

Natália Moraes Cordeiro

CLASSE NORMAL UNIVARIADA GERADA A PARTIR DE DUAS
DISTRIBUIÇÕES-BASE

Recife
2021



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA
APLICADA

CLASSE NORMAL UNIVARIADA GERADA A PARTIR DE DUAS
DISTRIBUIÇÕES-BASE

Dissertação apresentada ao Programa de Pós-Graduação em Biometria e Estatística Aplicada como exigência parcial para obtenção do título de Mestre.

Área de Concentração: Biometria e Estatística Aplicada

Natália Moraes Cordeiro

Orientador: Prof. Dr. Frank Sinatra Gomes da Silva

Coorientador: Prof. Dr. Cícero Carlos Ramos de Brito

Recife
2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

- C794c Cordeiro, Natália Moraes
CLASSE NORMAL UNIVARIADA GERADA A PARTIR DE DUAS DISTRIBUIÇÕES-BASE / Natália Moraes
Cordeiro. - 2021.
58 f. : il.
- Orientador: Frank Sinatra Gomes da Silva.
Coorientador: Cicero Carlos Ramos de Brito.
Inclui referências.
- Dissertação (Mestrado) - Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Biometria e Estatística Aplicada, Recife, 2021.
1. Classe Normal Generalizada. 2. Distribuição Normal. 3. Gerador. I. Silva, Frank Sinatra Gomes da, orient. II. Brito, Cicero Carlos Ramos de, coorient. III. Título

CDD 519.5

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA
APLICADA

CLASSE NORMAL UNIVARIADA GERADA A PARTIR DE DUAS
DISTRIBUIÇÕES-BASE

Natália Moraes Cordeiro

Dissertação julgada adequada para obtenção
do título de Mestre em Biometria e Estatística
Aplicada, defendida e aprovada em 24/02/2021
pela comissão examinadora.

Banca examinadora:

Prof. Dr. Frank Sinatra Gomes da Silva (Orientador)
(PPGBEA-UFRPE)

Prof. Dr. Cícero Carlos Ramos de Brito (Coorientador)
(IFPE-DAFFG)

Profa. Dra. Maria do Carmo Soares de Lima
(PPGE-UFPE)

Recife
2021

Com todo o meu amor, aos meus pais, Josefa Josineide e Carlos Cordeiro, à minha irmã, Naiara Moraes e às minhas avós, Iraci e Maria, pela paciência, cuidado e carinho que a mim dedicam, por serem a minha base e por estarem sempre comigo.

Agradecimentos

Agradeço primeiramente a Deus, pois sem Ele e a Sua infinita bondade para comigo, nada disso seria possível. Aos meus pais, por me educarem e amarem sem nunca medir esforços. A eles, devo quem sou e tudo que até hoje realizei. À minha irmã, Naiara Moraes, pela paciência e apoio incondicional.

Às minhas avós, Vó Iraci e Vó Maria, por me colocarem constantemente em suas orações, e por todo carinho e cuidado que a mim elas dedicam, em especial Vó Iraci, que é, além de Vó, minha confidente e conselheira, e sempre me acalenta com seu carinho e palavras afetuosas. À minha tia Joseane que, ainda que de longe, se faz presente acompanhando cada passo meu.

Ao meu primo, Leandro Dias, por todo apoio, incentivo e afeto que tem por mim. Aos meus amigos, José Isaac que alegra os meus dias e supera a distância física, dividindo das lágrimas às melhores gargalhadas, amo você bem muito. A Epaminondas Soares por me ouvir, acalmar, aconselhar, por acreditar em mim e sempre me tirar um sorriso. A Jonnathan Felipe por me apoiar incondicionalmente, por me incentivar e por nunca me deixar. A Lucas Lima, pelas longas conversas, os conselhos, as risadas e até os choros.

À Daniela Neves que ilumina a minha vida desde que nos conhecemos e fui presenteada com a dádiva da sua amizade. A Taís Cristina, Ana Maria e Géssica Simões, que apesar de todos os quilômetros que nos separam, estão sempre pertinho no meu coração. A Marciele Lima que dividiu comigo o curso, as angústias, as alegrias, a casa e a vida. Obrigada por tudo e por tanto, você fez os meus momentos difíceis ficarem mais fáceis. Amo você. Ao meu amigo, Lucas Amaral pela sua generosidade e coração enorme. Amo a sua amizade.

Aos "véi" do Deinfo, Eduardo, Henrique e, de modo especial, João, que é o meu exemplo de pessoa incrível. Nunca serei grata o suficiente por todo afeto. Pela paciência em me ouvir, aconselhar, corrigir meus trabalhos um milhão de vezes, e, principalmente por me aproximar de Deus, você é parte fundamental de tudo isso. Obrigada por todo carinho e cuidado. À Camila pela paciência e doçura que tem comigo, pela boa vontade em ajudar, você é luz na minha vida, é mais uma prova do amor de Deus para comigo. À Gabi, pela sua amizade, as caronas, conversas, risadas, pela energia boa que você transmite.

A todos os Professores do Programa de Pós-graduação em Biometria e Estatística Aplicada, em especial, ao meu orientador prof. Dr. Frank Sinatra Gomes da Silva, ao prof. Dr. Jader da Silva Jale pelas contribuições, e ao meu coorientador prof. Dr. Cícero Carlos Ramos de Brito por ter acreditado em mim, bem como pela constante assistência, atenção, motivação e ensinamentos que me fizeram seguir adiante.

À Universidade Federal Rural de Pernambuco, em especial ao Departamento de Estatística e Informática e à Coordenação do Programa de Pós-Graduação em Biometria e Estatística Aplicada, por ter dado todas as condições necessárias para o desenvolvimento do curso. À CNPq pelo apoio financeiro.

À profa. Dra. Maria do Carmo Soares de Lima por fazer parte da banca, como avaliadora externa e por dedicar parte de seu tempo para a leitura e contribuições para este trabalho.

Ao prof. Dr Luciano Souza e ao colega Renan Fernandes pela disponibilidade e boa vontade que tiveram para comigo. Sou grata em tê-los conhecido.

E, por fim, a todos que até aqui conservaram esperança, acreditaram e torceram por mim.

*Não temas, porque eu sou contigo; não te assombres,
porque eu sou o teu Deus; eu te esforço, e te ajudo,
e te sustento com a destra da minha justiça.
- Isaías 41:10.*

Resumo

Tomando a distribuição normal, foi proposto um novo gerador de distribuições contínuas, a partir das funções monotônicas $\text{tg} \left(\frac{\pi}{2} G_1 \right)$ e $\log(1 - G_2)$, em que G_1 e G_2 são as distribuições-base para a classe. A partir da ideia proposta, realizou-se o estudo de identificabilidade, desenvolveu-se as expansões para a função de distribuição acumulada e para a função densidade de probabilidade. Além disso, foram mostradas propriedades de caracterização da classe e suas respectivas expansões, como função de risco, momentos e momentos centrais de ordem m , função geradora de momentos, função característica, coeficiente geral e os desvios, médio e quantílico. Assim como as derivadas da função de log-verossimilhança e o estudo do suporte. Fez-se também uma aplicação teórica da Classe, seguida de uma aplicação a dados simulados e a um conjunto de dados reais, comparando o modelo proposto com outros modelos já existentes verificando assim, também, o seu potencial em relação aos demais. Ressaltamos a vantagem da nova classe acoplar em suas distribuições-base, distribuições que podem ser contínuas ou não, além de apresentar bimodalidade e não possuir problemas de identificabilidade.

Palavras-chaves: Classe Normal Generalizada; Distribuição Normal; Gerador.

Abstract

Taking the normal distribution, a new generator of continuous distributions was proposed, starting from the monotonic functions $\text{tg} \left(\frac{\pi}{2} G_1 \right)$ and $\log(1 - G_2)$, where G_1 and G_2 are the base distributions for the class. From the proposed idea, the identifiability study was carried out, the expansions for the cumulative distribution function and for the probability density function were developed. In addition, properties of class characterization and their respective expansions were shown, such as risk function, central moments and moments of order m , moment generating function, characteristic function, general coefficient and deviations, mean and quantile. As well as those derived from the log-likelihood function and the study of support. A theoretical application of the Class was also made, followed by an application to simulated data and a set of real data, comparing the proposed model with other existing models, thus also verifying its potential in relation to the others. We emphasize the advantage of the new class coupling in its base distributions, distributions that can be continuous or not, in addition to presenting bimodality and not having identifiability problems.

Keywords: Generalized Normal Class; Generator; Normal Distribution.

Lista de Figuras

1	Função densidade de probabilidade da distribuição Normal	15
2	Função densidade de probabilidade da distribuição Weibull	16
3	Função densidade de probabilidade da distribuição gama	17
4	Função densidade de probabilidade da distribuição Gama Inversa	18
5	Função densidade de probabilidade da distribuição de Gompertz	19
6	Comportamento do suporte para a Classe Normal $(\text{tg}(\frac{\pi}{2}G_1), \log(1 - G_2))$	43
7	Distribuição Normal $(\text{tg}(\frac{\pi}{2}G_0), \log(1 - G_1))$	46
8	Histograma	51

Lista de Tabelas

1	Viés e EQM para as estimativas do modelo Normal $(\text{tg}(\frac{\pi}{2}G_0), \log(1 - G_aI))$	47
2	Dados definidos por Nichols e Padgett (2006) composto por 100 observações de quebra de fibras de carbono por stress (em Gba).	48
3	Estatísticas Descritivas	48
4	Estimativas e respectivos erros-padrão (entre parênteses)	49
5	Critérios de informação e Estatísticas de teste	50

Sumário

1	Introdução	11
2	Objetivos	13
2.1	Objetivo Geral	13
2.2	Objetivos Específicos:	13
3	Revisão de Literatura	14
3.1	Introdução	14
3.2	Modelos Probabilísticos	14
3.2.1	Modelo Normal	14
3.2.2	Modelo Weibull	15
3.2.3	Modelo Gama	17
3.2.4	Modelo Gama Inversa	18
3.2.5	Modelo Gompertz	19
3.3	Modelos de misturas	20
3.4	Método de máxima verossimilhança	21
3.5	Momentos e momentos centrais de ordem m	21
3.6	Momentos Ponderados de Probabilidade	22
3.7	Função geradora de momentos	22
3.8	Função característica	22
3.9	Critérios para selecionar modelos	23
3.9.1	Critério de informação de Akaike	23
3.9.2	Critério de informação de Akaike corrigido	24
3.9.3	Critério de informação bayesiano	24
3.9.4	Critério de informação de Hannan-Quinn	24
3.10	Testes de Anderson-Darling e Cramér Von Mises	24
3.10.1	Teste de Anderson-Darling	25
3.10.2	Teste de Cramér Von Mises	25
3.11	Método de Monte Carlo	25
3.12	Suporte para as classes de distribuições probabilísticas	26
3.13	Classes de distribuições generalizadas	27
3.13.1	Classes de distribuição exponencializada	27
3.13.2	Classe de distribuição Beta Generalizada	27
3.13.3	Classe de distribuição McDonald	28
3.13.4	Classes de Distribuições Kumaraswamy	28
3.13.5	Classes de distribuições da Gama-G generalizada	28
3.14	Algoritmo EM	29
3.14.1	Descrição do EM	29
4	Classe Normal $(\text{tg}(\frac{\pi}{2}G_1), \log(1 - G_2))$	30
4.1	Introdução	30
4.2	Obtenção da Classe Normal $(\text{tg}(\frac{\pi}{2}G_1), \log(1 - G_2))$	30
4.2.1	Modelo funcional da classe	30

4.2.2	Expansão da função de distribuição acumulada	31
4.2.3	Expansão da função densidade de probabilidade	35
4.3	Propriedades de Caracterização da Classe	35
4.3.1	Função de risco	35
4.3.2	Estudo de identificabilidade	36
4.3.3	Expansão para os momentos de ordem m	36
4.3.4	Expansão para os momentos centrais de ordem m	37
4.3.5	Expansão para a função geradora de momentos e para a função característica	38
4.3.6	Expansão para o coeficiente geral	39
4.3.7	Expansão para o Desvio Médio e Desvio Quantílico	40
4.3.8	Gerenciamento de algumas expressões	41
4.4	Derivadas da função log-verossimilhança em relação aos parâmetros	42
4.5	Estudo do suporte para a Classe Normal $(\text{tg}(\frac{\pi}{2}\mathbf{G}_1), \mathbf{log}(\mathbf{1} - \mathbf{G}_2))$	42
5	Aplicações da Classe Normal $(\text{tg}(\frac{\pi}{2}\mathbf{G}_0), \mathbf{log}(\mathbf{1} - \mathbf{G}_1))$	45
5.1	Introdução	45
5.2	Distribuição Normal $(\text{tg}(\frac{\pi}{2}\mathbf{G}_0), \mathbf{log}(\mathbf{1} - \mathbf{G}_1))$	45
5.3	Distribuição Normal $(\text{tg}(\frac{\pi}{2}\mathbf{G}_0), \mathbf{log}(\mathbf{1} - \mathbf{G}_1))$ aplicada a dados simulados	46
5.4	Distribuição Normal $(\text{tg}(\frac{\pi}{2}\mathbf{G}_1), \mathbf{log}(\mathbf{1} - \mathbf{G}_2))$ aplicada a dados de quebra de fibras de carbono por stress (em Gba)	47
6	Considerações Finais	52
7	Referências	53

1 Introdução

As distribuições de probabilidade são importantes ferramentas para descrever fenômenos, posto que fornecem modelos eficazes que ajudam a representar problemas reais. Elas podem ser utilizadas em um vasto campo de aplicações, com diferentes propósitos. Na área da saúde (VELOSO & MALIK, 2010; SANTOS & CUNHA, 2016), no campo ambiental (BERNARDI *et al.*, 2001; SABINO *et al.*, 2014), social (DRACHLER *et al.*, 2003; ALVES *et al.*, 2010), nas ciências agrárias (SANTOS, 2019; SILVEIRA, 2019) dentre outras.

Nos últimos anos, diversos métodos geradores de novas distribuições e, conseqüentemente, novas distribuições, têm sido propostos a fim de fornecer uma boa flexibilidade e adequação a conjuntos de dados reais. Remetemos o leitor a Lee *et al.* (2013) para uma revisão de alguns métodos e Tahir e Nadarajah (2015) para um levantamento detalhado sobre classes generalizadas de distribuições contínuas que são amplamente utilizadas, dentre as quais destacamos a beta normal de Eugene *et al.* (2002); a McDonald-G (Mc-G) de Alexander *et al.* (2012); a família exponencializada exponencial-Poisson de Ristic e Nadarajah (2014); a logistic-X e a nova Weibull-G de Tahir *et al.* (2015); a beta-G de Souza (2016) e a Normal-G de Silveira (2019). Além destes, existe o gerador T-X definido por Alzaatreh *et al.* (2013), que origina diversas classes generalizadas, através de distribuições contínuas.

