}{1-p(1-i)} + \frac{p^2 \dot{\beta}_k \dot{I}_a}{(1-p(1-i))^2} \right] + \sum_{i=r+1}^n \left[-\frac{\ddot{\beta}_{ka}}{1-i} - \frac{\dot{\beta}_k \dot{I}_a}{(1-i)^2} \right. \\
&\quad \left. - \frac{p \ddot{\beta}_{ka}}{1-p(1-i)} + \frac{p^2 \dot{\beta}_k \dot{I}_a}{(1-p(1-i))^2} \right], \\
L_{\beta_k b} &= \sum_{i=1}^r \left[-\frac{\dot{\beta}_k}{1-u} - \frac{2p \ddot{\beta}_{kb}}{1-p(1-i)} + \frac{2p^2 \dot{\beta}_k \dot{I}_b}{(1-p(1-i))^2} \right] + \sum_{i=r+1}^n \left[-\frac{\ddot{\beta}_{kb}}{1-i} - \frac{\dot{\beta}_k \dot{I}_b}{(1-i)^2} \right. \\
&\quad \left. - \frac{p \ddot{\beta}_{kb}}{1-p(1-i)} + \frac{p^2 \dot{\beta}_k \dot{I}_b}{(1-p(1-i))^2} \right], \\
L_{\beta_k p} &= \sum_{i=1}^r \left[-\frac{2\dot{\beta}_k}{1-p(1-i)} + \frac{2p \dot{\beta}_k (-1+i)}{(1-p(1-i))^2} \right] + \sum_{i=r+1}^n \left[-\frac{\dot{\beta}_k}{1-p(1-i)} + \frac{p \dot{\beta}_k (-1+i)}{(1-p(1-i))^2} \right], \\
L_{\sigma \sigma} &= \sum_{i=1}^r \left[\frac{\sqrt{2}(y-x_i^t \beta)}{\sigma^3} - \frac{\sqrt{2}(y-x_i^t \beta) e^{\frac{\sqrt{2}(y-x_i^t \beta)}{\sigma}}}{\sigma^3} - \frac{(y-x_i^t \beta)^2}{\sigma^4} e^{\frac{\sqrt{2}(y-x_i^t \beta)}{\sigma}} \right. \\
&\quad + 2 \frac{(a-1) \ddot{\sigma}}{2u-1} - 4 \frac{(a-1) (\dot{\sigma})^2}{(2u-1)^2} - \frac{(b-1) \ddot{\sigma}}{1-u} - \frac{(b-1) (\dot{\sigma})^2}{(1-u)^2} - \frac{2p \ddot{\sigma}}{1-p(1-i)} \\
&\quad \left. + 2 \frac{p^2 (\dot{\sigma})^2}{(1-p(1-i))^2} \right] + \sum_{i=r+1}^n \left[-\frac{\ddot{\sigma}}{1-i} - \frac{(\dot{\sigma})^2}{(1-i)^2} - \frac{p \ddot{\sigma}}{1-p(1-i)} + \frac{p^2 (\dot{\sigma})^2}{(1-p(1-i))^2} \right], \\
L_{\sigma a} &= \sum_{i=1}^r \left[\frac{2\dot{\sigma}}{2u-1} - \frac{2p \ddot{\sigma} a}{1-p(1-i)} + \frac{2p^2 \dot{\sigma} \dot{I}_a}{(1-p(1-i))^2} \right] + \sum_{i=r+1}^n \left[-\frac{\ddot{\sigma} a}{1-i} - \frac{\dot{\sigma} \dot{I}_a}{(1-i)^2} \right. \\
&\quad \left. - \frac{p \ddot{\sigma} a}{1-p(1-i)} + \frac{p^2 \dot{\sigma} \dot{I}_a}{(1-p(1-i))^2} \right], \\
L_{\sigma b} &= \sum_{i=1}^r \left[-\frac{\dot{\sigma}}{1-u} - \frac{2p \ddot{\sigma} b}{1-p(1-i)} + \frac{2p^2 \dot{\sigma} \dot{I}_b}{(1-p(1-i))^2} \right] + \sum_{i=r+1}^n \left[-\frac{\ddot{\sigma} b}{1-i} - \frac{\dot{\sigma} \dot{I}_b}{(1-i)^2} \right. \\
&\quad \left. - \frac{p \ddot{\sigma} b}{1-p(1-i)} + \frac{p^2 \dot{\sigma} \dot{I}_b}{(1-p(1-i))^2} \right],
\end{aligned}$$

$$\begin{aligned}
L_{\sigma p} &= \sum_{i=1}^r \left[-\frac{2\dot{I}_\sigma}{1-p(1-i)} + \frac{2p\dot{I}_\sigma(-1+i)}{(1-p(1-i))^2} \right] + \sum_{i=r+1}^n \left[-\frac{\dot{I}_\sigma}{1-p(1-i)} + \frac{p\dot{I}_\sigma(-1+i)}{(1-p(1-i))^2} \right], \\
L_{aa} &= -r[\Psi(1, a) - \Psi(1, a+b)] + \sum_{i=1}^r \left[-\frac{2p\ddot{I}_{aa}}{1-p(1-i)} + \frac{2p^2(\dot{I}_a)^2}{(1-p(1-i))^2} \right] \\
&\quad + \sum_{i=r+1}^n \left[-\frac{\ddot{I}_{aa}}{1-i} - \frac{(\dot{I}_a)^2}{(1-i)^2} - \frac{p\ddot{I}_{aa}}{1-p(1-i)} + \frac{p^2(\dot{I}_a)^2}{(1-p(1-i))^2} \right], \\
L_{ab} &= r\Psi(1, a+b) + \sum_{i=1}^r \left[-\frac{2p\ddot{I}_{ab}}{1-p(1-i)} + \frac{2p^2\dot{I}_a\dot{I}_b}{(1-p(1-i))^2} \right] + \sum_{i=r+1}^n \left[-\frac{\ddot{I}_{ab}}{1-i} - \frac{\dot{I}_a\dot{I}_b}{(1-i)^2} \right. \\
&\quad \left. - \frac{p\ddot{I}_{ab}}{1-p(1-i)} + \frac{p^2\dot{I}_a\dot{I}_b}{(1-p(1-i))^2} \right], \\
L_{ap} &= \sum_{i=1}^r \left[-\frac{2\dot{I}_a}{1-p(1-i)} + \frac{2p\dot{I}_a(-1+i)}{(1-p(1-i))^2} \right] + \sum_{i=r+1}^n \left[-\frac{\dot{I}_a}{1-p(1-i)} + \frac{p\dot{I}_a(-1+i)}{(1-p(1-i))^2} \right], \\
L_{bb} &= -r[\Psi(1, b) - \Psi(1, a+b)] + \sum_{i=1}^r \left[-\frac{2p\ddot{I}_{bb}}{1-p(1-i)} + \frac{2p^2(\dot{I}_b)^2}{(1-p(1-i))^2} \right] \\
&\quad + \sum_{i=r+1}^n \left[-\frac{\ddot{I}_{bb}}{1-i} - \frac{(\dot{I}_b)^2}{(1-i)^2} - \frac{p\ddot{I}_{bb}}{1-p(1-i)} + \frac{p^2(\dot{I}_b)^2}{(1-p(1-i))^2} \right], \\
L_{bp} &= \sum_{i=1}^r \left[-\frac{2\dot{I}_b}{1-p(1-i)} + \frac{2p\dot{I}_b(-1+i)}{(1-p(1-i))^2} \right] + \sum_{i=r+1}^n \left[-\frac{\dot{I}_b}{1-p(1-i)} + \frac{p\dot{I}_b(-1+i)}{(1-p(1-i))^2} \right], \\
L_{pp} &= -\frac{n}{(1-p)^2} + \sum_{i=1}^r \left[\frac{2(-1+i)^2}{(1-p(1-i))^2} \right] + \sum_{i=r+1}^n \left[\frac{(-1+i)^2}{(1-p(1-i))^2} \right].
\end{aligned}$$

