



Universidade Estadual de Maringá

Departamento de Estatística

Programa de Pós-graduação em Bioestatística



JEAN CARLOS CARDOSO

## Extensões da Distribuição Half-Normal

Maringá - Paraná  
Fevereiro de 2018

**JEAN CARLOS CARDOSO**

# **Extensões da Distribuição Half-Normal**

Dissertação apresentada ao programa de Pós-graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para obtenção do título de Mestre em Bioestatística.

Orientador: Prof. Dr. Josmar Mazucheli

**Maringá - Paraná  
Fevereiro de 2018**

**Dados Internacionais de Catalogação na Publicação (CIP)**  
**(Biblioteca Central - UEM, Maringá, PR, Brasil)**

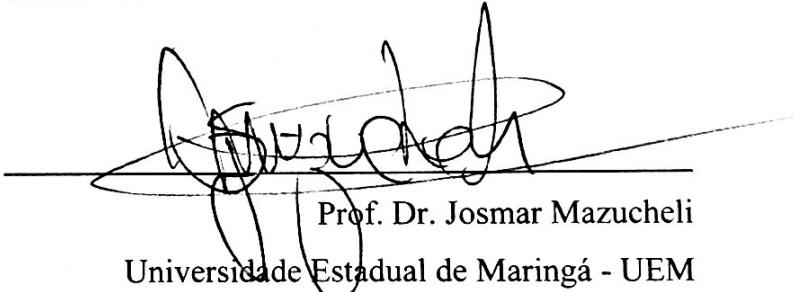
C268e	<p>Cardoso, Jean Carlos Extensões da distribuição Half-Normal / Jean Carlos Cardoso. -- Maringá, 2018. 63 f. : il., figs., tabs.</p> <p>Orientador: Prof. Dr. Josmar Mazucheli. Dissertação (mestrado) - Universidade Estadual de Maringá, Departamento de Estatística, Programa de Pós-Graduação em Bioestatística, 2018.</p> <p>1. Discretização. 2. Distribuição Half-Normal generalizada. 3. Método dos momentos. 4. Simulação Monte Carlo. 5. Verossimilhança. 6. Distribuição Half-Normal. 7. Estimação Paramétrica. 8. Transmutação. I. Mazucheli, Josmar, orient. II. Universidade Estadual de Maringá. Departamento de Estatística. Programa de Pós-Graduação em Bioestatística. III. Título.</p>
	<p>CDD 21.ed.570.15195 ECSL-1202/9</p>

**JEAN CARLOS CARDOSO**

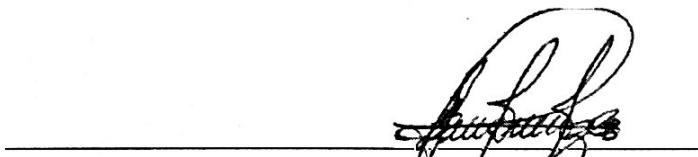
**Extensões da Distribuição Half-Normal**

Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para a obtenção do título de Mestre em Bioestatística.

**BANCA EXAMINADORA**



Prof. Dr. Josmar Mazucheli  
Universidade Estadual de Maringá - UEM



Profª. Drª. Eniuce Menezes de Souza  
Universidade Estadual de Maringá – PCS/UEM



Prof. Dr. Fernando Antonio Moala  
Universidade Estadual Paulista – Júlio de Mesquita Filho – UNESP

Maringá, 19 de fevereiro de 2018.

*A minha amada esposa  
Priscila Friedemann Cardoso*

---

## AGRADECIMENTOS

Em primeiro lugar agradeço a Deus, por ter me auxiliado em mais essa etapa de minha vida, me dando saúde, coragem e perseverança para continuar e concluir-la.

Quero agradecer o apoio incondicional da minha esposa, Priscila Friedemann Cardoso, que esteve sempre ao meu lado, ouvindo por diversas vezes explicações sobre algo que ela nem entendia.

Quero também agradecer ao Professor Josmar Mazucheli, que esteve sempre presente quando precisei, me deu inúmeras ideias, muitos conselhos e é responsável por boa parte do meu conhecimento em estatística e em especial na área de análise de sobrevivência.

Um agradecimento aos meus pais, Josué e Diva Cardoso, ao meu irmão, Renan Caio Cardoso, aos meus sogros, Osnildo e Liziane Friedemann e também ao meu cunhado Daniel Filipe Friedemann pelo apoio e pelas orações.

Agradeço também a três amigos que estiveram unidos comigo desde o início, João Pedro Serenini, Aline Edlaine de Medeiros e Guilherme Lima Desoler, fica aqui meu muito obrigado a vocês pela companhia e pela parceria não só nos estudos, mas também na vida.

Agradeço ao departamento de estatística da UEM, em especial ao programa de pós-graduação em bioestatística, assim como a todos os professores do corpo docente, bem como aos mestrandos das turmas de 2016 e 2017.

Por fim, agradeço à CAPES pelos meses de apoio financeiro, me proporcionando um envolvimento total em meus estudos.

*“A obtenção de um resultado novo em pesquisa é, para o cientista, uma fonte de intenso prazer, ligado intimamente ao instinto de criação e eternidade, pois, independente da importância da contribuição no contexto da ciência, ou de sua utilização, representa algo acrescentado ao conhecimento humano que marca sua existência na terra.”*

*Pierre Curie (Físico)*

---

## RESUMO

Neste trabalho propomos duas novas distribuições derivadas, respectivamente, das distribuições Half-Normal generalizada e Half-Normal. Na primeira proposta, a partir do método de discretização de Nakagawa e Osaki (1975), formulamos a versão discreta da distribuição Half-Normal generalizada. Algumas de suas propriedades são apresentadas e via simulações Monte Carlo são avaliados o viés e a acurácia das estimativas obtidas pelo método da máxima verossimilhança e pelo método dos momentos. Dois exemplos utilizando dados da literatura ilustram a aplicação desta distribuição. Na segunda proposta introduzimos a distribuição Half-Normal transmutada a partir da metodologia de Shaw e Buckley (2009) tomando como base a distribuição Half-Normal. As expressões analíticas explícitas da esperança, variância, assimetria, curtose, função geradora de momentos, entropia de Shannon, desvios médios, curvas de Bonferroni e Lorenz, estatística de ordem e confiabilidade são deduzidas e apresentadas. Os parâmetros da distribuição são estimados usando o método da máxima verossimilhança. O viés e a acurácia dos estimadores são estimados via simulações Monte Carlo. Duas aplicações, a primeira sem a presença de covariáveis e a segunda na presença de covariáveis, são usadas com propósitos ilustrativos.

**Palavras-chave:** Discretização, Distribuição Half-Normal Generalizada, Método dos Momentos, Simulação Monte Carlo, Verossimilhança, Distribuição Half-Normal, Estimação Paramétrica, Transmutação.

---

## ABSTRACT

In this work we propose two new distributions derived, respectively, from the generalized Half-Normal and Half-Normal distributions. In the first proposal, from the discretization method of Nakagawa and Osaki (1975), we formulated the discrete version of the generalized Half-Normal distribution. Some of its properties are presented and the bias and the root mean square error of estimates obtained by maximum likelihood and by method of moments are evaluated by Monte Carlo simulation. Two examples using data of literature illustrate the application of the distribution. In the second proposal we introduce the half-normal distribution transmuted from the methodology of Shaw and Buckley (2009) based on the Half-Normal distribution. We provide explicit expressions for the moments, moment-generating function, Shannon's entropy, mean deviations, Bonferroni and Lorenz curves, order statistics, and reliability. The estimation of the parameters is approached by the maximum likelihood method. The bias and accuracy of the estimators are investigated by the Monte Carlo simulations. Two real data sets show that the proposed distribution is useful and may be a good alternative to several extensions of the Half-Normal distribution.

**Keywords:** Discretization, Generalized Half-Normal Distribution, Methods of Moments, Monte Carlo Simulation, Likelihood, Half-Normal Distribution, Parametric Estimation, Transmutation.

---

## LISTA DE FIGURAS

2.1	Comportamento da função de probabilidade da distribuição HNGD.	21
2.2	Comportamento das funções de risco da distribuição HNGD.	22
2.3	Comportamento da média, variância, assimetria e curtose da distribuição HNGD. $\theta = 1.0$ (—), $\theta = 2.0$ (---), $\theta = 5$ (···) e $\theta = 10.0$ (----).	24
3.1	Comportamento da função densidade de probabilidade da distribuição HNT.	36
3.2	Comportamento da função de risco da distribuição HNT.	37
3.3	Estimativa do vício e do erro quadrático médio de $\theta$ ( $\theta = 1, 1 : \lambda = -0.9, 2 : \lambda = -0.7, 3 : \lambda = -0.5, 4 : \lambda = -0.3, 5 : \lambda = 0, 6 : \lambda = 0.3, 7 : \lambda = 0.5, 8 : \lambda = 0.7$ e $9 : \lambda = 0.9$ ).	49
3.4	Estimativa do vício e do erro quadrático médio de $\lambda$ ( $\theta = 1, 1 : \lambda = -0.9, 2 : \lambda = -0.7, 3 : \lambda = -0.5, 4 : \lambda = -0.3, 5 : \lambda = 0, 6 : \lambda = 0.3, 7 : \lambda = 0.5, 8 : \lambda = 0.7$ e $9 : \lambda = 0.9$ ).	49
3.5	TTT-plot.	51
3.6	Média empírica e estimada para cada período de dez dias.	57

---

## LISTA DE TABELAS

2.1	Índice de dispersão da distribuição HNGD. . . . .	23
2.2	Estimativas do viés e da raiz do erro quadrático médio (REQM) dos estimadores de máxima verossimilhança (MV) e dos estimadores obtidos pelo método dos momentos (MM) para $\theta = 1.0$ . . . . .	27
2.3	Estimativas do viés e da raiz do erro quadrático médio (REQM) dos estimadores de máxima verossimilhança (MV) e dos estimadores obtidos pelo método dos momentos (MM) para $\theta = 10$ . . . . .	28
2.4	Distribuição do número de gols “em casa” observados e esperados por partida na Premier League. . . . .	29
2.5	Distribuição do número de gols “fora de casa” observados e esperados por partida na Premier League. . . . .	30
2.6	Tempo de vida em dias de 43 pacientes que foram diagnosticados com Leucemia. . . . .	30
2.7	Estimativas de máxima verossimilhança (erros padrão), valor da estatística do teste qui-quadrado de aderência e os valores de AIC e BIC. . . . .	30
3.1	Medidas resumo. . . . .	51
3.2	Estimativas de máxima verossimilhança (erros padrão). . . . .	55
3.3	Valores de -2Log, AIC, AICc, BIC, KS, AD e CvM. . . . .	55
3.4	Resultados dos testes de aderência. . . . .	56
3.5	Estimativas de parâmetros e intervalos de confiança de 95%. . . . .	56
3.6	Médias empíricas e estimadas para cada período de dez dias. . . . .	57
3.7	Medidas de seleção de modelo. . . . .	57

---

# SUMÁRIO

<b>1 Resumo Geral</b>	<b>12</b>
1.1 Introdução . . . . .	12
1.2 Justificativa . . . . .	13
1.3 Objetivos . . . . .	14
1.3.1 Objetivo Geral . . . . .	14
1.3.2 Objetivos Específicos . . . . .	14
1.4 Metodologia . . . . .	14
1.5 Resultados . . . . .	15
1.6 Conclusão . . . . .	15
1.7 Propostas Futuras . . . . .	15
<b>2 A Distribuição Half-Normal Generalizada Discreta</b>	<b>16</b>
2.1 Introdução . . . . .	17
2.2 Discretização via Função de Sobrevivência . . . . .	18
2.3 A Distribuição Half-Normal Generalizada . . . . .	19
2.3.1 Versão Contínua . . . . .	19
2.3.2 Versão Discreta . . . . .	20
2.3.3 Função Geradora de Momentos e Algumas Propriedades . . . . .	22
2.4 Métodos de Estimação . . . . .	24
2.4.1 Estimação Via Método da Máxima Verossimilhança . . . . .	24
2.4.2 Estimação Via Método dos Momentos . . . . .	25
2.5 Estudo de Simulação . . . . .	26
2.6 Aplicações . . . . .	29
2.7 Conclusão . . . . .	31
<b>3 A Distribuição Half-Normal Transmutada</b>	<b>32</b>
3.1 Introdução . . . . .	33

3.2 A Distribuição Half-Normal . . . . .	34
3.3 A Distribuição Half-Normal Transmutada . . . . .	35
3.4 Funções Úteis na Análise de Dados de Sobrevidência . . . . .	37
3.5 Propriedades da Distribuição HNT . . . . .	39
3.5.1 Momento de Ordem $k$ . . . . .	39
3.5.2 Função Geradora de Momentos . . . . .	40
3.5.3 Entropia Diferencial . . . . .	41
3.5.4 Desvios Médios . . . . .	42
3.5.5 Curvas de Bonferroni e Lorenz . . . . .	43
3.5.6 Estatística de Ordem . . . . .	44
3.5.7 Confiabilidade . . . . .	45
3.6 Estimação Via Máxima Verossimilhança . . . . .	47
3.7 Simulação . . . . .	48
3.8 Aplicações . . . . .	50
3.9 Conclusão . . . . .	57
<b>Referências</b>	<b>59</b>

---

# CAPÍTULO 1

---

## RESUMO GERAL

### 1.1 Introdução

Subjacente a todo procedimento paramétrico de inferência existe uma distribuição de probabilidade, usada para descrever o comportamento de uma variável aleatória na população. Nos últimos anos, um incontável número de distribuições de probabilidade, em sua maioria com suporte nos números reais, foram propostas na literatura. Muitas estratégias são usadas na geração de uma nova distribuição de probabilidade, e grande parte consiste em adicionar um ou mais parâmetros a alguma distribuição base (Normal, Gumbel, Exponencial, Weibull, Gama, Log-Normal, entre tantas outras). Em geral, introduzir um parâmetro extra trás mais flexibilidade as distribuições base, e isso pode ser muito útil para fins de análise de dados.

Um recente levantamento dos principais métodos utilizados para gerar uma nova distribuição de probabilidade pode ser encontrado em Nadarajah e Rocha (2016), Tahir e Nadarajah (2015), Brito, Rêgo e Oliveira (2015), Aljarrah, Lee e Famoye (2014), Lee, Famoye e Alzaatreh (2013), Lai (2011), Gupta e Kundu (2009) e referências citadas nos mesmos.

É importante enfatizar que a transformação de uma variável aleatória  $Z$  em outra  $X$ , da forma  $X = g(Z)$ , é a forma mais simples de se gerar uma nova distribuição de probabilidade. Um exemplo de uma distribuição obtida por transformação é a distribuição Half-Normal uniparamétrica (HN), resultante da transformação  $X = \theta |Z|$ , em que  $Z$  tem distribuição Normal padrão e  $\theta > 0$ . Nos últimos anos várias extensões da distribuição Half-Normal uniparamétrica foram propostas. Entre essas extensões cita-se na literatura: a distribuição Half-Normal geral (PEWSEY, 2002, 2004), a distribuição Half-Normal generalizada (COORAY; ANANDA, 2008), a distribuição Beta (log-Beta) Half-Normal generalizada (PESCIM et al., 2010; CORDEIRO et al., 2013; PESCIM et

al., 2013), a distribuição Kumaraswamy Half-Normal generalizada (CORDEIRO; PES-CIM; ORTEGA, 2012), a distribuição Beta Half-Normal generalizada geométrica (RAMIRES et al., 2013), a extensão da distribuição Half-Normal generalizada (OLMOS et al., 2014), a distribuição Half-Normal potência (GÓMEZ; BOLFARINE, 2015), a distribuição Half-Normal generalizada estendida (SANCHEZ; FREITAS; CORDEIRO, 2016) e a distribuição odd log-logistica Half-Normal generalizada (CORDEIRO et al., 2017).

Neste trabalho propomos duas novas distribuições derivadas, respectivamente, das distribuições Half-Normal generalizada e Half-Normal. Na primeira proposta, a partir do método de discretização de Nakagawa e Osaki (1975), formulamos a versão discreta da distribuição Half-Normal generalizada. Na segunda proposta introduzimos a distribuição Half-Normal transmutada a partir da metodologia de Shaw e Buckley (2009) tomando como base a distribuição Half-Normal.

No Capítulo 2 apresentamos o método de discretização utilizado na geração da distribuição Half-Normal generalizada discreta juntamente com algumas de suas propriedades matemáticas. Via simulação Monte Carlo estudamos o viés e a raiz quadrada do erro médio quadrático dos estimadores obtidos pelos métodos de máxima verossimilhança e pelo método dos momentos. Duas aplicações utilizando dados da literatura compara a distribuição proposta com as distribuições Weibull discreta, Binomial negativa e Poisson.

No Capítulo 3 formulamos a distribuição Half-Normal transmutada e estudamos algumas de suas propriedades matemáticas. O viés e a precisão dos estimadores de máxima verossimilhança são estudados via simulação Monte Carlo. Duas aplicações, a primeira sem a presença de covariáveis e a segunda na presença de covariáveis, são usadas com propósitos ilustrativos.

## 1.2 Justificativa

Em geral, dados estritamente discretos, obtidos por processo de contagem ou por truncamento são modelados pelas distribuições Geométrica, Poisson ou Binomial negativa. Estas distribuições tem aplicabilidade limitada quando se trata de tempo de falha e confiabilidade, pois suas funções densidade e risco não proporcionam flexibilidade o suficiente para a análise de dados dessa natureza. Tendo isto em vista, neste trabalho propomos a distribuição Half-Normal generalizada discreta, para a análise de dados discretos.

Usando a recente metodologia introduzida por Shaw e Buckley (2009), a qual consiste em introduzir assimetria ou curtose em uma distribuição base (simétrica ou não), propomos a distribuição contínua Half-Normal transmutada. Um dos benefícios de se trabalhar com esta distribuição é que ela possui propriedades matemáticas interessantes, como por exemplo, uma forma analítica simples para a esperança, proporcionando introduzir regressão. Outra vantagem se deve ao fato da mesma apresentar baixo viés e baixo erro quadrático médio, mesmo para pequenos tamanhos de amostra.

## 1.3 Objetivos

### 1.3.1 Objetivo Geral

Propor duas novas distribuições de probabilidade, a distribuição Half-Normal generalizada discreta, para modelar dados de natureza estritamente discreta, e a distribuição contínua Half-Normal transmutada para análise de dados contínuos.

### 1.3.2 Objetivos Específicos

- Propor uma análoga discreta a distribuição Half-Normal generalizada, que é obtida a partir da aplicação do método de discretização via sobrevivência, e a versão transmutada da distribuição Half-Normal;
- Caracterizar e estudar algumas propriedades destas novas distribuições;
- Avaliar via simulação Monte Carlo o vício e a raiz quadrada do erro médio quadrático das estimativas obtidas pelo método da máxima verossimilhança e pelo método dos momentos, para a distribuição Half-Normal generalizada discreta;
- Avaliar via simulação Monte Carlo o vício e o erro médio quadrático das estimativas de máxima verossimilhança da distribuição Half-Normal transmutada;
- Apresentar aplicações em dados da literatura e em dados reais, para verificar a qualidade de ajuste de ambas distribuições, Half-Normal generalizada discreta e Half-Normal transmutada, em comparação com as outras distribuições presentes na literatura.

## 1.4 Metodologia

Realizamos um estudo das principais propriedades matemáticas de ambas as distribuições, Half-Normal generalizada discreta e Half-Normal transmutada, caracterizando os comportamentos das funções densidade ou de probabilidade, de risco e momentos. Para o estudo das propriedades dos estimadores da distribuição Half-Normal generalizada discreta, obtidos pelos métodos da máxima verossimilhança e dos momentos utilizamos simulação Monte Carlo, considerando os valores do vício e da raiz quadrada do erro médio quadrático. Para a distribuição Half-Normal transmutada consideramos o vício e o erro médio quadrático dos estimadores de máxima verossimilhança.

## 1.5 Resultados

Foi possível verificar que a distribuição Half-Normal generalizada discreta possui boas propriedades para seus estimadores de máxima verossimilhança, pois em todos os cenários estudados verificamos que seus estimadores são não viciados assintoticamente. Além disso, ao ser aplicada em conjuntos de dados da literatura, mostrou uma boa qualidade de ajuste, visto que, a mesma possui muita flexibilidade em termos de comportamento, tanto para a função de probabilidade quanto para a função de risco, podendo assumir nesta última as formas decrescente, crescente e banheira. De semelhante forma, a distribuição Half-Normal transmutada mostrou excelentes propriedades matemáticas, tendo em sua forma analítica a esperança, variância, assimetria e curtose. Seus estimadores são não viciados assintoticamente, e além disso, ela mostrou um excelente ajuste em dados climatológicos, se comparada com as demais extensões da distribuição Half-Normal.

## 1.6 Conclusão

A partir dos estudos realizados sobre a distribuição Half-Normal generalizada discreta conclui-se que ela oferece um ajuste significativo para a análise de dados estritamente discretos quando comparado aos modelos tradicionais usados na análise de dados dessa natureza, pois em termos de comportamento, sua função de probabilidade e de risco mostraram ter bastante flexibilidade. Já a distribuição Half-Normal transmutada possui boas propriedades para seus estimadores, além de expressões fechadas para esperança, variância, assimetria e curtose, o que a torna uma forte concorrente para as demais distribuições de probabilidade testadas, sendo uma boa alternativa na modelagem de dados climatológicos.

## 1.7 Propostas Futuras

Com o objetivo de dar continuidade ao estudo das distribuições Half-Normal generalizada discreta e Half-Normal transmutada, estabelecemos como propostas para um trabalho futuro:

- Aplicar a distribuição Half-Normal generalizada discreta no contexto de risco competitivos, longa-duração e em dados inflacionados de zeros;
- Avaliar a performance de diferentes métodos de estimação dos parâmetros da distribuição Half-Normal transmutada.