Neste trabalho, apresentamos uma nova classe geradora de distribuições chamada Classe Normal $(\text{tg}(\frac{\pi}{2}G_1), \log(1 - G_2))$, dependente de duas funções disjuntas e monótonas de G_1 e G_2 , generalizada a partir da distribuição normal através do método gerador de métodos geradores de distribuições de probabilidades e classes de distribuição de probabilidades apresentado nos trabalhos de Brito *et al.* (2019) e Brito (2014). A classe aqui proposta apresenta vantagens sobre os geradores citados anteriormente, pois permite que as distribuições-base, G_1 e G_2 contemplem distribuições contínuas e não contínuas, diferentemente, do que ocorre com o gerador T-X. Além disso, como veremos mais adiante, não possui problemas de identificabilidade, o que acontece, por exemplo, com as misturas convexas (TEICHER, 1961), e apresenta bimodalidade. A bimodalidade caracteriza distribuições que apresentam duas curvas, com duas modas diferentes. Estas aparecem como picos distintos (máximo local) na função densidade de probabilidade (fdp), garantindo maior flexibilidade na modelagem de dados, em situações em que estão presentes, por exemplo, duas amostras diferentes.

Nas linhas que seguem, detalhamos a estrutura desta dissertação, que apresenta caráter teórico e contempla seis capítulos, incluindo este. No segundo capítulo apresentamos os objetivos, geral e específicos deste trabalho, o terceiro trata do referencial teórico. Nele mostraremos alguns modelos probabilísticos contínuos e discutiremos sobre os modelos de misturas. Faremos uma revisão sobre o método de estimação de máxima verossimilhança, bem como algumas estruturas matemáticas, tais como função geradora de momentos, momentos de ordem m , momentos centrais de ordem m e função característica. Além disso, apresentaremos os critérios estatísticos utilizados para comparar o ajuste dos modelos nas aplicações a dados reais, a simulação de Monte Carlo, que será usada para avaliar o comportamento assintótico das estimações e por fim, apresentaremos o estudo do suporte para as classes de distribuições probabilísticas e o algoritmo de expectativa e maximização (EM), usado nas aplicações para as misturas convexas.

No quarto capítulo iremos apresentar a Classe Normal $(\text{tg}(\frac{\pi}{2}G_1), \log(1 - G_2))$, logo após faremos o estudo de identificabilidade, e calcularemos as expansões para a função de distribuição acumulada (fda) e para a fdp. Além disso, mostraremos algumas propriedades de caracterização da classe e suas respectivas expansões, tais como função de risco, momentos e momentos centrais de ordem m , função geradora de momentos, função característica, coeficiente geral e os desvios, médio e quantílico. Além das derivadas da função de log-verossimilhança da classe e o estudo do suporte.

No capítulo cinco traremos uma aplicação teórica da Classe, seguida de uma aplicação em dados simulados, bem como a um conjunto de dados reais, relacionados à tensão de ruptura de fibras de carbono (em Gba). Faremos a comparação do modelo proposto com outros modelos já existentes e às misturas convexas normal-normal (NN) e weibull-weibull (WW), verificando assim, também, o seu potencial em relação aos demais. No sexto e último capítulo, exibiremos as considerações finais, baseadas nas análises dos resultados adquiridos, e, adicionalmente, algumas sugestões para possíveis pesquisas futuras.

2 Objetivos

2.1 Objetivo Geral

O presente trabalho tem por objetivo geral desenvolver uma nova classe de distribuições de probabilidade, dependente de duas funções monótonas de G_1 e G_2 , denominada Classe Normal $(\text{tg}(\frac{\pi}{2}G_1), \log(1 - G_2))$.

2.2 Objetivos Específicos:

- Estudar estruturas matemáticas da classe, tais como função de risco, momentos e momentos centrais de ordem m , função geradora de momentos, função característica e o coeficiente geral;
- Verificar a identificabilidade da classe;
- Realizar as estimativas de máxima verossimilhança;
- Fazer o estudo do Suporte da Classe;
- Aplicar o novo modelo a um conjunto de dados reais comparando os resultados com outros modelos já existentes verificando assim, também, o seu potencial em relação aos demais.

3 Revisão de Literatura

3.1 Introdução

Alguns tópicos se fizeram necessários para o desenvolvimento deste trabalho como um todo, neste sentido, no presente capítulo apresentaremos alguns deles, de modo a revisar conceitos e definições que irão facilitar a compreensão do leitor nos capítulos que sucedem. Assim, discutiremos sobre modelos probabilísticos, modelos de misturas e o método de máxima verossimilhança.

Falaremos dos momentos e momentos centrais de ordem m , os momentos ponderados de probabilidade e a função característica. Além dos critérios para selecionar modelos e os testes de Anderson-Darling e Cramér Von Mises e o método de Monte Carlo. Apresentaremos o suporte para as classes de distribuições probabilísticas, algumas classes de distribuições generalizadas e por fim, o algoritmo EM.

3.2 Modelos Probabilísticos

Um modelo probabilístico é uma expressão matemática que descreve fenômenos tanto naturais quanto sociais, permitindo, assim, realizar previsões e fazer interpretações a respeito dos mesmos com um considerável nível de confiabilidade. Estes podem ser utilizados nas mais variadas áreas do conhecimento. A seguir, será apresentado brevemente, alguns dos principais modelos.

3.2.1 Modelo Normal

O modelo normal (N), também chamado de modelo gaussiano, foi primeiro introduzido por Abraham de Moivre (1667-1754), um matemático francês, num artigo reimpresso na segunda edição do seu livro “A doutrina do acaso” de 1738. Mais tarde seus resultados foram estendidos pelo francês Pierre Simon de Laplace (1748-1827) em seu livro “Teoria Analítica da Probabilidade” de 1812, num resultado que hoje é conhecido como Teorema de Moivre-Laplace. De forma independente os matemáticos alemães Robert Adrian (1775-1843) em 1808 e Carl Friedrich Gauss (1777-1855) em 1809 desenvolveram a equação da distribuição e mostraram que ela modelava bastante bem os erros de observações astronômicas.

Ele é um dos mais importantes entre os modelos probabilísticos contínuos (SOUZA e POPOVIC *et al.*, 2017), posto que vários dos fenômenos naturais apresentam sua distribuição de probabilidade tão proximoamente normal, que ela se ajusta como se fosse.

Seu grande destaque na área de estatística se deve ao Teorema do Limite Central, pois ele garante que o comportamento probabilístico de vários estimadores possa ser descrito com boa aproximação pela distribuição normal. Além disso, muitas variáveis encontradas na natureza tendem a apresentar comportamento normal (BITTENCOURT, 2006), a medida que o número de dados aumente, como consequência deste teorema.

A sua fdp é dada por (JOHNSON *et al.*, 1995):

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2},$$

em que μ se refere a média das observações na série de dados e σ^2 a variância. Seu gráfico é representado pela Figura 1 a seguir.

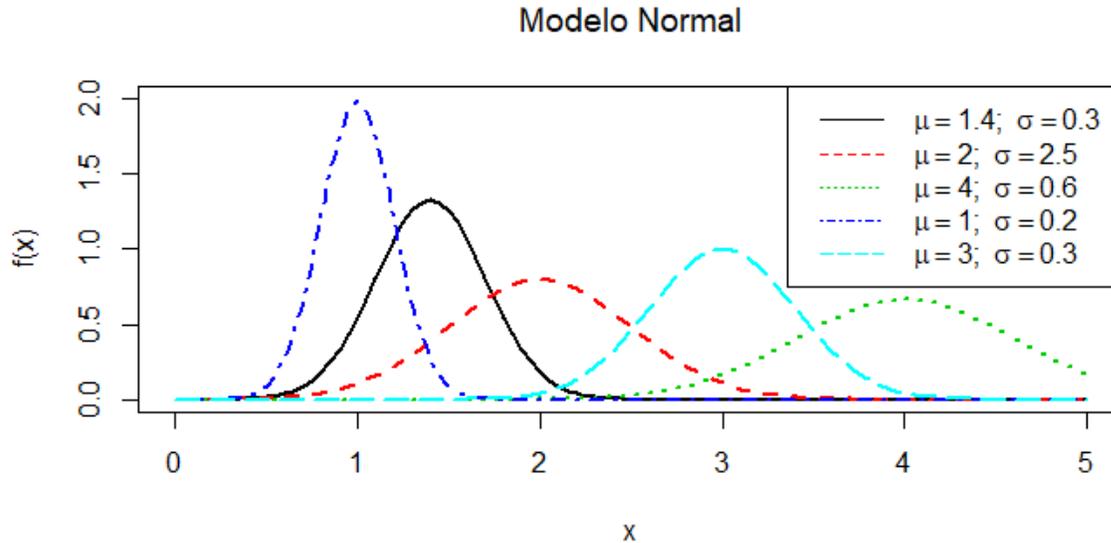


Figura 1: Função densidade de probabilidade da distribuição Normal

Note que a curva tem forma de "sino", e a medida que aumentamos o valor de μ , acontece o achatamento da curva. Além disso, quando $\mu = 0$ e $\sigma^2 = 1$ ela é chamada de distribuição normal padrão.

Sua fda é representada por

$$F(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx,$$

para todo $x \in \mathbb{R}$.

3.2.2 Modelo Weibull

Proposto por Waladdi Weibull (1954), o modelo probabilístico Weibull (W) é comumente usado para modelar o tempo de vida médio e a taxa de falha em função do tempo da população analisada. Uma das suas características mais significativas é que ele assume várias formas, todas com uma propriedade básica, sua função taxa de falha é monótona. Isto é, ela pode ser crescente, decrescente ou constante.

Sua fdp é dada por (JOHNSON *et al.*, 1995):

$$f(x) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^\alpha},$$

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

O gráfico é representado pela Figura 2 a seguir.

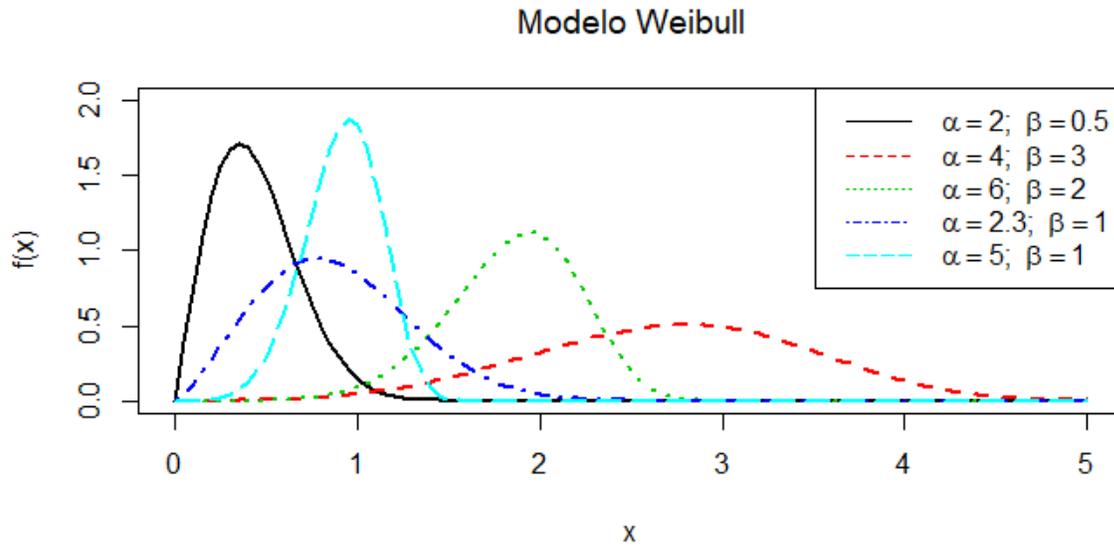


Figura 2: Função densidade de probabilidade da distribuição Weibull

Observe que a medida que variamos os valores para α e β percebemos as diversas formas que ela pode assumir.

Sua fda é representada por

$$F(x) = 1 - e^{-\left(\frac{x}{\beta}\right)^\alpha}.$$

A distribuição Weibull não tem uma forma fechada para a função geradora de momentos, entretanto, é possível encontrá-la fazendo

$$M_X(t) = \mathbb{E}(X^k) = \beta^k \Gamma\left(1 + \frac{k}{\alpha}\right),$$

em que $\Gamma(\cdot)$ é a função gama, e pode ser definida da seguinte forma $\Gamma(\alpha) = (\alpha - 1)!$.

Sua média e variância podem ser expressas como

$$\mathbb{E}(X) = \beta \Gamma\left(1 + \frac{1}{\alpha}\right)$$

e

$$Var(X) = \beta^2 \left(1 + \frac{2}{\alpha}\right) - \left[\beta \Gamma\left(1 + \frac{1}{\alpha}\right)\right]^2.$$

3.2.3 Modelo Gama

Geralmente utilizado em estudos de sobrevivência e confiabilidade, o modelo gama (Ga) é um dos modelos probabilísticos contínuos mais gerais quando comparado aos demais. Sua fdp é dada por (JOHNSON *et al.*, 1995):

$$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)},$$

em que $x \geq 0$, $\alpha > 0$ é o parâmetro de forma e $\beta > 0$ é o parâmetro de escala.

O gráfico é representado pela Figura 3 a seguir.

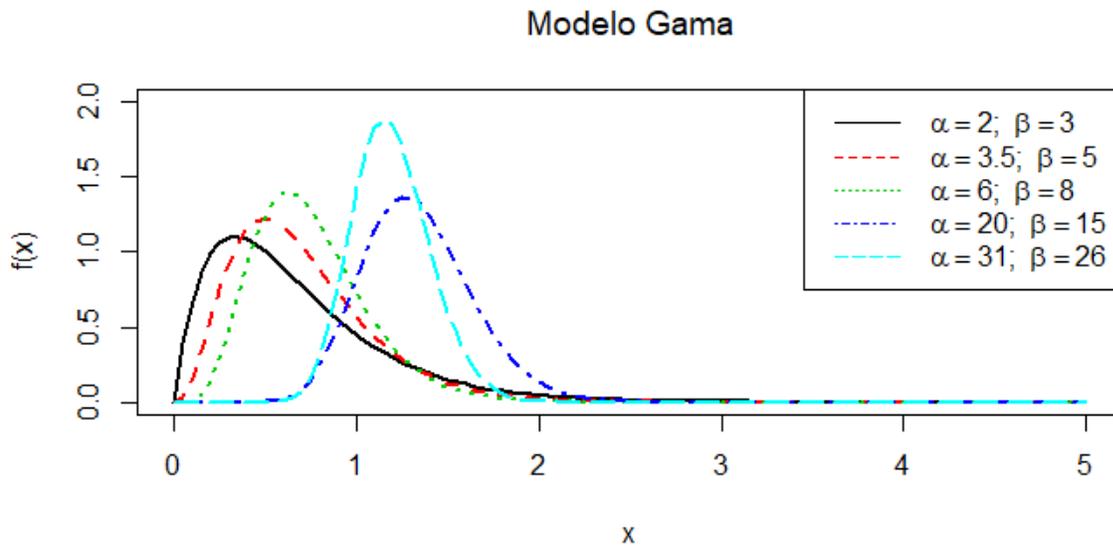


Figura 3: Função densidade de probabilidade da distribuição gama

O parâmetro α indica o formato geral da curva. Assim, analisando a Figura 3, é possível notar que quanto maior o valor de α , mais a distribuição tende a se aproximar da normal. Além disso, quando $\alpha = 1$, ela se transforma numa distribuição exponencial, e quando $\alpha = \frac{n}{2}$ e $\beta = 1$ ela define a distribuição qui-quadrado.