D.1 Sintaxe do *software* R para apresentação de formas da distribuição BSNGG

```
#####Distribuição Generalizada Half-Normal.
g<-function(x,alpha,theta)
{sqrt(2/pi)*(alpha/x)*(x/theta)^alpha*exp((-1/2)*(x/theta)^(2*alpha))}
G<-function(x,alpha,theta){2*pnorm((x/theta)^alpha)-1}
###Distribuição beta generalizada. Half-Normal
f<-function(x,alpha,theta,a,b)
{(g(x,alpha,theta)*(G(x,alpha,theta))^(a-1)*(1-G(x,alpha,theta))^(b-1))/(beta(a,b))}
F<-function(x,alpha,theta,a,b){(pbeta(G(x,alpha,theta),a,b))}
###Distribuição Geometrica Beta Generalizada Half-Normal.
n<-function(x,alpha,theta,a,b,p)
{(f(x,alpha,theta,a,b)*(1-p))/(1-p*(1-F(x,alpha,theta,a,b)))^2}
N<-function(x,alpha,theta,a,b,p){
(F(x,alpha,theta,a,b))/(1-p*(1-F(x,alpha,theta,a,b)))}
S<-function(x,alpha,theta,a,b,p){1-N(x,alpha,theta,a,b,p)}
h<-function(x,alpha,theta,a,b,p){n(x,alpha,theta,a,b,p)/S(x,alpha,theta,a,b,p)}

# Plots
v<-seq(0,150,0.005)

#Densidades
plot(v,n(v,0.5,1.5,5,3,.1),lwd=2,lty=1,ylim=c(0,2),
      typ="l",main="",xlab="x",ylab="f(x)",xlim=c(0,3),col="4")
lines(v,n(v,1.5,1.5,4,3,.3),lty=1,lwd=2,col="2")
lines(v,n(v,2.5,1.5,3,3,.5),lty=1,lwd=2,col="7")
lines(v,n(v,3.5,1.5,3,3,.7),lty=1,lwd=2,col="1")
lines(v,n(v,4.5,1.5,1,3,.9),lty=1,lwd=2,col="3")
legend(1.5,2,border="red",expression(
  paste(alpha,"=0,5;",a,"=5,0;",p,"=0,1"),
  paste(alpha,"=1,5;",a,"=4,0;",p,"=0,3"),
  paste(alpha,"=2,5;",a,"=3,0;",p,"=0,5"),
  paste(alpha,"=3,5;",a,"=3,0;",p,"=0,7"),
  paste(alpha,"=4,5;",a,"=1,0;",p,"=0,9")
),lty=c(1,1,1,1,1),col=c("4","2","7","1","3")
  ,bty="n",y.intersp=.6,seg.len=1)

plot(v,n(v,2.5,5.5,5,.9,.1),lwd=2,lty=1,ylim=c(0,1.5),
      typ="l",main="",xlab="x",ylab="f(x)",xlim=c(0,10),col="4")
lines(v,n(v,2.5,4.5,5,.7,.3),lty=1,lwd=2,col="2")
lines(v,n(v,2.5,3.5,5,.5,.5),lty=1,lwd=2,col="7")
lines(v,n(v,2.5,2.5,5,.3,.7),lty=1,lwd=2,col="1")
lines(v,n(v,2.5,1.5,5,.1,.9),lty=1,lwd=2,col="3")
legend(4.5,1.5,border="red",expression(
  paste(theta,"=5,5;",b,"=0,9;",p,"=0,1"),
  paste(theta,"=4,5;",b,"=0,7;",p,"=0,3"),
  paste(theta,"=3,5;",b,"=0,5;",p,"=0,5"),
  paste(theta,"=2,5;",b,"=0,3;",p,"=0,7"),
  paste(theta,"=1,5;",b,"=0,1;",p,"=0,9")
),lty=c(1,1,1,1,1),col=c("4","2","7","1","3")
  ,bty="n",y.intersp=.6,seg.len=1)
```



```

plot(v,n(v,1.5,5.5,1.5,1.5,.1),lwd=2,lty=1,ylim=c(0,1.5),
     typ="l",main="",xlab="x",ylab="f(x)",xlim=c(0,7),col="4")
lines(v,n(v,1.5,4.5,2.5,1.5,.3),lty=1,lwd=2,col="2")
lines(v,n(v,1.5,3.5,3.5,1.5,.5),lty=1,lwd=2,col="7")
lines(v,n(v,1.5,2.5,4.5,1.5,.7),lty=1,lwd=2,col="1")
lines(v,n(v,1.5,1.5,5.5,1.5,.9),lty=1,lwd=2,col="3")
legend(3,1.5,border="red",expression(
  paste(theta,"=5,5;" ,b,"=0,9;" ,p,"=0,1") ,
  paste(theta,"=4,5;" ,b,"=0,7;" ,p,"=0,3") ,
  paste(theta,"=3,5;" ,b,"=0,5;" ,p,"=0,5") ,
  paste(theta,"=2,5;" ,b,"=0,3;" ,p,"=0,7") ,
  paste(theta,"=1,5;" ,b,"=0,1;" ,p,"=0,9")
),lty=c(1,1,1,1,1),col=c("4","2","7","1","3")
, bty="n",y.intersp=.6,seg.len=1)

```

#Risco

```

plot(v,h(v,1.5,55,.1,2,.9),lwd=2,lty=1,ylim=c(.03,.15),
     typ="l",main="",xlab="x",ylab="h(x)",xlim=c(0,60),col="4")
lines(v,h(v,1.5,55,.2,2,.7),lty=1,lwd=2,col="2")
lines(v,h(v,1.5,55,.3,2,.5),lty=1,lwd=2,col="7")
lines(v,h(v,1.5,55,.4,2,.3),lty=1,lwd=2,col="1")
lines(v,h(v,1.5,55,.5,2,.1),lty=1,lwd=2,col="3")
legend(12,.14,border="red",expression(
  paste(a,"=0,1;" ,p,"=0,9") ,
  paste(a,"=0,2;" ,p,"=0,7") ,
  paste(a,"=0,3;" ,p,"=0,5") ,
  paste(a,"=0,4;" ,p,"=0,3") ,
  paste(a,"=0,5;" ,p,"=0,1")
),lty=c(1,1,1,1,1),col=c("4","2","7","1","3")
, bty="n",y.intersp=.6,seg.len=1)

```

```

plot(v,h(v,0.5,5.5,15,.9,.95),lwd=2,lty=1,ylim=c(.0,.23),
     typ="l",main="",xlab="x",ylab="h(x)",xlim=c(0,60),col="4")
lines(v,h(v,0.5,5.5,15,.7,.90),lty=1,lwd=2,col="2")
lines(v,h(v,0.5,5.5,15,.5,.85),lty=1,lwd=2,col="7")
lines(v,h(v,0.5,5.5,15,.3,.80),lty=1,lwd=2,col="1")
lines(v,h(v,0.5,5.5,15,.1,.75),lty=1,lwd=2,col="3")
legend(25,.22,border="red",expression(
  paste(b,"=0,9;" ,p,"=0,95") ,
  paste(b,"=0,7;" ,p,"=0,90") ,
  paste(b,"=0,5;" ,p,"=0,85") ,
  paste(b,"=0,3;" ,p,"=0,80") ,
  paste(b,"=0,1;" ,p,"=0,75")
),lty=c(1,1,1,1,1),col=c("4","2","7","1","3")
, bty="n",y.intersp=.6,seg.len=1)

```

```

plot(v,h(v,1,60,1.5,2,.95),lwd=2,lty=1,ylim=c(.0,.15),
     typ="l",main="",xlab="x",ylab="h(x)",xlim=c(0,120),col="4")
lines(v,h(v,1,58,1.5,2,.90),lty=1,lwd=2,col="2")
lines(v,h(v,1,56,1.5,2,.85),lty=1,lwd=2,col="7")
lines(v,h(v,1,54,1.5,2,.80),lty=1,lwd=2,col="1")
lines(v,h(v,1,52,1.5,2,.75),lty=1,lwd=2,col="3")
legend(40,.13,border="red",expression(
  paste(theta,"=60;",p,"=0,95"),
  paste(theta,"=58;",p,"=0,90"),
  paste(theta,"=56;",p,"=0,85"),
  paste(theta,"=54;",p,"=0,80"),
  paste(theta,"=52;",p,"=0,75")
),lty=c(1,1,1,1,1),col=c("4","2","7","1","3")
, bty="n",y.intersp=.6,seg.len=1)