---

---

## CAPÍTULO 2

---

# A DISTRIBUIÇÃO HALF-NORMAL GENERALIZADA DISCRETA

### Resumo

Em geral dados obtidos por processos de contagem, estritamente discretos ou discretizados (provenientes de truncamentos e/ou arredondamentos), são modelados, sem sermos exaustivos, pelas distribuições Geométrica, Logarítmica, Poisson e Binomial negativa. Nos últimos anos, a partir de procedimentos de discretização de variáveis aleatórias contínuas, um grande número de distribuições discretas vêm sendo propostas na literatura. Dada uma variável aleatória  $X$  contínua, é possível obter a versão discreta da distribuição de  $X$ . Muitos dos métodos de discretização preservam uma ou mais características da versão contínua, sendo a proposta de Nakagawa e Osaki (1975) a mais utilizada. Neste artigo, a partir desta metodologia, a qual faz uso da função de sobrevivência, é proposta a versão discreta da distribuição Half-Normal generalizada a qual foi introduzida na literatura por Cooray e Ananda (2008). Algumas de suas propriedades são apresentadas e via simulações Monte Carlo são avaliados o viés e a acurácia das estimativas obtidas pelo método da máxima verossimilhança e pelo método dos momentos. Dois exemplos utilizando dados da literatura ilustram a aplicação da distribuição proposta.

**Palavras-chave:** Discretização, Distribuição Half-Normal Generalizada, Método dos Momentos, Simulação Monte Carlo, Verossimilhança.

## Abstract

Generally, obtained data by count processes, strictly discrete or discretized, obtained by truncation and/or rounding are modeled by Geometric, Poisson and Binomial negative distributions. In the last years, a big number of new discrete distributions has been proposed in literature. These new distributions are obtained from procedures of discretization of random continuous variable. Given a random continuous variable  $X$ , it is possible to obtain a discrete version of distribution of  $X$ . Many of discretization methods preserve one or more characteristics of continuous version, being the proposal from Nakagawa e Osaki (1975) the most used. In this paper, from this methodology which makes use of survival function, we propose the discrete version of generalized Half-Normal distribution, introduced in literature by Cooray e Ananda (2008). Some of its properties are presented and the bias and the root mean square error of estimates obtained by maximum likelihood and by method of moments are evaluated by Monte Carlo simulation. Two examples using data of literature illustrate the application of the proposed distribution.

**Keywords:** Discretization, Generalized Half-Normal Distribution, Methods of Moments, Monte Carlo Simulation, Likelihood.

## 2.1 Introdução

O número de vezes que um dispositivo eletrônico é ligado antes de apresentar algum defeito, o número de dias que um paciente fica em observação na UTI antes de receber alta e o número de semanas, meses ou anos que um transplantado sobrevive, são apenas alguns exemplos de dados de sobrevivência discretos que na maioria das vezes são analisados por meio de distribuições contínuas. Klein e Moeschberger (1997) traz em seu primeiro capítulo alguns exemplos de dados de sobrevivência, entretanto, analisados no decorrer do texto via distribuições contínuas. O mesmo ocorre nos vários exemplos apresentados, entre outros, em Collett (2003), Lawless (2003) e Lee e Wang (2003).

Nos últimos anos várias distribuições discretas têm sido propostas na literatura, e, apesar de suas atraentes potencialidades, não são ainda muito usadas na análise de dados de sobrevivência discretos/discretizados ou em dados de contagem. Um dos primeiros, senão o primeiro artigo científico que propõe uma distribuição discreta (derivado da distribuição Weibull) data da década de 1970. Desde então uma infinidade de novas distribuições discretas foram propostas. Uma revisão dos vários métodos de discretização e das distribuições discretizadas por eles estão presentes em Chakraborty (2015), Lai (2013) e Bracquemond e Gaudoin (2003). Nesses trabalhos são discutidos alguns métodos: o baseado na função de sobrevivência (NAKAGAWA; OSAKI, 1975), o baseado em uma série infinita (GOOD, 1953), o baseado na função de risco (STEIN; DATTERO, 1984), o

composto por dois estágios (CHAKRABORTY, 2015), o baseado na função de risco reversa (GHOSH; ROY; CHANDRA, 2013), o via função de distribuição acumulada (ROY; DASGUPTA, 2001) e o método análogo ao de Pearson (PEARSON, 1895).

O objetivo deste artigo é, a partir do método de discretização baseado na função de sobrevivência, introduzir a distribuição discreta análoga a distribuição Half-Normal generalizada contínua, introduzida por Cooray e Ananda (2008). Na Seção 2.2 apresentamos uma breve revisão a respeito do método de Nakagawa e Osaki (1975). A Seção 2.3 apresenta a versão contínua da distribuição Half-Normal generalizada (HNG) e logo em seguida é introduzida a versão discreta. Na Seção 2.4 o método da máxima verossimilhança e o método dos momentos são apresentados. Um estudo de simulação Monte Carlo na Seção 2.5 é usado para avaliar o viés e a acurácia dos estimadores obtidos por esses dois métodos. A Seção 2.6 apresenta três aplicações considerando dados reais da literatura e a Seção 2.7 encerra este artigo com algumas conclusões.

## 2.2 Discretização via Função de Sobrevidência

Um dos principais métodos propostos na literatura para se obter um análogo discreto de uma variável aleatória contínua é o baseado na função de sobrevidência. Este método foi proposto por Nakagawa e Osaki (1975) e tem como principal característica, a preservação da função de sobrevidência original em sua parte inteira (KEMP, 2004; CHAKRABORTY, 2015). Alguns exemplos de distribuições discretas geradas por este método são: a distribuição Weibull discreta (NAKAGAWA; OSAKI, 1975), a distribuição Weibull geométrica discreta (BRACQUEMOND; GAUDOIN, 2003), a distribuição Gumbel discreta (CHAKRABORTY; CHAKRAVARTY, 2014), a distribuição gama discreta (CHAKRABORTY; CHAKRAVARTY, 2012), a distribuição Lindley discreta (GÓMEZ-DÉNIZ; CALDERÍN-OJEDA, 2011; BAKOUCH; JAZI; NADARAJAH, 2014), a distribuição Rayleigh inversa discreta (HUSSAIN; AHMAD, 2014), a distribuição Weibull inversa discreta (JAZI; LAI; ALAMATSAZ, 2010), a distribuição Maxwell discreta (KRISHNA; PUNDIR, 2007) e as distribuições Burr e Pareto discretas (KRISHNA; PUNDIR, 2009).

Segundo a convenção de Kemp (2004), o processo de discretização a partir da função de sobrevidência utiliza a seguinte definição.

**Definição 2.2.1.** *Seja  $S_X(x) = P[X \geq x]$  a função de sobrevidência de uma variável aleatória contínua  $X$ . Definimos a função probabilidade de  $Y = \lfloor X \rfloor$  por:*

$$P(Y = k) = S_X(k) - S_X(k + 1) \quad (2.1)$$

em que  $k = 0, 1, \dots$  e  $\lfloor \cdot \rfloor$  indica a parte inteira menor ou igual ao argumento.

Uma propriedade importante deste método é que o mesmo mantém a forma da função de sobrevidência da versão contínua, ou seja  $S_Y(k) = S_X(k)$ . Dado que, a função de

sobrevidência mantém sua forma para o caso discreto, a função de distribuição acumulada definida como  $F_Y(k) = 1 - S_Y(k)$  também mantém sua forma para o caso discreto, ou seja,  $F_Y(k) = F_X(k)$ . Essa propriedade é útil para se gerar valores pseudo-aleatórios da função de probabilidade definida em (2.1).

Tendo em vista que a função de risco pode ser obtida pela razão entre a função de probabilidade e a função de sobrevidência, temos então a seguinte definição:

**Definição 2.2.2.** *Seja  $Y$  uma variável aleatória discreta, com função de probabilidade e de sobrevidência  $p(k)$  e  $S_Y(k)$ , respectivamente. A função de risco de  $Y$  é definida como a probabilidade condicional de que a falha é observada em  $k$ , dado que esta não ocorreu antes de  $k$ , e é escrita na forma:*

$$h_Y(k) = \frac{p(k)}{S_Y(k)} = \frac{S_Y(k) - S_Y(k+1)}{S_Y(k)} = 1 - \frac{S_Y(k+1)}{S_Y(k)} \quad k \in \mathbb{Z}. \quad (2.2)$$

É importante observar que, por esta definição,  $h_Y(k)$  é sempre restrita ao intervalo zero-um, ou seja  $0 \leq h_Y(k) \leq 1$ .

## 2.3 A Distribuição Half-Normal Generalizada

### 2.3.1 Versão Contínua

Cooray e Ananda (2008) introduziram a distribuição Half-Normal generalizada (HNG) de dois parâmetros que tem como caso particular a distribuição Half-Normal (HN), desenvolvida por Daniel (1959).

Para uma variável aleatória  $X$  não negativa que segue a distribuição HNG com parâmetro de escala  $\theta > 0$  e parâmetro de forma  $\alpha > 0$  temos que a função densidade de probabilidade (fdp) e a função distribuição acumulada (fda) são escritas, respectivamente, como:

$$f(x | \alpha, \theta) = \frac{2\alpha}{\theta} \left(\frac{x}{\theta}\right)^{\alpha-1} \phi\left[\left(\frac{x}{\theta}\right)^\alpha\right] \quad \text{e} \quad F(x | \alpha, \theta) = 2\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right] - 1 \quad (2.3)$$

em que  $\phi(\cdot)$  e  $\Phi(\cdot)$  são, respectivamente, a fdp e a fda de uma variável aleatória Normal padrão.

Mostra-se que a fdp (2.3) é monotonicamente decrescente para  $\alpha < 1$  e unimodal para  $\alpha \geq 1$  com ponto de inflexão em  $\theta \left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{2\alpha}}$ . A função de risco, escrita como:

$$h(x | \alpha, \theta) = \frac{\alpha}{\theta} \left(\frac{x}{\theta}\right)^{\alpha-1} \phi\left[\left(\frac{x}{\theta}\right)^\alpha\right] \Phi\left[-\left(\frac{x}{\theta}\right)^\alpha\right]^{-1} \quad (2.4)$$

pode ser monotonicamente crescente, monotonicamente decrescente ou assumir a forma de banheira (COORAY; ANANDA, 2008).

É evidente que se  $\alpha = 1$  a distribuição HNG tem como caso particular a distribuição HN com parâmetro de escala  $\theta$ . Outras características da distribuição HNG são discutidas em Cooray e Ananda (2008). Estes autores também apresentaram o método da máxima verossimilhança e estudaram, por meio de simulações, as probabilidades de cobertura dos intervalos de confiança assintóticos. Recentemente, (MAZUCHELI; DEY, 2017) deduziram, pela metodologia sugerida por Cox e Snell (1968), as expressões analíticas para a correção do viés dos estimadores de máxima verossimilhança de  $\alpha$  e  $\theta$ .

### 2.3.2 Versão Discreta

Para a geração da distribuição Half-Normal generalizada discreta (HNGD), foi considerada a distribuição HNG e a metodologia que faz uso da função de sobrevivência.

**Definição 2.3.1.** *Uma variável aleatória  $X$  que toma valores no conjunto de inteiros  $\mathbb{Z}_+$  é dita seguir uma distribuição HNGD com parâmetros  $\alpha$  e  $\theta$  se sua função de probabilidade é escrita na forma:*

$$P(X = x | \alpha, \theta) = 2\Phi\left[\left(\frac{x+1}{\theta}\right)^\alpha\right] - 2\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right] \quad (2.5)$$

em que  $\alpha > 0$  e  $\theta > 0$  são os parâmetros de forma e escala, respectivamente.

Ilustramos na Figura 2.1 alguns comportamentos da função de probabilidade considerando diferentes valores de  $\alpha$  e  $\theta$ . Vale ressaltar que o comportamento de (2.5) é o mesmo que no caso contínuo ou seja, monótona descrescente para  $\alpha < 1$  e unimodal caso contrário.

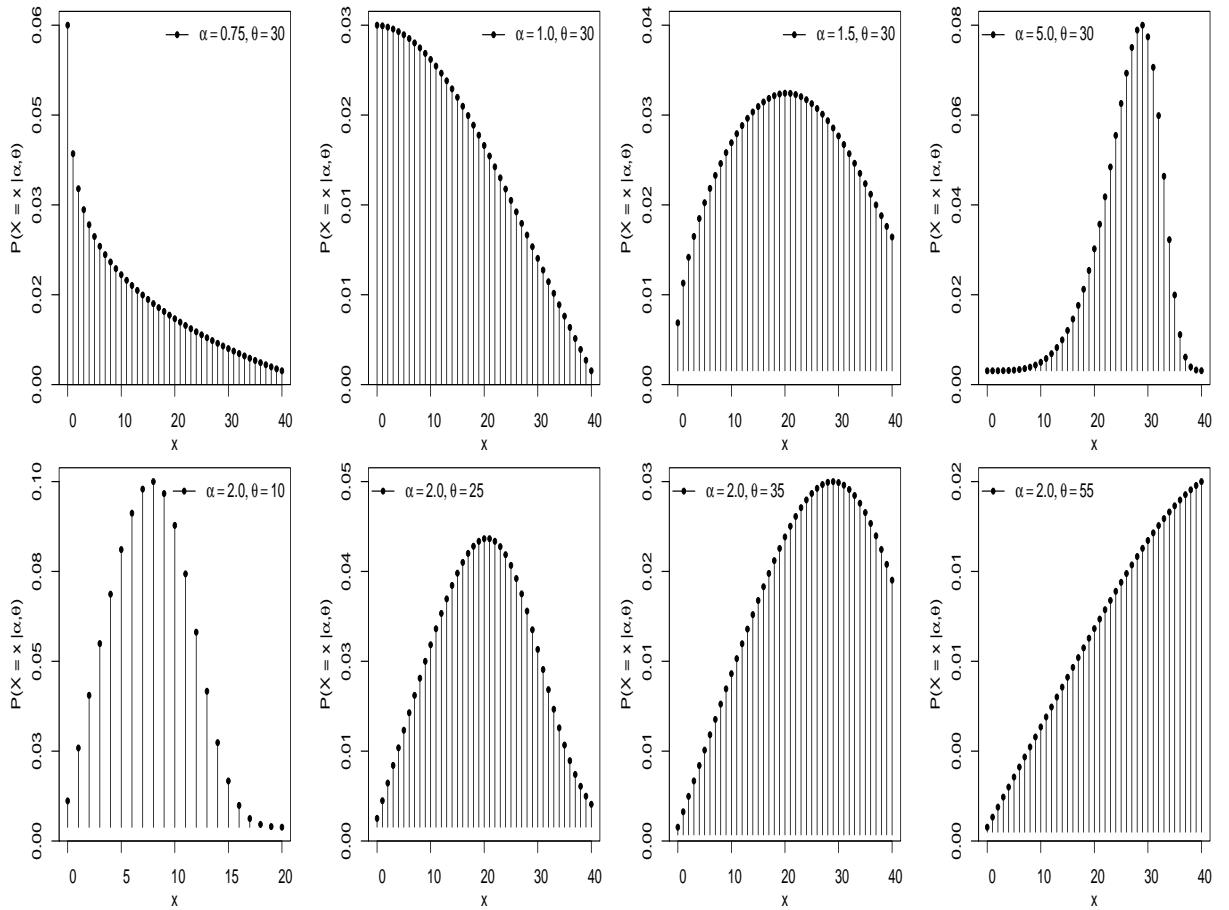


Figura 2.1: Comportamento da função de probabilidade da distribuição HNGD.

As funções de distribuição acumulada e de sobrevivência são dadas, respectivamente, por:

$$F(x | \alpha, \theta) = 2\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right] - 1 \quad \text{e} \quad S(x | \alpha, \theta) = 2\Phi\left[-\left(\frac{x}{\theta}\right)^\alpha\right]. \quad (2.6)$$

É importante mencionar que para uma variável aleatória discreta a função de risco possui alguns equívocos, (HUSSAIN; AHMAD, 2014), como por exemplo, o fato dela ser limitada,  $h(x) < 1$ , implica que ela não é equivalente a  $-\log S(x)$  (como no caso contínuo). Para contornar esse problema foi proposta a chamada de segunda função de risco, dada por:

$$h^*(x | \alpha, \theta) = \log\left[\frac{S(x | \alpha, \theta)}{S(x+1 | \alpha, \theta)}\right] \quad x = 0, 1, 2, \dots \quad (2.7)$$

Para a distribuição HNGD temos:

$$h^*(x | \alpha, \theta) = \log\left\{\Phi\left[-\left(\frac{x}{\theta}\right)^\alpha\right]\right\} - \log\left\{\Phi\left[-\left(\frac{x+1}{\theta}\right)^\alpha\right]\right\}. \quad (2.8)$$

O comportamento da função de risco (painel superior) e da segunda função de risco (painel inferior), considerando vários valores de  $\alpha$  e  $\theta$ , é ilustrado na Figura 2.2. É fácil perceber que os comportamentos de ambas as funções são bastante similares e úteis em situações de risco descrescente, crescente e em forma de banheira.

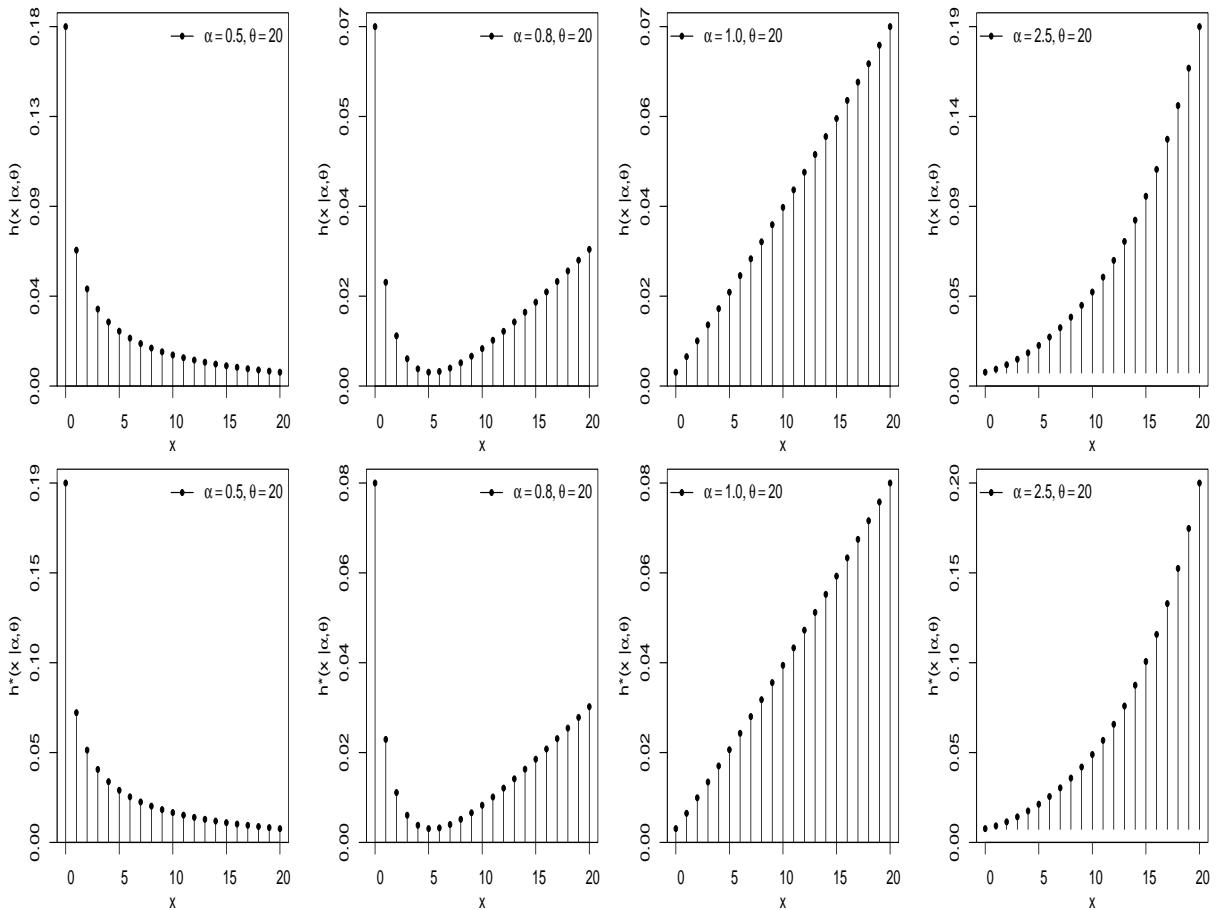


Figura 2.2: Comportamento das funções de risco da distribuição HNGD.

### 2.3.3 Função Geradora de Momentos e Algumas Propriedades

Algumas funções auxiliares desempenham importante papel na obtenção de diversas características interessantes das variáveis aleatórias. Elas surgem a partir do estudo da esperança matemática ou valor esperado de uma variável aleatória. Entre estas funções citamos a função geradora de momentos, definida como segue:

**Definição 2.3.2.** Seja  $X$  uma variável aleatória discreta que assume valores em  $\mathbb{Z}$ . A função geradora de momentos de  $X$ , quando existe, é então calculada da seguinte forma:

$$M_X(t) = \mathbb{E}(e^{tX}) = \sum_x e^{tX} P(X = x) \quad (2.9)$$

em que  $P(X = x)$  é a função de probabilidade da variável aleatória  $X$ .

A função geradora de momentos da distribuição HNGD pode ser escrita como:

$$M_X(t) = \mathbb{E}(e^{tX}) = \sum_{i=0}^{\infty} 2e^{tx_i} \left\{ \Phi\left[\left(\frac{x_i+1}{\theta}\right)^{\alpha}\right] - \Phi\left[\left(\frac{x_i}{\theta}\right)^{\alpha}\right] \right\}. \quad (2.10)$$

Como sabemos, o primeiro momento coincide com o valor esperado de  $X$ , logo:

$$\mathbb{E}(X) = \sum_{i=0}^{\infty} 2x_i \left\{ \Phi\left[\left(\frac{x_i+1}{\theta}\right)^{\alpha}\right] - \Phi\left[\left(\frac{x_i}{\theta}\right)^{\alpha}\right] \right\}$$

enquanto que a variância é dada por:

$$\begin{aligned} Var(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \sum_{i=0}^{\infty} 2x_i^2 \left\{ \Phi\left[\left(\frac{x_i+1}{\theta}\right)^{\alpha}\right] - \Phi\left[\left(\frac{x_i}{\theta}\right)^{\alpha}\right] \right\} + \\ &\quad - \left\{ \sum_{i=0}^{\infty} 2x_i \left[ \Phi\left[\left(\frac{x_i+1}{\theta}\right)^{\alpha}\right] - \Phi\left[\left(\frac{x_i}{\theta}\right)^{\alpha}\right] \right] \right\}^2. \end{aligned}$$

Podemos observar que tanto a esperança quanto a variância não possuem formas fechadas o que impossibilita a verificação da subdispersão ou superdispersão algebricamente. Tendo isso em vista, é apresentado na Tabela 2.1 o índice de dispersão, calculado numericamente. Nota-se que distribuição HNGD pode ser usada para dados com subdispersão ou superdispersão. Dados com superdispersão são comumente encontrados na literatura, e existem várias distribuições que podem ser usadas para modelá-los, porém, há poucos modelos que comportam a subdispersão, logo a distribuição HNGD pode ser uma boa alternativa para este tipo de problema, visto que a grande parte de suas combinações paramétricas apresentam subdispersão.