A expressão da fda é representada por

$$F(x) = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)},$$

em que $\gamma(\alpha, \beta x)$ é a função gama incompleta inferior.

A sua função geradora de momentos é dada por:

$$M_X(t) = \left(\frac{\beta}{\beta - t} \right)^\alpha.$$

A sua média por:

$$\mathbb{E}(X) = \frac{\alpha}{\beta},$$

e a variância é definida como:

$$\text{Var}(X) = \frac{\alpha}{\beta^2}.$$

3.2.4 Modelo Gama Inversa

A distribuição gama inversa (GaI) é usada, especialmente, na estatística Bayesiana, onde ela surge como a distribuição marginal posterior para a variância desconhecida de uma distribuição normal (WITKOVSKY, 2001). Sua fdp é dada por:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{x}\right)^{\alpha+1} e^{-\frac{\beta}{x}},$$

com $x > 0$. com parâmetro de forma $\alpha > 0$ e parâmetro de escala $\beta > 0$. Seu gráfico é representado na Figura 4 a seguir.

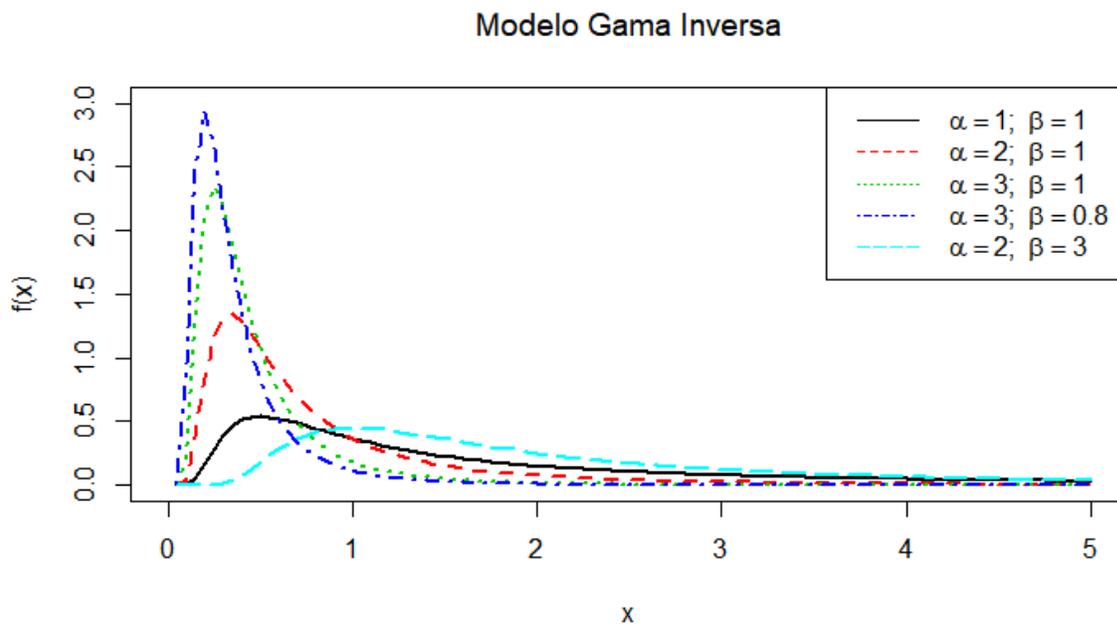


Figura 4: Função densidade de probabilidade da distribuição Gama Inversa

Perceba o achatamento da curva a medida que se aproxima de zero.

Sua fda é representada por

$$F(x) = \frac{\Gamma(\alpha, \frac{\beta}{x})}{\Gamma(\alpha)}.$$

A distribuição gama inversa não possui função geradora de momentos. A sua média é dada por:

$$\mathbb{E}(X) = \frac{\beta}{\alpha - 1},$$

para $\alpha > 1$, e a sua variância pela expressão a seguir:

$$Var(X) = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)}.$$

3.2.5 Modelo Gompertz

Apresentada por Benjamin Gompertz (1825), a distribuição de Gompertz (Go) é uma distribuição de probabilidade contínua. De acordo com Mazzini *et al.* (2005), a distribuição de Gompertz foi usada, inicialmente, para descrever a taxa de mortalidade de uma população. No entanto, posteriormente, sua aplicação em modelos sigmodais foi sugerida por Duarte, em 1975 (MENDES, 2011).

Sua fdp é dada por (JOHNSON *et al.*, 1995):

$$f(x) = \theta e^{\lambda x} e^{-\frac{\theta}{\lambda}(e^{\lambda x} - 1)},$$

em que $x \geq 0$, $\theta > 0$ é o parâmetro de escala e $\lambda > 1$ é o parâmetro de forma.

O seu gráfico é representado pela Figura 5 a seguir.

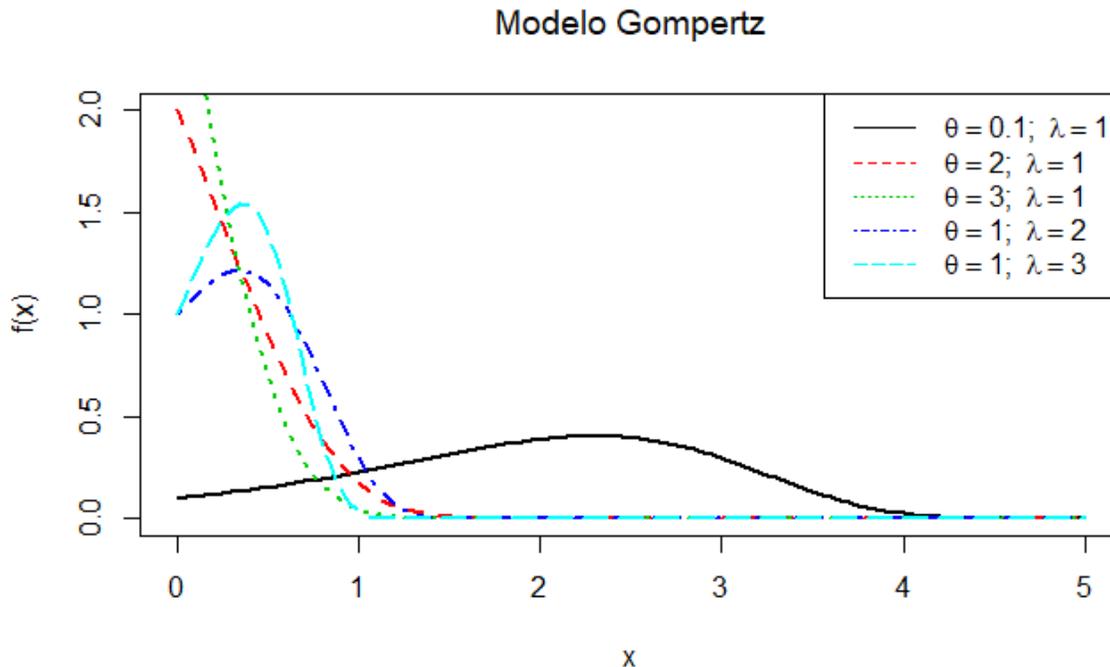


Figura 5: Função densidade de probabilidade da distribuição de Gompertz

Note que quanto menor o valor de θ , mais achatada será a curva, isto é, maior o coeficiente de curtose.

E a sua fda é definida pela seguinte expressão:

$$F(x) = 1 - e^{-\frac{\theta}{\lambda}(e^{\lambda x} - 1)}.$$

A sua função geradora de momentos é obtida fazendo:

$$M_X(t) = \lambda e^\sigma E_{t/\theta}(\lambda),$$

em que,

$$E_{t/\theta}(\lambda) = \int_1^\infty e^{-\lambda v} v^{-t/\theta} dv,$$

com $t > 0$.

E a sua média é dada por

$$\mathbb{E}(X) = \frac{1}{\theta} e^\lambda E_i(-\lambda),$$

em que

$$E_i(z) = \int_z^\infty \frac{e^v}{v} dv.$$

3.3 Modelos de misturas

Um modelo estatístico permite descrever conjuntos de dados em termos de suas características. No entanto, estes dados podem pertencer a uma população composta de subpopulações, fazendo da modelagem um processo difícil. Neste sentido, pode-se utilizar várias distribuições de probabilidade, em que cada uma descreve o comportamento de cada subpopulação. Estas distribuições, geralmente, fazem parte de uma única família, formando assim, o modelo de misturas de distribuições.

McLachlan e Peel (2004) definem mistura de distribuições para uma variável aleatória X como

$$f(x; p, \theta) = \sum_{k=1}^K p_k f_k(x_k; \theta_k),$$

em que $f_k(x_k; \theta_k)$ é a função densidade do k -ésimo componente da mistura, θ_k são os parâmetros da k -ésima componente, $p_k > 0$ a proporção de cada componente com restrição $\sum_{k=1}^K p_k = 1$, $0 \leq p_k \leq 1$, para $k = 1, \dots, K$ e K a quantidade de componentes do modelo.

Pearson (1894) foi o precursor da mistura NN, ou gaussianas, como também são conhecidas. Desde então, estas ocupam um lugar de destaque no estudo de misturas, devido a importância dessa distribuição dentro da estatística. Além disso, a mistura WW também é bastante utilizada, posto que esta generaliza as misturas de exponenciais, conforme referido em Everitt e Hand (1981).

3.4 Método de máxima verossimilhança

O Estimador de Máxima Verossimilhança (EMV) é um método para estimar os parâmetros de um modelo utilizando as estimativas que tornam máximo o valor da função de verossimilhança. Sendo apresentado nas linhas seguir.

Seja $X = (X_1, \dots, X_n)$ uma amostra aleatória de tamanho n da variável aleatória X com fdp dada por $f(x|\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$, em que Θ é o espaço paramétrico. A função de verossimilhança de $\boldsymbol{\theta}$ correspondente à amostra aleatória observada $x = (x_1, \dots, x_n)$ é dada por (CASELLA & BERGER, 2002):

$$L(\boldsymbol{\theta}; x) = \prod_{i=1}^n f(x_i|\boldsymbol{\theta}),$$

em que $\boldsymbol{\theta} \in \Theta$.

Perceba que a função de verossimilhança é a fdp conjunta de X avaliada na amostra observada x , vista como função do parâmetro $\boldsymbol{\theta}$.

Os estimadores de máxima verossimilhança (EMV) são os valores de $\boldsymbol{\theta}$ que maximizam $L(\boldsymbol{\theta}; x)$. Em geral, o logaritmo da função de verossimilhança $\ell(\boldsymbol{\theta}; x) = \log(L(\boldsymbol{\theta}; x))$, denominado função de log-verossimilhança, é usado com esse objetivo. Então, como a função logaritmo é monótona crescente, maximizar $\ell(\boldsymbol{\theta}; x)$ ou $\log(L(\boldsymbol{\theta}; x))$ em Θ são atividades equivalentes.

Portanto, a função de log-verossimilhança, é dada por:

$$\log(L(\boldsymbol{\theta}; x)) = \sum_{i=1}^n \log(f(x_i|\boldsymbol{\theta})).$$

3.5 Momentos e momentos centrais de ordem m

Para que uma variável aleatória seja completamente descrita, é preciso que se conheça a sua função de distribuição. Para isto, a depender da complexidade do problema ou da falta de informações a respeito do mesmo, se faz necessário uma descrição parcial através de indicadores capazes de caracterizar os principais aspectos da distribuição da variável. Esses indicadores são os chamados momentos.

O m -ésimo momento ou momento de ordem m de uma variável aleatória X cuja função densidade é dada por f , é definido por (CASELLA BERGER, 2002):

$$\mu_m = \mathbb{E}[X^m] = \int_{-\infty}^{\infty} x^m f(x) dx,$$

em que \mathbb{E} indica a operação de valor esperado.

Além disso, para cada número inteiro m , o momento central de ordem m de uma variável aleatória X , é definido como (CASELLA BERGER, 2002):

$$\mu_m = \mathbb{E}[X - \mathbb{E}(X)]^m.$$

3.6 Momentos Ponderados de Probabilidade

Apresentado por Greenwood *et al.* (1979), os Momentos Ponderados de Probabilidade (MPP) são uma generalização dos momentos habituais para as distribuições de probabilidade. Os MPP são uma alternativa para realizar estimações quando não é possível através dos métodos dos momentos e da máxima verossimilhança.

Os MPP são definidos por (GREENWOOD *et al.*, 1979):

$$\tau_{i,j,k} = \mathbb{E} \{ X^i [F(x)]^j [1 - F(x)]^k \} = \int_0^1 \mathbb{Q}[F(x)]^i F(x)^j [1 - F(x)]^k dF(x),$$

em que $i, j, k \in \mathbb{R}$ e $\mathbb{Q}(\cdot)$ representa a função quantílica de $F(\cdot)$.

Perceba que fazendo $j = k = 0$ e i um inteiro não negativo, $\tau_{i,0,0}$ é a generalização dos momentos usuais. Portanto, se $\tau_{i,0,0}$ for finito, $\tau_{i,j,k}$ é bem definido para todo $i, j, k \in \mathbb{R}$. A vantagem de calcular os momentos em termos dos MPP é que a maioria das distribuições, ou pelo menos as principais delas, já tem esse resultado definido na literatura (GAUSS; NADARAJAH, 2011).

3.7 Função geradora de momentos

A função geradora de momentos (fgm) possui esse nome pois, a partir dela, pode-se encontrar todos os momentos da variável aleatória X , quando estes existem. Para isto, a função precisa estar definida em uma vizinhança do ponto zero, pois os momentos serão obtidos através de sucessivas diferenciações aplicadas em zero. Neste sentido, seja X uma variável aleatória, sua fgm é definida como (CASELLA & BERGER, 2002):

$$M_X(t) = \mathbb{E}(e^{tX}),$$

desde que $\mathbb{E}(e^{tX})$ exista em algum intervalo do tipo $(-a, a)$ para algum número real $a > 0$.

Como $\mathbb{E}(e^{tX})$ está definido para $t \in (-a, a)$ para algum $h \in \mathbb{R}$, tem-se que

$$M_X^{(n)}(t) = \frac{d^n}{dt^n} \mathbb{E}(e^{tX}) = \mathbb{E} \left(\frac{d^n}{dt^n} e^{tX} \right) = \mathbb{E}(X^n e^{tX}).$$

Além disso, considerando X uma variável aleatória contínua, sua fgm, $M_X(t)$ é obtida fazendo-se (CASELLA & BERGER, 2002):

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

em que $f(x)$ é a fdp de X .

3.8 Função característica

Também chamada de transformada de Fourier, a função característica, tem aplicações nas mais diversas áreas do conhecimento científico, como por exemplo, no processamento de sinais e imagens, na física quântica, entre outros. Neste sentido é um operador relevante dentro da área de

probabilidade.