```

D.2 Sintaxe do *software* R para assimetria e curtose da distribuição BSNGG

```

vs <- list()
for(s in 1:4) {
  k<-c()
  F<-c()
  p<-c()
  a<-c()
  for (a in c(0.5,1,1.5,2)){
    for (p in c(seq(0,.99,0.01))){
      alpha<-1
      theta<-55
      b<-2
      x<-function(x){(x^s)*((((sqrt(2/pi)*(alpha/x)*(x/theta)^alpha*exp((-1/2)*
        (x/theta)^(2*alpha)))*((2*pnorm((x/theta)^alpha)-1))^(a-1)*
        (1-(2*pnorm((x/theta)^alpha)-1))^(b-1))/(beta(a,b)))*(1-p))/(1-p*
        (1-((pbeta((2*pnorm((x/theta)^alpha)-1),a,b))))^2))}
      i1<-integrate(x,lower=0,upper=Inf)
      f1<-i1$value
      F<-cbind(F,f1)}}
  v<-c(F)
  vs[[s]] <- matrix(0, ncol=301, nrow=100)
  for(i in c(1,101,201,301)) {
    vs[[s]][,i] <- v[i:(i+99)]}
  for(i in 1:4) {
    vs[[i]] <- vs[[i]][,c(1,101,201,301)]}

#=====PLOTS=====
var<-list()
for (j in 1:4){var[[j]]<-vs[[2]][,j]-(vs[[1]][,j])^2}
assi<-list()
for (j in 1:4){ assi[[j]]<-(vs[[3]][,j]-(3*vs[[1]][,j]*vs[[2]][,j]))+(
  2*((vs[[1]][,j])^3))/((var[[j]]^(3/2)))}
curt<-list()
for (j in 1:4){
  curt[[j]]<-((vs[[4]][,j]-(4*vs[[1]][,j]*vs[[3]][,j]))+(

```

```

6*(vs[[2]][,j])*((vs[[1]][,j])^2))-3*((vs[[1]][,j])^4)
)/((var[[j]])^2)}

A<-c(seq(0,.99,0.01)) # eixo x <- variação de p
#Assimetria vari. A
plot(c(0,1),c(0,5), type="n", xlab="p",
      ylab="Assimetria",main="",xlim=c(0,1.5))
for(j in 1:4){
  lines(A,assi[[j]],lty=1,lwd=2,col=j)}
legend(1,4,c("a=0.5","a=1.0","a=1.5","a=2.0"),
      lty=c(1,1,1,1),col=c("1","2","3","4"),seg.len=1,y.intersp=1,bty="n")
#Curtose varia. A
plot(c(0,1),c(5,40), type="n", xlab="p",
      ylab="Curtose",main="",xlim=c(0,1.5))
for(j in 1:4){
  lines(A,curt[[j]],lty=1,lwd=2,col=j)}
legend(1.1,35,c("a=0.5","a=1.0","a=1.5","a=2.0"),
      lty=c(1,1,1,1),col=c("1","2","3","4"),seg.len=1,y.intersp=1,bty="n")

##### Variando B #####
rm(list=ls(all=TRUE))

vs <- list()
for(s in 1:4) {
  k<-c()
  F<-c()
  p<-c()
  a<-c()
  for (b in c(1,1.5,2,2.5)){
    for (p in c(seq(0,.99,0.01))){
      alpha<-1
      theta<-55
      a<- .5
      x<-function(x){(x^s)*((((sqrt(2/pi)*(alpha/x)*(x/theta)^alpha*exp((-1/2)*
(x/theta)^(2*alpha))))*((2*pnorm((x/theta)^alpha)-1))^(a-1)*
(1-(2*pnorm((x/theta)^alpha)-1))^(b-1))/(beta(a,b)))*(1-p))/(1-p*
(1-((pbeta((2*pnorm((x/theta)^alpha)-1),a,b))))^2)}
      i1<-integrate(x,lower=0,upper=Inf)
      f1<-i1$value
      F<-cbind(F,f1)}}
  v<-c(F)
  vs[[s]] <- matrix(0, ncol=301, nrow=100)
  for(i in c(1,101,201,301)) {
    vs[[s]][,i] <- v[i:(i+99)]}
  for(i in 1:4) {
    vs[[i]] <- vs[[i]][,c(1,101,201,301)]}

#=====PLOTS=====
var<-list()
for (j in 1:4){var[[j]]<-vs[[2]][,j]-(vs[[1]][,j])^2}
assi<-list()

```

```

for (j in 1:4){ assi[[j]]<-(vs[[3]][,j]-(3*vs[[1]][,j]*vs[[2]][,j]))+(
  2*((vs[[1]][,j])^3))/((var[[j]])^(3/2))}
curt<-list()
for (j in 1:4){
  curt[[j]]<-((vs[[4]][,j]-(4*vs[[1]][,j]*vs[[3]][,j]))+(
    6*(vs[[2]][,j])*((vs[[1]][,j])^2))-3*((vs[[1]][,j])^4)
  )/((var[[j]])^2)}

A<-c(seq(0,.99,0.01)) # eixo x <- variação de p
setwd("C:\\Users\\Thiago\\Desktop\\plots")

#Assimetria vari. A
plot(c(0,1),c(1,4), type="n", xlab="p",
      ylab="Assimetria",main="",xlim=c(0,1.5))
for(j in 1:4){
  lines(A,assi[[j]],lty=1,lwd=2,col=j)}
legend(1,4,c("b=1.0","b=1.5","b=2.0","b=2.5"),
      lty=c(1,1,1,1),col=c("1","2","3","4"),seg.len=1,y.intersp=1,bty="n")

#Curtose varia. A
plot(c(0,1),c(5,40), type="n", xlab="p",
      ylab="Curtose",main="",xlim=c(0,1.5))
for (j in 1:4){
  lines(A,curt[[j]],lty=1,lwd=2,col=j)}
legend(1,1,35,c("b=1.0","b=1.5","b=2.0","b=2.5"),
      lty=c(1,1,1,1),col=c("1","2","3","4"),seg.len=1,y.intersp=1,bty="n")

```

D.3 Sintaxe do *software* R para curvas de Bonferroni e Lorenz utilizando a distribuição GBGSN

```

#gera o valor de Q |  $F^{-1}(\rho)=q$ 
aleatorio2<-function(rho,aalpha,ttheta,aa,bb,ppp){
  ttheta*(qnorm(p=((1+qbeta(p=((rho*(1-ppp))/(1-ppp*rho))),
    shape1=aa,shape2=bb))/2)))^(1/aalpha)}

#bonferroni
bonf<-function(rho,aalpha,ttheta,aa,bb,ppp){
  alpha<-aalpha
  theta<-ttheta
  a<-aa
  b<-bb
  pp<-ppp
  sss<-seq(0,1000,0.1)
  #fdp
  fdp<-function(x){(x)*((((sqrt(2/pi)*(alpha/x)*(x/theta)^alpha*exp((-1/2)*
    (x/theta)^(2*alpha))))*((2*pnorm((x/theta)^alpha)-1))^(a-1)*
    (1-(2*pnorm((x/theta)^alpha)-1))^(b-1))/(beta(a,b)))*(1-pp)/(1-pp*
    (1-((pbeta((2*pnorm((x/theta)^alpha)-1),a,b))))^2))}
  #Maximo valor de x
  FX <-fdp(sss)
  FXmax<-FX[!is.na(FX) & !is.infinite(FX)]