Tabela 2.1: Índice de dispersão da distribuição HNGD.

$\theta \backslash \alpha$	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0
0.5	<b>1.66</b>	0.96	1.00	1.00	0.61	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.0	<b>2.68</b>	0.91	0.71	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.68
1.5	<b>3.70</b>	<b>1.07</b>	0.66	0.47	0.35	0.26	0.20	0.16	0.13	0.10	0.09	0.07
2.0	<b>4.72</b>	<b>1.27</b>	0.72	0.50	0.37	0.30	0.26	0.23	0.21	0.20	0.19	0.18
2.5	<b>5.74</b>	<b>1.48</b>	0.81	0.54	0.40	0.30	0.24	0.19	0.16	0.13	0.11	0.10
3.0	<b>6.75</b>	<b>1.69</b>	0.91	0.60	0.43	0.33	0.27	0.23	0.20	0.18	0.16	0.15
3.5	<b>7.77</b>	<b>1.92</b>	<b>1.02</b>	0.66	0.47	0.36	0.29	0.24	0.20	0.16	0.14	0.12
4.0	<b>8.78</b>	<b>2.14</b>	<b>1.13</b>	0.73	0.52	0.39	0.31	0.26	0.22	0.19	0.17	0.15
4.5	<b>9.78</b>	<b>2.36</b>	<b>1.24</b>	0.80	0.57	0.43	0.34	0.28	0.23	0.20	0.17	0.15
5.0	<b>10.79</b>	<b>2.59</b>	<b>1.35</b>	0.87	0.62	0.46	0.37	0.30	0.25	0.21	0.18	0.16
5.5	<b>11.80</b>	<b>2.81</b>	<b>1.46</b>	0.94	0.66	0.50	0.39	0.32	0.27	0.23	0.19	0.17
6.0	<b>12.80</b>	<b>3.04</b>	<b>1.58</b>	<b>1.01</b>	0.71	0.54	0.42	0.34	0.28	0.24	0.21	0.18

O comportamento da média, variância, assimetria e curtose, em função de  $\alpha$  é ilustrado na Figura 2.3. Pelo gráfico do coeficiente de assimetria temos que o mesmo pode ser positivo ou negativo, logo útil na análise de dados com assimetria negativa ou positiva.

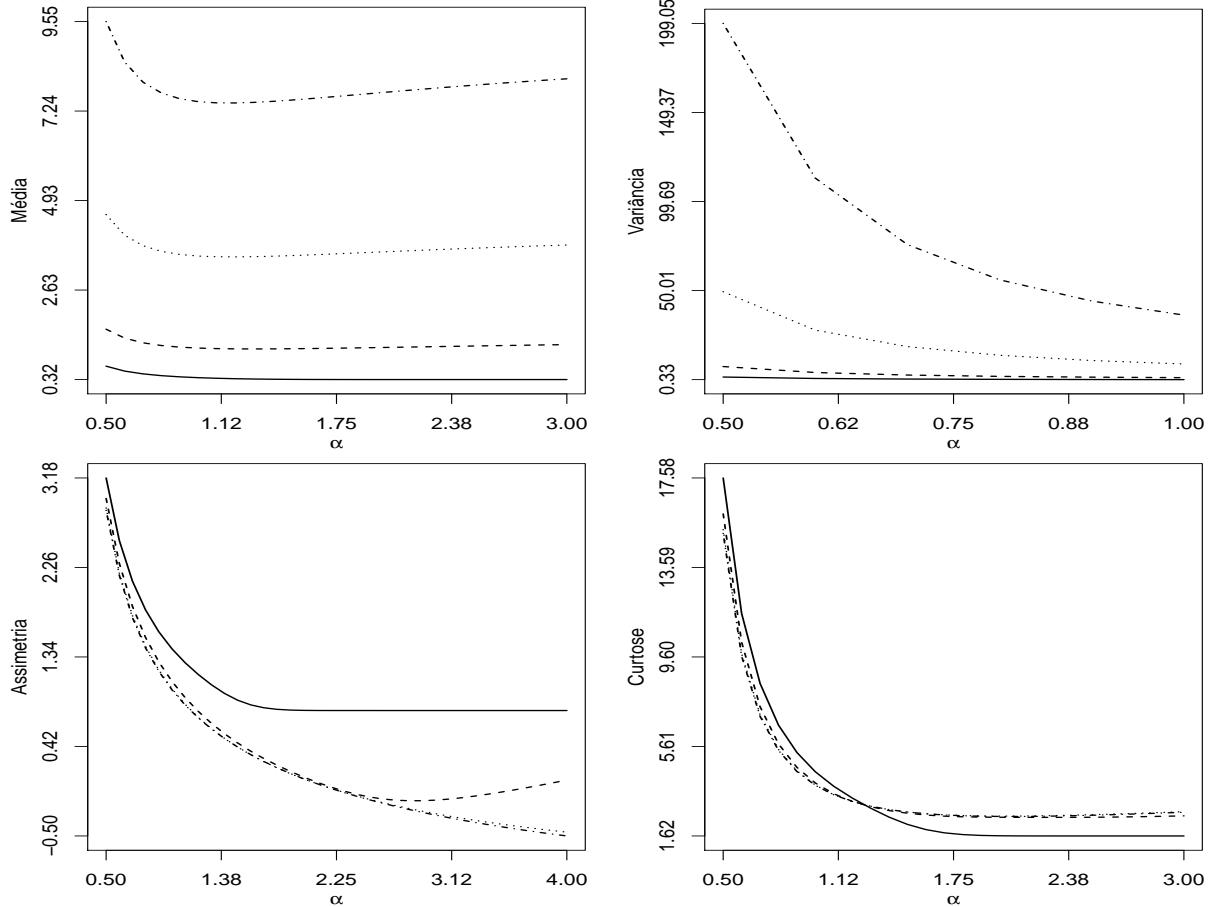


Figura 2.3: Comportamento da média, variância, assimetria e curtose da distribuição HNGD.  $\theta = 1.0$  (—),  $\theta = 2.0$  (---),  $\theta = 5$  (···) e  $\theta = 10.0$  (----).

## 2.4 Métodos de Estimação

### 2.4.1 Estimação Via Método da Máxima Verossimilhança

Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da distribuição HNGD com parâmetros  $\alpha$  e  $\theta$  e função de probabilidade dada por (2.5). A função de verossimilhança pode ser escrita na forma:

$$L(\alpha, \theta | \mathbf{x}) = \prod_{i=1}^n 2\Phi\left[\left(\frac{x_i + 1}{\theta}\right)^\alpha\right] - 2\Phi\left[\left(\frac{x_i}{\theta}\right)^\alpha\right]. \quad (2.11)$$

Aplicando a função logarítmica em (2.11) tem-se que o logaritmo da função de veros-

similarhança é dado por:

$$\ell(\alpha, \theta | \mathbf{x}) = \sum_{i=1}^n \log \left[ 2\Phi \left[ \left( \frac{x_i + 1}{\theta} \right)^\alpha \right] - 2\Phi \left[ \left( \frac{x_i}{\theta} \right)^\alpha \right] \right]. \quad (2.12)$$

As funções escores para os parâmetros  $\alpha$  e  $\theta$ , são dadas respectivamente por:

$$\begin{aligned} \frac{\partial \ell(\alpha, \theta | \mathbf{x})}{\partial \alpha} &= \sum_{i=1}^n \frac{\phi \left[ \left( \frac{x_i + 1}{\theta} \right)^\alpha \right] \left( \frac{x_i + 1}{\theta} \right)^\alpha \log \left( \frac{x_i + 1}{\theta} \right) - \phi \left[ \left( \frac{x_i}{\theta} \right)^\alpha \right] \left( \frac{x_i}{\theta} \right)^\alpha \log \left( \frac{x_i}{\theta} \right)}{\Phi \left[ \left( \frac{x_i + 1}{\theta} \right)^\alpha \right] - \Phi \left[ \left( \frac{x_i}{\theta} \right)^\alpha \right]} \\ \frac{\partial \ell(\alpha, \theta | \mathbf{x})}{\partial \theta} &= \sum_{i=1}^n \frac{\phi \left[ \left( \frac{x_i}{\theta} \right)^\alpha \right] \left( \frac{x_i}{\theta} \right)^\alpha \frac{\alpha}{\theta} - \phi \left[ \left( \frac{x_i + 1}{\theta} \right)^\alpha \right] \left( \frac{x_i + 1}{\theta} \right)^\alpha \frac{\alpha}{\theta}}{\Phi \left[ \left( \frac{x_i + 1}{\theta} \right)^\alpha \right] - \Phi \left[ \left( \frac{x_i}{\theta} \right)^\alpha \right]}, \end{aligned}$$

em que  $\phi(\cdot)$  e  $\Phi(\cdot)$  denotam, respectivamente, a fdp e a fda da Normal padrão.

Os estimadores de máxima verossimilhança de  $\alpha$  e  $\theta$  são as soluções das equações não-lineares  $\frac{\partial}{\partial \alpha} \ell(\alpha, \theta | \mathbf{x}) = 0$  e  $\frac{\partial}{\partial \theta} \ell(\alpha, \theta | \mathbf{x}) = 0$ . Observa-se que para obter as estimativas de máxima verossimilhança é necessário a utilização de métodos numéricos.

#### 2.4.2 Estimação Via Método dos Momentos

Para aplicar o método dos momentos para estimar  $\alpha$  e  $\theta$  temos que equacionar os momentos da população com os momentos amostrais e resolver duas equações simultaneamente. Como os momentos populacionais são expressos em uma série infinita as equações não podem ser resolvidas por exemplo, pelo método da substituição. Denotemos por  $\mu_1$  e  $\mu_2$  o primeiro e segundo momento sobre a origem, respectivamente, temos:

$$\mu_1 = 2 \sum_{i=0}^{\infty} x_i \left\{ \Phi \left[ \left( \frac{x_i + 1}{\theta} \right)^\alpha \right] - \Phi \left[ \left( \frac{x_i}{\theta} \right)^\alpha \right] \right\}, \quad (2.13)$$

$$\mu_2 = 2 \sum_{i=0}^{\infty} x_i^2 \left\{ \Phi \left[ \left( \frac{x_i + 1}{\theta} \right)^\alpha \right] - \Phi \left[ \left( \frac{x_i}{\theta} \right)^\alpha \right] \right\}. \quad (2.14)$$

Os momentos amostrais correspondentes aos momentos populacionais, são dados por:

$$M_1 = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{e} \quad M_2 = \frac{1}{n} \sum_{i=1}^n x_i^2. \quad (2.15)$$

Para obtermos as estimativas de  $\alpha$  e  $\theta$ , podemos utilizar a proposta em Khan, Khalique

e Abouammoh (1989), que consiste em minimizar, em  $\alpha$  e  $\theta$ , a equação:

$$(M_1 - \mu_1)^2 + (M_2 - \mu_2)^2. \quad (2.16)$$

## 2.5 Estudo de Simulação

Nesta seção são apresentados os resultados de um estudo de simulação Monte Carlo usado para avaliar o viés e o erro quadrático médio das estimativas obtidas pelo método da máxima verossimilhança e pelo método dos momentos. Foram geradas amostras de tamanho  $n = 20, 50, \dots, 170, 200$  com  $\alpha = 0.5, 1.0, 3.0, 5.0, 10.0$  e  $\theta = 1.0$  e  $10.0$ . Para cada uma das combinações de  $n, \alpha$  e  $\theta$  foram simuladas, pelo método da transformação inversa,  $N = 10000$  amostras pseudo-aleatórias da distribuição HNGD. Foram estimados o viés e a raiz do erro quadrático médio. Estas estimativas são apresentadas nas Tabelas 2.2 e 2.3.

Ao observarmos o vício de  $\hat{\alpha}$ , percebemos que  $\theta$  exerce influência sobre o mesmo. Para  $\theta = 1.0$  e  $\alpha > 1$ , temos oscilações na estimativa do vício de  $\alpha$  para ambos os métodos de estimação, fazendo com que o vício se torne instável, isto é, decresça e cresça a cada acréscimo amostral. De forma semelhante, o vício de  $\hat{\theta}$  também oscila, entretanto, a amplitude dessa oscilação é pequena e centrada em zero. Mediante isto, podemos perceber que a influência de  $\alpha$  em  $\theta$  é mínima.

Já para  $\theta = 1.0$  e  $\alpha > 1$ , o REQM de  $\hat{\alpha}$  possui oscilações em suas estimativas. Para  $\theta = 10$  e  $\alpha > 1$ , podemos observar que a convergência do REQM para zero é mais demorada. No entanto, se  $\theta = 10$  e  $\alpha \leq 1$  o REQM de  $\hat{\theta}$  não convergiu diretamente para zero, mesmo com tamanho amostral 200.

Em geral, o comportamento do vício de  $\hat{\alpha}$  foi o mesmo em ambos os métodos, tendo como única diferença as amplitudes. Apesar do vício de  $\hat{\theta}$  ser menor no método da máxima verossimilhança, temos mais cenários convergindo para zero no métodos dos momentos. O REQM de  $\hat{\alpha}$  tem comportamento idêntico em ambos os métodos, sendo a convergência para zero mais rápida no método da máxima verossimilhança. Para o REQM de  $\hat{\theta}$ , novamente, temos um comportamento muito semelhante entre os dois métodos, sendo a convergência para zero pelo método da máxima verossimilhança pouca coisa mais rápida em quase todos os cenários.

Tabela 2.2: Estimativas do viés e da raiz do erro quadrático médio (REQM) dos estimadores de máxima verossimilhança (MV) e dos estimadores obtidos pelo método dos momentos (MM) para  $\theta = 1.0$ .

$\alpha$	$n$	MV				MM			
		Viés		REQM		Viés		REQM	
		$\alpha$	$\theta$	$\alpha$	$\theta$	$\alpha$	$\theta$	$\alpha$	$\theta$
0.5	20	0.0747	0.0282	0.0407	0.1451	0.1139	0.0960	0.0417	0.1319
	50	0.0317	0.0150	0.0129	0.0689	0.0526	0.0511	0.0141	0.0689
	80	0.0189	0.0093	0.0070	0.0436	0.0333	0.0336	0.0080	0.0451
	110	0.0135	0.0067	0.0048	0.0320	0.0247	0.0252	0.0057	0.0337
	140	0.0106	0.0057	0.0037	0.0252	0.0197	0.0205	0.0044	0.0270
	170	0.0087	0.0045	0.0030	0.0207	0.0164	0.0169	0.0036	0.0224
	200	0.0075	0.0041	0.0025	0.0176	0.0141	0.0147	0.0030	0.0193
1.0	20	-0.0134	0.0300	0.0716	0.0532	0.0145	0.0278	0.0794	0.0494
	50	0.0321	0.0024	0.0389	0.0186	0.0326	0.0030	0.0380	0.0181
	80	0.0355	-0.0010	0.0289	0.0114	0.0349	-0.0009	0.0283	0.0113
	110	0.0308	-0.0012	0.0222	0.0083	0.0302	-0.0012	0.0218	0.0082
	140	0.0257	-0.0012	0.0176	0.0065	0.0251	-0.0013	0.0173	0.0064
	170	0.0215	-0.0009	0.0142	0.0054	0.0209	-0.0011	0.0140	0.0053
	200	0.0180	-0.0008	0.0118	0.0046	0.0175	-0.0009	0.0116	0.0045
3.0	20	0.0619	-0.0029	0.0095	0.0046	0.0716	0.0011	0.0150	0.0052
	50	0.0175	0.0007	0.0013	0.0021	0.0175	0.0007	0.0013	0.0021
	80	0.0618	0.0004	0.0092	0.0012	0.0618	0.0004	0.0092	0.0012
	110	-0.0096	0.0004	0.0003	0.0010	-0.0112	0.0004	0.0004	0.0010
	140	0.0126	-0.0017	0.0102	0.0006	0.0114	0.0011	0.0097	0.0007
	170	0.0491	0.0002	0.0057	0.0006	0.0492	0.0002	0.0057	0.0006
	200	0.0139	0.0002	0.0006	0.0005	0.0137	0.0002	0.0006	0.0005
5.0	20	-0.0006	0.0005	0.0005	0.0022	-0.0005	0.0005	0.0005	0.0022
	50	-0.0160	0.0005	0.0072	0.0008	-0.0160	0.0005	0.0072	0.0008
	80	-0.0079	0.0001	0.0002	0.0005	-0.0079	0.0001	0.0002	0.0005
	110	0.0494	0.0001	0.0066	0.0003	0.0494	0.0001	0.0066	0.0003
	140	0.0148	0.0001	0.0007	0.0003	0.0148	0.0001	0.0007	0.0003
	170	-0.0211	0.0001	0.0011	0.0002	-0.0211	0.0001	0.0011	0.0002
	200	-0.0260	0.0003	0.0082	0.0002	-0.0260	0.0003	0.0082	0.0002
10.0	20	-0.0002	0.0001	0.0001	0.0005	-0.0002	0.0001	0.0001	0.0005
	50	-0.0301	0.0001	0.0024	0.0002	-0.0301	0.0001	0.0024	0.0002
	80	-0.0042	0.0001	0.0000	0.0001	-0.0042	0.0001	0.0000	0.0001
	110	0.0270	0.0000	0.0022	0.0001	0.0270	0.0000	0.0022	0.0001
	140	0.0078	0.0002	0.0002	0.0001	0.0078	0.0002	0.0002	0.0001
	170	-0.0114	0.0000	0.0004	0.0001	-0.0114	0.0000	0.0004	0.0001
	200	-0.0309	0.0000	0.0028	0.0000	-0.0311	0.0000	0.0028	0.0000

Tabela 2.3: Estimativas do viés e da raiz do erro quadrático médio (REQM) dos estimadores de máxima verossimilhança (MV) e dos estimadores obtidos pelo método dos momentos (MM) para  $\theta = 10$ .

$\alpha$	$n$	EMV				EMM			
		Viés		REQM		Viés		REQM	
		$\alpha$	$\theta$	$\alpha$	$\theta$	$\alpha$	$\theta$	$\alpha$	$\theta$
0.5	20	0.0477	-0.0243	0.0148	0.3315	0.0818	-0.0123	0.0210	0.3284
	50	0.0173	-0.0252	0.0044	0.3225	0.0361	-0.0101	0.0065	0.3232
	80	0.0108	-0.0252	0.0025	0.3206	0.0238	-0.0074	0.0038	0.3196
	110	0.0077	-0.0282	0.0018	0.3147	0.0179	-0.0059	0.0027	0.3141
	140	0.0058	-0.0224	0.0014	0.3087	0.0142	-0.0040	0.0021	0.3120
	170	0.0046	-0.0243	0.0011	0.3045	0.0116	-0.0054	0.0017	0.3039
	200	0.0038	-0.0244	0.0009	0.2992	0.0098	-0.0063	0.0015	0.3021
1.0	20	0.0817	-0.0201	0.0503	0.3182	0.0825	-0.0178	0.0484	0.3168
	50	0.0309	-0.0210	0.0157	0.2987	0.0319	-0.0202	0.0158	0.2992
	80	0.0188	-0.0157	0.0093	0.2781	0.0195	-0.0161	0.0095	0.2788
	110	0.0133	-0.0146	0.0066	0.2587	0.0139	-0.0134	0.0068	0.2602
	140	0.0103	-0.0139	0.0052	0.2417	0.0107	-0.0142	0.0054	0.2428
	170	0.0082	-0.0131	0.0042	0.2256	0.0086	-0.0128	0.0044	0.2263
	200	0.0069	-0.0104	0.0036	0.2106	0.0072	-0.0101	0.0038	0.2117
3.0	20	0.0830	-0.0431	0.2107	0.2194	0.0458	-0.0433	0.2228	0.2220
	50	0.0726	-0.0214	0.1278	0.1261	0.0585	-0.0248	0.1410	0.1282
	80	0.0543	-0.0122	0.0863	0.0831	0.0482	-0.0144	0.0986	0.0846
	110	0.0403	-0.0085	0.0624	0.0608	0.0372	-0.0100	0.0726	0.0619
	140	0.0321	-0.0070	0.0487	0.0479	0.0301	-0.0082	0.0570	0.0487
	170	0.0267	-0.0051	0.0397	0.0392	0.0252	-0.0060	0.0468	0.0399
	200	0.0223	-0.0042	0.0334	0.0333	0.0210	-0.0050	0.0394	0.0339
5.0	20	0.0187	-0.0497	0.2859	0.1139	-0.0028	-0.0524	0.2991	0.1219
	50	0.0368	-0.0228	0.2284	0.0487	0.0147	-0.0242	0.2507	0.0517
	80	0.0464	-0.0133	0.1854	0.0307	0.0309	-0.0148	0.2131	0.0323
	110	0.0444	-0.0092	0.1524	0.0225	0.0341	-0.0106	0.1812	0.0235
	140	0.0422	-0.0064	0.1275	0.0177	0.0365	-0.0079	0.1564	0.0184
	170	0.0380	-0.0047	0.1083	0.0146	0.0348	-0.0060	0.1360	0.0152
	200	0.0342	-0.0037	0.0933	0.0124	0.0331	-0.0046	0.1189	0.0129
10.0	20	-0.0084	-0.0370	0.3083	0.0363	-0.0073	-0.0333	0.3097	0.0384
	50	0.0000	-0.0127	0.3091	0.0140	-0.0134	-0.0143	0.3108	0.0159
	80	-0.0029	-0.0082	0.2933	0.0088	0.0034	-0.0088	0.3118	0.0098
	110	0.0072	-0.0058	0.2804	0.0064	-0.0010	-0.0060	0.2966	0.0072
	140	0.0145	-0.0048	0.2677	0.0050	0.0059	-0.0050	0.2876	0.0056
	170	0.0171	-0.0040	0.2549	0.0042	0.0080	-0.0040	0.2774	0.0047
	200	0.0200	-0.0036	0.2417	0.0036	0.0107	-0.0035	0.2699	0.0040

## 2.6 Aplicações

Nesta seção, utilizando dados da literatura, verificamos a primazia da distribuição HNGD em relação às tradicionais distribuições Poisson, Binomial negativa (Bin. Neg.) e Weibull discreta (DWeibull) como distribuições candidatas à análise de dados de contagem. São considerados três conjuntos de dados e os parâmetros das distribuições candidatas estimados pelo método de máxima verossimilhança. Para comparar os desempenhos das várias distribuições, além do teste  $\chi^2$  de aderência, também é utilizado o AIC (Akaike Information Criterion) e o BIC (Bayesian Information Criterion), dados, respectivamente, por:  $AIC = -2 \log(L) + 2m$  e  $BIC = -2 \log(L) + m \log(n)$ , em que  $m$  é o número de parâmetros,  $\log(L)$  é o valor da função log-verossimilhança maximizada e  $n$  é o tamanho da amostra.