Assim, seja X uma variável aleatória, então a sua função característica é uma função $\mathbb{R} \rightarrow \mathbb{C}$ definida por (CASELLA BERGER, 2002):

$$\varphi(t) = \varphi_X(t) = \mathbb{E}(e^{itX}) = \int_{\mathbb{R}} e^{itx} dF_X(x),$$

em que

$$\mathbb{E}(e^{itX}) = \mathbb{E}(\cos(tX)) + i \mathbb{E}(\text{sen}(tX)),$$

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

3.9 Critérios para selecionar modelos

A escolha do modelo apropriado, do ponto de vista estatístico, é fundamental na análise de dados (BOZDANGAN, 1987). Busca-se o modelo mais parcimonioso, isto é, o modelo que melhor explique o comportamento do fenômeno estudado. Neste sentido, muitos procedimentos comuns na literatura podem ser utilizados para seleção de modelos. Estes devem levar em consideração a complexidade do modelo no critério de seleção.

O EMV se torna assintoticamente eficiente e a função de verossimilhança tende a ser um critério com maior sensibilidade a pequenos desvios dos parâmetros, caso o tamanho da amostra seja grande e as condições de regularidade forem mantidas. Assim, dentre os modelos a serem selecionados, aquele que obtiver o menor valor, em qualquer um dos critérios apresentados a seguir, será indicado como o que melhor se ajustará aos dados.

Na literatura há uma divisão de pensamentos quanto ao uso de medidas de bondade de ajustes. Alguns autores como Cox (1962), acreditam que estes critérios não são indicados para modelos não encaixados, isto é, modelos que não podem ser obtidos de outro por meio de restrições sobre os parâmetros que os indexam. No entanto, outros, como Pinheiro e Bates (2000), acreditam que também seja válida a comparação usando esses critérios. Em vista disso, resolvemos, neste trabalho, usar todas as medidas de bondade de ajuste, assim como as estatísticas de Anderson-Darling (A^*) e Cramér Von Mises (W^*).

3.9.1 Critério de informação de Akaike

Akaike (1974) propôs utilizar a informação de Kullback-Leibler para a seleção de modelos. Ele estabeleceu uma relação entre a máxima verossimilhança e a informação de Kullback-Leibler desenvolvendo então um critério para estimar a informação de Kullback-Leibler, posteriormente chamado Critério de Informação de Akaike (AIC).

Considerando uma amostra aleatória X_1, \dots, X_n de tamanho n e o vetor de parâmetros $\hat{\theta}$, a estatística AIC pode ser calculada por:

$$AIC = -2 \log(L(\hat{\theta})) + 2(p),$$

em que $\log(L(\hat{\theta}))$ é a função de log-verossimilhança e p é o número de parâmetros a serem estimados.

3.9.2 Critério de informação de Akaike corrigido

O critério de informação de Akaike corrigido (AICc) foi proposto por Liang e Zou (2008). De acordo com este, se X_1, \dots, X_n é uma amostra aleatória de tamanho n e o vetor de parâmetros $\hat{\theta}$, então a estatística AICc é dada por :

$$AICc = AIC + \frac{2(p+2)(p+3)}{n-p-3},$$

em que p é o número de parâmetros a serem estimados e n é o número de observações da amostra.

3.9.3 Critério de informação bayesiano

O critério de informação bayesiano (BIC) foi proposto por Schwarz (1978). É um procedimento de avaliação de modelos definido em termos da probabilidade a posteriori, e é conhecido também como critério de Schwarz, posto que Schwarz forneceu um argumento bayesiano para prová-lo.

Dada uma amostra aleatória X_1, \dots, X_n de tamanho n e o vetor de parâmetros $\hat{\theta}$, então a estatística BIC é obtida através da expressão

$$BIC = -2 \log(L(\hat{\theta})) + p \log n,$$

em que $\log(L(\hat{\theta}))$ é a função de log-verossimilhança, p é o número de parâmetros e n é o número de observações da amostra.

3.9.4 Critério de informação de Hannan-Quinn

Proposto por Hannan e Quinn (1979), o critério de informação de Hannan-Quinn (HQIC) é um critério de seleção de modelos alternativo ao AIC e BIC. É definido como:

$$HQIC = -2p \log(\log n) - 2 \log(L(\hat{\theta})),$$

em que $\log(L(\hat{\theta}))$, p e n são respectivamente a função de log-verossimilhança, o número de parâmetros e o número de observações da amostra.

Este critério tem pouco uso prático pois, a maioria dos conjuntos de dados tem poucas observações. Assim como os critérios anteriores, a seleção do modelo é através da comparação do HQIC de todos os modelos possíveis, selecionando o de menor valor (BURNHAM; ANDERSON, 2002).

3.10 Testes de Anderson-Darling e Cramér Von Mises

Seja $x = (x_1, \dots, x_n)$ uma amostra aleatória observada da função de distribuição F_X . A função de distribuição associada a X^* que atribui uma mesma probabilidade aos valores da variável,

$$P[X^* = x_i] = \frac{1}{n}$$

é dita função de distribuição empírica, e é denotada por F_n .

Dito isto, outros dois critérios para selecionar modelos são as modificações das estatísticas de Cramér Von Mises (W^*) e Anderson-Darling (A^*). Estes, são testes baseados na função de distribuição empírica dos dados, e apresentam vantagens sobre o teste de aderência qui-quadrado, incluindo maior poder e invariância em relação aos pontos médios dos intervalos escolhidos (CHEN; BALAKRISHNAN, 1995). Quanto menor for as estatísticas (W^*) e (A^*), melhor será o ajuste dos dados.

3.10.1 Teste de Anderson-Darling

Proposto por Anderson e Darling (1952), e posteriormente, modificado por Chen e Balakrishnan (1995), o teste de Anderson-Darling modificado é mais utilizado quando o tamanho da amostra não é superior a 25.

Considere $\delta_i = F(x_{(i)}; \theta)$ uma *fda*, com $x_{(i)}$ em ordem crescente. Faça $y_{(i)} = \Phi^{-1}(\delta_i)$ em que Φ representa a distribuição acumulada da normal padrão.

Seja $p_{(i)} = \Phi\left(\frac{[y_{(i)} - \bar{y}]}{S_y}\right)$, em que \bar{y} é a média e S_y , o desvio padrão dos $y_{(i)}$.

Assim, a expressão da estatística de teste é dada por:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n [(2i - 1) \log(p_{(i)}) + (2n + 12i) \log(1 - p_{(i)})]$$

em que as quantidades $p_{(i)}$ são percentis ordenados da distribuição normal padrão. A estatística de Anderson-Darling é calculada fazendo $A^* = A^2(1 + \frac{0,75}{n} + \frac{2,25}{n^2})$.

3.10.2 Teste de Cramér Von Mises

O teste de Cramér Von Mises foi proposto por Darling (1957). A modificação desta estatística de teste, proposta por Chen e Balakrishnan (1995), tem a seguinte expressão

$$W^2 = \frac{1}{12n} + \sum_{i=1}^n \left(p_{(i)} - \frac{2i - 1}{2n} \right)^2,$$

em que $p_{(i)}$ é definido de modo análogo a seção anterior. Assim, a estatística de Cramér Von Mises é dada por $W^* = W^2(1 + \frac{0,5}{n})$.

3.11 Método de Monte Carlo

Lustosa *et al.* (2010) definem que “o Método de Monte Carlo, consiste numa técnica que utiliza a geração de números aleatórios para atribuir valores às variáveis do sistema que se deseja investigar.” Isto é, repetindo sucessivas simulações um elevado número de vezes, para que se possa calcular as probabilidades que estão sendo estudadas.

Assim, suponha que se quer estimar θ , o valor esperado de alguma variável aleatória X é $\theta = \mathbb{E}(X)$, além disso, suponha que possam ser gerados valores de variáveis aleatórias independentes com a mesma distribuição de probabilidade de X . Cada vez que for gerado um novo valor, diz-se que uma simulação foi concluída. Suponha que vão ser realizadas n simulações, assim, serão gerados X_1, X_2, \dots, X_n .

Se $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ for sua média, então pela lei forte dos grandes números, será usado como um estimador para θ . Seu valor esperado e sua variância são dados a seguir.

Para o valor esperado tem-se

$$\mathbb{E}(X) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \theta.$$

Agora, fazendo $\sigma^2 = \text{Var}(X)$, tem-se

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

Além disso, decorre do Teorema Central do Limite que, para n grande, \bar{X} terá uma distribuição normal aproximada. Assim, se $\frac{\sigma}{\sqrt{n}}$ é pequeno, então \bar{X} tende a estar próximo de θ e, quando n for grande, \bar{X} será um bom estimador para θ .

3.12 Suporte para as classes de distribuições probabilísticas

Neste tópico comentaremos sobre o suporte da classe, estudo apresentado em Brito *et al.* (2019) e Brito (2014), para as classes e as distribuições originadas pelo método gerador por eles proposto.

Por definição de suporte de distribuição de probabilidade, tem-se que para qualquer distribuição F , o seu suporte, S_F é dado por:

$$S_F = \{x \in R : F(x) - F(x - \epsilon) > 0, \forall \epsilon > 0\}.$$

Sabe-se que

$$F_{G_1, \dots, G_m}(x) = \int_{\mu_1(G_1, \dots, G_m)}^{\ell_1(G_1, \dots, G_m)} \phi(t) dt, \quad -\infty < t < +\infty.$$

Considerando o Teorema 3.4 (BRITO, 2014, p. 60) que mostra condições em que $S_{H_{G_1, \dots, G_m}} = \bigcup_{j=1}^m S_{G_j}$, temos que

1. S_F for um conjunto convexo;

2. (a) $\mu_n(1, \dots, 1) = \sup \{x \in \mathbb{R} : F(x) < 1\}$, $\ell_1(1, \dots, 1) = \inf \{x \in \mathbb{R} : F(x) > 0\}$;
 $\mathcal{U}(\cdot)(x) > 0, \forall x \in \mathbb{R}$, e $\mu_j(\cdot)(x)$ ou $\ell_j(\cdot)(x)$, para algum $j = 1, 2, \dots, n$, forem, estritamente monotônicas ou $v_n(0, \dots, 0) = \sup \{x \in \mathbb{R} : \varphi < 1\}$;
- (b) $m_1(0, \dots, 0) = \inf \{x \in \mathbb{R} : \varphi > 0\}$, $\vartheta(\cdot)(x) > 0, \forall x \in \mathbb{R}$, e $v_j(\cdot)(x)$ ou $m_j(\cdot)(x)$, para algum $j = 1, 2, \dots, n$, forem estritamente monotônicas ;

então $S_{F_{G_1, \dots, G_m}} = U_{j=1}^m S_G$. Isto é, o suporte da nova distribuição é a união dos suportes das distribuições-base.

3.13 Classes de distribuições generalizadas

Os estudos em busca de novos geradores ou classes de distribuições generalizadas têm aumentado significativamente nos últimos anos. A forma generalizada de uma distribuição flexibiliza, de maneira razoável, a modelagem de dados que apresentam assimetria. Isto se deve a capacidade que essas distribuições possuem de englobar um grande número de submodelos.

Neste sentido, será apresentado a seguir, algumas das principais distribuições generalizadas já existentes.

3.13.1 Classes de distribuição exponencializada

A classe exponencializada passou a existir quando Gompertz (1825), elevou a função de distribuição acumulada da exponencial a um expoente $\alpha > 0$. Isto é,

$$F(x) = (1 - e^{-\lambda x})^\alpha,$$

para $x > 0$, $\lambda > 0$ e $\alpha > 0$ são parâmetros de forma.

Para definir a distribuição exponencializada generalizada, tem-se que:

Seja $G(x)$ uma distribuição-base arbitrária, diz-se que $X \sim Exp - G(\alpha)$, com $\alpha > 0$, se a fda e a fdp de X são, respectivamente

$$G_\alpha(x) = G^\alpha(x)$$

e

$$g_\alpha(x) = \alpha g(x) G^{\alpha-1}(x),$$

em que $g(x) = \frac{dG(x)}{dx}$.

3.13.2 Classe de distribuição Beta Generalizada

Eugene, Lee et al. (2002) foram os propulsores dos estudos sobre a generalização da classe de distribuição Beta, a Beta-G (beta tipo 1). Eles definiram a distribuição beta-normal inserindo a acumulada da normal no limite superior da integral da distribuição clássica da beta. Obtendo,

desta feita, uma distribuição mais vantajosa que a normal.

Assim, dado uma variável aleatória contínua X com fda $G(x)$, define-se a classe Beta-G como:

$$F(x) = \frac{1}{B(a,b)} \int_0^{G(x)} t^{\alpha-1} (1-t)^{b-1} dt,$$

em que $a > 0$ e $b > 0$ são parâmetros de forma e $B(a,b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$ é a função beta.

3.13.3 Classe de distribuição McDonald

McDonald (1984) apresentou duas distribuições da Beta-G, a Beta Generalizada do tipo 1 e a Beta Generalizada do tipo 2. Posteriormente, Alexander et al. (2012) trocou o limite superior da integral na equação da beta generalizada (tipo 1), por $G^c(x)$, em que $c > 0$ é o terceiro parâmetro adicional de forma. A partir daí, definiram uma nova classe de distribuição, denominada beta-gerada generalizada ou McDonald-G (Mc-G) cuja fda é dada por

$$F(x) = I_{G^c(x)}(a,b) = \frac{1}{B(a,b)} \int_0^{G^c(x)} t^{\alpha-1} (1-t)^{b-1} dt,$$

e a fdp correspondente definida como

$$f(x) = F'(x) = \frac{c}{B(a,b)} g(x) G^{ac-1}(x) [1 - G^c(x)]^{b-1} \quad x > 0,$$

em que $a, b, c > 0$ são parâmetros que definem a forma da distribuição.

3.13.4 Classes de Distribuições Kumaraswamy

Proposta por Cordeiro e de Castro (2011), a classe de distribuição Kumaraswamy generalizada (Kw-G), tem por base a distribuição Kumaraswamy (Kw), apresentada por Kumaraswamy (1980). Sua fda é dada por:

$$F(x) = 1 - [1 - G(x)^\alpha]^\beta,$$

em que $\alpha > 0$ e $\beta > 0$, são os parâmetros de forma.

Sua fdp é dada por:

$$f(x) = \alpha\beta g(x) G(x)^{\alpha-1} [1 - G(x)^\alpha]^{\beta-1}.$$

Quando $\alpha = \beta = 1$ tem-se a densidade da distribuição-base.

3.13.5 Classes de distribuições da Gama-G generalizada

A família gama-generalizada de distribuições foi introduzida por Zografos e Balakrishnan (2009). Sua fda (para $x \in \mathbb{R}$) é definida por:

$$F(x) = \frac{1}{\Gamma(\delta)} \int_0^{-\log[S(x)]} t^{\delta-1} e^{-t} dt,$$

em que $S(x) = 1 - G(x)$ é a função de sobrevivência da distribuição-base e $\Gamma(\delta) = \int_0^\infty x^{\delta-1} e^{-x} dx$ é a função gama.

3.14 Algoritmo EM

Apresentado por Dempster *et al.* (1977) o algoritmo de expectativa e maximização, ou simplesmente, algoritmo EM é um método iterativo usado para estimar parâmetros em modelos estatísticos, quando este depende de variáveis não observadas (variáveis latentes). A iteração EM alterna entre executar uma etapa de expectativa (E), e uma de maximização (M). A etapa de expectativa cria uma função para a expectativa da verossimilhança logarítmica usando a estimativa atual para os parâmetros. A etapa de maximização (M), calcula parâmetros para maximizar a verossimilhança logarítmica encontrada na etapa E. Essas estimativas de parâmetro são usadas para determinar a distribuição das variáveis latentes na próxima etapa E, e o algoritmo se repete várias vezes (por isso é chamado iterativo).