```

```

max<- sss[length(FXmax)] # escolha o max de x
min<-sss[1]
#bonferroni
(1/((integrate(fdp,lower=0.1,upper=max)$value)*rho))*
(integrate(fdp,lower=.1,upper=aleatorio2(rho,alpha,theta,a,b,pp))$value)
}

v<-c()
for(jj in c(0.1,0.4,0.7,0.9)){
for (i in c(seq(0.01,0.99,0.01))){
  v<-c(v,bonf(i,3,10,.3,.3,jj))
}}

ro1<-v[1:99]
ro4<-v[100:198]
ro7<-v[199:(199+98)]
ro9<-v[298:(298+98)]

ps<-seq(0.01,0.99,0.01)
plot(ps,ro1,type="l",
      ylab=expression(B(rho)),
      xlab=expression(rho),
      lwd=2)
lines(ps,ro4,lwd=2,col="2")
lines(ps,ro7,lwd=2,col="3")
lines(ps,ro9,lwd=2,col="4")
legend(0.0,0.9,c("p=0,1","p=0,4","p=0,7","p=0,9"),
       lwd=c(2,2,2,2),col=c(1,2,3,4),,seg.len=1,y.intersp=1,bty="n")

#lorenz
lorenz<-function(rho,aalpha,ttheta,aa,bb,ppp){
  alpha<-aalpha
  theta<-ttheta
  a<-aa
  b<-bb
  pp<-ppp
  sss<-seq(0,1000,0.1)
  #fdp
  fdp<-function(x){(x)*((((sqrt(2/pi)*(alpha/x)*(x/theta)^alpha*exp((-1/2)*
    (x/theta)^(2*alpha))))*((2*pnorm((x/theta)^alpha)-1))^(a-1)*
    (1-(2*pnorm((x/theta)^alpha)-1))^(b-1))/(beta(a,b)))*(1-pp))/(1-pp*
    (1-((pbeta((2*pnorm((x/theta)^alpha)-1),a,b))))^2))}
  #Maximo valor de x
  FX <-fdp(sss)
  FXmax<-FX[!is.na(FX) & !is.infinite(FX)]
  max<- sss[length(FXmax)] # escolha o max de x
  #lorenz

```

```

(1/(integrate(fdp,lower=0,upper=max)$value))*
  integrate(fdp,lower=0.1,upper=aleatorio2(rho,alpha,theta,a,b,pp))$value}

v<-c()
for(jj in c(0.1,0.4,0.7,0.9)){
  for (i in c(seq(0.01,0.99,0.01))){
    v<-c(v,lorenz(i,3,10,.3,.3,jj))
  }}

ro1<-v[1:99]
ro4<-v[100:198]
ro7<-v[199:(199+98)]
ro9<-v[298:(298+98)]

ps<-seq(0.01,0.99,0.01)
plot(ps,ro1,type="l",
      ylab=expression(L(rho)),
      xlab=expression(rho),
      lwd=2)
lines(ps,ro4,lwd=2,col="2")
lines(ps,ro7,lwd=2,col="3")
lines(ps,ro9,lwd=2,col="4")
legend(0.0,0.9,c("p=0,1","p=0,4","p=0,7","p=0,9"),
      lwd=c(2,2,2,2),col=c(1,2,3,4),,seg.len=1,y.intersp=1,bty="n")

```

D.4 Sintaxe do *software* R para apresentação de formas da distribuição LGBGSN

#####Distribuição Generalizada Half-Normal.

```

g<-function(x,alpha,theta)
{sqrt(2/pi)*(alpha/x)*(x/theta)^alpha*exp((-1/2)*(x/theta)^(2*alpha))}
G<-function(x,alpha,theta){2*pnorm((x/theta)^alpha)-1}

```

###Distribuição beta generalizada. Half-Normal

```

f<-function(x,alpha,theta,a,b){
  (g(x,alpha,theta)*(G(x,alpha,theta))^(a-1) *(1-G(x,alpha,theta))^(b-1))/
  (beta(a,b))}
F<-function(x,alpha,theta,a,b){
  (pbeta(G(x,alpha,theta),a,b))}

```

###Distribuição Geometrica Beta Generalizada Half-Normal.

```

n<-function(x,alpha,theta,a,b,p){
  (f(x,alpha,theta,a,b)*(1-p))/
  (1-p*(1-F(x,alpha,theta,a,b)))^2}
N<-function(x,alpha,theta,a,b,p){
  (F(x,alpha,theta,a,b))/
  (1-p*(1-F(x,alpha,theta,a,b)))}
S<-function(x,alpha,theta,a,b,p){
  1-N(x,alpha,theta,a,b,p)}

```

```

h<-function(x,alpha,theta,a,b,p){
  n(x,alpha,theta,a,b,p)/
  S(x,alpha,theta,a,b,p)}
##### Log Geo Beta Generalizada Half-Normal

#Densidade
l<-function(y,mu,sig,a,b,p){
  (n(exp(y),(sqrt(2)/(2*sig)),exp(mu),a,b,p))*abs(exp(y))}
#Acumulada
L<-function(y,mu,sig,a,b,p){
  ((N(exp(y),(sqrt(2)/(2*sig)),exp(mu),a,b,p)))}
#Sobrevivencia
SL<-function(y,mu,sig,a,b,p){1-L(y,mu,sig,a,b,p)}
#Risco
HL<-function(y,mu,sig,a,b,p){(l(y,mu,sig,a,b,p))/
  SL(y,mu,sig,a,b,p)}

#Plots
#densidade
v<-seq(-100,100,0.01)
plot(v,l(v,0,1,2.5,.5,0.1),lty=1,ylim=c(0,0.8),
  xlim=c(-4,2),lwd=2,typ="l",main="",ylab="f(y)",xlab="y")
lines(v,l(v,0,1,2.5,1.5,0.3),lty=1,lwd=2,col="2")
lines(v,l(v,0,1,2.5,2.5,0.5),lty=1,lwd=2,col="3")
lines(v,l(v,0,1,2.5,3.5,0.7),lty=1,lwd=2,col="4")
lines(v,l(v,0,1,2.5,4.5,0.9),lty=1,lwd=2,col="6")
legend(-.5,0.85,expression(
  paste(b,"=0,5;" ,p,"=0,1"),
  paste(b,"=1,5;" ,p,"=0,1"),
  paste(b,"=2,5;" ,p,"=0,1"),
  paste(b,"=3,5;" ,p,"=0,1"),
  paste(b,"=4,5;" ,p,"=0,1")
),lty=c(1,1,1,1,1),col=c("1","2","3","4","6"),bty="n",
  seg.len=1,y.intersp=.6)

plot(v,l(v,0,.9,4,.5,0.9),lty=1,ylim=c(0,1.3),
  xlim=c(-2,3),lwd=2,typ="l",main="",ylab="f(y)",xlab="y")
lines(v,l(v,0,.7,1.5,.5,0.7),lty=1,lwd=2,col="2")
lines(v,l(v,0,.5,2,.5,0.5),lty=1,lwd=2,col="3")
lines(v,l(v,0,.5,2.5,.5,0.3),lty=1,lwd=2,col="4")
lines(v,l(v,0,.4,2,.5,0.1),lty=1,lwd=2,col="6")
legend(.7,1.1,expression(
  paste(sigma,"=0,9;" ,a,"=4,0;" ,p,"=0,9"),
  paste(sigma,"=0,7;" ,a,"=1,5;" ,p,"=0,7"),
  paste(sigma,"=0,5;" ,a,"=2,0;" ,p,"=0,5"),
  paste(sigma,"=0,5;" ,a,"=2,5;" ,p,"=0,3"),
  paste(sigma,"=0,4;" ,a,"=2,0;" ,p,"=0,1")
),lty=c(1,1,1,1,1),col=c("1","2","3","4","6"),
  bty="n",seg.len=1,y.intersp=.6)

```

```

plot(v,l(v,-2,.9,1.5,1.5,0.1),lty=1,ylim=c(0,0.8),
      xlim=c(-5,4),lwd=2,typ="l",main="",ylab="f(y)",xlab="y")
lines(v,l(v,-1,.7,1.5,1.5,0.3),lty=1,lwd=2,col="2")
lines(v,l(v,0,.5,1.5,1.5,0.5),lty=1,lwd=2,col="3")
lines(v,l(v,1,.7,1.5,1.5,0.7),lty=1,lwd=2,col="4")
lines(v,l(v,2,.9,1.5,1.5,0.9),lty=1,lwd=2,col="6")
legend(0,0.84,expression(
  paste(mu,"=-2;",sigma,"=0,9;",p,"=0,1"),
  paste(mu,"=-1;",sigma,"=0,7;",p,"=0,3"),
  paste(mu,"=0;",sigma,"=0,5;",p,"=0,5"),
  paste(mu,"= 1;",sigma,"=0,7;",p,"=0,7"),
  paste(mu,"= 2;",sigma,"=0,9;",p,"=0,9")
),lty=c(1,1,1,1,1),col=c("1","2","3","4","6"),
      bty="n",seg.len=1,y.intersp=.6)