O primeiro conjunto de dados foi obtido de Oliveira (2017) e se refere ao número de gols por partida considerando os jogos “em casa” e os jogos “fora de casa” na temporada 2013–2014 da Premier League. As Tabelas 2.4 e 2.5 mostram a distribuição do número de gols observado em comparação com o número de gols esperados segundo as distribuições Half-Normal generalizada discreta, Weibull Discreta, Poisson e Binomial Negativa. A média e a variância do número de gols são, respectivamente, 1.56 gols e 1.77 gols<sup>2</sup> para os jogos “em casa” e 1.19 gols e 1.42 gols<sup>2</sup> para os jogos “fora de casa”.

Note que os valores esperados segundo a distribuição HNGD são bastante próximos dos valores observados. Além disso, de acordo com os valores de AIC e BIC a distribuição HNGD fornece o melhor ajuste.

Tabela 2.4: Distribuição do número de gols “em casa” observados e esperados por partida na Premier League.

Número de Gols	Frequência Observada	Frequência Esperada			
		HNGD	DWeibull	Bin. Neg.	Poisson
0	95	95.1	90.2	90.0	80.4
1	113	110.6	122.4	121.7	125.3
2	85	89.1	88.8	90.3	97.6
3	49	53.8	48.6	48.7	50.7
4	28	23.8	21.8	21.3	19.9
5	10	7.6	8.2	8.0	6.1
Total	380	380	380	380	380
$\chi^2$		1.4	3.3	3.6	9.4
valor-p		0.705	0.352	0.305	0.051
AIC		1221.6	1226.1	1226.8	1228.2
BIC		1229.5	1234.0	1234.7	1232.2

Tabela 2.5: Distribuição do número de gols “fora de casa” observados e esperados por partida na Premier League.

Número de Gols	Frequência Observada	Frequência Esperada			
		HNGD	DWeibull	Bin. Neg.	Poisson
0	137	133.7	131.0	130.7	115.5
1	114	117.3	126.4	126.6	137.7
2	66	75.6	72.3	73.2	82.1
3	49	36.6	33.3	32.7	32.6
4	10	13.3	12.6	12.5	9.7
5	4	3.5	4.4	4.3	2.4
Total	380	380	380	380	380
$\chi^2$		6.3	11.3	11.6	19.6
valor-p		0.052	0.003	0.003	<0.01
AIC		1106.8	1112.4	1112.8	1117.7
BIC		1114.7	1120.2	1120.6	1121.7

O segundo exemplo utiliza os dados retirados de Almalki e Nadarajah (2014) que foram coletados pelo Ministério da Saúde Hospitalar na Arábia Saudita. Os mesmos indicam o tempo de vida em dias de 43 pacientes que foram diagnosticados com Leucemia, ver Tabela 2.6. A média e variância são dadas, respectivamente, por 1191.63 e 256679. Na Tabela 2.7 são apresentados os valores provenientes dos ajustes das distribuições HNGD, WD, Binomial negativa e Poisson.

Tabela 2.6: Tempo de vida em dias de 43 pacientes que foram diagnosticados com Leucemia.

115	181	255	418	441	461	516	739	743	789	807
865	924	983	1025	1062	1063	1165	1191	1222	1222	1251
1277	1290	1357	1369	1408	1455	1478	1519	1578	1578	1599
1603	1605	1696	1735	1799	1815	1852	1899	1925	1965	

Tabela 2.7: Estimativas de máxima verossimilhança (erros padrão), valor da estatística do teste qui-quadrado de aderência e os valores de AIC e BIC.

Distribuições	$\alpha$	$\theta$	$\chi^2$	valor-p	AIC	BIC
HNGD	2.15 (0.29)	1431.70 (75.88)	3.25	0.661	657.99	661.51
DWeibull	2.56 (0.33)	1335.05 (82.53)	5.26	0.385	662.83	666.36
Bin. Neg.	3.61 (0.75)	1191.47 (95.82)	8.06	0.153	671.66	675.19
Poisson	1191.63 (5.26)		Inf.	<0.001	11069.55	11071.31

Da Tabela 2.7 é notório perceber a utilidade da distribuição HNGD proposta neste trabalho. Os valores de AIC e BIC são os menores entre as distribuições concorrentes. O mesmo é verdade para o valor da estatística do teste qui-quadrado de aderência.

## 2.7 Conclusão

Neste artigo foi apresentada a distribuição HNGD formulada a partir do método baseado na função de sobrevivência o qual foi proposto por Nakagawa e Osaki (1975). Algumas características e propriedades matemáticas da distribuição proposta foram estudadas. Verificamos também que distribuição HNGD pode ser utilizada na análise de dados com subdispersão e superdispersão. Um estudo de simulação Monte Carlo mostrou que os parâmetros são eficientemente estimados pelo método da máxima verossimilhança e pelo método dos momentos. Em duas aplicações a discriminação entre as distribuições Weibull discreta, Binomial negativa e Poisson foi realizada utilizando os critérios AIC e BIC. A qualidade do ajuste foi avaliada pela estatística do teste qui-quadrado de aderência. Os resultados obtidos mostraram que a distribuição HNGD foi superior às distribuições supracitadas o que evidencia sua utilização na análise de dados de contagem e de sobrevivência.

---

---

## CAPÍTULO 3

---

# A DISTRIBUIÇÃO HALF-NORMAL TRANSMUTADA

### Resumo

A distribuição Half-Normal, a qual pode ser obtida pela transformação de uma variável aleatória Normal padrão, vem sendo intensivamente estendida nos últimos anos. Uma revisão na literatura mostrou pelo menos 10 extensões introduzidas entre 2008 e 2017. Estas extensões ampliam o comportamento da função de densidade e de risco, restrito ao comportamento monótono decrescente e monótono crescente, respectivamente. Neste artigo propomos uma nova extensão e a denominamos de distribuição Half-Normal transmutada, uma vez que a mesma é desenvolvida usando os mapas de transmutação de classificação quadrática, introduzidos por Shaw e Buckley (2009). Uma descrição abrangente das propriedades matemáticas da nova distribuição é apresentada. As expressões analíticas explícitas da esperança, variância, assimetria, curtose, função geradora de momentos, entropia de Shannon, desvios médios, curvas de Bonferroni e Lorenz, estatística de ordem e confiabilidade são deduzidas e apresentadas. Os parâmetros da distribuição proposta são estimados usando o método da máxima verossimilhança. O viés e a acurácia dos estimadores são estimados via simulação Monte Carlo. Dois exemplos utilizando dados reais de precipitação são usados para mostrar que a distribuição proposta é útil e pode ser uma boa alternativa à várias extensões da distribuição Half-Normal recentemente propostas na literatura.

**Palavras-chave:** Distribuição Half-Normal, Momentos, Estimação Paramétrica, Transmutação.

## Abstract

The Half-Normal distribution has been intensively extended in the recent years. A review in the literature showed that at least 10 extensions were introduced between 2008 and 2016. These generalizations extended the behavior of the density and hazard functions, which are restricted to monotonous decreasing and monotonically increasing, respectively. In this paper we propose a new extension called the transmuted Half-Normal distribution since it is obtained using the quadratic rank transmutation map, introduced by Shaw e Buckley (2009). A comprehensive account of mathematical properties of the new distribution is presented. We provide explicit expressions for the moments, moment-generating function, Shannon's entropy, mean deviations, Bonferroni and Lorenz curves, order statistics, and reliability. The estimation of the parameters is approached by the maximum likelihood method. The bias and accuracy of the estimators are investigated by the Monte Carlo simulations. Two real data sets show that the proposed distribution is useful and may be a good alternative to several extensions of the Half-Normal distribution.

**Keywords:** Half-Normal Distribution, Moments, Parametric Estimation, Transmutation.

### 3.1 Introdução

Subjacente a todo procedimento paramétrico de inferência existe uma distribuição de probabilidade usada para descrever o comportamento de uma variável aleatória na população. Nos últimos anos, um incontável número de distribuições de probabilidade, em sua maioria com suporte nos números reais positivos, foram propostas na literatura. Muitas estratégias podem ser usadas para a geração/extensão de uma distribuição de probabilidade. A maior parte destas estratégias adiciona um ou mais parâmetros à alguma distribuição base (Normal, Gumbel, Exponencial, Weibull, Gama, Log-Normal, entre tantas outras). Em geral, introduzir um ou mais parâmetros trás maior flexibilidade no comportamento das funções de densidade e de risco das distribuições tomadas como base.

Um recente levantamento dos principais métodos utilizados para estender uma distribuição base estão disponíveis, por exemplo, em Nadarajah e Rocha (2016), Tahir e Nadarajah (2015), Brito, Rêgo e Oliveira (2015), Aljarrah, Lee e Famoye (2014), Lee, Famoye e Alzaatreh (2013), Lai (2011) e Gupta e Kundu (2009).

É importante enfatizar que a transformação de uma variável aleatória  $Z$  em outra  $X$ , da forma  $X = g(Z)$ , é a forma mais simples de se gerar ou estender uma distribuição de probabilidade base. Um exemplo de uma distribuição obtida por transformação é a distribuição Half-Normal uniparamétrica (HN), resultante da transformação  $X = \theta |Z|$ , em que  $Z$  tem distribuição Normal padrão e  $\theta > 0$  é um parâmetro de escala. Nos últimos anos várias extensões da distribuição Half-Normal uniparamétrica foram propostas. Entre estas extensões cita-se: a distribuição Half-Normal geral (PEWSEY, 2002, 2004), a distribuição

Half-Normal generalizada (COORAY; ANANDA, 2008), a distribuição Beta (log-Beta) Half-Normal generalizada (PESCIM et al., 2010; CORDEIRO et al., 2013; PESCIM et al., 2013), a distribuição Kumaraswamy Half-Normal generalizada (CORDEIRO; PESCIM; ORTEGA, 2012), a distribuição Beta Half-Normal generalizada geométrica (RAMIRES et al., 2013), a extensão da distribuição Half-Normal generalizada (OLMOS et al., 2014), a distribuição Half-Normal potência (GÓMEZ; BOLFARINE, 2015), a distribuição Half-Normal generalizada estendida (SANCHEZ; FREITAS; CORDEIRO, 2016) e a distribuição Odd Log-Logistica Half-Normal generalizada (CORDEIRO et al., 2017).

Este artigo introduz a distribuição Half-Normal transmutada (HNT), oriunda da distribuição HN uniparamétrica. A extensão de uma distribuição base via transmutação foi proposta por Shaw e Buckley (2009) tomando como base a distribuição Weibull. Em Tahir e Cordeiro (2016) são enumeradas 51 distribuições estendidas por transmutação. É importante ressaltar que a distribuição HN uniparamétrica não faz parte da relação das distribuições tomadas como base. Na Seção 3.2 apresentamos a distribuição HN usada como distribuição base a ser estendida. A sua versão transmutada é apresentada na Seção 3.3. Na Seção 3.4 são deduzidas as expressões de várias medidas úteis na análise de dados de sobrevivência, como por exemplo, a função de sobrevivência, a função de risco e a vida média residual. As expressões analíticas dos momentos de ordem  $k$ , entropia de Shannon, desvios médios, curvas de Bonferroni e Lorenz, estatística de ordem e confiabilidade são deduzidas na Seção 3.5. O método da máxima verossimilhança, usado para estimar os parâmetros, é apresentado na Seção 3.6 enquanto que na Seção 3.7 estimamos, via simulações Monte Carlo, o viés e a acurácia dos estimadores de máxima verossimilhança. Duas aplicações considerando a distribuição proposta e todas as suas extensões são apresentadas na Seção 3.8. Na Seção 3.9 finalizamos o artigo com algumas conclusões.

## 3.2 A Distribuição Half-Normal

Se uma variável aleatória não negativa  $X$  tem distribuição HN com parâmetro de escala  $\theta > 0$ , então a função densidade de probabilidade (fdp) e a função distribuição acumulada (fda) são escritas, respectivamente, como:

$$g(x | \theta) = \frac{2}{\theta} \phi\left(\frac{x}{\theta}\right) \text{ e } G(x | \theta) = 2\Phi\left(\frac{x}{\theta}\right) - 1 \quad (3.1)$$

em que  $\phi(\cdot)$  e  $\Phi(\cdot)$  denotam, respectivamente, a fdp e fda de distribuição de uma variável aleatória com distribuição Normal padrão. A correspondente função de risco é escrita como:

$$h(x | \theta) = \frac{1}{\theta} \phi\left(\frac{x}{\theta}\right) \Phi\left(-\frac{x}{\theta}\right)^{-1}$$

que é monotonicamente crescente para todo  $\theta$ . Vale lembrar que a fdp é monotonicamente decrescente para todo  $\theta$ .

Formalmente, se uma variável aleatória  $Z$  é normalmente distribuída com média igual a zero e variância igual a um, então a distribuição HN é a distribuição de  $X = \theta |Z|$ . Para  $Z$  normalmente distribuída com média igual a  $\mu$  e variância igual a um, a transformação  $X = \theta |Z|$  leva à distribuição folded-Normal (LEONE; NELSON; NOTTINGHAM, 1961; TSAGRIS; BENEKI; HASSANI, 2014). A distribuição folded-Normal é um caso especial da distribuição Normal sempre que o sinal da variável medida for desconhecido, perdido ou não relevante. Em outras palavras, ela é utilizada quando se está interessado no tamanho da variável medida e não na direção ou no sinal (CHOU; LIU, 1998). É evidente que a distribuição HN é um caso particular da distribuição folded-Normal quando  $\mu = 0$ , que também coincide com a distribuição Normal truncada em zero (NADARAJAH; KOTZ, 2006b). Além disso, as expressões em (3.1) podem ser obtidas como caso particular da distribuição Rayleigh generalizada (VODĀ, 1976a, 1976b), da distribuição Gama generalizada (STACY, 1962) e da raiz quadrada de uma variável aleatória Qui-Quadrado com um grau de liberdade (JOHNSON; KOTZ; BALAKRISHNAN, 1994).

Para  $X \sim HN(\theta)$  o  $k$ -ésimo momento de  $X$  é  $E(X^k) = \frac{1}{\sqrt{\pi}} 2^{\frac{k}{2}} \theta^k \Gamma\left(\frac{k+1}{2}\right)$ . Assim, para  $k = 1$  temos  $E(X) = \sqrt{\frac{2}{\pi}} \theta$  e para  $k = 2$ ,  $E(X^2) = \theta^2$ . O terceiro e quarto momentos padronizados (assimetria e curtose, respectivamente) não dependem de  $\theta$  e são escritos como:

$$\sqrt{\beta_1} = \frac{(4 - \pi) \sqrt{2}}{(\pi - 2)^{\frac{3}{2}}} = 0.9952 \quad \text{e} \quad \beta_2 = \frac{3\pi^2 - 4\pi - 12}{(\pi - 2)^2} = 3.8691.$$

A função quantil pode ser escrita na forma:

$$Q(p | \theta) = G^{-1}(p | \theta) = \theta \Phi^{-1}\left(\frac{p+1}{2}\right) \quad (3.2)$$

em que  $0 < p < 1$  e  $\Phi^{-1}(\cdot)$  é a função quantil de uma variável aleatória Normal padrão. Note que a partir de (3.2) podemos gerar valores pseudo-aleatórios de  $X$ . Uma estratégia alternativa pode ser encontrada em Singh (1994) como consequência da aplicação da transformação Box-Muller (SINGH, 1995).

### 3.3 A Distribuição Half-Normal Transmutada

Motivados pela necessidade de distribuições com densidade e risco mais versáteis, Shaw e Buckley (2009) propuseram uma estratégia que se têm mostrado útil pra se estender uma distribuição base. A distribuição obtida por este processo recebe o nome da distribuição base mais o termo “transmutada”, por exemplo distribuição Weibull transmutada. Nesta seção, a partir do procedimento de transmutação, introduziremos a distribuição HN transmutada (HNT).

Utilizando a estratégia proposta por Shaw e Buckley (2009), uma variável aleatória  $X$  é dita ter uma distribuição transmutada se sua fda é escrita na forma:

$$F(x | \theta, \lambda) = (1 + \lambda)G(x | \theta) - \lambda [G(x | \theta)]^2, \quad |\lambda| \leq 1.$$

Consequentemente a fdp é dada por:

$$f(x | \theta, \lambda) = g(x | \theta)[(1 + \lambda) - 2\lambda G(x | \theta)]$$

em que  $G(x | \theta)$  e  $g(x | \theta)$ , são respectivamente, a fda e fdp da distribuição base, indexadas por um vetor de parâmetros  $\theta$ . Para  $\lambda = 0$ , temos a distribuição da variável aleatória base como caso particular.

Portanto, uma variável aleatória não negativa  $X$  tem distribuição HNT, com parâmetro de escala  $\theta > 0$  e forma  $-1 \leq \lambda \leq 1$ , se suas respectivas fdp e fda são escritas como:

$$f(x | \theta, \lambda) = \frac{2}{\theta} \phi\left(\frac{x}{\theta}\right) \left[1 + \lambda - 2\lambda \left(2\Phi\left(\frac{x}{\theta}\right) - 1\right)\right] \quad (3.3)$$

e

$$F(x | \theta, \lambda) = (1 + \lambda) \left(2\Phi\left(\frac{x}{\theta}\right) - 1\right) - \lambda \left(2\Phi\left(\frac{x}{\theta}\right) - 1\right)^2. \quad (3.4)$$

A Figura 3.1 ilustra como o parâmetro  $\lambda$  influencia o comportamento de (3.3). Uma vez que  $\theta$  é parâmetro de escala, o mesmo está fixado em  $\theta = 1$ .

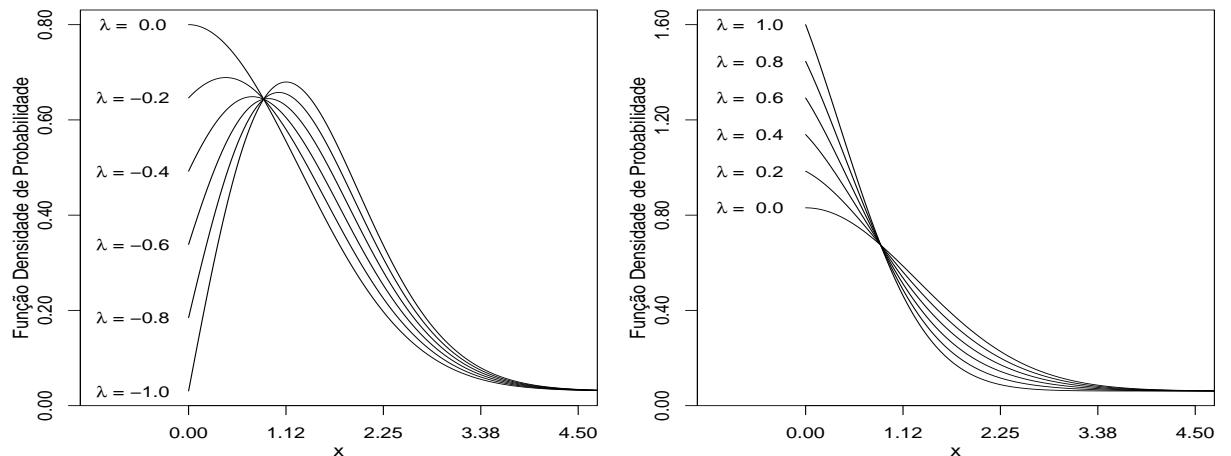


Figura 3.1: Comportamento da função densidade de probabilidade da distribuição HNT.

### 3.4 Funções Úteis na Análise de Dados de Sobrevivência

Nesta seção são apresentadas algumas funções intrínsecas a uma variável aleatória com distribuição HNT — função de sobrevivência  $S(x | \theta, \lambda)$ , função de risco  $h(x | \theta, \lambda)$  e a vida média residual  $m(x | \theta, \lambda)$ . Estas funções são úteis na descrição, por exemplo, do tempo até a ocorrência de algum evento.

A função de sobrevivência  $S(x | \theta, \lambda)$  descreve a probabilidade, por exemplo, de um item ou indivíduo sobreviver ao tempo  $x$ . Para a distribuição HNT temos:

$$S(x | \theta, \lambda) = 1 - (1 + \lambda) \left( 2 \Phi \left( \frac{x}{\theta} \right) - 1 \right) + \lambda \left( 2 \Phi \left( \frac{x}{\theta} \right) - 1 \right)^2. \quad (3.5)$$

A função de risco descreve a probabilidade instantânea de ocorrência do evento “sobrevivência” até um tempo arbitrário. Matematicamente, temos  $h(x | \theta, \lambda) = \frac{f(x | \theta, \lambda)}{S(x | \theta, \lambda)}$  e, consequentemente:

$$h(x | \theta, \lambda) = \frac{\frac{2}{\theta} \phi \left( \frac{x}{\theta} \right) [1 + \lambda - 2 \lambda (2 \Phi \left( \frac{x}{\theta} \right) - 1)]}{1 - (1 + \lambda) (2 \Phi \left( \frac{x}{\theta} \right) - 1) + \lambda (2 \Phi \left( \frac{x}{\theta} \right) - 1)^2}. \quad (3.6)$$

A Figura 3.2 ilustra como o parâmetro  $\lambda$  influencia o comportamento de (3.6). Uma vez que  $\theta$  é parâmetro de escala, o mesmo está fixado em  $\theta = 1$ .