3.14.1 Descrição do EM

Dados o modelo estatístico que gera um conjunto X dos dados observados, um conjunto de dados não observados Z , um vetor de parâmetros desconhecidos θ , e uma função de verossimilhança $L(\theta; x)$.

Os dados gerados de $Y = (X, Z)$ são chamados dados completos. Se Z é latente, ele pode ser visto como tendo sido removido do Y completo por meio da aplicação de uma função $X = M(Y)$. Sejam $f_X(x|\theta)$ e $f_Y(y|\theta)$ as densidades dos dados observados e os dados completos, respectivamente, a densidade condicional dos dados faltantes dado os dados observados é dada por:

$$f_{Z|X}(z|x, \theta) = \frac{f_Y(y|\theta)}{f_X(x|\theta)}.$$

O algoritmo EM visa, iterativamente, maximizar $L(\theta; x)$ com respeito a θ . Assim, seja $\theta^{(t)}$ a estimativa de máximo na iteração t , para $t = 0, 1, \dots$, e $Q(\theta|\theta^{(t)})$ a esperança da função de log-verossimilhança para dados completos, condicional aos dados observados $X = x$, definida por:

$$Q(\theta|\theta^{(t)}) = E \left[\log L(\theta; Y) | x, \theta^{(t)} \right] = \int \log f_Y(y|\theta) f_{Z|X}(z|x, \theta^{(t)}) dz.$$

O algoritmo é inicializado com $\theta^{(0)}$ e alterna entre passos de esperança e maximização. Portanto, pode ser descrito como:

1. Iniciar $\theta^{(0)}$ e fazer $t = 0$;
2. Passo E: Calcular $Q(\theta|\theta^{(t)})$;
3. Passo M: Obter $\theta^{(t+1)}$ ao maximizar $Q(\theta|\theta^{(t)})$ com respeito a θ e fazer $t = t + 1$;
4. Repetir passos E e M até obter convergência para a sequência de estimativas $\theta^{(t)}$, segundo algum critério de parada pré-definido.

4 Classe Normal $\left(\text{tg}\left(\frac{\pi}{2}G_1\right), \log(1 - G_2)\right)$

4.1 Introdução

O avanço presente nos métodos de análise de dados, aponta a necessidade de distribuições probabilísticas mais sensíveis a massa de dados, isto é, que descrevam o mais fiel possível o comportamento do fenômeno estudado. Por conseguinte, a proposta de novas distribuições de probabilidade tem sido um tema de pesquisa considerado por diversos autores motivados a propor distribuições que tenham mais flexibilidade e que, portanto, se ajustem a uma variedade maior de dados.

Assim, tendo em vista a grande quantidade de fenômenos que podem ser descritos por novas distribuições de probabilidade, o respectivo trabalho tem como objetivo propor uma generalização da classe normal, a Classe Normal $\left(\text{tg}\left(\frac{\pi}{2}G_1\right), \log(1 - G_2)\right)$, em que G_1 e G_2 são suas distribuições-base, as quais podem ser contínuas ou não.

4.2 Obtenção da Classe Normal $\left(\text{tg}\left(\frac{\pi}{2}G_1\right), \log(1 - G_2)\right)$

A generalização aqui proposta, é uma aplicação do método gerador de distribuições e classes de distribuições probabilísticas apresentado por Brito *et al.* (2019) e Brito (2014). Este método é um teorema com sete corolários, que amplia o processo de construções de distribuições de probabilidades, permitindo que as classes de distribuições sejam construídas partindo de distribuições conhecidas e funções monotônicas univariadas pré-definidas.

A partir do funcional construtor

$$F_{G_1, \dots, G_m}(x) = \int_{\ell_1(\cdot)(x)}^{\mu_1(\cdot)(x)} dF(t),$$

em que $\mu_1(\cdot)(x) = \text{tg}\left(\frac{\pi}{2}G_1(x)\right)$ e $\ell_1(\cdot)(x) = \log(1 - G_2(x))$ determinar-se-á a nova classe de distribuição Normal Generalizada.

4.2.1 Modelo funcional da classe

Considerando as funções monotônicas $\mu_1(G_1, G_2, \dots, G_m)(x) = \text{tg}\left(\frac{\pi}{2}G_1(x)\right)$, $\ell_1(G_1, G_2, \dots, G_m)(x) = \log(1 - G_2(x))$ e a fdp da normal padrão, e fazendo uso do método gerador de distribuições e classes de distribuições probabilísticas mencionado anteriormente, tem-se que para quaisquer distribuições-base, G_1, G_2 a fda da classe será dada por:

$$H_{G_1, G_2}(x) = \int_{\log(1 - G_2(x))}^{\text{tg}\left(\frac{\pi}{2}G_1(x)\right)} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt. \quad (1)$$

Uma vez que, $\Phi(x) = \int_{-\infty}^x \phi(t)dt$ e $\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$, pode-se escrever a Equação (1) como

$$H_{G_1, G_2}(x) = \Phi \left[\text{tg} \left(\frac{\pi}{2} G_1(x) \right) \right] - \Phi [\log(1 - G_2(x))]. \quad (2)$$

Fazendo referência à nomenclatura estabelecida em (BRITO, 2014, p. 64), esta nova classe será denominada como Classe Normal $(\text{tg}(\frac{\pi}{2}G_1), \log(1 - G_2))$. As funções $\text{tg}(\frac{\pi}{2}G_1)$ e $\log(1 - G_2)$ nesta denominação fazem alusão, respectivamente, aos limites superior e inferior da Equação (1).

Além disso, sejam $g_1(x)$ e $g_2(x)$ as fdp's de $G_1(x)$ e $G_2(x)$, nessa ordem, então a fdp de H_{G_1, G_2} é dada por:

$$\begin{aligned} h_{G_1, G_2}(x) &= H'_{G_1, G_2}(x) = \Phi' \left[\text{tg} \left(\frac{\pi}{2} G_1(x) \right) \right] - \Phi' [\log(1 - G_2(x))] \\ &= \frac{\pi}{2} g_1(x) \sec^2 \left(\frac{\pi}{2} G_1(x) \right) \phi \left[\text{tg} \left(\frac{\pi}{2} G_1(x) \right) \right] + \frac{g_2(x)}{1 - G_2(x)} \phi [\log(1 - G_2(x))]. \end{aligned} \quad (3)$$

Utiliza-se a normal padrão no integrando da acumulada geradora definida na Equação (1), em virtude de ganhos de parcimônia. Além disso, quando $G_1 = G_2$ na acumulada (1) recaímos em um gerador completamente novo. Para registro, neste trabalho, considera-se o caso em que ambas distribuições-base são contínuas, embora o gerador proposto, como já mencionado, também possa contemplar distribuições-base não contínuas, diferentemente, por exemplo, do que ocorre com o gerador T-X proposto por Alzaatreh et al. (2013). Vale ressaltar que para cada par de distribuições-base temos um submodelo novo e, além disso, quaisquer que sejam G_1 e G_2 , $H_{G_1, G_2} \neq H_{G_2, G_1}$ o que não ocorre para as misturas convexas, fato este que conduz a problemas de identificabilidade.

A partir de agora, a menos que seja dito o contrário, $X \sim \text{Normal}(\text{tg}(\frac{\pi}{2}G_1), \log(1 - G_2))$.

4.2.2 Expansão da função de distribuição acumulada

Com o objetivo de simplificar a expressão definida na Equação (1) e os cálculos que seguirão, faremos uso de expansões em séries de potência, posto que obter as expressões de forma fechada, para este trabalho, é inviável.

A fda da normal está relacionada à função erro (erf) da seguinte forma:

$$\Phi(z) = \frac{1}{2} \left[1 + \text{erf} \left(\frac{z}{\sqrt{2}} \right) \right], \quad (4)$$

em que $\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-\frac{s^2}{2}} ds$. Desde que $\text{erf}(\frac{z}{\sqrt{2}})$ possa ser representado linearmente por:

$$\begin{aligned} \text{erf} \left(\frac{z}{\sqrt{2}} \right) &= \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n \left(\frac{z}{\sqrt{2}} \right)^{2n+1}}{n! (2n+1)} \\ &= \sqrt{\frac{2}{\pi}} \sum_{n=0}^{\infty} \left(-\frac{1}{2} \right)^n \frac{z^{2n+1}}{n! (2n+1)}. \end{aligned} \quad (5)$$

Substituindo a Equação (5) na Equação (4), obtem-se

$$\Phi(z) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \left(-\frac{1}{2} \right)^n \frac{z^{(2n+1)}}{n! (2n+1)}.$$

Deste modo, tem-se que

$$\Phi \left[\operatorname{tg} \left(\frac{\pi}{2} G_1(x) \right) \right] = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \left(-\frac{1}{2} \right)^n \frac{\left[\operatorname{tg} \left(\frac{\pi}{2} G_1(x) \right) \right]^{2n+1}}{n! (2n+1)}. \quad (6)$$

Note que,

$$\operatorname{tg}(y) = \sum_{k=0}^{\infty} \frac{(-1)^k 2^{2k+2} (2^{2k+2} - 1) B_{2k+2}}{(2k+2)!} y^{2k+1}, \quad (7)$$

em que $|y| < \frac{\pi}{2}$ e B_p são números de Bernoulli em que $B_0 = 1$ e $B_p = \sum_{l=0}^{p-1} \binom{p}{l} \frac{B_l}{k-l+1}$. Portanto, substituindo a Equação (7) em (6), tem-se

$$\begin{aligned} \Phi \left[\operatorname{tg} \left(\frac{\pi}{2} G_1(x) \right) \right] &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \left(-\frac{1}{2} \right)^n \\ &\quad \times \frac{\left[\sum_{k=0}^{\infty} \frac{(-1)^k 2^{2k+2} (2^{2k+2} - 1) B_{2k+2}}{(2k+2)!} \left[\frac{\pi}{2} G_1(x) \right]^{2k+1} \right]^{2n+1}}{n! (2n+1)} \\ &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \left(-\frac{1}{2} \right)^n \left(\frac{\pi}{2} \right)^{4kn+2k+2n+1} \\ &\quad \times \frac{[G_1(x)]^{2n+1} \left[\sum_{k=0}^{\infty} \frac{(-1)^k 2^{2k+2} (2^{2k+2} - 1) B_{2k+2}}{(2k+2)!} [G_1(x)]^{2k} \right]^{2n+1}}{n! (2n+1)} \\ &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \left(-\frac{1}{2} \right)^n \left(\frac{\pi}{2} \right)^{4kn+2k+2n+1} \\ &\quad \times \frac{[G_1(x)]^{2n+1} \left[\sum_{k=0}^{\infty} \frac{(-1)^k 2^{2k+2} (2^{2k+2} - 1) B_{2k+2}}{(2k+2)!} [G_1(x)]^{2k} \right]^{2n+1}}{n! (2n+1)}. \quad (8) \end{aligned}$$

Um resultado conhecido, para uma série de potências elevada a um número inteiro positivo N é a equação definida por Gradshteyn e Ryzhik (2007), como

$$\left[\sum_{j=0}^{\infty} a_j y^j \right]^N = \sum_{k=0}^{\infty} c_k y^k,$$

em que $c_0 = a_0^N$ e $c_j = \frac{1}{j a_0} \sum_{s=1}^j (sN - j + s) a_s c_{j-s}$, para $j \geq 1$ e $N \in \mathbb{N}$. Desta forma, a Equação

(8) pode ser escrita da seguinte forma:

$$\begin{aligned}
\Phi \left[\operatorname{tg} \left(\frac{\pi}{2} G_1(x) \right) \right] &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n \pi^{4kn+2k+2n+1}}{2^{4kn+2k+3n+1}} \frac{[G_1(x)]^{2n+1} \sum_{k=0}^{\infty} c_k \{ [G_1(x)]^2 \}^k}{n! (2n+1)} \\
&= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n \pi^{4kn+2k+2n+1} [G_1(x)]^{2n+1} \sum_{k=0}^{\infty} c_k \{ [G_1(x)]^2 \}^k}{2^{4kn+2k+3n+1} n! (2n+1)} \\
&= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \frac{(-1)^n \pi^{4nk+2k+2n+1} c_k}{n! (2n+1) 2^{4kn+2k+3n+1}} [G_1(x)]^{2n+2k+1}. \tag{9}
\end{aligned}$$

em que $c_0 = a_0^{2n+1}$, $c_k = \frac{1}{ka_0} \sum_{s=1}^k (s(2n+1) - k + s) a_s c_{k-s}$, $a_0 = 1$ e $a_s = \frac{(-1)^k 2^{2k+2} (2^{2k+2} - 1) B_{2k+2}}{(2k+2)!}$, para $k \geq 1$ e $n \in \mathbb{N}$.

De maneira inteiramente análoga, tem-se que

$$\Phi [\log(1 - G_2(x))] = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \left(-\frac{1}{2} \right)^n \frac{\log(1 - G_2(x))^{2n+1}}{n! (2n+1)}. \tag{10}$$

Além disso, a expressão $\log(1 - y) = -\sum_{j=0}^{\infty} \frac{y^{j+1}}{j+1}$, para $|y| < 1$. Logo, a Equação (10) pode ser escrita como:

$$\begin{aligned}
\Phi [\log(1 - G_2(x))] &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \left(-\frac{1}{2} \right)^n \frac{\left[-\sum_{j=0}^{\infty} \frac{[G_2(x)]^{j+1}}{j+1} \right]^{2n+1}}{n! (2n+1)} \\
&= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^{3n+1}}{n! (2n+1) 2^n} \left[\sum_{j=0}^{\infty} \frac{[G_2(x)]^{j+1}}{j+1} \right]^{2n+1}. \tag{11}
\end{aligned}$$

Usando mais uma vez o resultado de Gradshteyn e Ryzhik (2007) em que

$$\left[\sum_{k=0}^{\infty} a_k y^k \right]^N = \sum_{q=0}^{\infty} d_q y^q,$$

com $d_0 = a_0^N$ e $d_k = \frac{1}{ka_0} \sum_{s=1}^k (sN - k + s) a_s d_{k-s}$, para $k \geq 1$ e $N \in \mathbb{N}$, tem-se que a Equação (11) pode ser expressa como

$$\begin{aligned}
\Phi [\log (1 - G_2(x))] &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^{3n+1}}{n! (2n+1) 2^n} [G_2(x)]^{2n+1} \left[\sum_{j=0}^{\infty} \frac{[G_2(x)]^j}{j+1} \right] \\
&= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^{3n+1}}{n! (2n+1) 2^n} [G_2(x)]^{2n+1} \sum_{j=0}^{\infty} d_j [G_2(x)]^j \\
&= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^{3n+1}}{n! (2n+1) 2^n} \sum_{j=0}^{\infty} d_j [G_2(x)]^{2n+j+1} \\
&= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{j=0}^{\infty} \frac{(-1)^{3n+1}}{n! (2n+1) 2^n} d_j [G_2(x)]^{2n+j+1}, \tag{12}
\end{aligned}$$

em que $d_0 = 1$ e $d_j = \frac{1}{j} \sum_{s=1}^j (2s(n+1) - j) a_s d_{j-s}$ e $a_s = \frac{1}{2s+1}$ para $j \geq 1$ e $N \in \mathbb{N}$.