```

D.5 Sintaxe do *software* R para assimetria e curtose da distribuição LGBGSN

```

#####Distribuição Generalizada Half-Normal.
g<-function(x,alpha,theta){
  sqrt(2/pi)*(alpha/x)*(x/theta)^alpha*exp((-1/2)*(x/theta)^(2*alpha))}
G<-function(x,alpha,theta){
  2*pnorm((x/theta)^alpha)-1}
###Distribuição beta generalizada. Half-Normal
f<-function(x,alpha,theta,a,b){
  (g(x,alpha,theta)* (G(x,alpha,theta))^(a-1) * (1-G(x,alpha,theta))^(b-1))/
  (beta(a,b))}
FF<-function(x,alpha,theta,a,b){
  (pbeta(G(x,alpha,theta),a,b))}
###Distribuição Geométrica Beta Generalizada Half-Normal.
n<-function(x,alpha,theta,a,b,p){
  (f(x,alpha,theta,a,b)*(1-p))/
  (1-p*(1-FF(x,alpha,theta,a,b)))^2}
N<-function(x,alpha,theta,a,b,p){
  (FF(x,alpha,theta,a,b))/
  (1-p*(1-FF(x,alpha,theta,a,b)))}
##### Log Geo Beta Generalizada Half-Normal
l<-function(x){
  (n(exp(x),(sqrt(2)/(2*sig)),exp(mu),a,b,p))*exp(x)}

# mu= 1, sigma= 50 b=2
vs <- list()
for(s in 1:4) {
  k<-c()
  F<-c()
  p<-c()
  a<-c()
  for (a in c(0.3,.5,.9,1.4)){
    for (p in c(seq(0,.99,0.01))){
      sig<-13
      mu<-5

```



```

b<-.3
x<-function(x){(x^s)*l(x)}
i1<-integrate(x,lower=-20,upper=40)
f1<-i1$value
F<-cbind(F,f1)}}
v<-c(F)
vs[[s]] <- matrix(0, ncol=301, nrow=100)
for(i in c(1,101,201,301)) {
  vs[[s]][,i] <- v[i:(i+99)]}}
for(i in 1:4) {
  vs[[i]] <- vs[[i]][,c(1,101,201,301)]}
#=====PLOTS=====
var<-list()
for (j in 1:4){var[[j]]<-vs[[2]][,j]-(vs[[1]][,j])^2}
assi<-list()
for (j in 1:4){ assi[[j]]<-(vs[[3]][,j]-(3*vs[[1]][,j]*vs[[2]][,j])+
  2*((vs[[1]][,j]^3)))/((var[[j]]^(3/2)))}
curt<-list()
for (j in 1:4){
  curt[[j]]<-((vs[[4]][,j]-(4*vs[[1]][,j]*vs[[3]][,j]))+(
    6*(vs[[2]][,j]*((vs[[1]][,j]^2))-3*((vs[[1]][,j]^4)
    )/((var[[j]]^2)}

A<-c(seq(0,.99,0.01)) # eixo x <- variação de p

#Assimetria vari. A
plot(c(0,1),c(-1,3), type="n", xlab="p",
      ylab="Assimetria",main="",xlim=c(0,1.5))
for(j in 1:4){
  lines(A,assi[[j]],lty=1,lwd=2,col=j)}
legend(1,2,c("a=0.3","a=0.5","a=0.9","a=1.4"),
      lty=c(1,1,1,1),col=c("1","2","3","4"),seg.len=1,y.intersp=1,bty="n")

plot(c(0,1),c(0,15), type="n", xlab="p",
      ylab="Curtose",main="",xlim=c(0,1.5))
for (j in 1:4){
  lines(A,curt[[j]],lty=1,lwd=2,col=j)}
legend(1.1,12,c("a=0.3","a=0.5","a=0.9","a=1.4"),
      lty=c(1,1,1,1),col=c("1","2","3","4"),seg.len=1,y.intersp=1,bty="n")

#####
##### Variando B #####
#####

vs <- list()
for(s in 1:4) {
  k<-c()
  F<-c()
  p<-c()
  a<-c()
  for (b in c(0.3,.5,.9,1.4)){

```

```

for (p in c(seq(0,.99,0.01))){
  sig<-13
  mu<-5
  a<-.3
  x<-function(x){(x^s)*l(x)}
  i1<-integrate(x,lower=-20,upper=40)
  f1<-i1$value
  F<-cbind(F,f1)}}
v<-c(F)
vs[[s]] <- matrix(0, ncol=301, nrow=100)
for(i in c(1,101,201,301)) {
  vs[[s]][,i] <- v[i:(i+99)]}}
for(i in 1:4) {
  vs[[i]] <- vs[[i]][,c(1,101,201,301)]}}

#=====PLOTS=====

var<-list()
for (j in 1:4){var[[j]]<-vs[[2]][,j]-(vs[[1]][,j])^2}
assi<-list()
for (j in 1:4){ assi[[j]]<-(vs[[3]][,j]-(3*vs[[1]][,j]*vs[[2]][,j])+(
  2*((vs[[1]][,j])^3)))/((var[[j]])^(3/2))}
curt<-list()
for (j in 1:4){
  curt[[j]]<-((vs[[4]][,j]-(4*vs[[1]][,j]*vs[[3]][,j]))+(
    6*(vs[[2]][,j])*((vs[[1]][,j])^2))-3*((vs[[1]][,j])^4)
  )/((var[[j]])^2)}

A<-c(seq(0,.99,0.01)) # eixo x <- variação de p

#Assimetria vari. B
plot(c(0,1),c(-5,8), type="n", xlab="p",
      ylab="Assimetria",main="",xlim=c(0,1.3))
for(j in 1:4){
  lines(A,assi[[j]],lty=1,lwd=2,col=j)}
legend(1,4,c("b=0.3","b=0.5","b=0.9","b=1.4"),
      lty=c(1,1,1,1),col=c("1","2","3","4"),seg.len=1,y.intersp=1,bty="n")

plot(c(0,1),c(0,15), type="n", xlab="p",
      ylab="Curtose",main="",xlim=c(0,1.3))
for (j in 1:4){
  lines(A,curt[[j]],lty=1,lwd=2,col=j)}

legend(0.9,10,c("b=0.3","b=0.5","b=0.9","b=1.4"),
      lty=c(1,1,1,1),col=c("1","2","3","4"),seg.len=1,y.intersp=1,bty="n")

```

D.6 Sintaxe do *software* R para apresentação de formas da distribuição GBGSN com fração de cura

```
#####Distribuição Generalizada Half-Normal.
```

```

g<-function(x,alpha,theta)
{sqrt(2/pi)*(alpha/x)*(x/theta)^alpha*exp((-1/2)*(x/theta)^(2*alpha))}
G<-function(x,alpha,theta){2*pnorm((x/theta)^alpha)-1}
SG<-function(x,alpha,theta){1-G(x,alpha,theta)}
hg<-function(x,alpha,theta){g(x,alpha,theta)/SG(x,alpha,theta)}
###Distribuição beta generalizada. Half-Normal
f<-function(x,alpha,theta,a,b)
{(g(x,alpha,theta)*(G(x,alpha,theta))^(a-1)*(1-G(x,alpha,theta))^(b-1))/(beta(a,b))}
F<-function(x,alpha,theta,a,b){(pbeta(G(x,alpha,theta),a,b))}
SB<-function(x,alpha,theta,a,b){1-F(x,alpha,theta,a,b)}
hb<-function(x,alpha,theta,a,b){f(x,alpha,theta,a,b/SB(x,alpha,theta,a,b))}
###Distribuição Geométrica Beta Generalizada Half-Normal.
n<-function(x,alpha,theta,a,b,p)
{(f(x,alpha,theta,a,b)*(1-p))/(1-p*(1-F(x,alpha,theta,a,b)))^2}
N<-function(x,alpha,theta,a,b,p){(F(x,alpha,theta,a,b))/(1-p*(1-F(x,alpha,theta,a,b)))}
S<-function(x,alpha,theta,a,b,p){1-N(x,alpha,theta,a,b,p)}
h<-function(x,alpha,theta,a,b,p){n(x,alpha,theta,a,b,p)/S(x,alpha,theta,a,b,p)}
#Fração de cura Geométrica beta generalizada Semi-Normal
Spop<-function(x,alpha,theta,a,b,p,k){k+(1-k)*S(x,alpha,theta,a,b,p)}
Fpop<-function(x,alpha,theta,a,b,p,k){1-Spop(x,alpha,theta,a,b,p,k)}
fpop<-function(x,alpha,theta,a,b,p,k){(1-k)*n(x,alpha,theta,a,b,p)}
hpop<-function(x,alpha,theta,a,b,p,k){fpop(x,alpha,theta,a,b,p,k)
/Spop(x,alpha,theta,a,b,p,k)}