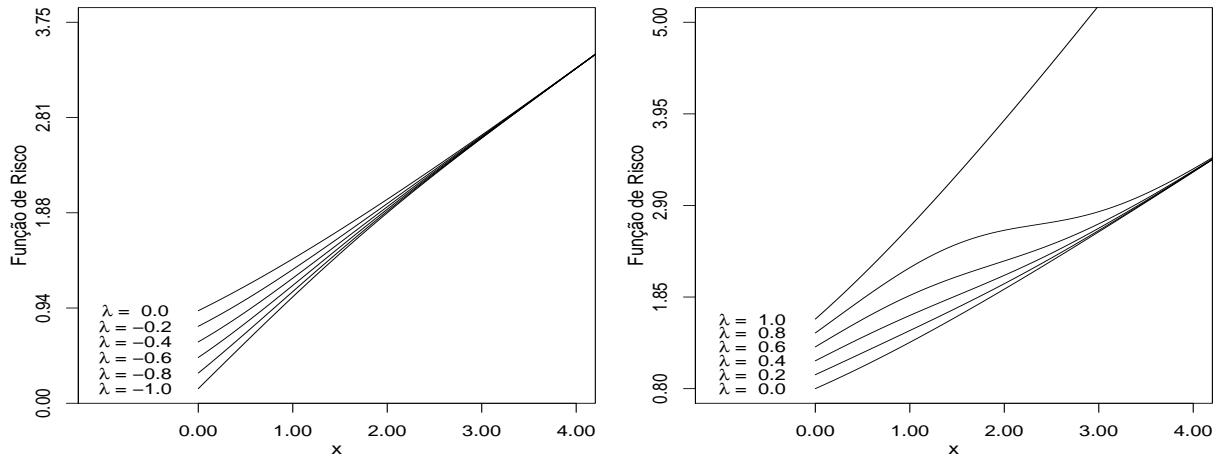


Figura 3.2: Comportamento da função de risco da distribuição HNT.

Outra função útil na análise de dados de sobrevivência é a vida média residual. A vida média residual descreve, para indivíduos com idade  $t$ , o tempo médio restante de vida e corresponde a razão entre a área sob a curva de sobrevivência à direita de  $t$  e  $S(x | \theta, \lambda)$ .

Matematicamente, a partir de (3.5), temos:

$$m(t | \theta, \lambda) = \mathbb{E}(X - t | X > t) = \frac{1}{S(t | \theta, \lambda)} \int_t^\infty S(x | \theta, \lambda) dx. \quad (3.7)$$

**Teorema 3.4.1.** A Vida média residual da distribuição HNT é dada por:

$$\begin{aligned} m(t | \theta, \lambda) &= \frac{-\pi^{-\frac{1}{2}}}{S(t | \theta, \lambda)} \left\{ 2^{\frac{3}{2}} \theta \lambda e^{\frac{-t^2}{2\theta^2}} \gamma(t) - (\lambda + 1) \theta \sqrt{2} e^{\frac{-t^2}{2\theta^2}} - 2t\sqrt{\pi} \left[ \Phi\left(\frac{t}{\theta}\right) - 1 \right] + \right. \\ &\quad \left. - 4\theta\lambda \left[ \Phi\left(\frac{t\sqrt{2}}{\theta}\right) - 1 \right] + \lambda t \sqrt{\pi} \gamma(t) (\gamma(t) - 1) \right\} \\ \text{em que } \gamma(t) &= 2\Phi\left(\frac{t}{\theta}\right) - 1. \end{aligned}$$

*Demonstração.* De fato, substituindo a equação (3.5) em (3.7), obtemos

$$m(t | \theta, \lambda) = \frac{1}{S(t | \theta, \lambda)} \int_t^\infty 1 - (1 + \lambda) \left[ 2\Phi\left(\frac{x}{\theta}\right) - 1 \right] + \lambda \left[ 2\Phi\left(\frac{x}{\theta}\right) - 1 \right]^2 dx.$$

Resolvendo a integral, obtemos:

$$\begin{aligned} m(t | \theta, \lambda) &= \frac{-\pi^{-\frac{1}{2}}}{S(t | \theta, \lambda)} \left\{ 2^{\frac{3}{2}} \theta \lambda e^{\frac{-t^2}{2\theta^2}} \left[ 2\Phi\left(\frac{t}{\theta}\right) - 1 \right] - 2\theta\lambda \left[ 2\Phi\left(\frac{t\sqrt{2}}{\theta}\right) - 1 \right] + \right. \\ &\quad + \lambda t \sqrt{\pi} \left[ 2\Phi\left(\frac{t}{\theta}\right) - 1 \right]^2 + t\sqrt{\pi} - t\sqrt{\pi} \left[ 2\Phi\left(\frac{t}{\theta}\right) - 1 \right] + \\ &\quad \left. - \lambda t \sqrt{\pi} \left[ 2\Phi\left(\frac{t}{\theta}\right) - 1 \right] - \lambda\theta \sqrt{2} e^{\frac{-t^2}{2\theta^2}} - \theta\sqrt{2} e^{\frac{-t^2}{2\theta^2}} + 2\lambda\theta \right\}. \end{aligned}$$

A partir de manipulações algébricas chegamos a seguinte expressão:

$$\begin{aligned} m(t | \theta, \lambda) &= \frac{-\pi^{-\frac{1}{2}}}{S(t | \theta, \lambda)} \left\{ 2^{\frac{3}{2}} \theta \lambda e^{\frac{-t^2}{2\theta^2}} \left[ 2\Phi\left(\frac{t}{\theta}\right) - 1 \right] - (\lambda + 1) \theta \sqrt{2} e^{\frac{-t^2}{2\theta^2}} - 2t\sqrt{\pi} \left[ \Phi\left(\frac{t}{\theta}\right) - 1 \right] + \right. \\ &\quad \left. - 4\theta\lambda \left[ \Phi\left(\frac{t\sqrt{2}}{\theta}\right) - 1 \right] + \lambda t \sqrt{\pi} \left[ 2\Phi\left(\frac{t}{\theta}\right) - 1 \right] \left[ \left( 2\Phi\left(\frac{t}{\theta}\right) - 1 \right) - 1 \right] \right\}. \end{aligned}$$

Substituindo  $2\Phi\left(\frac{t}{\theta}\right) - 1 = \gamma(t)$ , obtemos o resultado.  $\square$

## 3.5 Propriedades da Distribuição HNT

### 3.5.1 Momento de Ordem $k$

Os momentos são medidas capazes de caracterizar uma distribuição de probabilidade, sendo os quatro primeiros momentos indicativos de tendência central, dispersão, assimetria e curtose, nessa ordem. Diferentemente de muitas extensões da distribuição HN, o momento de ordem  $k$  da distribuição HNT pode ser obtido analiticamente.

**Teorema 3.5.1.** *Se  $X$  segue a distribuição HNT, temos que o momento de ordem  $k$  pode ser escrito da seguinte forma:*

$$E(X^k) = \frac{\theta^k 2^{\frac{k}{2}}}{\pi^{3/2}} \left\{ (\lambda + 1) \pi \Gamma\left(\frac{k+1}{2}\right) - 2\lambda k \sqrt{\pi} Hgeo\left[\frac{1}{2}, \frac{k}{2} + 1; \frac{3}{2}; -1\right] \Gamma\left(\frac{k}{2}\right) \right\}$$

em que  $Hgeo$  é a distribuição Hipergeométrica (FELLER, 1968).

*Demonstração.* A partir da definição de momento de ordem  $k$  de uma variável aleatória contínua, temos:

$$\begin{aligned} E(X^k) &= \int_0^\infty x^k \frac{2}{\theta} \phi\left(\frac{x}{\theta}\right) \left[ 1 + \lambda - 2\lambda \left( 2\Phi\left(\frac{x}{\theta}\right) - 1 \right) \right] dx \\ &= (1 + \lambda) \int_0^\infty \frac{2}{\theta} x^k \phi\left(\frac{x}{\theta}\right) dx - 2\lambda \int_0^\infty \frac{2}{\theta} x^k \phi\left(\frac{x}{\theta}\right) \left( 2\Phi\left(\frac{x}{\theta}\right) - 1 \right) dx \end{aligned}$$

que pode ser escrito como:

$$E(X^k) = \frac{\theta^k 2^{\frac{k}{2}}}{\pi^{3/2}} \left\{ (\lambda + 1) \pi \Gamma\left(\frac{k+1}{2}\right) - 2\lambda k \sqrt{\pi} Hgeo\left[\frac{1}{2}, \frac{k}{2} + 1; \frac{3}{2}; -1\right] \Gamma\left(\frac{k}{2}\right) \right\}. \quad (3.8)$$

□

A partir de (3.8), temos:

$$E(X) = \frac{\theta}{\sqrt{\pi}} \left[ (1 + \lambda) \sqrt{2} - 2\lambda \right], \quad (3.9)$$

e

$$Var(X) = \frac{\theta^2}{\sqrt{\pi}} \left[ (\pi - 2) + (4\sqrt{2} - 6)(\lambda - \lambda^2) \right].$$

As expressões do coeficiente de variação, assimetria e curtose são obtidas a partir das seguintes relações:

$$CV(X) = \frac{\sqrt{Var(X)}}{E(X)} = \frac{[(\pi - 2) + (4\sqrt{2} - 6)(\lambda - \lambda^2)]^{1/2}}{\sqrt{2}(\lambda - 1) - 2\lambda},$$

$$\begin{aligned}
Sk(X) &= \frac{E(X - E(X))^3}{Var(X)^{3/2}} \\
&= \left[ \sqrt{2}(4 - \pi) + \lambda \left( \sqrt{2}(18 - \pi) + \pi - 24 \right) + \lambda^2 \left( 42\sqrt{2} - 60 \right) \right. \\
&\quad \left. + \lambda^3 \left( 28\sqrt{2} - 40 \right) \right] \times \left[ (\pi - 2) + (4\sqrt{2} - 6)(\lambda - \lambda^2) \right]^{-3/2}
\end{aligned}$$

e

$$\begin{aligned}
Cur(X) &= \frac{E(X - E(X))^4}{Var(X)^2}, \\
&= \left[ (3\pi^2 - 4\pi - 12) + \lambda \left( 4\pi(3\sqrt{2} - 4) - 48\sqrt{2} - 72 \right) \right. \\
&\quad + \lambda^2 \left( 4\pi(3\sqrt{2} - 5) + 192\sqrt{2} - 264 \right) + \lambda^3 \left( 288\sqrt{2} - 408 \right) \\
&\quad \left. + \lambda^4 \left( 144\sqrt{2} - 204 \right) \right] \times \left[ (\pi - 2) + (4\sqrt{2} - 6)(\lambda - \lambda^2) \right]^{-2}.
\end{aligned}$$

### 3.5.2 Função Geradora de Momentos

**Teorema 3.5.2.** Se  $X$  tem distribuição HNT, então a função geradora de momentos de  $X$ ,  $M_X(t)$ , é dada por:

$$\begin{aligned}
M_X(t) &= -2\lambda \sum_{n=0}^{\infty} \frac{t^n}{n!} Hgeo \left[ \frac{1}{2}, \frac{n}{2} + 1; \frac{3}{2}; -1 \right] \Gamma \left( \frac{n}{2} + 1 \right) \frac{\theta^n 2^{n/2+1}}{\pi} + \\
&\quad + (1 + \lambda) \sum_{n=0}^{\infty} \frac{t^n}{n!} \Gamma \left( \frac{n+1}{2} \right) \frac{\theta^n 2^{n/2}}{\sqrt{\pi}}.
\end{aligned}$$

*Demonstração.* Seja  $X$  uma variável aleatória com distribuição HNT, então a função geradora de momentos de  $X$  é dada por:

$$M_X(t) = \int_0^\infty e^{xt} \frac{2}{\theta} \phi \left( \frac{x}{\theta} \right) \left[ 1 + \lambda - 2\lambda \left( 2\Phi \left( \frac{x}{\theta} \right) - 1 \right) \right] dx. \quad (3.10)$$

A partir da expansão em série de Taylor da função  $e^{tx}$ , temos que (3.10) pode ser escrita na forma:

$$\begin{aligned}
M_x(t) &= (1 + \lambda) \sum_{n=0}^{\infty} \frac{t^n}{n!} \int_0^\infty \frac{2}{\theta} x^n \phi \left( \frac{x}{\theta} \right) dx + \\
&\quad - 2\lambda \sum_{n=0}^{\infty} \frac{t^n}{n!} \int_0^\infty \frac{2}{\theta} x^n \phi \left( \frac{x}{\theta} \right) \left( 2\Phi \left( \frac{x}{\theta} \right) - 1 \right) dx.
\end{aligned}$$

Resolvendo as integrais, obtemos:

$$\begin{aligned} M_X(t) = & -2\lambda \sum_{n=0}^{\infty} \frac{t^n}{n!} \text{Hgeo}\left[\frac{1}{2}, \frac{n}{2} + 1; \frac{3}{2}; -1\right] \Gamma\left(\frac{n}{2} + 1\right) \frac{\theta^n 2^{n/2+1}}{\pi} + \\ & + (1+\lambda) \sum_{n=0}^{\infty} \frac{t^n}{n!} \Gamma\left(\frac{n+1}{2}\right) \frac{\theta^n 2^{n/2}}{\sqrt{\pi}}. \end{aligned} \quad (3.11)$$

□

### 3.5.3 Entropia Diferencial

Vamos introduzir agora o conceito de entropia diferencial, que é a entropia de uma variável aleatória contínua. Essa medida mede a variação da incerteza e um grande valor de entropia indica uma maior incerteza nos dados. Uma das medidas mais populares é a entropia de Shannon (SHANNON, 1951).

O conceito de entropia de Shannon se refere a incerteza de uma distribuição de probabilidade. Um fato importante é que essa entropia não é uma função da variável aleatória  $X$ , mas sim da distribuição de probabilidade dessa variável.

**Definição 3.5.1.** A entropia de Shannon  $\mathcal{H}_{Sh}$  de uma variável aleatória contínua  $X$  com uma função densidade de probabilidade  $f(x)$  é definida como

$$\mathcal{H}_{Sh}(X) = - \int_S f(x) \log f(x) dx,$$

em que  $S$  é o conjunto suporte da variável aleatória.

**Teorema 3.5.3.** A entropia de Shannon para uma variável aleatória contínua  $X$  com distribuição HNT é dada por:

$$\begin{aligned} \mathcal{H}_{Sh}(X) = & -\log\left(\frac{2}{\theta}\right) + \frac{\pi(1-\log 2\pi) - 2\lambda}{2\pi} + \\ & + 2\lambda^2 \{1 - \log[(1+\lambda)(1-\lambda)]\} - \lambda(\lambda^2 + 1) \log\left(\frac{1+\lambda}{1-\lambda}\right). \end{aligned}$$

*Demonstração.* De fato, segue diretamente de (3.3) e da definição anterior que

$$\begin{aligned} \mathcal{H}_{Sh}(X) = & - \int_0^\infty \frac{2}{\theta} \phi\left(\frac{x}{\theta}\right) \left[1 + \lambda - 2\lambda \left(2\Phi\left(\frac{x}{\theta}\right) - 1\right)\right] \times \\ & \times \log \left\{ \frac{2}{\theta} \phi\left(\frac{x}{\theta}\right) \left[1 + \lambda - 2\lambda \left(2\Phi\left(\frac{x}{\theta}\right) - 1\right)\right] \right\} dx. \end{aligned}$$

Utilizando as propriedades distributivas da função logarítmica, temos:

$$\begin{aligned}\mathcal{H}_{Sh}(X) &= - \int_0^\infty \frac{2}{\theta} \phi\left(\frac{x}{\theta}\right) \left[ 1 + \lambda - 2\lambda \left( 2\Phi\left(\frac{x}{\theta}\right) - 1 \right) \right] \times \\ &\quad \times \left\{ \log\left(\frac{2}{\theta}\right) + \log\left(\phi\left(\frac{x}{\theta}\right)\right) + \log\left[ 1 + \lambda - 2\lambda \left( 2\Phi\left(\frac{x}{\theta}\right) - 1 \right) \right] \right\} dx \\ &= - \int_0^\infty \log\left(\frac{2}{\theta}\right) \left\{ \frac{2}{\theta} \phi\left(\frac{x}{\theta}\right) \left[ 1 + \lambda - 2\lambda \left( 2\Phi\left(\frac{x}{\theta}\right) - 1 \right) \right] \right\} + \\ &\quad + \log\left(\phi\left(\frac{x}{\theta}\right)\right) \left\{ \frac{2}{\theta} \phi\left(\frac{x}{\theta}\right) \left[ 1 + \lambda - 2\lambda \left( 2\Phi\left(\frac{x}{\theta}\right) - 1 \right) \right] \right\} + \\ &\quad + \log\left(1 + \lambda - 2\lambda \left( 2\Phi\left(\frac{x}{\theta}\right) - 1 \right)\right) \left\{ \frac{2}{\theta} \phi\left(\frac{x}{\theta}\right) \left[ 1 + \lambda - 2\lambda \left( 2\Phi\left(\frac{x}{\theta}\right) - 1 \right) \right] \right\} dx.\end{aligned}$$

Como as funções em questão são todas integráveis, temos pelas propriedades de integração que

$$\begin{aligned}\mathcal{H}_{Sh}(X) &= - \int_0^\infty \log\left(\frac{2}{\theta}\right) \left\{ \frac{2}{\theta} \phi\left(\frac{x}{\theta}\right) \left[ 1 + \lambda - 2\lambda \left( 2\Phi\left(\frac{x}{\theta}\right) - 1 \right) \right] \right\} dx + \\ &\quad - \int_0^\infty \log\left(\phi\left(\frac{x}{\theta}\right)\right) \left\{ \frac{2}{\theta} \phi\left(\frac{x}{\theta}\right) \left[ 1 + \lambda - 2\lambda \left( 2\Phi\left(\frac{x}{\theta}\right) - 1 \right) \right] \right\} dx + \\ &\quad - \int_0^\infty \log\left(1 + \lambda - 2\lambda \left( 2\Phi\left(\frac{x}{\theta}\right) - 1 \right)\right) \left\{ \frac{2}{\theta} \phi\left(\frac{x}{\theta}\right) \left[ 1 + \lambda - 2\lambda \left( 2\Phi\left(\frac{x}{\theta}\right) - 1 \right) \right] \right\} dx.\end{aligned}$$

Resolvendo as integrais, segue que a entropia de Shannon para a distribuição HNT é dada por:

$$\begin{aligned}\mathcal{H}_{Sh}(X) &= - \log\left(\frac{2}{\theta}\right) + \frac{\pi(1 - \log 2\pi) - 2\lambda}{2\pi} + \\ &\quad + 2\lambda^2 \{1 - \log[(1 + \lambda)(1 - \lambda)]\} - \lambda(\lambda^2 + 1) \log\left(\frac{1 + \lambda}{1 - \lambda}\right).\end{aligned}$$

□

### 3.5.4 Desvios Médios

A quantidade de dispersão em uma população pode ser medida pelos desvios em torno da média e da mediana, definidas por:

$$\delta_1 = \int_0^\infty |x - \mu| f(x) dx \quad \text{e} \quad \delta_2 = \int_0^\infty |x - M| f(x) dx,$$

respectivamente, em que  $\mu$  denota a média e  $M$  denota a mediana.

Para calcular essas medidas, podemos usar as seguintes relações apresentadas em Na-

darajah e Kotz (2006a):

$$\delta_1 = 2\mu F(\mu) - 2 \int_0^\mu x f(x) dx \quad \text{e} \quad \delta_2 = \mu + 2MF(M) - M - 2 \int_0^M x f(x) dx,$$

em que  $F(\mu)$  e  $F(M)$  podem ser calculadas segundo a equação (3.4). Tomando a  $f(x)$  como a fdp da distribuição HNT, temos:

$$\begin{aligned} \int_0^\mu x f(x | \theta, \lambda) dx &= \frac{\theta}{\sqrt{\pi}} \left\{ \sqrt{2}(1+\lambda) + 2\lambda \left[ 2\Phi\left(\frac{\mu\sqrt{2}}{\theta}\right) - 1 \right] \right\} + \\ &\quad - \theta^2 \left\{ \frac{2}{\theta} \phi\left(\frac{\mu}{\theta}\right) \left[ 1 + \lambda - 2\lambda \left( 2\Phi\left(\frac{\mu}{\theta}\right) - 1 \right) \right] \right\}, \end{aligned}$$

De maneira análoga obtemos  $\int_0^M x f(x | \theta, \lambda) dx$ .

### 3.5.5 Curvas de Bonferroni e Lorenz

As curvas de Bonferroni e Lorenz, propostas por Bonferroni (1930), são aplicações dos desvios médios. Podemos ver seu uso em áreas como confiabilidade, demografia, economia, medicina e seguro. Além disso, são consideradas por economistas como uma medida de desigualdade social, pois relacionam as porcentagens acumuladas de renda e população. Formalmente, as curvas de Bonferroni e Lorenz são definidas como:

**Definição 3.5.2.** Suponha que  $X$  seja uma variável aleatória não negativa com função densidade de probabilidade  $f(x)$  e função de distribuição acumulada  $F(x)$ . As curvas de Bonferroni e Lorenz denotadas por  $B(p)$  e  $L(p)$ , respectivamente, são definidas como:

$$B(p) = \frac{1}{p\mu} \int_0^q x f(x) dx \quad \text{e} \quad L(p) = \frac{1}{\mu} \int_0^q x f(x) dx,$$

em que  $\mu = \mathbb{E}(X)$ ,  $q = F^{-1}(p)$  e  $p \in [0, 1]$ .