Assim, substituindo as Equações (9) e (12) na Equação (2), escreve-se

$$\begin{aligned}
H_{G_1, G_2}(x) &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \frac{(-1)^n \pi^{4nk+2k+2n+1} c_k}{n! (2n+1) 2^{4kn+2k+3n+1}} [G_1(x)]^{2n+2k+1} \\
&\quad - \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{j=0}^{\infty} \frac{(-1)^{3n+1} d_j}{n! (2n+1) 2^n} [G_2(x)]^{2n+j+1} \\
&= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \frac{(-1)^n \pi^{4nk+2k+2n+1} c_k}{n! (2n+1) 2^{4kn+2k+3n+1}} [G_1(x)]^{2n+2k+1} \\
&\quad - \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{j=0}^{\infty} \frac{(-1)^{3n+1} d_j}{n! (2n+1) 2^n} [G_2(x)]^{2n+j+1}. \tag{13}
\end{aligned}$$

Fazendo-se $\beta_{1,n,k} = \frac{(-1)^n \pi^{4nk+2k+2n+1} c_k}{n! (2n+1) 2^{4kn+2k+3n+1}}$ e $\beta_{2,n,j} = \frac{(-1)^{3n+1} d_j}{n! (2n+1) 2^n}$, a Equação (13) passa a ser expressa como:

$$H_{G_1, G_2}(x) = \frac{1}{\sqrt{2\pi}} \left\{ \sum_{n,k=0}^{\infty} \beta_{1,n,k} [G_1(x)]^{2n+2k+1} - \sum_{n,j=0}^{\infty} \beta_{2,n,j} [G_2(x)]^{2n+j+1} \right\}, \tag{14}$$

que é a expressão equivalente a expansão para a fda da classe, a qual é dada em termos das distribuições-base exponencializadas, descrita na seção 3.14.1. Isto é justificado pelo fato de que as G's exponencializadas foram as primeiras famílias propostas, assim, encontramos diversas delas na literatura, e, conseqüentemente, suas respectivas propriedades. Nesse sentido, se escrevemos a nossa proposta em função das G's exponencializadas, obteremos suas propriedades sem maiores dificuldades a partir delas.

4.2.3 Expansão da função densidade de probabilidade

A fdp é uma equação que representa a distribuição de probabilidade de uma variável. Neste caso, a fdp de X é representada pela Equação (3). Para que sua expansão seja obtida, tome $g_1(x)$ e $g_2(x)$ como as fdp's de $G_1(x)$ e $G_2(x)$ respectivamente, e considere

$$g_{1,2n+2k+1}(x) = [2n + 2k + 1]g_1(x)[G_1(x)]^{2n+2k} \text{ e } g_{2,2n+j+1}(x) = [2n + j + 1]g_2(x)[G_2(x)]^{2n+j}.$$

A fdp da $H_{G_1, G_2}(x)$, na forma expandida é dada por :

$$h_{G_1, G_2}(x) = H'_{G_1, G_2}(x). \quad (15)$$

Assim, substituindo a Equação (14) na Equação (15), obtem-se

$$\begin{aligned} h_{G_1, G_2}(x) &= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \beta_{1,n,k} (2n + 2k + 1) g_1(x) [G_1(x)]^{2n+2k} \\ &\quad - \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{j=0}^{\infty} \beta_{2,n,j} (2n + j + 1) g_2(x) [G_2(x)]^{2n+j} \\ &= \frac{1}{\sqrt{2\pi}} \left\{ \sum_{n,k=0}^{\infty} \beta_{1,n,k} g_{1,2n+2k+1}(x) - \sum_{n,j=0}^{\infty} \beta_{2,n,j} g_{2,2n+j+1}(x) \right\}, \end{aligned} \quad (16)$$

que é a forma expandida para a fdp da classe.

4.3 Propriedades de Caracterização da Classe

Nesta seção serão apresentadas algumas propriedades matemáticas de caracterização da classe proposta.

4.3.1 Função de risco

A função de risco ou taxa de falha descreve a probabilidade instantânea do evento "sobrevivência" ocorrer até um determinado tempo t . Isto é, ela descreve a taxa condicional de ocorrência do evento dado sobrevivência até o momento imediatamente antes do tempo t .

Sabe-se que

$$\mathcal{R}_{H_{G_1, G_2}}(x) = \frac{h_{G_1, G_2}(x)}{1 - H_{G_1, G_2}(x)}. \quad (17)$$

Substituindo os valores das Equações (14) e (16), a Equação (17) passa a ser dada por

$$\mathcal{R}_{H_{G_1, G_2}}(x) = \frac{\frac{1}{\sqrt{2\pi}} \left\{ \sum_{n,k=0}^{\infty} \beta_{1,n,k} g_{1,2n+2k+1}(x) - \sum_{n,j=0}^{\infty} \beta_{2,n,j} g_{2,2n+j+1}(x) \right\}}{1 - \frac{1}{\sqrt{2\pi}} \left\{ \sum_{n,k=0}^{\infty} \beta_{1,n,k} [G_1(x)]^{2n+2k+1} - \sum_{n,j=0}^{\infty} \beta_{2,n,j} [G_2(x)]^{2n+j+1} \right\}}.$$

4.3.2 Estudo de identificabilidade

O objetivo desta seção é fazer um estudo de identificabilidade da classe proposta, a fim de mostrar que a mesma não possui problemas desse tipo. Neste sentido, iremos usar o Teorema 1 do artigo (em processo de submissão) de Silveira (2021), enunciado a seguir.

Teorema da Identificabilidade para a Classe Normal $(tg(\frac{\pi G_1}{2}), \ln(1 - G_2))$

Sejam $G_1(x|\theta_1)$ e $G_2(x|\theta_2)$ as baselines da Classe Normal $(tg(\frac{\pi G_1}{2}), \ln(1 - G_2))$ definida pela expressão $H_{G_1, G_2}(x) = \Phi(tg(\frac{\pi}{2}G_1(x))) - \Phi(\ln(1 - G_2(x)))$. Além disso, $\theta_1 = (\theta_{11}, \dots, \theta_{1n}) \in \Theta_1$, $\theta_2 = (\theta_{21}, \dots, \theta_{2m}) \in \Theta_2$ e $\theta = (\theta_{11}, \dots, \theta_{1n}, \theta_{21}, \dots, \theta_{2m}) \in \Theta$, em que Θ_1, Θ_2 e Θ são espaços paramétricos associados a G_1, G_2 e H_{G_1, G_2} , respectivamente. Se G_1 e G_2 são identificáveis, então H_{G_1, G_2} é identificável.

Diante desse resultado, propomos a seguinte demonstração:

Tome $\Phi(tg(\frac{\pi}{2}G_1(x|\theta_1))) = \Phi(tg(\frac{\pi}{2}G_1(x|\theta_1^*)))$, em que $\{\theta_1, \theta_1^*\} \subset \Theta_1$ e $\theta_1 \neq \theta_1^*$. Como Φ é injetiva, $G_1(x|\theta_1) = G_1(x|\theta_1^*)$, mas isto é uma contradição, posto que nega a identificabilidade de G_1 . Deste modo, se $\theta_1 \neq \theta_1^*$, então $\Phi(tg(\frac{\pi}{2}G_1(x|\theta_1))) \neq \Phi(tg(\frac{\pi}{2}G_1(x|\theta_1^*)))$.

Analogamente, verificamos isso para $\theta_2, \theta_2^* \subset \Theta_2$. Se $\theta_2 \neq \theta_2^*$, então $\Phi(\ln[1 - G_2(x|\theta_2)]) \neq \Phi(\ln[1 - G_2(x|\theta_2^*)])$. Agora considere $\{\theta, \theta^*\} \subset \Theta$ e suponha que $H_{G_1, G_2}(x|\theta) = H_{G_1, G_2}(x|\theta^*)$. Se $\theta_1 = \theta_1^*$ e $\theta_2 \neq \theta_2^*$, então podemos inferir da equação $H_{G_1, G_2}(x) = \Phi(tg(\frac{\pi}{2}G_1(x))) - \Phi(\ln(1 - G_2(x)))$ que $G_2(x|\theta_2) = G_2(x|\theta_2^*)$, a saber, um absurdo. Da mesma forma se $\theta_1 \neq \theta_1^*$ e $\theta_2 = \theta_2^*$, chegaremos a uma contradição semelhante, em que $G_1(x|\theta_2) = G_1(x|\theta_2^*)$. Se $\theta_1 \neq \theta_1^*$ e $\theta_2 \neq \theta_2^*$ então a suposição falha desde $H_{G_1, G_2}(x|\theta) \neq H_{G_1, G_2}(x|\theta^*)$ para quase todos os valores de x pertencentes ao suporte. Logo, podemos concluir que H_{G_1, G_2} é identificável. ■

4.3.3 Expansão para os momentos de ordem m

Os momentos são muito importantes em estatística, posto que caracterizam as distribuições de probabilidade determinando as medidas de tendência central, dispersão, assimetria e curtose da distribuição. Neste sentido, a seguir será introduzida a expressão dos momentos probabilisticamente ponderados e em seguida o desenvolvimento dos cálculos da expansão para os momentos de ordem m para a Classe Normal $(tg(\frac{\pi}{2}G_1), \log(1 - G_2))$.

Sabe-se que

$$\mu_m = \mathbb{E}(X^m) = \int_{-\infty}^{+\infty} x^m dH_{G_1, G_2}(x). \quad (18)$$

Logo, substituindo a Equação (14) na Equação (18) tem-se:

$$\begin{aligned} \mu_m &= \int_{-\infty}^{+\infty} x^m \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \beta_{1,n,k}(2n + 2k + 1) g_1(x) [G_1(x)]^{2n+2k} dx \\ &\quad - \int_{-\infty}^{+\infty} x^m \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{j=0}^{\infty} \beta_{2,n,j}(2n + j + 1) g_2(x) [G_2(x)]^{2n+j} dx. \end{aligned} \quad (19)$$

Fazendo $[2n + 2k + 1]g_1(x)[G_1(x)]^{2n+2k}dx = \tau_{m,2n+2k,0}$ e $[2n + j + 1]g_2(x)[G_2(x)]^{2n+j}dx = \tau_{m,2n+j,0}$, a Equação (19) passa a ser expressa por:

$$\mu_m = \frac{1}{\sqrt{2\pi}} \sum_{n,k=0}^{\infty} (2n + 2k + 1)\beta_{1,n,k}\tau_{m,2n+2k,0} + \frac{1}{\sqrt{2\pi}} \sum_{n,j=0}^{\infty} \beta_{2,n,j}(2n + j + 1)\tau_{m,2n+j,0}, \quad (20)$$

em que

$$\tau_{m,\eta,r} = E\{X^m H_{G_1,G_2}(X)^\eta [1 - H_{G_1,G_2}(X)]^r\} = \int_{-\infty}^{+\infty} x^m H_{G_1,G_2}(x)^\eta [1 - H_{G_1,G_2}(x)]^r dH_{G_1,G_2}(x).$$

Em particular, tem-se a seguinte expansão para média da Classe Normal $(\text{tg}(\frac{\pi}{2}G_1), \log(1 - G_2))$:

Como $\mu = \mu_1$, a Equação (20) passa a ser escrita como:

$$\mu = \frac{1}{\sqrt{2\pi}} \sum_{n,k=0}^{\infty} (2n + 2k + 1)\beta_{1,n,k}\tau_{1,2n+2k,0} + \frac{1}{\sqrt{2\pi}} \sum_{n,j=0}^{\infty} (2n + j + 1)\beta_{2,n,j}\tau_{1,2n+j,0}.$$

4.3.4 Expansão para os momentos centrais de ordem m

Para os momentos centrais de ordem m , tem-se que

$$\mu'_m = E[(X - \mu)^m] = \int_{-\infty}^{+\infty} (X - \mu)^m dH(x),$$

e

$$\mu'_m = \sum_{r=0}^m \binom{m}{r} (-1)^r \mu^r \mu_{m-r}. \quad (21)$$

Mas,

$$\begin{aligned} \mu_{m-r} &= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} (2n + 2k + 1)\beta_{1,n,k}\tau_{m-r,2n+2k,0} \\ &\quad - \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{j=0}^{\infty} (2n + j + 1)\beta_{2,n,j}\tau_{m-r,2n+j,0}. \end{aligned} \quad (22)$$

Logo, substituindo a Equação (22) em (21), obtém-se a seguinte expressão:

$$\begin{aligned} \mu'_m &= \sum_{r=0}^m \binom{m}{r} (-1)^r \mu^r \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} (2n + 2k + 1)\beta_{1,n,k}\tau_{m-r,2n+2k,0} \\ &\quad - \sum_{r=0}^m \binom{m}{r} (-1)^r \mu^r \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{j=0}^{\infty} (2n + j + 1)\beta_{2,n,j}\tau_{m-r,2n+j,0} \\ &= \frac{1}{\sqrt{2\pi}} \sum_{r=0}^m \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \binom{m}{r} (-1)^r \mu^r (2n + 2k + 1)\beta_{1,n,k}\tau_{m-r,2n+2k,0} \\ &\quad - \frac{1}{\sqrt{2\pi}} \sum_{r=0}^m \sum_{n=0}^{\infty} \sum_{j=0}^{\infty} \binom{m}{r} (-1)^r \mu^r (2n + j + 1)\beta_{2,n,j}\tau_{m-r,2n+j,0}. \end{aligned} \quad (23)$$

Além disso, sejam

$$\delta_{1,m,r,n,k} = \binom{m}{r} (-1)^r \mu^r (2n + 2k + 1) \beta_{1,n,k} \text{ e } \delta_{2,m,r,n,j} = \binom{m}{r} (-1)^r \mu^r (2n + j + 1) \beta_{2,n,j},$$

a Equação (23) passará a ser escrita como

$$\mu'_m = \frac{1}{\sqrt{2\pi}} \sum_{r=0}^m \sum_{n,k=0}^{\infty} \delta_{1,m,r,n,k} \tau_{m-r,2n+2k,0} - \frac{1}{\sqrt{2\pi}} \sum_{r=0}^m \sum_{n,j=0}^{\infty} \delta_{2,m,r,n,j} \tau_{m-r,2n+j,0}. \quad (24)$$

Note que se $m = 1$ a Equação (24) descreve a expansão da variância para a Classe Normal $(\text{tg}(\frac{\pi}{2}G_1), \log(1 - G_2))$, que é dada por:

$$\sigma^2 = \mu'_2 = \frac{1}{\sqrt{2\pi}} \sum_{r=0}^2 \sum_{n,k=0}^{\infty} \delta_{1,2,r,n,k} \tau_{2-r,2n+2k,0} - \frac{1}{\sqrt{2\pi}} \sum_{r=0}^2 \sum_{n,j=0}^{\infty} \delta_{2,2,r,n,j} \tau_{2-r,2n+j,0}.$$

4.3.5 Expansão para a função geradora de momentos e para a função característica

A função geradora de momentos é bastante útil, mas nem sempre existe. Por este motivo, pode-se recorrer a função característica, que sempre existe. Neste sentido, nas linhas que seguem será mostrado o desenvolvimento dos cálculos da expansão para a função geradora de momentos e para a função característica da classe.