```

```
v<-seq(0,150,0.01)
```

```

#Distribuição fdp
plot(v,fpop(v,1.5,40,1,1,0.01,.9),lwd=2,lty=1,ylim=c(0,0.035),
     typ="l",main="",xlab="t",ylab="fpop(t) ")
lines(v,fpop(v,1.5,40,1,1,0.01,.5),lty=2,lwd=2,col="1")
lines(v,fpop(v,1.5,40,1,1,0.01,.1),lty=3,lwd=2,col="1")
lines(v,fpop(v,1.5,40,3,0.9,0.01,.9),lty=1,lwd=2,col="2")
lines(v,fpop(v,1.5,40,3,0.9,0.01,.5),lty=2,lwd=2,col="2")
lines(v,fpop(v,1.5,40,3,0.9,0.01,.1),lty=3,lwd=2,col="2")
lines(v,fpop(v,1.55,3,0.9,0.01,.9),lty=1,lwd=2,col="3")
lines(v,fpop(v,1.55,3,0.9,0.01,.5),lty=2,lwd=2,col="3")
lines(v,fpop(v,1.55,3,0.9,0.01,.1),lty=3,lwd=2,col="3")
lines(v,fpop(v,2.5,70,1,.1,0.01,.9),lty=1,lwd=2,col="4")
lines(v,fpop(v,2.5,70,1,.1,0.01,.5),lty=2,lwd=2,col="4")
lines(v,fpop(v,2.5,70,1,.1,0.01,.1),lty=3,lwd=2,col="4")
legend(70,0.035,border="red",expression(
  paste(alpha,"=1.5;",theta,"=40;",a,"=1;",b,"=1.0;",p,"=0.01;",k),
  paste(alpha,"=1.5;",theta,"=40;",a,"=3;",b,"=0.9;",p,"=0.01;",k),
  paste(alpha,"=1.0;",theta,"=55;",a,"=3;",b,"=0.9;",p,"=0.01;",k),
  paste(alpha,"=2.5;",theta,"=70;",a,"=1;",b,"=0.1;",p,"=0.01;",k),
  paste(k,"=0.9"),paste(k,"=0.5"),paste(k,"=0.1")
),horiz=FALSE,lty=c(1,1,1,1,2,3),col=c("1","2","3","4","1","1","1"),
     bty="n",y.intersp=1,seg.len=1)

```

```

plot(v,fpop(v,8.5,80,0.07,0.05,0.3,0.9),lty=1,
     ylim=c(0,0.06),,xlim=c(0,96),lwd=2,typ="l",main="",ylab="fpop(t)",xlab="t")
lines(v,fpop(v,8.5,80,0.07,0.05,0.3,0.5),lwd=2,lty=2,col="1")
lines(v,fpop(v,8.5,80,0.07,0.05,0.3,0.1),lwd=2,lty=3,col="1")
lines(v,fpop(v,8.0,80,0.20,0.3,0.01,.9),lwd=2,lty=1,col="2")
lines(v,fpop(v,8.0,80,0.20,0.3,0.01,.5),lwd=2,lty=2,col="2")
lines(v,fpop(v,8.0,80,0.20,0.3,0.01,.1),lwd=2,lty=3,col="2")
lines(v,fpop(v,2.5,70,0.002,0.02,0.01,.9),lty=1,lwd=2,col="4")
lines(v,fpop(v,2.5,70,0.002,0.02,0.01,.5),lty=2,lwd=2,col="4")
lines(v,fpop(v,2.5,70,0.002,0.02,0.01,.1),lty=3,lwd=2,col="4")
lines(v,fpop(v,1.5,40,0.002,0.02,0.01,.9),lty=1,lwd=2,col="3")
lines(v,fpop(v,1.5,40,0.002,0.02,0.01,.5),lty=2,lwd=2,col="3")
lines(v,fpop(v,1.5,40,0.002,0.02,0.01,.1),lty=3,lwd=2,col="3")
legend(40,0.057,expression(
  paste(alpha,"=8.5;",theta,"=80;",a,"=0.07;",b,"=0.05;",p,"=0.30;",k),
  paste(alpha,"=8.0;",theta,"=80;",a,"=0.20;",b,"=0.30;",p,"=0.01;",k),
  paste(alpha,"=2.5;",theta,"=70;",a,"=2.00;",b,"=2.00;",p,"=0.01;",k),
  paste(alpha,"=1.5;",theta,"=40;",a,"=2.00;",b,"=2.00;",p,"=0.01;",k),
  paste(k,"=0.9"),
  paste(k,"=0.5"),
  paste(k,"=0.1")
),lty=c(1,1,1,1,2,3),col=c("1","2","4","3","1","1","1"),
     bty="n",seg.len=1,y.intersp=1)

#Sobrevivencia
plot(v,Spop(v,1.5,40,1,1,0.9,.3),lty=1,
     lwd=2,typ="l",main="",xlab="t",ylab="Spop(t)",xlim=c(0,140),ylim=c(0,1))
lines(v,Spop(v,1.5,40,3,0.9,0.9,.3),lty=1,lwd=2,col="2")
lines(v,Spop(v,1.5,40,0.5,2,0.9,.3),lty=1,lwd=2,col="3")
lines(v,Spop(v,2.5,90,3,0.9,0.01,.3),lty=1,lwd=2,col="4")
lines(v,Spop(v,7,104,0.1,0.3,0.01,.3),lwd=2,col="6")
legend(60,.3,expression(
  paste(alpha,"=1.5;",theta,"= 40;",a,"=1.0;",b,"=1.0;",p,"=0.90;",k,"=0.3"),
  paste(alpha,"=1.5;",theta,"= 40;",a,"=3.0;",b,"=0.9;",p,"=0.90;",k,"=0.3"),
  paste(alpha,"=1.5;",theta,"= 40;",a,"=0.5;",b,"=2.0;",p,"=0.90;",k,"=0.3"),
  paste(alpha,"=2.5;",theta,"= 90;",a,"=3.0;",b,"=0.9;",p,"=0.01;",k,"=0.3"),
  paste(alpha,"=7.0;",theta,"=104;",a,"=0.1;",b,"=0.3;",p,"=0.01;",k,"=0.3")
),lty=c(1,1,1,1,1),col=c("1","2","3","4","6"),bty="n",
     seg.len=1,y.intersp=1)

#Risco
plot(v,hpop(v,0.5,70,0.2,0.9,0.9,.2),lty=1,ylim=c(0,0.10),
     lwd=2,typ="l",main="",xlab="h(t)",ylab="")
lines(v,hpop(v,6.2,110,0.1,0.9,0.9,.3),lty=1,lwd=2,col="2")
lines(v,hpop(v,2.5,70,0.5,2,0.9,.2),lty=1,lwd=2,col="3")
lines(v,hpop(v,2.5,90,3,0.9,0.01,.2),lty=1,lwd=2,col="4")
legend(60,.08,expression(
  paste(alpha,"=0.5;",theta,"= 70;",a,"=0.2;",b,"=0.9;",p,"=0.90;",k,"=0.3"),
  paste(alpha,"=6.2;",theta,"=110;",a,"=0.1;",b,"=0.9;",p,"=0.90;",k,"=0.3"),
  paste(alpha,"=2.5;",theta,"= 70;",a,"=0.5;",b,"=2.0;",p,"=0.90;",k,"=0.3"),

```

```
paste(alpha,"=2.5;",theta,"= 90;",a,"=3.0;",b,"=0.9;",p,"=0.01;",k,"=0.3")
),lty=c(1,1,1,1),col=c("1","2","3","4"),bty="n",seg.len=1,y.intersp=1)
```