Em particular, para a distribuição HNT temos:

**Teorema 3.5.4.** Seja  $X$  uma variável aleatória contínua não negativa que segue a distribuição HNT. As curvas de Bonferroni e Lorenz são dadas por:

$$\begin{aligned} B(p) &= \frac{-\theta^2}{p\mu} \left\{ \frac{2}{\theta} \phi\left(\frac{q}{\theta}\right) \left[ 1 + \lambda - 2\lambda \left( 2\Phi\left(\frac{q}{\theta}\right) - 1 \right) \right] + \right. \\ &\quad \left. - \frac{1}{\theta\sqrt{\pi}} \left[ \sqrt{2}(1+\lambda) + 2\lambda \left( 2\Phi\left(\frac{q\sqrt{2}}{\theta}\right) - 1 \right) \right] \right\}. \end{aligned}$$

e

$$\begin{aligned} L(p) = & \frac{-\theta^2}{\mu} \left\{ \frac{2}{\theta} \phi\left(\frac{q}{\theta}\right) \left[ 1 + \lambda - 2\lambda \left( 2\Phi\left(\frac{q}{\theta}\right) - 1 \right) \right] + \right. \\ & \left. - \frac{1}{\theta\sqrt{\pi}} \left[ \sqrt{2}(1+\lambda) + 2\lambda \left( 2\Phi\left(\frac{q\sqrt{2}}{\theta}\right) - 1 \right) \right] \right\}. \end{aligned}$$

*Demonstração.* De fato, aplicando a distribuição HNT na definição anterior temos

$$\begin{aligned} B(p) &= \frac{1}{p\mu} \int_0^q x \frac{2}{\theta} \phi\left(\frac{x}{\theta}\right) \left[ 1 + \lambda - 2\lambda \left( 2\Phi\left(\frac{x}{\theta}\right) - 1 \right) \right] dx \\ &= \frac{-\theta^2}{p\mu} \left[ \frac{2}{\theta} \phi\left(\frac{q}{\theta}\right) 2\lambda \left( 2\Phi\left(\frac{q}{\theta}\right) - 1 \right) - \lambda \frac{2}{\theta} \phi\left(\frac{q}{\theta}\right) + \right. \\ &\quad \left. - \frac{2}{\theta} \phi\left(\frac{q}{\theta}\right) - \frac{2\lambda}{\theta\sqrt{\pi}} \left( 2\Phi\left(\frac{q\sqrt{2}}{\theta}\right) - 1 \right) + \frac{\sqrt{2}\lambda + \sqrt{2}}{\theta\sqrt{\pi}} \right] \\ &= \frac{-\theta^2}{p\mu} \left\{ \frac{2}{\theta} \phi\left(\frac{q}{\theta}\right) \left[ 1 + \lambda - 2\lambda \left( 2\Phi\left(\frac{q}{\theta}\right) - 1 \right) \right] + \right. \\ &\quad \left. - \frac{1}{\theta\sqrt{\pi}} \left[ \sqrt{2}(1+\lambda) + 2\lambda \left( 2\Phi\left(\frac{q\sqrt{2}}{\theta}\right) - 1 \right) \right] \right\}. \end{aligned}$$

□

De forma análoga obtemos a curva de Lorenz.

### 3.5.6 Estatística de Ordem

As estatísticas de ordem, assim como os momentos amostrais, desempenham um importante papel na inferência estatística (DAVID; NAGARAJA, 2003). Momentos de estatísticas de ordem desempenham um papel importante nos testes de controle de qualidade e confiabilidade, para prever o fracasso de itens futuros com base nos tempos de algumas falhas iniciais.

Sejam  $X_{(1)}, \dots, X_{(n)}$  as estatísticas de ordem de uma amostra aleatória  $X_1, \dots, X_n$  obtidas de uma população que segue uma distribuição  $f(x)$ , então a função de distribuição de probabilidade da  $j$ -ésima estatística de ordem é dada por:

$$f_j(x) = \frac{n!}{(j-1)!(n-j)!} f(x) [F(x)]^{j-1} [1 - F(x)]^{n-j} \quad (3.12)$$

para  $j = 1, \dots, n$ .

Aplicando as equações (3.3) e (3.4) em (3.12) obtemos a densidade da  $j$ -ésima estatís-

tica de ordem da HNT.

$$f_j(x | \theta, \lambda) = \frac{n!}{(j-1)!(n-j)!} \left\{ \frac{2}{\theta} \phi\left(\frac{x}{\theta}\right) \left[ 1 + \lambda - 2\lambda \left( 2\Phi\left(\frac{x}{\theta}\right) - 1 \right) \right] \right\} \times \\ \times \left\{ (1+\lambda) \left( 2\Phi\left(\frac{x}{\theta}\right) - 1 \right) - \lambda \left( 2\Phi\left(\frac{x}{\theta}\right) - 1 \right)^2 \right\}^{j-1} \times \\ \times \left\{ 1 - (1+\lambda) \left( 2\Phi\left(\frac{x}{\theta}\right) - 1 \right) + \lambda \left( 2\Phi\left(\frac{x}{\theta}\right) - 1 \right)^2 \right\}^{n-j}$$

Considerando a substituição  $\gamma(x) = 2\Phi\left(\frac{x}{\theta}\right) - 1$  temos:

$$f_j(x | \theta, \lambda) = \frac{n!}{(j-1)!(n-j)!} \left[ \frac{2}{\theta} \phi\left(\frac{x}{\theta}\right) (1 + \lambda - 2\lambda\gamma(x)) \right] \times \\ \times [(1+\lambda)\gamma(x) - \lambda\gamma(x)^2]^{j-1} [1 - (1+\lambda)\gamma(x) + \lambda\gamma(x)^2]^{n-j}.$$

Reordenando os termos, obtemos a fórmula geral para a  $j$ -ésima estatística de ordem da distribuição HNT:

$$f_j(x | \theta, \lambda) = \frac{n!}{(j-1)!(n-j)!} \frac{2}{\theta} \phi\left(\frac{x}{\theta}\right) \gamma(x)^{j-1} (1 + \lambda - 2\lambda\gamma(x)) \times \\ \times (1 + \lambda - \lambda\gamma(x))^{j-1} (\gamma(x)\lambda - 1)^{n-j} (\gamma(x) - 1)^{n-j}. \quad (3.13)$$

Podemos usar a distribuição HNT para modelar eventos máximos ou mínimos. Portanto, a seguir estão calculadas a primeira e a  $n$ -ésima estatística de ordem. Para a estatística de ordem  $n$ , devemos substituir  $j = n$  na fórmula geral (3.13). Portanto, a  $n$ -ésima estatística de ordem, que representa a distribuição do máximo da distribuição HNT é dada por:

$$f_n(x) = n \frac{2}{\theta} \phi\left(\frac{x}{\theta}\right) \gamma(x)^{n-1} (1 + \lambda - 2\lambda\gamma(x)) (1 + \lambda - \lambda\gamma(x))^{n-1}.$$

Para a estatística de ordem 1, substituímos na fórmula geral  $j = 1$ . O resultado gera a distribuição do mínimo da distribuição HNT dada por:

$$f_1(x) = n \frac{2}{\theta} \phi\left(\frac{x}{\theta}\right) (1 + \lambda - 2\lambda\gamma(x)) (\gamma(x)\lambda - 1)^{n-1} (\gamma(x) - 1)^{n-1}.$$

### 3.5.7 Confiabilidade

A estimativa da confiabilidade é de grande interesse na engenharia, sendo utilizada em modelos de estresse-força ou como medida de performance em sistemas elétricos e eletrônicos. Entretanto, também pode ser aplicada em outras áreas, pois permite uma medida geral das diferenças entre duas populações (ASGHARZADEH; VALIOLLAHI; RAQAB, 2011).

**Teorema 3.5.5.** Seja  $R = P(X_2 < X_1)$ , em que  $X_1$  e  $X_2$  são variáveis aleatórias independentes com distribuição HNT e parâmetros de escala e transmutação  $(\theta_1, \lambda_1)$  e  $(\theta_2, \lambda_2)$ , respectivamente, então:

$$\begin{aligned}
R &= P(X_2 < X_1) = \int_0^\infty f_1(x | \theta, \lambda) F_2(x | \theta, \lambda) dx \\
&= (1 + \lambda_1)(1 + \lambda_2) \frac{4}{\theta_1 \pi} \sum_{j=1}^{\infty} \frac{(-1)^{j+1} \left(\frac{\sqrt{2}}{2\theta_2}\right)^{2j+1} \theta^{2j+2} 2^{-\frac{1}{2}+j} \Gamma(j+1)}{(2j-1)(j-1)!} + \\
&\quad - (1 + \lambda_1)\lambda_2 \frac{8}{\theta_1 \pi^{\frac{3}{2}}} \sum_{j=1}^{\infty} \sum_{l=1}^{\infty} \frac{(-1)^{j+l+2} \left(\frac{\sqrt{2}}{2\theta_2}\right)^{2j+2l+2} \theta^{2j+2l+3} 2^{j+l} \Gamma(j+l+\frac{3}{2})}{(2j-1)(j-1)!(2l-1)(l-1)!} + \\
&\quad - \lambda_1(1 + \lambda_2) \frac{16}{\theta_1 \pi^{\frac{3}{2}}} \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \frac{(-1)^{k+j+2} \left(\frac{\sqrt{2}}{2\theta_1}\right)^{2k+1} \left(\frac{\sqrt{2}}{2\theta_2}\right)^{2j+1} \theta^{2k+2l+3} 2^{k+l} \Gamma(k+l+\frac{3}{2})}{(2k-1)(k-1)!(2j-1)(j-1)!} + \\
&\quad + \lambda_1\lambda_2 \frac{32}{\theta_1 \pi^2} \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \sum_{l=1}^{\infty} \frac{(-1)^{k+j+l+3} \left(\frac{\sqrt{2}}{2\theta_1}\right)^{2k+1} \left(\frac{\sqrt{2}}{2\theta_2}\right)^{2j+2l+2} \theta^{2k+2l+2j+4} 2^{\frac{1}{2}+k+l+j}}{(2k-1)(k-1)!(2j-1)(j-1)!(2l-1)(l-1)!} \times \\
&\quad \times \Gamma(k+l+j+2)
\end{aligned}$$

*Demonstração.* Utilizando a fdp e a fda definidas em (3.3) e (3.4) temos:

$$\begin{aligned}
R &= P(X_2 < X_1) = \int_0^\infty \frac{2}{\theta_1} \phi\left(\frac{x}{\theta_1}\right) \left[ 1 + \lambda_1 - 2\lambda_1 \left( 2\Phi\left(\frac{x}{\theta_1}\right) - 1 \right) \right] \times \\
&\quad \times \left[ (1 + \lambda_2) \left( 2\Phi\left(\frac{x}{\theta_2}\right) - 1 \right) - \lambda_2 \left( 2\Phi\left(\frac{x}{\theta_2}\right) - 1 \right)^2 \right] dx.
\end{aligned}$$

Substituindo  $2\Phi\left(\frac{x}{\theta}\right) - 1$  por  $\frac{2}{\sqrt{\pi}} \sum_{k=1}^{\infty} \frac{(-1)^{k+1} \left(\frac{\sqrt{2}}{2\theta}\right)^{2k+1} x^{2k+1}}{(2k-1)(k-1)!}$ , temos que

$$\begin{aligned}
R &= \int_0^\infty \frac{2}{\theta_1} \phi\left(\frac{x}{\theta_1}\right) \left\{ 1 + \lambda_1 - 2\lambda_1 \left[ \frac{2}{\sqrt{\pi}} \sum_{k=1}^{\infty} \frac{(-1)^{k+1} \left(\frac{\sqrt{2}}{2\theta_1}\right)^{2k+1} x^{2k+1}}{(2k-1)(k-1)!} \right] \right\} \times \\
&\quad \times \left\{ (1 + \lambda_2) \left( 2\Phi\left(\frac{x}{\theta_2}\right) - 1 \right) - \lambda_2 \left[ \frac{2}{\sqrt{\pi}} \sum_{j=1}^{\infty} \frac{(-1)^{j+1} \left(\frac{\sqrt{2}}{2\theta_2}\right)^{2j+1} x^{2j+1}}{(2j-1)(j-1)!} \right]^2 \right\} dx.
\end{aligned}$$

Aplicando a distributiva, usando as propriedades de integração e isolando os termos

que não dependem de  $x$ , obtemos:

$$\begin{aligned}
 R = & (1 + \lambda_1)(1 + \lambda_2) \frac{4}{\theta_1 \sqrt{\pi}} \sum_{j=1}^{\infty} \frac{(-1)^{j+1} \left(\frac{\sqrt{2}}{2\theta_2}\right)^{2j+1} \int_0^{\infty} \phi\left(\frac{x}{\theta_1}\right) x^{2j+1} dx}{(2j-1)(j-1)!} + \\
 & - (1 + \lambda_1)\lambda_2 \frac{8}{\theta_1 \pi} \sum_{j=1}^{\infty} \sum_{l=1}^{\infty} \frac{(-1)^{j+l+2} \left(\frac{\sqrt{2}}{2\theta_2}\right)^{2j+2l+2} \int_0^{\infty} \phi\left(\frac{x}{\theta_1}\right) x^{2j+2l+2} dx}{(2j-1)(j-1)!(2l-1)(l-1)!} + \\
 & - \lambda_1(1 + \lambda_2) \frac{16}{\theta_1 \pi} \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \frac{(-1)^{k+j+2} \left(\frac{\sqrt{2}}{2\theta_1}\right)^{2k+1} \left(\frac{\sqrt{2}}{2\theta_2}\right)^{2j+1} \int_0^{\infty} \phi\left(\frac{x}{\theta_1}\right) x^{2k+2j+2} dx}{(2k-1)(k-1)!(2j-1)(j-1)!} + \\
 & + \lambda_1\lambda_2 \frac{32}{\theta_1 \pi^{\frac{3}{2}}} \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \sum_{l=1}^{\infty} \frac{(-1)^{k+j+l+3} \left(\frac{\sqrt{2}}{2\theta_1}\right)^{2k+1} \left(\frac{\sqrt{2}}{2\theta_2}\right)^{2j+2l+2} \int_0^{\infty} \phi\left(\frac{x}{\theta_1}\right) x^{2k+2j+2l+3} dx}{(2k-1)(k-1)!(2j-1)(j-1)!(2l-1)(l-1)!}
 \end{aligned}$$

Resolvendo as integrais, concluímos que:

$$\begin{aligned}
 R = & (1 + \lambda_1)(1 + \lambda_2) \frac{4}{\theta_1 \pi} \sum_{j=1}^{\infty} \frac{(-1)^{j+1} \left(\frac{\sqrt{2}}{2\theta_2}\right)^{2j+1} \theta^{2j+2} 2^{-\frac{1}{2}+j} \Gamma(j+1)}{(2j-1)(j-1)!} + \\
 & - (1 + \lambda_1)\lambda_2 \frac{8}{\theta_1 \pi^{\frac{3}{2}}} \sum_{j=1}^{\infty} \sum_{l=1}^{\infty} \frac{(-1)^{j+l+2} \left(\frac{\sqrt{2}}{2\theta_2}\right)^{2j+2l+2} \theta^{2j+2l+3} 2^{j+l} \Gamma(j+l+\frac{3}{2})}{(2j-1)(j-1)!(2l-1)(l-1)!} + \\
 & - \lambda_1(1 + \lambda_2) \frac{16}{\theta_1 \pi^{\frac{3}{2}}} \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \frac{(-1)^{k+j+2} \left(\frac{\sqrt{2}}{2\theta_1}\right)^{2k+1} \left(\frac{\sqrt{2}}{2\theta_2}\right)^{2j+1} \theta^{2k+2l+3} 2^{k+l} \Gamma(k+l+\frac{3}{2})}{(2k-1)(k-1)!(2j-1)(j-1)!} + \\
 & + \lambda_1\lambda_2 \frac{32}{\theta_1 \pi^2} \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \sum_{l=1}^{\infty} \frac{(-1)^{k+j+l+3} \left(\frac{\sqrt{2}}{2\theta_1}\right)^{2k+1} \left(\frac{\sqrt{2}}{2\theta_2}\right)^{2j+2l+2} \theta^{2k+2l+2j+4} 2^{\frac{1}{2}+k+l+j}}{(2k-1)(k-1)!(2j-1)(j-1)!(2l-1)(l-1)!} \times \\
 & \times \Gamma(k+l+j+2).
 \end{aligned}$$

□

Observe que para o caso em que  $\theta_1 = \theta_2$ , segue imediatamente que

$$R = \frac{1}{6}(\lambda_2 - \lambda_1) + \frac{1}{2}.$$

### 3.6 Estimação Via Máxima Verossimilhança

Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da distribuição HNT com parâmetros  $\theta$  e  $\lambda$  e função de probabilidade dada por (3.3). A função de log-verossimilhança

pode ser escrita na forma:

$$\ell(\theta, \lambda | \mathbf{x}) = n \log \left( \frac{2}{\theta} \right) + \sum_{i=1}^n \log \left( \phi \left( \frac{x_i}{\theta} \right) \right) + \sum_{i=1}^n \log \left( -4 \lambda \Phi \left( \frac{x_i}{\theta} \right) + 3 \lambda + 1 \right). \quad (3.14)$$

Diferenciando (3.14) com respeito a  $\theta$  e  $\lambda$ , obtemos as seguintes equações:

$$\begin{cases} \frac{\partial \ell(\theta, \lambda | \mathbf{x})}{\partial \lambda} = \sum_{i=1}^n \frac{-4 \Phi \left( \frac{x_i}{\theta} \right) + 3}{-4 \lambda \Phi \left( \frac{x_i}{\theta} \right) + 3 \lambda + 1} \\ \frac{\partial \ell(\theta, \lambda | \mathbf{x})}{\partial \theta} = -\frac{n}{\theta} - \frac{1}{\theta^2} \sum_{i=1}^n \frac{\phi' \left( \frac{x_i}{\theta} \right)}{\phi \left( \frac{x_i}{\theta} \right)} x_i + \frac{4\lambda}{\theta^2} \sum_{i=1}^n \frac{\phi \left( \frac{x_i}{\theta} \right)}{-4 \lambda \Phi \left( \frac{x_i}{\theta} \right) + 3 \lambda + 1} x_i \end{cases}$$

em que  $\phi' \left( \frac{x_i}{\theta} \right) = \frac{d}{d\theta} \phi \left( \frac{x_i}{\theta} \right)$ .

Os estimadores de máxima verossimilhança (EMV) de  $\theta$  e  $\lambda$  são as soluções das equações não-lineares  $\frac{\partial \ell(\theta, \lambda | \mathbf{x})}{\partial \lambda} = 0$  e  $\frac{\partial \ell(\theta, \lambda | \mathbf{x})}{\partial \theta} = 0$ . Observa-se que para obter as estimativas de máxima verossimilhança é necessário a utilização de métodos numéricos.

Dado o vetor de parâmetros desconhecidos  $\Theta = (\theta, \lambda)$ , para amostras grandes é conhecido que a distribuição assintótica do EMV  $\Theta$  é

$$(\hat{\Theta} - \Theta) \rightarrow N_2(0, I^{-1}(\Theta)).$$

A inversa da matriz de informação de Fisher observada,  $I^{-1}(\Theta)$ , é dada por:

$$I^{-1}(\Theta) = \begin{pmatrix} \text{var}(\hat{\theta}_{EMV}) & \text{cov}(\hat{\theta}_{EMV}, \hat{\lambda}_{EMV}) \\ \text{cov}(\hat{\lambda}_{EMV}, \hat{\theta}_{EMV}) & \text{var}(\hat{\lambda}_{EMV}) \end{pmatrix}$$

Portanto, a abordagem acima é usada para obter os intervalos de confiança de 100%(1− $\tau$ ) aproximados dos parâmetros  $\Theta = (\theta, \lambda)$ , como nas seguintes formas

$$\hat{\theta}_{EMV} \pm z_{\frac{\tau}{2}} \sqrt{\text{var}(\hat{\theta}_{MLE})} \quad ; \quad \hat{\lambda}_{EMV} \pm z_{\frac{\tau}{2}} \sqrt{\text{var}(\hat{\lambda}_{MLE})}$$

em que, tem-se o  $z_{\frac{\tau}{2}}$ -ésimo percentil da distribuição Normal padrão.

## 3.7 Simulação

Nesta seção são apresentados os resultados de um estudo de simulação Monte Carlo usado para avaliar o viés e o erro quadrático médio das estimativas obtidas pelo método da máxima verossimilhança. Foram geradas amostras de tamanho  $n = 20, 50, \dots, 170, 200$  com  $\lambda = -0.9, -0.7, -0.5, -0.3, 0, 0.3, 0.5, 0.7, 0.9$  e  $\theta = 1.0$ . Para cada uma das combi-

nações de  $n, \theta$  e  $\lambda$ , foram simuladas, pelo método da transformação inversa,  $N = 10000$  amostras pseudo-aleatórias da distribuição HNT. As estimativas do viés e do erro quadrático médio são apresentadas nas Figuras 3.3 e 3.4.

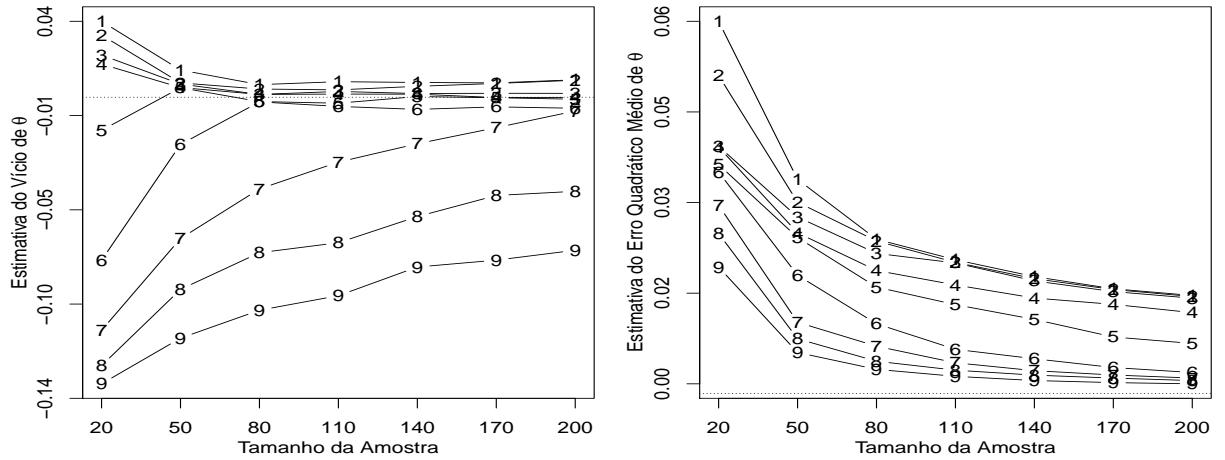


Figura 3.3: Estimativa do vício e do erro quadrático médio de  $\theta$  ( $\theta = 1$ ,  $1 : \lambda = -0.9$ ,  $2 : \lambda = -0.7$ ,  $3 : \lambda = -0.5$ ,  $4 : \lambda = -0.3$ ,  $5 : \lambda = 0$ ,  $6 : \lambda = 0.3$ ,  $7 : \lambda = 0.5$ ,  $8 : \lambda = 0.7$  e  $9 : \lambda = 0.9$ ).