Para a função geradora de momentos, seja

$$M_X(t) = \mathbb{E}(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx} dH_{G_1, G_2}(x). \quad (25)$$

Substituindo a Equação (14) na Equação (25), tem-se que

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{+\infty} e^{tx} \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} (2n + 2k + 1) \beta_{1,n,k} g_1(x) [G_1(x)]^{2n+2k} dx \\ &\quad - \int_{-\infty}^{+\infty} e^{tx} \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{j=0}^{\infty} (2n + j + 1) \beta_{2,n,j} g_2(x) [G_2(x)]^{2n+j} dx \\ &= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} (2n + 2k + 1) \beta_{1,n,k} \int_{-\infty}^{+\infty} e^{tx} g_1(x) [G_1(x)]^{2n+2k} dx \\ &\quad - \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{j=0}^{\infty} (2n + j + 1) \beta_{2,n,j} \int_{-\infty}^{+\infty} e^{tx} g_2(x) [G_2(x)]^{2n+j} dx. \end{aligned} \quad (26)$$

Mas,

$$e^{tx} = \sum_{m=0}^{\infty} \frac{t^m x^m}{m!}. \quad (27)$$

Logo, inserindo o valor da Equação (27) na Equação (26), obtem-se

$$\begin{aligned}
M_X(t) &= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} (2n+2k+1)\beta_{1,n,k} \int_{-\infty}^{+\infty} \sum_{m=0}^{\infty} \frac{t^m x^m}{m!} g_1(x) [G_1(x)]^{2n+2k} dx \\
&- \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{j=0}^{\infty} (2n+j+1)\beta_{2,n,j} \int_{-\infty}^{+\infty} \sum_{m=0}^{\infty} \frac{t^m x^m}{m!} g_2(x) [G_2(x)]^{2n+j} dx \\
&= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} \frac{(2n+2k+1)\beta_{1,n,k} t^m}{m!} \int_{-\infty}^{+\infty} x^m g_1(x) [G_1(x)]^{2n+2k} dx \\
&- \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \sum_{j=0}^{\infty} \sum_{m=0}^{\infty} \frac{(2n+j+1)\beta_{2,n,j} t^m}{m!} \int_{-\infty}^{+\infty} x^m g_2(x) [G_2(x)]^{2n+j} dx. \tag{28}
\end{aligned}$$

Fazendo $\int_{-\infty}^{+\infty} x^m g_1(x) [G_1(x)]^{2n+2k} dx = \tau_{m,2n+2k,0}$ e $\int_{-\infty}^{+\infty} x^m g_2(x) [G_2(x)]^{2n+j} dx = \tau_{m,2n+j,0}$, a Equação (28) passa a ser escrita como

$$M_X(t) = \frac{1}{\sqrt{2\pi}} \left\{ \sum_{k,m,n=0}^{\infty} \frac{(2n+2k+1)\beta_{1,n,k} t^m}{m!} \tau_{m,2n+2k,0} - \sum_{j,m,n=0}^{\infty} \frac{(2n+j+1)\beta_{2,n,j} t^m}{m!} \tau_{m,2n+j,0} \right\},$$

que corresponde a forma expandida para a função geradora de momentos.

De maneira inteiramente análoga, seguindo os mesmos passos do cálculo anterior obtém-se a expansão para a função característica, que é dada por

$$\varphi_X(t) = \frac{1}{\sqrt{2\pi}} \left\{ \sum_{k,m,n=0}^{\infty} \frac{(2n+2k+1)\beta_{1,n,k} i^m t^m}{m!} \tau_{m,2n+2k,0} - \sum_{j,m,n=0}^{\infty} \frac{(2n+j+1)\beta_{2,n,j} i^m t^m}{m!} \tau_{m,2n+j,0} \right\}.$$

4.3.6 Expansão para o coeficiente geral

O coeficiente geral é uma nova generalização que estende as medidas de assimetria e curtose. Sua expressão é representada por

$$C_g(m) = \frac{E[(X - \mu)^m]}{\sqrt{E[(X - \mu)^2]^m}} = \frac{E[(X - \mu)^m]}{\sigma^m} = \frac{\mu'_m}{\sigma^m}. \tag{29}$$

Portanto, Substituindo a Equação (24) na Equação (29), tem-se que

$$C_g(m) = \frac{\frac{1}{\sqrt{2\pi}} \left(\sum_{r=0}^m \sum_{n,k=0}^{\infty} \delta_{1,m,r,n,k} \tau_{m-r,2n+2k,0} - \sum_{r=0}^m \sum_{n,j=0}^{\infty} \delta_{2,m,r,n,j} \tau_{m-r,2n+j,0} \right)}{\left[\frac{1}{\sqrt{2\pi}} \left(\sum_{r=0}^2 \sum_{n,k=0}^{\infty} \delta_{1,2,r,n,k} \tau_{2-r,2n+2k,0} - \sum_{r=0}^2 \sum_{n,j=0}^{\infty} \delta_{2,2,r,n,j} \tau_{2-r,2n+j,0} \right) \right]^{\frac{m}{2}}}.$$

Em particular, como $C_a = C_g(3)$ tem-se que a expansão para o coeficiente de assimetria da classe é dada por:

$$C_a = \frac{\frac{1}{\sqrt{2\pi}} \left(\sum_{r=0}^3 \sum_{n,k=0}^{\infty} \delta_{1,3,r,n,k} \tau_{3-r,2n+2k,0} - \sum_{r=0}^3 \sum_{n,j=0}^{\infty} \delta_{2,3,r,n,j} \tau_{3-r,2n+j,0} \right)}{\left[\frac{1}{\sqrt{2\pi}} \left(\sum_{r=0}^2 \sum_{n,k=0}^{\infty} \delta_{1,2,r,n,k} \tau_{2-r,2n+2k,0} - \sum_{r=0}^2 \sum_{n,j=0}^{\infty} \delta_{2,2,r,n,j} \tau_{2-r,2n+j,0} \right) \right]^{\frac{3}{2}}}.$$

De maneira análoga, $C_c = C_g(4)$, assim a expansão para o coeficiente de curtose é expressa como

$$C_c = \frac{\frac{1}{\sqrt{2\pi}} \left(\sum_{r=0}^4 \sum_{n,k=0}^{\infty} \delta_{1,4,r,n,k} \tau_{4-r,2n+2k,0} - \sum_{r=0}^4 \sum_{n,j=0}^{\infty} \delta_{2,4,r,n,j} \tau_{4-r,2n+j,0} \right)}{\left[\frac{1}{\sqrt{2\pi}} \left(\sum_{r=0}^2 \sum_{n,k=0}^{\infty} \delta_{1,2,r,n,k} \tau_{2-r,2n+2k,0} - \sum_{r=0}^2 \sum_{n,j=0}^{\infty} \delta_{2,2,r,n,j} \tau_{2-r,2n+j,0} \right) \right]^2}.$$

4.3.7 Expansão para o Desvio Médio e Desvio Quantílico

O desvio em relação à média (μ) e o desvio em relação ao quantil (q) são duas estatísticas que medem a dispersão da variável aleatória X . As expansões para os desvios médio e quantílico da classe são definidos, respectivamente, como

$$\begin{aligned} d_1(X) &= \int_{-\infty}^{+\infty} |x - \mu| dH_{G_1, G_2}(x) \\ &= \int_{-\infty}^{+\infty} |x - \mu| \frac{1}{\sqrt{2\pi}} \left\{ \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \beta_{1,n,k} g_{1,2n+2k+1}(x) - \sum_{n=0}^{\infty} \sum_{j=0}^{\infty} \beta_{2,n,j} g_{2,2n+j+1}(x) \right\}, \end{aligned}$$

e

$$\begin{aligned} d_2(X) &= \int_{-\infty}^{+\infty} |x - q| dH_{G_1, G_2}(x) \\ &= \frac{1}{\sqrt{2\pi}} \left\{ \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \beta_{1,n,k} g_{1,2n+2k+1}(x) - \sum_{n=0}^{\infty} \sum_{j=0}^{\infty} \beta_{2,n,j} g_{2,2n+j+1}(x) \right\}. \end{aligned}$$

As medidas $d_1(X)$ e $d_2(X)$, expressas em Cordeiro e Lemonte (2012), são encontradas fazendo-se

$$\begin{aligned} d_1(X) &= 2\mu H_{G_1, G_2}(\mu) - 2J(\mu) \\ &= 2\mu \frac{1}{\sqrt{2\pi}} \left\{ \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \beta_{1,n,k} [G_1(x)]^{2n+2k+1} - \sum_{n=0}^{\infty} \sum_{j=0}^{\infty} \beta_{2,n,j} [G_2(x)]^{2n+j+1} \right\} - 2J(\mu), \end{aligned}$$

e

$$d_2(X) = 2qJ(q) - q + \mu - 2J(q),$$

em que

$$\begin{aligned} J(q) &= \int_{-\infty}^{\alpha} x dH_{G_1, G_2}(x) \\ &= \int_{-\infty}^{\alpha} x \frac{1}{\sqrt{2\pi}} \left\{ \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \beta_{1,n,k} g_{1,2n+2k+1}(x) - \sum_{n=0}^{\infty} \sum_{j=0}^{\infty} \beta_{2,n,j} g_{2,2n+j+1}(x) \right\}. \end{aligned}$$

Quando $q = m$ for a mediana, tem-se

$$d_2(X) = \mu - 2J(m).$$

4.3.8 Gerenciamento de algumas expressões

A maioria das expressões vistas são facilmente gerenciáveis posto que encontram-se, definidos na literatura, os MPPs de algumas distribuições de probabilidade (Cordeiro e Nadarajah, 2011). Embora, muitas vezes, esses MPPs dependam de funções especiais, sua manipulação acaba sendo facilitada a partir de softwares de linguagem algébrica tais como MAPLE, MATHEMATICA e MATLAB. Por exemplo, para ilustração, os momentos ordinários de $X \sim \text{Normal}(\text{tg}(\frac{\pi}{2}\text{Be}), \log(1 - \text{Ga}))$, em que Be e Ga representam as distribuições Beta e Gama com vetores parâmetros $\theta_1 = (a, b)$ e $\theta_2 = (c, d)$, respectivamente, são dados pela seguinte expressão,

$$\begin{aligned} \mu_m &= \frac{1}{\sqrt{2\pi}} \sum_{n,k=0}^{\infty} (2n+2k+1)\beta_{1,n,k} \tau_{m,2n+2k,0} + \frac{1}{\sqrt{2\pi}} \sum_{n,j=0}^{\infty} \beta_{2,n,j} (2n+j+1) \tau_{m,2n+j,0} \\ &= \frac{1}{\sqrt{2\pi}} \sum_{n,k=0}^{\infty} (2n+2k+1)\beta_{1,n,k} a^{-j} B(a,b)^{-(j+1)} B(b, m+a(2n+2k+1)) \\ &\times F_{1:1}^{1:2} \left((m+a(2n+2k+1)) : (1-b,a); \dots; (1-b,a) : (b+m+a(2n+2k+1)) \right. \\ &: (a+1); \dots; (a+1); 1, \dots, 1 \Big) + \frac{1}{\sqrt{2\pi}} \sum_{n,j=0}^{\infty} \beta_{2,n,j} (2n+j+1) d^{-m} e^{-2n-j} \\ &\times \frac{\Gamma(m+c(2n+j+1))}{\Gamma(c)^{2n+j+1}} F_A^{(j)}(r+c(j+1); c, \dots, c; c+1, \dots, c+1, -1, \dots, -1). \end{aligned}$$

As funções Kampé de Fériet e Lauricella do tipo A (Exton, 1978; Aarts, 2000) que surgem em μ_m , são definidas, respectivamente, pelas seguintes expressões

$$\begin{aligned} F_{C:D}^{A:B}((a) : (b_1); \dots; (b_n); (c) : (d_1); \dots; (d_n); x_1, \dots, x_n) \\ = \sum_{m=1=0}^{\infty} \dots \sum_{m_n=0}^{\infty} \frac{((a))_{m_1+\dots+m_n} ((b_1))_{m_1} \dots ((b_n))_{m_n} x_1^{m_1} \dots x_n^{m_n}}{((c))_{m_1+\dots+m_n} ((d_1))_{m_1} \dots ((d_n))_{m_n} m_1! \dots m_n!}, \end{aligned}$$

e

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

em que $a = (a_1, \dots, a_A)$, $b_i = (b_{i,1}, \dots, b_{i,B})$, $c = (c_1, \dots, c_C)$, $d_i = (d_{i,1}, \dots, d_{i,D})$ para $i = 1, \dots, n$ e $(a)_i$ é o fatorial ascendente definido (por convenção $(a)_0 = 1$)

$$(a)_i = a(a+1) \dots (a+i-1),$$

com

$$((f))_k = ((f_1, \dots, f_p))_k = (f_1)_k \dots (f_p)_k$$

denotando o produto de fatoriais ascendentes. Portanto, os momentos da Normal $(\text{tg}(\frac{\pi}{2}\text{Be}), \log(1 - \text{Ga}))$, podem ser escritos como uma soma ponderada infinita das funções Kampé de Fériet e Lauricella do tipo A.

4.4 Derivadas da função log-verossimilhança em relação aos parâmetros

O método de máxima verossimilhança é um dos métodos mais comuns para estimar parâmetros, graças às suas propriedades assintóticas como eficiência e consistência. As estimativas são pontos que maximizam a função de probabilidade sobre o domínio do espaço paramétrico. Portanto, como a função logarítmica está aumentando, realizar a maximização da função de probabilidade de log, além de ser uma tarefa mais conveniente, também fornece a estimativa de máxima verossimilhança.