D.7 Sintaxe do *software* R para modelagem dos dados Feigl

```
##### Dados Feigl #####
#####Distribuição Generalizada Half-Normal.
g<-function(x,alpha,theta)
{sqrt(2/pi)*(alpha/x)*(x/theta)^alpha*exp((-1/2)*(x/theta)^(2*alpha))}
G<-function(x,alpha,theta){2*pnorm((x/theta)^alpha)-1}
SG<-function(x,alpha,theta){1-G(x,alpha,theta)}
hg<-function(x,alpha,theta){g(x,alpha,theta)/SG(x,alpha,theta)}
###Distribuição beta generalizada. Half-Normal
f<-function(x,alpha,theta,a,b)
{(g(x,alpha,theta)* (G(x,alpha,theta))^(a-1) * (1-G(x,alpha,theta))^(b-1))/(beta(a,b))}
F<-function(x,alpha,theta,a,b){(pbeta(G(x,alpha,theta),a,b))}
SB<-function(x,alpha,theta,a,b){1-F(x,alpha,theta,a,b)}
hb<-function(x,alpha,theta,a,b){f(x,alpha,theta,a,b/SB(x,alpha,theta,a,b))}
###Distribuição Geométrica Beta Generalizada Half-Normal.
n<-function(x,alpha,theta,a,b,p)
{(f(x,alpha,theta,a,b)*(1-p))/(1-p*(1-F(x,alpha,theta,a,b)))^2}
N<-function(x,alpha,theta,a,b,p){(F(x,alpha,theta,a,b))/(1-p*(1-F(x,alpha,theta,a,b)))}
S<-function(x,alpha,theta,a,b,p){1-N(x,alpha,theta,a,b,p)}
h<-function(x,alpha,theta,a,b,p){n(x,alpha,theta,a,b,p)/S(x,alpha,theta,a,b,p)}

install.packages('SMIR',dep=T)
library(SMIR)
library(moments)
library(survival)

data(feigl)
attach(feigl)
hist(time,freq=FALSE)
cens<-rep(1,33)
feigl

t<-seq(0,160,0.1)
hist(time,freq=FALSE,xlab="t",ylab="f(t)",main="")
lines(t,f(t,0.3229,73.8221,3.2298,3.3246),lwd=1,lty=1)
lines(t,n(t,7.2631,132.29,0.1231,0.1907,0.8749),col="3",lwd=1)
lines(t,g(t,0.62,46.52),col="2",lwd=1,lty=1)
legend(100,0.023,c("BGHNG","BGHN","GHN"),
      lty=c(1,1,1),col=c("3","black","2"),
      cex=.8,y.intersp=1,seg.len=1,bty="n")

ekm<-survfit(Surv(time,cens)~1)
summary(ekm)
plot(ekm, conf.int=FALSE,ylab="S(t)",xlab="t")
lines(t,SB(t,0.3229,73.8221,3.2298,3.3246))
lines(t,S(t,7.2631,132.29,0.1231,0.1907,0.8749),col="3",lwd=1)
lines(t,SG(t,0.62,46.52),col="2",lwd=1,lty=1)
```

```

legend(100,0.9,c("BGHNG","BGHN","GHN"),
      lty=c(1,1,1),col=c("3","black","2"),
      cex=.8,y.intersp=1,seg.len=1,bty="n")

#risco
plot(t,hg(t,0.62,46.52),col="2",lwd=1,lty=1,type="l",ylim=c(0,0.15))
lines(t,hb(t,0.3229,73.8221,3.2298,3.3246),lwd=1,lty=1)
lines(t,h(t,7.2631,132.29,0.1231,0.1907,0.8749),col="3",lwd=1)
legend(100,0.14,c("BGHNG","BGHN","GHN"),
      lty=c(1,1,1),col=c("3","black","2"),
      cex=.8,y.intersp=1,seg.len=1,bty="n")

t<-seq(1.1,160,.01)
plot(t,h(t,7.2631,132.29,0.1231,0.1907,0.8749),col="blue",lwd=2,
      type="l",xlab="x",ylab="h(x)")

#ACUMULADAS
w=seq(0,160,.001)

plot(ecdf(time), pch=20, las=1, ylab="F(x)", xlab="x", main="",xlim=c(0,150))
lines(w,N(w,7.2631,132.29,0.1231,0.1907,0.8749),col="blue",lwd=2)
lines(w,G(w,0.62,46.52),col="green",lwd=2,lty=1)
legend(100,0.5,c('BGGSN','GHN'),bty='n',cex=1,
      col=c('blue','green'), lwd=c(2,2) , lty=c(1,1))

plot(ecdf(time), pch=20, las=1, ylab="F(x)", xlab="x", main="",xlim=c(0,150))
lines(w,N(w,7.2631,132.29,0.1231,0.1907,0.8749),col="blue",lwd=2)
lines(w,F(w,0.3229,73.8221,3.2298,3.3246),lwd=2,lty=1,col="red")
legend(100,0.5,c('BGHNG','BGHN'),bty='n',cex=1,
      col=c('blue','red'), lwd=c(2,2) , lty=c(1,1))

```

D.8 Sintaxe do *software* R para modelagem dos dados Feigl considerando covariáveis

```

#####Distribuição Log Generalizada Half-Normal.
lg<-function(y,b0,b1,b2,sig){g(exp(y),(sqrt(2)/(2*sig)),exp(b0+b1*x1+b2*x2))*exp(y)}
lG<-function(y,b0,b1,b2,sig){G(exp(y),(sqrt(2)/(2*sig)),exp(b0+b1*x1+b2*x2))}
lSG<-function(y,b0,b1,b2,sig){1-lG(y,b0,b1,b2,sig)}
##### Log Beta Generalizada Half-Normal
lf<-function(y,b0,b1,b2,sig,a,b){f(exp(y),(sqrt(2)/(2*sig)),exp(b0+b1*x1+b2*x2),a,b)*exp(y)}
lF<-function(y,b0,b1,b2,sig,a,b){F(exp(y),(sqrt(2)/(2*sig)),exp(b0+b1*x1+b2*x2),a,b)}
lS<-function(y,b0,b1,b2,sig,a,b){1-lF(y,b0,b1,b2,sig,a,b)}
lh<-function(y,b0,b1,b2,sig,a,b){ lf(y,b0,b1,b2,sig,a,b)/lS(y,b0,b1,b2,sig,a,b)}
##### Log Geo Beta Generalizada Half-Normal
l<-function(y,b0,b1,b2,sig,a,b,p){(n(exp(y),(sqrt(2)/(2*sig)),exp(b0+b1*x1+b2*x2),a,b,p))*e}
L<-function(y,b0,b1,b2,sig,a,b,p){((N(exp(y),(sqrt(2)/(2*sig)),exp(b0+b1*x1+b2*x2),a,b,p)))}
SL<-function(y,b0,b1,b2,sig,a,b,p){1-L(y,b0,b1,b2,sig,a,b,p)}
HL<-function(y,b0,b1,b2,sig,a,b,p){(l(y,b0,b1,b2,sig,a,b,p))/SL(y,b0,b1,b2,sig,a,b,p)}

#feigl
library(SMIR)
data(feigl)

```

```

attach(feigl)
dados<-feigl
feigl1<-dados[order(dados$time),]

attach(feigl1)
length(time)
cens<-rep(1,33)
y<-log(time)
se<-seq(min(y),max(y),0.001)
names(feigl1)
x1<-wbc
x2<-ifelse(ag=="+",1,0)

#Separando grupo
dd<-cbind(y,x1,x2)
dd<-dd[order(x2),]
dd
cbind(dd[,3],order(dd[,3]))
maxi<-max(dd[,2])
mini<-min(dd[,2])

#max x1
x1<-maxi
x2<-0
plot(se,SL(se,3.9608,-0.00428,0.8428,0.1015,0.1277,0.07477,0.8554),col="2",
      type="l",lwd=2,main="",xlab="y",ylab="")
x2<-1
lines(se,SL(se,3.9608,-0.00428,0.8428,0.1015,0.1277,0.07477,0.8554),col="4",
      type="l",lwd=2)
legend(4,0.8,expression(
  paste(x[2],"= -"),
  paste(x[2],"= +")),
      lty=c(1,1),col=c(2,4),bty="n",y.intersp=.6)

#min x1
x1<-mini
x2<-0
plot(se,SL(se,3.9608,-0.00428,0.8428,0.1015,0.1277,0.07477,0.8554),col="3",
      type="l",lwd=2,main="",xlab="y",ylab="")
x2<-1
lines(se,SL(se,3.9608,-0.00428,0.8428,0.1015,0.1277,0.07477,0.8554),col="6",
      type="l",lwd=2)
legend(4,0.8,expression(
  paste(x[2],"= -"),
  paste(x[2],"= +")),
      lty=c(1,1),col=c(3,6),bty="n",y.intersp=.6)