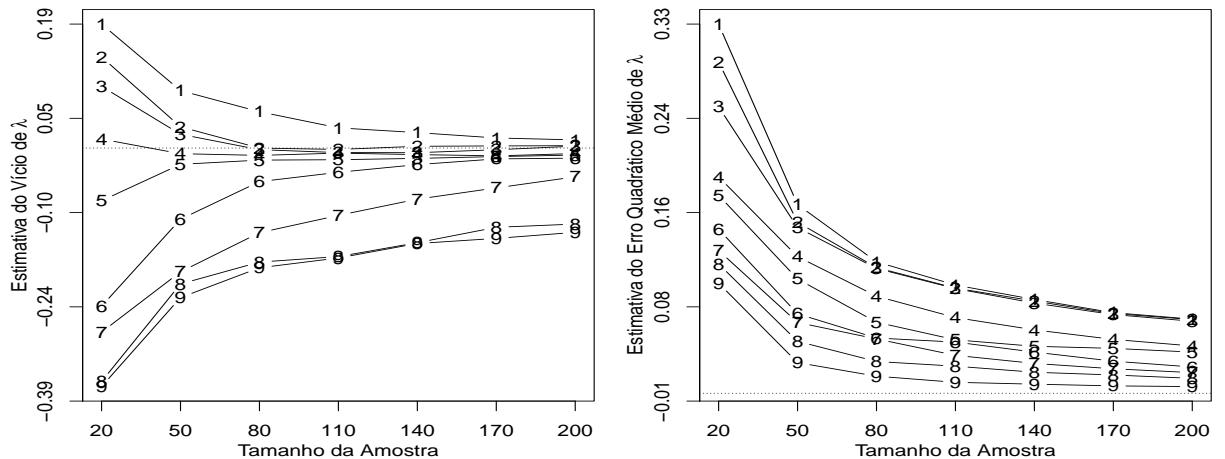


Figura 3.4: Estimativa do vício e do erro quadrático médio de  $\lambda$  ( $\theta = 1$ ,  $1 : \lambda = -0.9$ ,  $2 : \lambda = -0.7$ ,  $3 : \lambda = -0.5$ ,  $4 : \lambda = -0.3$ ,  $5 : \lambda = 0$ ,  $6 : \lambda = 0.3$ ,  $7 : \lambda = 0.5$ ,  $8 : \lambda = 0.7$  e  $9 : \lambda = 0.9$ ).

Ao analisarmos o viés de  $\hat{\theta}$ , podemos notar que para valores de  $\lambda = -0.9, -0.7, -0.5, -0.3, 0, 0.3, 0.5$ ,  $\hat{\theta}$  apresentou uma excelente estimativa convergindo para zero mesmo para amostras pequenas. Embora para  $\lambda = 0.7$  e  $0.9$  a convergência de  $\hat{\theta}$  seja um pouco mais demorada, com o acréscimo amostral o viés de estimativa é muito baixo, na casa dos 0.08.

Já o viés de  $\hat{\lambda}$  tem amplitudes maiores que o viés de  $\hat{\theta}$ , mostrando que o parâmetro de escala  $\theta$  exerce influência na estimativa do parâmetro de transmutação. Ainda assim, podemos notar que para os valores de  $\lambda = -0.9, -0.7, -0.5, -0.3, 0, 0.3, 0.5$ ,  $\hat{\lambda}$  é estimado

com precisão. Assim como para o caso anterior,  $\lambda = 0.7, 0.9$  tem uma convergência mais lenta, sendo o erro de estimativa próximo à 0.15.

Os erros quadráticos médios de  $\hat{\theta}$  são extremamente baixos, sendo os valores positivos de  $\lambda$  os que apresentam uma convergência mais precisa. Apesar dos parâmetros positivos de  $\lambda$  não convergirem diretamente para zero, podemos notar ao observar o gráfico que os erros na estimativa são inferiores a 0.03. Sendo assim, podemos concluir que os erros nas estimativas são praticamente insignificantes.

Por fim, ao observamos o erro quadrático médio do parâmetro  $\hat{\lambda}$ , percebemos que  $\theta$  exerce influência em sua estimação. De forma análoga à anterior, podemos ver que, para todos os cenários apresentados, os erros tendem a zero, sendo os mais afastados próximos a 0.1.

De modo geral, podemos concluir que os estimadores possuem a propriedade de não tendenciosidade, pois o vício tende a zero conforme  $n$  aumenta. Quanto ao erro quadrático médio, é possível observar a propriedade de consistência, pois quando o valor de  $n$  aumenta os erros tendem a zero.

## 3.8 Aplicações

Nesta seção aplicamos a distribuição HNT a dois conjuntos de dados. Nosso objetivo é avaliar o seu ajuste em relação a outras distribuições já apresentadas na literatura. Os dados utilizados foram compilados a partir das séries históricas de precipitação diária obtidas no portal do Instituto Nacional de Meteorologia (<http://www.inmet.gov.br>). No ajuste das distribuições, foram consideradas como resposta o total acumulado no mês. O total acumulado mensalmente, trimestralmente, e assim sucessivamente, é muito utilizado no cálculo do índice padronizado de precipitação (SPI: The Standardized Precipitation Index).

Para esta aplicação, consideramos os dados relativos ao mês de fevereiro de cada ano, de 1974 à 2016, da cidade de Chapecó no estado de Santa Catarina. É importante ressaltar que, devido a alguma falha, não ouve medição no mês de fevereiro de alguns dos anos no período considerado. Na Tabela 3.1 são apresentadas as medidas resumo a respeito dos dados utilizados. Observando o TTT-plot da Figura 3.5, é possível notar que o risco é crescente, um indicativo de que a distribuição HNT pode ser um modelo apropriado para ajuste.

<u>Tabela 3.1: Medidas resumo.</u>	
Número de Observações	30
Média	206.93 mm
Desvio-Padrão	128.55 mm
Mínimo	8.20 mm
Q25	107.18 mm
Mediana	191.80 mm
Q75	270.45 mm
Máximo	500.30 mm

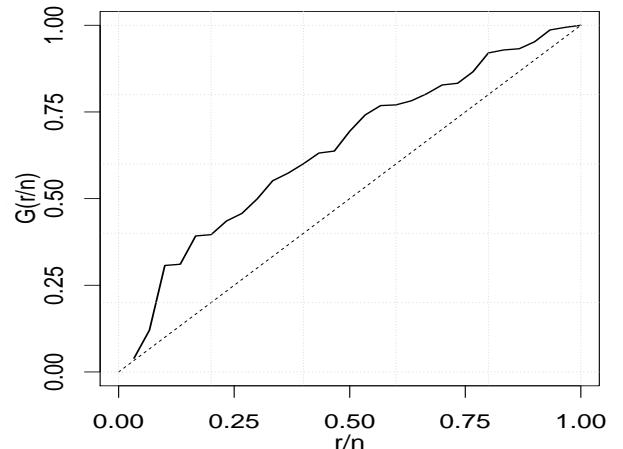


Figura 3.5: TTT-plot.

Assim, comparamos a distribuição HNT com as seguintes extensões da distribuição HN:

- Distribuição Half-Normal (HN) (DANIEL, 1959).

A função densidade da distribuição HN uniparamétrica é dada por:

$$f(x | \theta) = \frac{2}{\theta} \phi\left(\frac{x}{\theta}\right),$$

em que  $\theta > 0$  é o parâmetro de escala e  $\phi(\cdot)$  é a fdp da Normal padrão.

- Distribuição Half-Normal Geral (HNG) (PEWSEY, 2002, 2004).

A função densidade da distribuição HNG de dois parâmetros é dada por:

$$f(x | \theta, \alpha) = \frac{2}{\theta} \phi\left(\frac{x - \alpha}{\theta}\right),$$

em que  $\alpha = \min(x)$ ,  $\theta > 0$  é o parâmetro de escala e  $\phi(\cdot)$  é a fdp da Normal padrão.

- Distribuição Half-Normal Potência (HNP) (GÓMEZ; BOLFARINE, 2015).

A função densidade da distribuição HNP de dois parâmetros é dada por:

$$f(x | \theta, \alpha) = \frac{2\alpha}{\theta} \phi\left(\frac{x}{\theta}\right) \left(2\Phi\left(\frac{x}{\theta}\right) - 1\right)^{\alpha-1},$$

em que  $\theta > 0$  é o parâmetro de escala,  $\alpha > 0$  é o parâmetro de forma,  $\phi(\cdot)$  e  $\Phi(\cdot)$  são, respectivamente, a fdp e a fda da Normal padrão.

- Distribuição Half-Normal Generalizada (HNG II) (COORAY; ANANDA, 2008).

A função densidade da distribuição HNG II de dois parâmetros é dada por:

$$f(x | \theta\alpha) = \frac{2\alpha}{\theta} \left(\frac{x}{\theta}\right)^{\alpha-1} \phi\left[\left(\frac{x}{\theta}\right)^\alpha\right],$$

em que  $\theta > 0$  é o parâmetro de escala,  $\alpha > 0$  é o parâmetro de forma e  $\phi(\cdot)$  é a fdp da Normal padrão.

- Distribuição Gama Half-Normal (GMHN) (ALZAATREH; KNIGHT, 2013).

A função densidade da distribuição GMHN de três parâmetros é dada por:

$$f(x | \theta, \alpha, b) = \frac{\frac{2}{\theta}\phi\left(\frac{x}{\theta}\right)}{\Gamma(\alpha)b^\alpha} \left\{ -\log\left[2\Phi\left(\frac{-x}{\theta}\right)\right] \right\}^{\alpha-1} \left[2\Phi\left(\frac{-x}{\theta}\right)\right]^{\frac{1}{b}-1},$$

em que  $\theta > 0$  é um parâmetro de escala,  $\alpha > 0$  e  $b > 0$  são parâmetros adicionais,  $\phi(\cdot)$  e  $\Phi(\cdot)$  são, respectivamente, a fdp e a fda na Normal padrão.

- Distribuição Odd Log-Logistica Half-Normal Generalizada (OLLHNG) (CORDEIRO et al., 2017).

A função densidade da distribuição OLLHNG de três parâmetros é dada por:

$$f(x | \theta, \alpha, a) = \frac{a\sqrt{\left(\frac{2}{\pi}\right)}\left(\frac{\alpha}{x}\right)\left(\frac{x}{\theta}\right)^\alpha e^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}} \left\{ [2\Phi\left(\left(\frac{x}{\theta}\right)^\alpha\right) - 1] [2 - 2\Phi\left(\left(\frac{x}{\theta}\right)^\alpha\right)] \right\}^{a-1}}{\left\{ [2\Phi\left(\left(\frac{x}{\theta}\right)^\alpha\right) - 1]^a + [2 - 2\Phi\left(\left(\frac{x}{\theta}\right)^\alpha\right)]^a \right\}^2},$$

em que  $\theta > 0$  é um parâmetro de escala,  $\alpha > 0$  e  $a > 0$  são parâmetros de forma,  $\phi(\cdot)$  e  $\Phi(\cdot)$  são, respectivamente, a fdp e a fda na Normal padrão.

- Distribuição Half-Normal Generalizada Transmutada (HNGT).

A função densidade da distribuição HNGT de três parâmetros é dada por:

$$f(x | \theta, \alpha, \lambda) = \frac{2\alpha}{\theta} \left(\frac{x}{\theta}\right)^{\alpha-1} \phi\left(\left(\frac{x}{\theta}\right)^\alpha\right) \left[(1 + \lambda) - 2\lambda \left(2\Phi\left(\left(\frac{x}{\theta}\right)^\alpha\right) - 1\right)\right],$$

em que  $\theta > 0$  é o parâmetro de escala,  $\alpha > 0$  é o parâmetro de forma,  $|\lambda| \leq 1$  é o parâmetro de transmutação,  $\phi(\cdot)$  e  $\Phi(\cdot)$  são, respectivamente, a fdp e a fda da Normal padrão.

- Distribuição Beta Half-Normal Generalizada (BHNG) (PESCIM et al., 2010; CORDEIRO et al., 2013; PESCIM et al., 2013).

A função densidade da distribuição BHNG de quatro parâmetros é dada por:

$$f(x | \theta, \alpha, a, b) = \frac{2^b \alpha}{a \theta B(a, b)} \left(\frac{x}{\theta}\right)^{\alpha-1} \phi\left(\left(\frac{x}{\theta}\right)^\alpha\right) \left[2\Phi\left(\left(\frac{x}{\theta}\right)^\alpha\right) - 1\right]^{a-1} \left[\Phi\left(-\left(\frac{x}{\theta}\right)^\alpha\right)\right]^{b-1},$$

em que  $\theta, \alpha, a, b > 0$  e  $B(a, b)$  denota a função beta completa definida como  $\int_0^1 u^{a-1}(1-u)^{b-1} du$ .

- Distribuição Half-Normal Generalizada Estendida (HNGE) (SANCHEZ; FREITAS; CORDEIRO, 2016).

A função densidade da distribuição HNGE de quatro parâmetros é dada por:

$$f(x | \theta, \alpha, a, b) = ab\sqrt{\frac{2}{\pi}} \left(\frac{\alpha}{x}\right) \left(\frac{x}{\theta}\right)^\alpha \exp\left(-\frac{1}{2} \left(\frac{x}{\theta}\right)^{2\alpha}\right) \left[2 - 2\Phi\left(\left(\frac{x}{\theta}\right)^\alpha\right)\right]^{a-1} \times \\ \times \left\{1 - \left[2 - 2\Phi\left(\left(\frac{x}{\theta}\right)^\alpha\right)\right]^a\right\}^{b-1},$$

em que  $\theta > 0$  é o parâmetro de escala,  $\alpha, a, b > 0$  são parâmetros de forma e  $\Phi(\cdot)$  é a fda da Normal padrão.

- Distribuição Kumaraswamy Half-Normal Generalizada (KHNG) (CORDEIRO; PES-CIM; ORTEGA, 2012).

A função densidade da distribuição KHNG de quatro parâmetros é dada por:

$$f(x | \theta, \alpha, a, b) = ab\sqrt{\frac{2}{\pi}} \left(\frac{\alpha}{x}\right) \left(\frac{x}{\theta}\right)^\alpha \exp\left(-\frac{1}{2} \left(\frac{x}{\theta}\right)^{2\alpha}\right) \left[2\Phi\left(\left(\frac{x}{\theta}\right)^\alpha\right) - 1\right]^{a-1} \times \\ \times \left\{1 - \left[2\Phi\left(\left(\frac{x}{\theta}\right)^\alpha\right) - 1\right]^a\right\}^{b-1},$$

em que  $\theta > 0$  é o parâmetro de escala,  $\alpha, a, b > 0$  são parâmetros de forma e  $\Phi(\cdot)$  é a fda da Normal padrão.

- Distribuição Beta Half-Normal Generalizada Geométrica (BHNGG) (RAMIRES et al., 2013).

A função densidade da distribuição BHNGG de cinco parâmetros é dada por:

$$f(x | \theta, \alpha, a, b, p) = (1-p) \frac{2^{b-\frac{1}{2}} \left(\frac{\alpha}{x}\right) \left(\frac{x}{\theta}\right)^\alpha \exp\left(-\frac{1}{2} \left(\frac{x}{\theta}\right)^{2\alpha}\right)}{\sqrt{\pi} B(a, b)} \left[2\Phi\left(\left(\frac{x}{\theta}\right)^\alpha\right) - 1\right]^{a-1} \times \\ \times \left[1 - \Phi\left(\left(\frac{x}{\theta}\right)^\alpha\right)\right]^{b-1} \left[1 - p \left(1 - I_{2\Phi\left(\left(\frac{x}{\theta}\right)^\alpha\right)-1}(a, b)\right)\right]^{-2},$$

em que  $\alpha, a, b > 0$  e  $0 \leq p < 1$  são parâmetros de forma,  $\theta > 0$  é o parâmetro de escala e  $I_{2\Phi\left(\left(\frac{x}{\theta}\right)^\alpha\right)-1}(a, b) = B_{2\Phi\left(\left(\frac{x}{\theta}\right)^\alpha\right)-1}(a, b)/B(a, b)$  é a taxa da função beta incompleta.

A Tabela 3.2 mostra as estimativas de máxima verossimilhança (erros padrão) das 12 distribuições ajustadas. Todas as estimativas foram obtidas no SAS/PROC NLMI-

XED (SAS, 2010), aplicando a técnica de otimização de Newton-Raphson. Embora, para todas as distribuições, as matrizes de variância-covariância resultantes foram definidas positivamente e  $\max \left[ \frac{\partial}{\partial \Theta} \ell(\Theta | \mathbf{x}) \Big|_{\Theta=\hat{\Theta}} \right] < 0.000001$ , observamos erros padrão atípicos para alguns parâmetros nas distribuições GMHN, BGHN, HNGE e KHNG. Nossa palpite é que eles convergem para um mínimo local ou, muito provavelmente, que os parâmetros são funções lineares uns dos outros (ou quase colineares) em relação aos dados em questão. De fato, a partir das matrizes de correlação obtemos  $\text{corr}(\hat{\theta}, \hat{\alpha}) = 0.9594$ ,  $\text{corr}(\hat{b}, \hat{\alpha}) = -0.9707$  e  $\text{corr}(\hat{\theta}, \hat{b}) = -0.9984$  para a distribuição GMHN,  $\text{corr}(\hat{\alpha}, \hat{b}) = 0.7179$ ,  $\text{corr}(\hat{\alpha}, \hat{a}) = -0.8255$ ,  $\text{corr}(\hat{\theta}, \hat{\alpha}) = 0.8587$  e  $\text{corr}(\hat{\theta}, \hat{\alpha}) = 0.9711$  para a distribuição BHNG,  $\text{corr}(\hat{\theta}, \hat{b}) = 0.9817$ ,  $\text{corr}(\hat{a}, \hat{b}) = 0.9896$  e  $\text{corr}(\hat{\theta}, \hat{a}) = 0.9988$  para a distribuição HNGE,  $\text{corr}(\hat{\alpha}, \hat{a}) = -0.8899$ ,  $\text{corr}(\hat{\theta}, \hat{a}) = 0.8977$ ,  $\text{corr}(\hat{\alpha}, \hat{b}) = -0.9535$ ,  $\text{corr}(\hat{a}, \hat{b}) = 0.9568$ ,  $\text{corr}(\hat{\theta}, \hat{b}) = 0.9726$  e  $\text{corr}(\hat{\theta}, \hat{\alpha}) = -0.9807$  para a distribuição KHNG. Não foi possível estimar o erro padrão de todos os parâmetros da distribuição BHNGG e sua matriz de variância-covariância também não foi completamente preenchida, mas podemos observar que  $\text{corr}(\hat{\alpha}, \hat{p}) = 0.7391$ ,  $\text{corr}(\hat{\alpha}, \hat{a}) = -0.7480$ ,  $\text{corr}(\hat{\alpha}, \hat{b}) = -1.0000$  e  $\text{corr}(\hat{b}, \hat{p}) = -1.0000$ .

Para comparar as distribuições, consideramos as estatísticas baseadas na verossimilhança  $-2 \times \text{Log-like}$ , AIC, AICc e BIC e as medidas de bom ajuste KS, AD e CvM. O melhor modelo é aquele que fornece os valores mínimos desses critérios. A Tabela 3.3 mostra tais valores e o índice indica a classificação obtida para cada distribuição. Nós também temos na última coluna uma classificação total (soma das classificações) para cada distribuição. Uma vez que há grandes incertezas nas estimativas das distribuições GMHN, BHNG, HNGE, KHNG e BHNGG, elas não serão consideradas. Ao observarmos a tabela, podemos notar que a distribuição HNT ficou em primeiro lugar, seguida das distribuições HNGT e HNP. Ambas distribuições, HNT e HNP, possuem dois parâmetros. Já a distribuição HNGT possui três parâmetros, e justamente o parâmetro adicional ( $\lambda$ ) possui o zero em seu respectivo intervalo de confiança. Tendo isso em vista, podemos inferir que os modelos que obtiveram o melhor ajuste foram então HNT (17) e o HNP (24).

Tabela 3.2: Estimativas de máxima verossimilhança (erros padrão).

Distribuições	$\theta$	$\alpha$	$\lambda$	$a$	$b$	$p$
HN	242.4746 (31.3031)	.	.	.	.	.
HNG	235.5162 (30.4049)	.	.	.	.	.
HNP	211.7649 (28.9254)	1.4462 (0.3307)	.	.	.	.
HNT	202.6784 (24.8288)	.	-0.6767 (0.3097)	.	.	.
HNG II	260.1236 (28.3470)	1.2904 (0.1952)	.	.	.	.
HN GT	201.8721 (59.6638)	0.9952 (0.3247)	-0.6852 (0.6450)	.	.	.
OLLHNG	262.3615 (33.1280)	1.2245 (0.4707)	.	1.0701 (0.4822)	.	.
GMHN	120.00 (324.49)	1.2035 (0.9856)	.	.	2.4547 (12.5272)	.
BHNG	94.6287 (189.44)	1.0069 (0.4544)	.	1.1730 (0.8355)	0.2371 (0.6290)	.
HN GE	49.9935 (989.30)	0.9418 (0.4977)	.	0.0947 (3.3346)	1.0646 (4.4612)	.
KHNG	640.00 (298.26)	8.0006 (0.9535)	.	0.1852 (0.0772)	3.9998 (5.5428)	.
BHNGG	86.8732 (<0.001)	1.1999 (<0.001)	.	0.9728 (0.2289)	0.0806 (<0.001)	0.5568 (<0.001)

Tabela 3.3: Valores de -2Log, AIC, AICc, BIC, KS, AD e CvM.