```

E.1 Sintaxe do *software* SAS para as estimativas dos parâmetros das distribuições GBGSN, BGSN e GSN ajustadas ao conjunto de dados Feigl

```

data dados22;
input obs x1 t x2x$;

```

```
y=log(t); censur=1; if x2x="+" then x2=1; else x2=0;
cards;
```

```
1    2.30    65  +
2    0.75   156  +
3    4.30   100  +
4    2.60   134  +
5    6.00    16  +
6   10.50   108  +
7   10.00   121  +
8   17.00     4  +
9    5.40    39  +
10   7.00   143  +
11   9.40    56  +
12  32.00    26  +
13  35.00    22  +
14 100.00     1  +
15 100.00     1  +
16  52.00     5  +
17 100.00    65  +
18   4.40    56  -
19   3.00    65  -
20   4.00    17  -
21   1.50     7  -
22   9.00    16  -
23   5.30    22  -
24  10.00     3  -
25  19.00     4  -
26  27.00     2  -
27  28.00     3  -
28  31.00     8  -
29  26.00     4  -
30  21.00     3  -
31  79.00    30  -
32 100.00     4  -
33 100.00    43  -
;
```

```
/** Modelagem apenas com o tempo */
/**distribuição GSN**/
proc nlmixed cov data=dados22;
parms   alfa=0.01  teta=1 ;
bounds alfa>0;
bounds teta>0;
gx=(0.797894295)*(alfa/t)*((t/teta)**alfa)*exp(-0.5*((t/teta)**(2*alfa)));
GGx=2*PROBNORM((t/teta)**(alfa))-1 ;
logp=log(gx);
*logp=censur*log(gx) +(1-censur)*log(1-GGx);
model t ~ general(logp);
run;
```



```

/**distribuição BGSN**/
proc nlmixed cov data=dados22;
parms a=0.07688 b=0.264 alfa=0.7557 teta=73.6358;
bounds a>0;
bounds b>0;
bounds alfa>0;
bounds teta>0;
gx=(0.797894295)*(alfa/t)*((t/teta)**alfa)*exp(-0.5*((t/teta)**(2*alfa)));
GGx=2*PROBNORM((t/teta)**(alfa))-1 ;
fbb=(gx/BETA(a,b))*((GGx)**(a-1))*((1-GGx)**(b-1)) ;
FFbb=probbeta(GGx,a,b);
logp=log(fbb);
*logp=censur*log(fbb)+(1-censur)*log(1-FFbb);
model t ~ general(logp);
run;

```

```

/**distribuição GBGHN**/
proc nlmixed cov data=dados22;
parms a=0.07286 b=0.3137 alfa=7.2506 teta=132.32 p=0.5;
bounds a>0;
bounds b>0;
bounds alfa>0;
bounds teta>0;
bounds p>0;
gx=(0.797894295)*(alfa/t)*((t/teta)**alfa)*exp(-0.5*((t/teta)**(2*alfa)));
GGx=2*PROBNORM((t/teta)**(alfa))-1 ;
fbb=(gx/BETA(a,b))*((GGx)**(a-1))*((1-GGx)**(b-1)) ;
FFbb=probbeta(GGx,a,b);
fpp=(fbb*(1-p))*((1-p*(1-FFbb))**(-2));
FFpp=(FFbb)*((1-p*(1-FFbb))**(-1));
logp=log(fpp);
*logp=censur*log(fpp)+(1-censur)*log(1-FFpp);
model t ~ general(logp);
run;

```

/*Modelagem com covariáveis */

```

/**distribuição log GSN**/
proc nlmixed cov data=dados22;
parms b0=3 b1=3 b2=2 sigma=1.1389;
bounds sigma>0;
gx=(0.797894295)*((sqrt(2)/(2*sigma))/exp(y))*((exp(y)/exp((b0+b1*x1+b2*x2)))*
(sqrt(2)/(2*sigma)))*exp(-0.5*((exp(y)/exp((b0+b1*x1+b2*x2)))*
(2*(sqrt(2)/(2*sigma)))));
GGx=2*PROBNORM((exp(y)/exp((b0+b1*x1+b2*x2)))*((sqrt(2)/(2*sigma))))-1 ;
logp=log(gx);
*logp=censur*log(gx) +(1-censur)*log(1-GGx);
model y ~ general(logp);
run;

```

```

/**distribuição log BGSN**/
proc nlmixed cov data=dados22;
parms    b0=3.3118 b1=-0.00562 b2=1.0569  sigma=0.9456 a=0.2 b=1;
bounds sigma>0;
bounds a>0;
bounds b>0;
gx=(0.797894295)*((sqrt(2)/(2*sigma))/exp(y))*((exp(y)/exp((b0+b1*x1+b2*x2)))*
  (sqrt(2)/(2*sigma)))*exp(-0.5*((exp(y)/exp((b0+b1*x1+b2*x2)))*
  (2*(sqrt(2)/(2*sigma)))));
GGx=2*PROBNORM((exp(y)/exp((b0+b1*x1+b2*x2)))*((sqrt(2)/(2*sigma))))-1 ;
fbb=(gx/BETA(a,b))*((GGx)**(a-1))*((1-GGx)**(b-1)) ;
FFbb=probbeta(GGx,a,b);
logp=log(fbb);
*logp=censur*log(fbb)+(1-censur)*log(1-FFbb);
model y ~ general(logp);
run;

/**distribuição log GBGHN**/
proc nlmixed cov data=dados22;
parms    b0=4.0573 b1=-0.00435 b2=0.8235  sigma=0.09481 a=0.07032 b=0.2755 p=.9;
bounds sigma>0;
bounds a>0;
bounds b>0;
bounds p>=0;
gx=(0.797894295)*((sqrt(2)/(2*sigma))/exp(y))*((exp(y)/exp((b0+b1*x1+b2*x2)))*
  (sqrt(2)/(2*sigma)))*exp(-0.5*((exp(y)/exp((b0+b1*x1+b2*x2)))*
  (2*(sqrt(2)/(2*sigma)))));
GGx=2*PROBNORM((exp(y)/exp((b0+b1*x1+b2*x2)))*((sqrt(2)/(2*sigma))))-1 ;
fbb=(gx/BETA(a,b))*((GGx)**(a-1))*((1-GGx)**(b-1)) ;
FFbb=probbeta(GGx,a,b);
fpp=(fbb*(1-p))*((1-p*(1-FFbb))**(-2));
FFpp=(FFbb)*((1-p*(1-FFbb))**(-1));
logp=log(fpp);
*logp=censur*log(fpp)+(1-censur)*log(1-FFpp);
model y ~ general(logp);
run;

```

E.2 Sintaxe do *software* R para simulações

```

#Graficos
pontos<-seq(0,500,0.01)
comparacao<-function(n,alpha,theta,a,b,pp){
  aleatorio<-function(n,alpha,theta,a,b,pp){
    u<-runif(n=n,min=0,max=1)
    theta*(qnorm(p=((1+qbeta(p=((u*(1-pp))/(1-pp*u)),
                        shape1=a,
                        shape2=b))/2)))^(1/alpha)
  }
  tt<-aleatorio(n,alpha,theta,a,b,pp)

  plot(density(tt),col="2",lwd=2,xlim=c(0,max(tt)),ylim=c(0,max(density(tt)$y)))
  lines(pontos,n(pontos,alpha,theta,a,b,pp),typ="l",lwd=2)

```

```
hist(tt,freq=FALSE,xlab="x",ylab="f(x)",main="",
      ylim=c(0,max(density(tt)$y)))
lines(pontos,n(pontos,alpha,theta,a,b,pp),typ="l",lwd=2)

}
comparacao(100000,2,10,.5,.5,0.10)
legend(19,0.06,"p=0,10")
comparacao(100000,2,10,.5,.5,0.40)
legend(19,0.08,"p=0,40")
comparacao(100000,2,10,.5,.5,0.70)
legend(15,0.12,"p=0,70")
comparacao(100000,2,10,.5,.5,0.90)
legend(10,0.25,"p=0,90")
```