Distribuições	-2 × Log-Like	AIC	AICc	BIC	KS	AD	CvM	Total
HN	373.0016 <sup>7</sup>	375.0016 <sup>5</sup>	375.1444 <sup>5</sup>	376.4028 <sup>2</sup>	0.8602 <sup>7</sup>	0.7372 <sup>6</sup>	0.1312 <sup>7</sup>	39 <sup>7</sup>
HNG	371.2544 <sup>6</sup>	373.2544 <sup>1</sup>	373.3972 <sup>1</sup>	374.6556 <sup>1</sup>	0.7476 <sup>6</sup>	23.9874 <sup>7</sup>	0.0954 <sup>6</sup>	28 <sup>4</sup>
HNP	370.5842 <sup>5</sup>	374.5842 <sup>4</sup>	375.0287 <sup>4</sup>	377.3866 <sup>5</sup>	0.3984 <sup>3</sup>	0.1498 <sup>2</sup>	0.0167 <sup>1</sup>	24 <sup>3</sup>
HNT	370.4147 <sup>2</sup>	374.4147 <sup>2</sup>	374.8591 <sup>2</sup>	377.2170 <sup>3</sup>	0.3802 <sup>2</sup>	0.1500 <sup>3</sup>	0.0192 <sup>3</sup>	17 <sup>1</sup>
HNG II	370.4468 <sup>4</sup>	374.4468 <sup>3</sup>	374.8913 <sup>3</sup>	377.2492 <sup>4</sup>	0.4327 <sup>5</sup>	0.1721 <sup>5</sup>	0.0224 <sup>5</sup>	29 <sup>5</sup>
HN GT	370.4144 <sup>1</sup>	376.4144 <sup>6</sup>	377.3375 <sup>6</sup>	380.6180 <sup>6</sup>	0.3777 <sup>1</sup>	0.1488 <sup>1</sup>	0.0189 <sup>2</sup>	23 <sup>2</sup>
OLLHNG	370.4232 <sup>3</sup>	376.4232 <sup>7</sup>	377.3463 <sup>7</sup>	380.6268 <sup>7</sup>	0.4035 <sup>4</sup>	0.1589 <sup>4</sup>	0.0196 <sup>4</sup>	36 <sup>6</sup>

Tabela 3.4: Resultados dos testes de aderência.

Distribuição	KS (valor-p)	AD (valor-p)	CvM (valor-p)
HN	0.4589	0.5273	0.4544
HNG	0.6425	<0.001	0.6113
HNP	0.9982	0.9987	0.9993
HNT	0.9992	0.9987	0.9981
HNG II	0.9946	0.9963	0.9952
HNGT	0.9993	0.9987	0.9983
OLLHNG	0.9978	0.9979	0.9978

Para a próxima aplicação, vamos considerar os dados de precipitação acumulada de dez dias entre abril e julho na estação WMO 83498, localizada no estado da Bahia, Brasil. Utilizamos a série histórica de 1961 a 2017. Nesta aplicação, utilizaremos a seguinte função de ligação:

$$\log(\theta_i) = \beta_0 + \beta_1 x_i,$$

em que  $x_i$  denota a  $i$ -ésima observação associada ao  $t_i$  período de dez dias. Os períodos de dez dias foram considerados para os meses de abril, maio, junho e julho. Consideramos as distribuições HN, HNG II e HNT, uma vez que apenas elas possuem expressões analíticas fechadas para a média.

Na Tabela 3.5 são relatadas as estimativas dos parâmetros e os erros padrão. Uma vez que o  $\hat{\beta}_1$  tem um sinal negativo para todas as distribuições, temos um indicativo da diminuição da precipitação acumulada à medida que o tempo passa. As médias empíricas e estimadas (intervalos de confiança de 95%) são apresentadas na Tabela 3.6 e Figura 3.6. Além disso, na Tabela 3.7 estão presentes vários critérios para discriminar as distribuições HN, HNG II e HNT.

A partir desses resultados, podemos concluir que HNG II e HNT proporcionam melhor ajuste do que a distribuição HN. Também é observado que HNT tem os valores mais baixos desses critérios.

Tabela 3.5: Estimativas de parâmetros e intervalos de confiança de 95%.

HN		HNG II		HNT	
Estimativa	I.C.	Estimativa	I.C.	Estimativa	I.C.
$\beta_0$	5.0942 (4.7007, 5.4876)	5.0119 (4.5392, 5.4846)	5.4141 (4.9208, 5.9075)		
$\beta_1$	-0.0785 (-0.1064, -0.0507)	-0.0795 (-0.1129, -0.0462)	-0.0792 (-0.1123, -0.0462)		
$\alpha$	—	0.8348 (0.7700, 0.8996)	—	—	—
$\lambda$	—	—	—	0.7822 (0.5896, 0.9747)	

Tabela 3.6: Médias empíricas e estimadas para cada período de dez dias.

Decêndio	Empírica	HN	HNG II	HNT
10	51.6089	59.3420	55.1367	54.8366
11	48.1578	54.8614	50.9212	50.6590
12	49.1844	50.7191	47.0280	46.7996
13	41.0744	46.8896	43.4325	43.2342
14	41.1256	43.3492	40.1118	39.9404
15	37.8395	40.0761	37.0451	36.8976
16	32.3732	37.0502	34.2128	34.0866
17	31.9098	34.2528	31.5970	31.4897
18	26.2780	31.6665	29.1813	29.0907

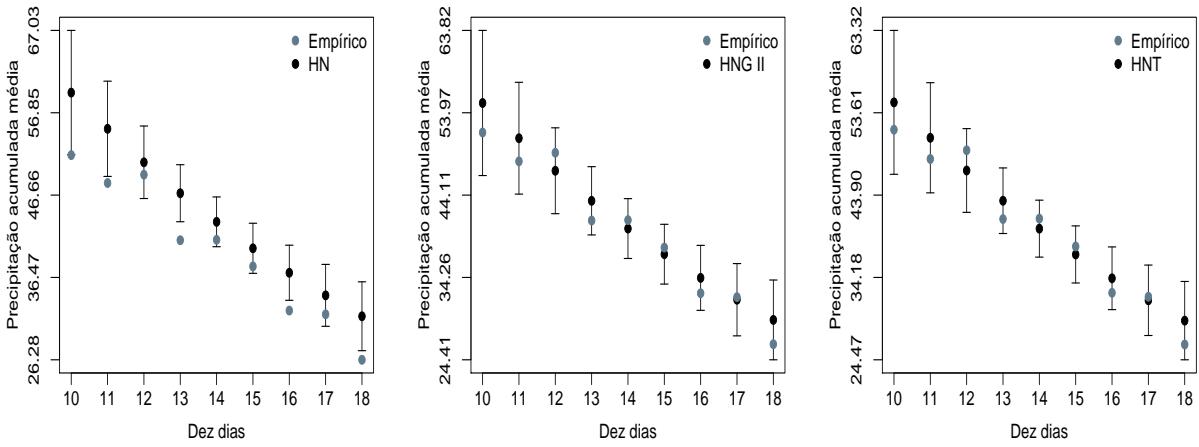


Figura 3.6: Média empírica e estimada para cada período de dez dias.

Tabela 3.7: Medidas de seleção de modelo.

Critério	HN	HNG II	HNT
$-2 \times \text{Log-Like}$	3631.9093	3608.9037	3597.1897
AIC	3635.9093	3614.9037	3603.1897
BIC	3643.8106	3626.7556	3615.0416
SSR	207.2568	43.8623	40.3417

## 3.9 Conclusão

Neste artigo foi apresentada a distribuição HNT formulada a partir dos mapas de transmutação de classificação quadrática propostos por Shaw e Buckley (2009). Algumas características e propriedades matemáticas da distribuição proposta foram estudadas. É importante observar que a função geradora de momentos, momento de ordem  $k$ , esperança, variância, assimetria e curtose apresentaram expressões analíticas explícitas, sendo as mesmas dependentes apenas dos parâmetros da distribuição HNT. Graças à simplicidade da distribuição, foi possível calcularmos medidas de incerteza como a entropia de

Shannon e os desvios médios. As curvas de Bonferroni e Lorenz assim como a estimativa da confiabilidade são explicitadas, dando a oportunidade de áreas como, por exemplo, a engenharia, se beneficiar com a utilização da distribuição proposta. Foi calculada a estatística de ordem para a distribuição HNT e também as respectivas expressões para a distribuição do máximo e do mínimo. Um estudo de simulação Monte Carlo mostrou que os parâmetros são eficientemente estimados pelo método da máxima verossimilhança, mostrando um baixo vício e uma alta acurácia mesmo para um pequeno tamanho de amostra, o que indica o potencial que a nova distribuição proporciona para a modelagem. Na primeira aplicação foram usados dados de precipitação diária, e 11 modelos propostos na literatura (derivados das distribuição HN) foram colocados em competição com a distribuição proposta nesse trabalho. Cinco modelos foram retirados da análise devido a inconsistência de suas estimativas e, ao observarmos as estatísticas baseadas na verossimilhança  $-2 \times \text{Log-like}$ , AIC, AICc e BIC bem como as medidas de bom ajuste KS, AD e CvM, temos que a distribuição HNT mostrou um melhor ajuste se comparada aos modelos utilizados. Como a distribuição apresentou uma expressão explícita e simples para a média, foi possível a utilização da mesma para uma aplicação usando regressão. Ao observarmos os critérios  $-2 \times \text{Log-like}$ , AIC, BIC e SSR notamos que a distribuição HNT apresentou um ajuste de ótima qualidade, reforçando sua primazia se comparada aos modelos aqui utilizados.

---

## REFERÊNCIAS

- ALJARRAH, M. A.; LEE, C.; FAMOYE, F. On generating T-X family of distributions using quantile functions. *Journal of Statistical Distributions and Applications*, v. 1, n. 1, p. 2, 2014.
- ALMALKI, S. J.; NADARAJAH, S. A new discrete modified Weibull distribution. *IEEE Transactions on Reliability*, v. 63, n. 1, p. 68–80, 2014.
- ALZAATREH, A.; KNIGHT, K. On the Gamma-Half Normal distribution and its applications. v. 12, n. 1, p. 103–119, 2013.
- ASGHARZADEH, A.; VALIOLLAHI, R.; RAQAB, M. Z. Stress-strength reliability of weibull distribution based on progressively censored samples. *SORT-Statistics and Operations Research Transactions*, v. 35, n. 2, p. 103–124, 2011.
- BAKOUCH, H. S.; JAZI, M. A.; NADARAJAH, S. A new discrete distribution. *Statistics*, v. 48, n. 1, p. 200–240, 2014.
- BONFERRONI, C. E. Elementi di statistica generale. *Libereria Seber Firenze*, 1930.
- BRACQUEMOND, C.; GAUDOIN, O. A survey on discrete lifetime distributions. *International Journal of Reliability, Quality and Safety Engineering*, v. 10, n. 1, p. 69–98, 2003.
- BRITO, C. C. R. de; RÉGO, L. C.; OLIVEIRA, W. R. de. Method for generating distributions and classes of probability distributions: The univariate case. *arXiv:1504.01062*, p. 1–50, 2015.
- CHAKRABORTY, S. Generating discrete analogues of continuous probability distributions - A survey of methods and constructions. *Journal of Statistical Distributions and Applications*, v. 2, n. 1, p. 6, 2015.
- CHAKRABORTY, S.; CHAKRAVARTY, D. Discrete gamma distributions: properties and parameter estimations. *Communications in Statistics - Theory and Methods*, v. 41, n. 18, p. 3301–3324, 2012.
- CHAKRABORTY, S.; CHAKRAVARTY, D. A discrete Gumbel distribution. *ArXiv e-prints*, 2014.

- CHOU, C. Y.; LIU, H. R. Properties of the Half-Normal distribution and its application to quality control. *Journal of Industrial Technology*, v. 14, n. 3, 1998.
- COLLETT, D. *Modelling Survival Data in Medical Research*. 2nd. ed. [S.l.]: Chapman and Hall, 2003.
- COORAY, K.; ANANDA, M. M. A. A generalization of the Half-Normal distribution with applications to lifetime data. *Communications in Statistics - Theory and Methods*, v. 37, n. 9, p. 1323–1337, 2008.
- CORDEIRO, G. M. et al. The odd log-logistic generalized Half-Normal lifetime distribution: Properties and applications. *Communications in Statistics - Theory and Methods*, v. 46, n. 9, p. 4195–4214, 2017.
- CORDEIRO, G. M.; PESCIM, R. R.; ORTEGA, E. M. M. The Kumaraswamy generalized Half-Normal distribution for skewed positive data. *Journal of Data Science*, v. 10, n. 2, p. 195–224, 2012.
- CORDEIRO, G. M. et al. The Beta generalized Half-Normal distribution: New properties. *Journal of Probability and Statistics*, v. 2013, p. 1–18, 2013.
- COX, D. R.; SNELL, E. J. A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 30, n. 2, p. 248–275, 1968.
- DANIEL, C. Use of Half-Normal plots in interpreting factorial two-level experiments. *Technometrics*, v. 1, n. 4, p. 311–341, 1959.
- DAVID, H. A.; NAGARAJA, H. N. *Order statistics*. 3nd. ed. [S.l.]: John Wiley, 2003.
- FELLER, W. *An introduction to probability theory and its applications*. 3nd. ed. [S.l.]: Wiley, 1968.
- GHOSH, T.; ROY, D.; CHANDRA, N. K. Reliability approximation through the discretization of random variables using reversed hazard rate function. *International Journal of Mathematical, Computational, Statistical, Natural and Physical Engineering*, v. 7, n. 4, p. 96–100, 2013.
- GÓMEZ-DÉNIZ, E.; CALDERÍN-OJEDA, E. The discrete Lindley distribution: properties and applications. *Journal of Statistical Computation and Simulation*, v. 81, n. 11, p. 1405–1416, 2011.
- GÓMEZ, Y. M.; BOLFARINE, H. Likelihood-based inference for the power Half-Normal distribution. *Journal of Statistical Theory and Applications*, v. 14, n. 4, 2015.
- GOOD, I. J. The population frequencies of species and the estimation of population parameters. *Biometrika*, v. 40, n. 3-4, p. 237–264, 1953.
- GUPTA, R. D.; KUNDU, D. Introduction of shape/skewness parameter(s) in a probability distribution. *Journal of Applied Statistical Science*, v. 7, n. 2, p. 153–171, 2009.
- HUSSAIN, T.; AHMAD, M. Discrete inverse Rayleigh distribution. *Pakistan Journal of Statistics*, v. 30, n. 2, p. 203–222, 2014.

- INMET, Instituto Nacional de Meteorologia. 2017. Disponível em: <<http://www.inmet.gov.br>>.
- JAZI, M. A.; LAI, C.-D.; ALAMATSAZ, M. H. A discrete inverse Weibull distribution and estimation of its parameters. *Statistical Methodology*, v. 7, n. 2, p. 121–132, 2010.
- JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. *Continuous univariate distributions*. Vol. 1. 2nd. ed. [S.l.]: John Wiley & Sons Inc., 1994.
- KEMP, A. W. Classes of discrete lifetime distributions. *Communications in Statistics - Theory and Methods*, v. 33, n. 12, p. 3069–3093, 2004.
- KHAN, M.; KHALIQUE, A.; ABOUAMMOH, A. On estimating parameters in a discrete Weibull distribution. *IEEE Transactions on Reliability*, v. 38, n. 3, p. 348–350, 1989.
- KLEIN, J. P.; MOESCHBERGER, M. L. *Survival Analysis: Techniques for Censored and Truncated Data*. 2nd. ed. [S.l.]: Springer-Verlag, 1997.
- KRISHNA, H.; PUNDIR, P. S. Discrete Maxwell distribution. *Interstat*, 2007.
- KRISHNA, H.; PUNDIR, P. S. Discrete Burr and discrete Pareto distributions. *Statistical Methodology*, v. 6, n. 2, p. 177–188, 2009.
- LAI, C. D. Issues concerning constructions of discrete lifetime models. *Quality Technology & Quantitative Management*, v. 10, n. 2, p. 251–262, 2013.
- LAI, D. C. Constructions and applications of lifetime distributions. *Applied Stochastic Models in Business and Industry*, v. 29, n. 2, 2011.
- LAWLESS, J. F. *Statistical Models and Methods for Lifetime Data*. 2nd. ed. [S.l.]: John Wiley & Sons, 2003.
- LEE, C.; FAMOYE, F.; ALZAATREH, A. Y. Methods for generating families of univariate continuous distributions in the recent decades. *Wiley Interdisciplinary Reviews: Computational Statistics*, v. 5, n. 3, p. 219–238, 2013.
- LEE, E. T.; WANG, J. W. *Statistical Methods for Survival Data Analysis*. 3rd. ed. [S.l.]: Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2003.
- LEONE, F. C.; NELSON, L. S.; NOTTINGHAM, R. B. The folded Normal distribution. *Technometrics*, v. 3, n. 4, p. 543–550, 1961.
- MAZUCHELI, J.; DEY, S. Bias-corrected maximum likelihood estimation of the parameters of the generalized Half-Normal distribution. *submitted to Journal of Statistical Computation and Simulation*, 2017.
- NADARAJAH, S.; KOTZ, S. The beta exponential distribution. *Reliability Engineering & System Safety*, v. 91, n. 6, p. 689–697, 2006.
- NADARAJAH, S.; KOTZ, S. R programs for truncated distributions. *Journal of Statistical Software*, v. 16, n. 1, p. 1–8, 2006.
- NADARAJAH, S.; ROCHA, R. Newdistns: An R package for new families of distributions. *Journal of Statistical Software*, v. 69, n. 1, p. 1–32, 2016.

- NAKAGAWA, T.; OSAKI, S. The discrete Weibull distribution. *IEEE Transactions on Reliability*, R-24, n. 5, p. 300–301, 1975.
- OLIVEIRA, S. B. *A distribuição Touchard e suas aplicações*. Dissertação (Mestrado) — UnB, Universidade de Brasília, Brasília, DF, 2017. Disponível em: <<http://repositorio.unb.br/handle/10482/22995>>.
- OLMOS, N. M. et al. An extension of the generalized Half-Normal distribution. *Statistical Papers*, v. 55, n. 4, p. 967–981, 2014.
- PEARSON, K. Contributions to the mathematical theory of evolution II. Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London. A*, v. 186, p. 343–414, 1895.
- PESCIM, R. R. et al. The Beta generalized Half-Normal distribution. *Computational Statistics & Data Analysis*, v. 54, n. 4, p. 945–957, 2010.
- PESCIM, R. R. et al. The Log-Beta generalized Half-Normal regression model. *Journal of Statistical Theory and Applications*, v. 12, n. 4, p. 330–347, 2013.
- PEWSEY, A. Large-sample inference for the general Half-Normal distribution. *Communications in Statistics - Theory and Methods*, v. 31, n. 7, p. 1045–1054, 2002.
- PEWSEY, A. Improved likelihood based inference for the general Half-Normal distribution. *Communications in Statistics - Theory and Methods*, v. 33, n. 2, p. 197–204, 2004.
- RAMIRES, T. G. et al. The Beta generalized Half-Normal geometric distribution. *Studia Scientiarum Mathematicarum Hungarica*, v. 50, n. 4, p. 523–554, 2013.
- ROY, D.; DASGUPTA, T. A discretizing approach for evaluating reliability of complex systems under stress-strength model. v. 50, n. 2, p. 145–150, 2001.
- SANCHEZ, J. J. D.; FREITAS, W. W. L.; CORDEIRO, G. M. The extended generalized Half-Normal distribution. *Brazilian Journal of Probability and Statistics*, v. 30, n. 3, 2016.
- SAS. *The NL MIXED Procedure, SAS/STAT® User's Guide, Version 9.22*. Cary, NC: SAS Institute Inc.: [s.n.], 2010. 4967–5062 p.
- SHANNON, C. E. Prediction and entropy of printed english. *Bell Labs Technical Journal*, v. 30, n. 1, p. 50–64, 1951.
- SHAW, W.; BUCKLEY, I. R. C. The alchemy of probability distributions: Beyond Gram-Charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map. *arXiv:0901.0434v1 [q-fin.ST]*, 2009.
- SINGH, R. Simulation of observations for the Half-Normal distribution. *Sankhyā: The Indian Journal of Statistics, Series B*, v. 56, n. 2, p. 137–139, 1994.
- SINGH, R. Editorial note on the paper “Simulation of observations for the Half-Normal distribution”. *Sankhyā: The Indian Journal of Statistics, Series B*, v. 57, n. 3, p. 461–461, 1995.
- STACY, E. W. A generalization of the Gamma distribution. *The Annals of Mathematical Statistics*, v. 33, n. 3, p. 1187–1192, 1962.

- STEIN, W. E.; DATTERO, R. A new discrete Weibull distribution. *IEEE Transactions on Reliability*, v. 33, n. 2, p. 196–197, 1984.
- TAHIR, M. H.; CORDEIRO, G. M. Compounding of distributions: a survey and new generalized classes. *Journal of Statistical Distributions and Applications*, v. 3, n. 1, p. 13, 2016.
- TAHIR, M. H.; NADARAJAH, S. Parameter induction in continuous univariate distributions: Well-established g families. *Journal of Probability and Statistics*, v. 87, n. 2, p. 539–568, 2015.
- TSAGRIS, M.; BENEKI, C.; HASSANI, H. On the folded Normal distribution. *Mathematics*, v. 2, n. 1, p. 12–28, 2014.
- VODĀ, V. G. Inferential procedures on a generalized Rayleigh variate I. *Aplikace Matematiky*, v. 21, n. 6, 1976.
- VODĀ, V. G. Inferential procedures on a generalized Rayleigh variate II. *Aplikace Matematiky*, v. 21, n. 6, 1976.