Seja $X = (X_1, \dots, X_n)$ uma amostra aleatória completa de tamanho n . Seja também $\boldsymbol{\theta}_1 = (\theta_{11}, \dots, \theta_{1s})$ o vetor paramétrico associado a $G_1(x) = G_1(x|\boldsymbol{\theta}_1)$, $\boldsymbol{\theta}_2 = (\theta_{21}, \dots, \theta_{2m})$ o vetor paramétrico associado a $G_2(x) = G_2(x|\boldsymbol{\theta}_2)$ e $f_{G_1, G_2}(x) = f_{G_1, G_2}(x|\boldsymbol{\theta})$, em que $\boldsymbol{\theta} = (\theta_{11}, \dots, \theta_{1s}, \theta_{21}, \dots, \theta_{2m})$, então a função log-verossimilhança de X pode ser escrita como:

$$\ell(\boldsymbol{\theta}; X) = \sum_{i=1}^n \left\{ \log \left\{ \frac{\pi}{2} g_1(x_i; \theta_1) \sec^2 \left[\frac{\pi}{2} G_1(x_i; \theta_1) \right] \phi \left[\frac{\pi}{2} G_1(x_i; \theta_1) \right] \right\} + \frac{g_2(x_i; \theta_2)}{1 - G_2(x_i; \theta_2)} \phi \{ \log[1 - G_2(x_i; \theta_2)] \} \right\}.$$

As estimativas de máxima verossimilhança para $\boldsymbol{\theta}$ também podem ser obtidas resolvendo-se o sistema $U(\boldsymbol{\theta}; X) = o_{s+m}$, em que o_{s+m} é o vetor de zeros, de tamanho $s + m$ e $U(\boldsymbol{\theta}; X) = \Delta_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}; X) = (u_j)_{1 \leq j \leq s+m}$ é o vetor de pontuação, com

$$u_j = \sum_{i=1}^n \left\{ \frac{\partial}{\partial \theta_{1j}} \log \left\{ \frac{\pi}{2} g_1(x_i; \theta_1) \sec^2 \left[\frac{\pi}{2} G_1(x_i; \theta_1) \right] \phi \left[\frac{\pi}{2} G_1(x_i; \theta_1) \right] \right\} + \frac{g_2(x_i; \theta_2)}{1 - G_2(x_i; \theta_2)} \phi \{ \log[1 - G_2(x_i; \theta_2)] \} \right\},$$

para $1 \leq j \leq s$, e

$$u_j = \sum_{i=1}^n \left\{ \frac{\partial}{\partial \theta_{2k}} \log \left\{ \frac{\pi}{2} g_1(x_i; \theta_1) \sec^2 \left[\frac{\pi}{2} G_1(x_i; \theta_1) \right] \phi \left[\frac{\pi}{2} G_1(x_i; \theta_1) \right] \right\} + \frac{g_2(x_i; \theta_2)}{1 - G_2(x_i; \theta_2)} \phi \{ \log[1 - G_2(x_i; \theta_2)] \} \right\},$$

para $1 \leq j \leq m$.

4.5 Estudo do suporte para a Classe Normal $(\text{tg}(\frac{\pi}{2} G_1), \log(1 - G_2))$

Até agora, tem-se, na literatura, classes com uma única distribuição-base, em que o suporte desta é o suporte daquela. Todavia, com o surgimento do método que gera classes com duas ou mais distribuições-base, é necessário estudar o suporte da classe em função do suporte das mesmas.

Lembre, inicialmente que,

$$H_{G_1, G_2}(x) = \int_{\text{tg}(\frac{\pi}{2} G_1)}^{\log[1 - G_2(x)]} \phi(t) dt, \quad -\infty < t < +\infty,$$

em que (vide notação estabelecida na Seção (3.12))

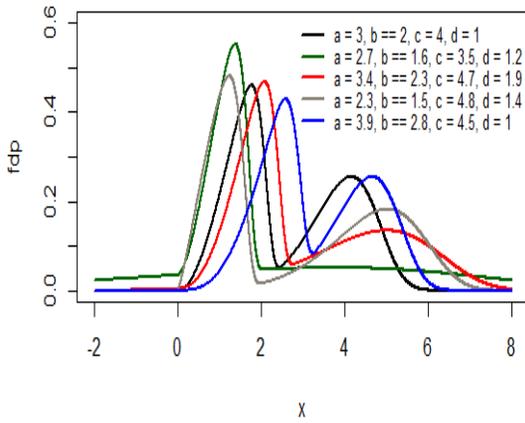
$$\mu_1[G_1(x), G_2(x)] = \text{tg}(\frac{\pi}{2} G_1) \text{ e } \ell_1[G_1(x), G_2(x)] = \log[1 - G_2(x)].$$

Logo, tem-se que,

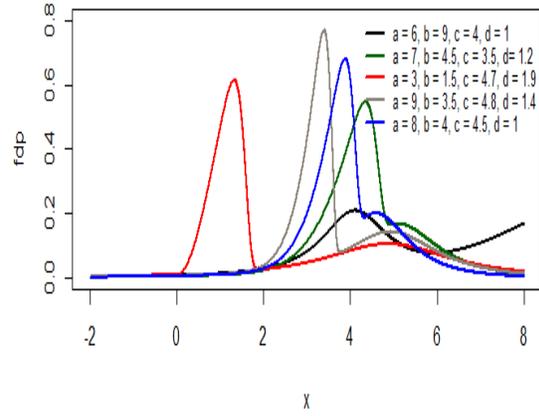
1. O Suporte da distribuição normal, $S_{\phi} = (-\infty, +\infty)$, é um conjunto convexo;

2. (a) $\mu_1[G_1(+\infty), G_2(+\infty)] = \mu_1(1,1) = \text{tg} \left[\frac{\pi}{2} G_1(+\infty) \right] = +\infty = \sup\{x \in \mathbb{R} : \Phi(x) < 1\}$;
- (b) $\ell_1[G_1(+\infty), G_2(+\infty)] = \ell_1(1,1) = \log[1 - G_2(+\infty)] = -\infty = \inf\{x \in \mathbb{R} : \Phi(x) > 0\}$;
- (c) $\mu_1[G_1(x), G_2(x)]$ e $\ell_1[G_1(x), G_2(x)] = \log[1 - G_2(x)]$ são funções monotônicas.

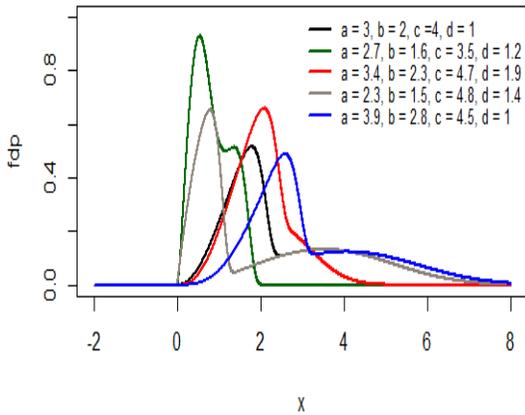
Estes itens satisfazem o conjunto de hipóteses estabelecidas no Teorema 3.4 (transcrito na Secção (3.12)) de Brito *et al.* (2019). Portanto, por este resultado, o suporte de $H_{G_1, G_2}(\cdot)$ resulta da união dos suportes de G_1 e G_2 , ou seja, $S_{H_{G_1, G_2}} = S_{G_1} \cup S_{G_2}$. A Figura 6 ao mesmo tempo que demonstra a flexibilidade da classe proposta para algumas escolhas de G_1 e G_2 , ilustram o importante resultado do suporte visto nesta secção.



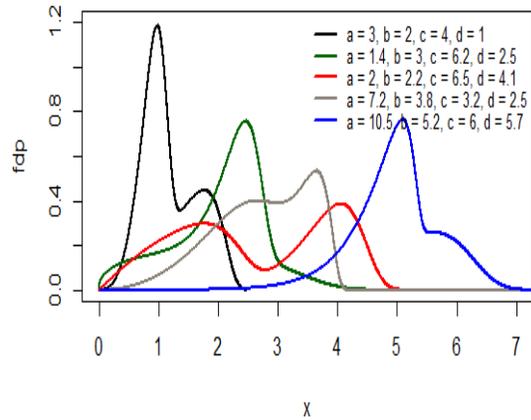
(a) Distribuição Normal $(\text{tg}(\frac{\pi}{2}W), \log(1 - N))$



(b) Distribuição Normal $(\text{tg}(\frac{\pi}{2}W), \log(1 - C))$



(c) Distribuição Normal $(\text{tg}(\frac{\pi}{2}W), \log(1 - G))$



(d) Distribuição Normal $(\text{tg}(\frac{\pi}{2}W), g(1 - W))$

Figura 6: Comportamento do suporte para a Classe Normal $(\text{tg}(\frac{\pi}{2}G_1), \log(1 - G_2))$

As distribuições-base escolhidas foram W, N, Cauchy (C) e Ga. Note que em alguns casos (Figs. 6a e 6b) $S_{G_1} \neq S_{G_2}$. Além disso, apontam a bimodalidade, o que é uma importante característica quando falamos em aplicação a dados reais.

5 Aplicações da Classe Normal $(\text{tg}(\frac{\pi}{2}\text{Go}), \log(1 - \text{GaI}))$

5.1 Introdução

Devido ao avanço tecnológico, predominante em todo mundo, a quantidade de informações disponíveis para análise, bem como a velocidade com que são geradas, vêm aumentando de forma significativa. Nesse contexto, generalizações das distribuições existentes têm sido propostas (RODRIGUES *et al.*, 2012), com a finalidade de proporcionar maior flexibilidade ao processo de modelagem, em relação às distribuições já existentes, fazendo com que o fenômeno em estudo seja melhor descrito. Além disso, com ferramentas computacionais, cada vez mais eficazes, como é o caso do software R, uma linguagem de programação voltada à manipulação, análise e visualização de dados, torna-se possível a análise de distribuições que possuam mais parâmetros para o melhor ajuste de massas de dados. Como é o caso da Classe Normal $(\text{tg}(\frac{\pi}{2}\text{G}_1), \log(1 - \text{G}_2))$ aqui apresentada.

Assim, nesta seção apresentaremos algumas aplicações desta classe.

5.2 Distribuição Normal $(\text{tg}(\frac{\pi}{2}\text{Go}), \log(1 - \text{GaI}))$

Realizamos estudos com G_1 e G_2 sendo a Go e a GaI , respectivamente, e, além destas, outras distribuições conhecidas na literatura. Entretanto, verificamos que a escolha $\text{G}_1 = \text{Go}$ e $\text{G}_2 = \text{GaI}$ proporcionou os melhores resultados que serão mostrados adiante. Ressaltamos que apesar destes terem sido os melhores, os outros também apresentaram resultados razoáveis quando comparados com as outras distribuições já conhecidas, porém, por questões de simplicidade, colocamos apenas esta aplicação.

Sejam G_1 e g_1 a fda e fdp (nessa ordem) da distribuição Go , e, analogamente G_2 e g_2 a fda e fdp da distribuição GaI . Então, a fda da Normal $(\text{tg}(\frac{\pi}{2}\text{Go}), \log(1 - \text{GaI}))$, obtida através da expressão definida na Seção 3.2.2 é dada por:

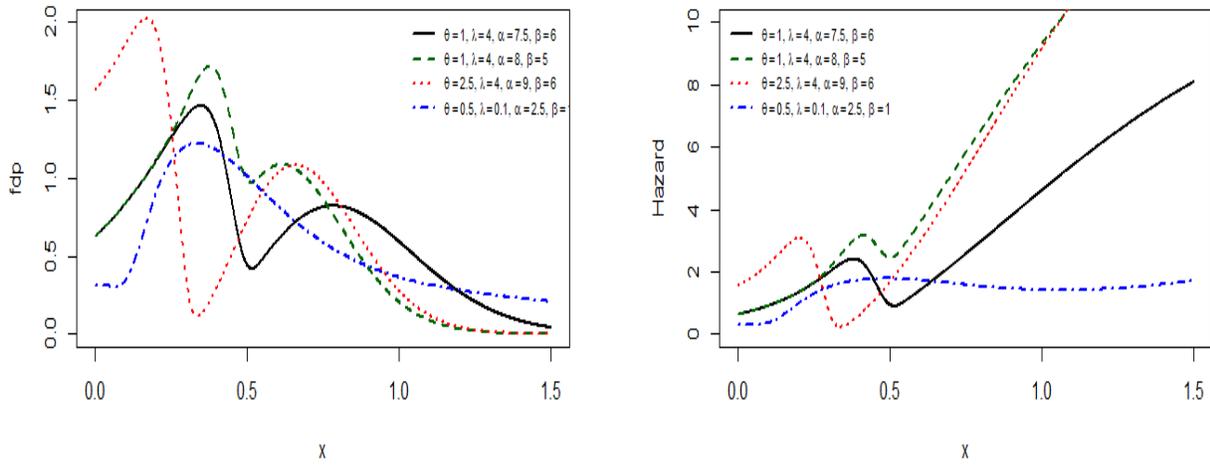
$$H_{\text{Go}, \text{GaI}}(x) = \int_{\log\left(1 - \frac{\Gamma(\alpha, \frac{\beta}{x})}{\Gamma(\alpha)}\right)}^{\text{tg}\left[\frac{\pi}{2}\left(1 - e^{-\frac{\theta}{\lambda}(e^{\lambda x} - 1)}\right)\right]} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

De maneira semelhante, a fdp da Normal $(\text{tg}(\frac{\pi}{2}\text{Go}), \log(1 - \text{GaI}))$, encontrada através da expressão obtida na Seção 3.2.3, é definida como

$$h_{\text{Go}, \text{GaI}}(x) = \frac{\pi}{2} \theta e^{\lambda x} e^{-\frac{\theta}{\lambda}(e^{\lambda x} - 1)} \sec^2\left(\frac{\pi}{2} \left(1 - e^{-\frac{\theta}{\lambda}(e^{\lambda x} - 1)}\right)\right) \phi\left[\text{tg}\left(\frac{\pi}{2} \left(1 - e^{-\frac{\theta}{\lambda}(e^{\lambda x} - 1)}\right)\right)\right] + \frac{\frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{x}\right)^{\alpha+1} e^{-\frac{\beta}{x}}}{1 - \frac{\Gamma(\alpha, \frac{\beta}{x})}{\Gamma(\alpha)}} \phi\left[\log\left(1 - \frac{\Gamma(\alpha, \frac{\beta}{x})}{\Gamma(\alpha)}\right)\right], \quad (30)$$

que é uma aplicação teórica da Classe.

A taxa de falha para o modelo Normal $(\text{tg}(\frac{\pi}{2}\text{Go}), \log(1 - \text{GaI}))$ pode ser obtida através da Equação (17), fazendo-se $\text{G}_1 = \text{Go}$ e $\text{G}_2 = \text{GaI}$. A representação gráfica para a fdp e para a taxa de falha, considerando alguns valores para os parâmetros θ , λ , α e β são apresentados na Figura (7) a seguir.



(a) Densidade

(b) Taxa de falha

Figura 7: Distribuição Normal $(\text{tg}(\frac{\pi}{2}\text{Go}), \log(1 - \text{GaI}))$

Note que a Figura 7a revela a unimodalidade e a bimodalidade da função de densidade da distribuição gerada pela classe, confirmando a flexibilidade que ela apresenta em suas formas. Além disso, a Figura 7b apresenta aumento-diminuição-aumento da taxa de falha.

5.3 Distribuição Normal $(\text{tg}(\frac{\pi}{2}\text{Go}), \log(1 - \text{GaI}))$ aplicada a dados simulados

Os experimentos computacionais executam um papel importante dentro da estatística. Graças a eles é possível examinar o desempenho de algo novo, verificar a validade de uma hipótese, ou demonstrar uma verdade já conhecida. Nesta seção, serão apresentadas as estimativas dos parâmetros para o modelo aqui proposto. Para realizar o estudo de simulação de Monte Carlo foi utilizado software R versão 3.4.4 (R Core Team, 2018). O número de repetições foi de 10.000. As amostras pseudo-aleatórias foram geradas usando os métodos Nelder Mead, SANN, BFGS e L-BFGS-B, No entanto, o L-BFGS-B, forneceu os melhores resultados.

Este procedimento requer apenas a fdp correspondente $y = f(x)$, um ponto de mínimo e um ponto de máximo para x e um ponto de máximo para y . Nesse caso, não é necessário implementar a função quantil. Consideramos quatro tamanhos de amostra, ou seja, $n = 50, 100, 200$ e 500 e cinco valores diferentes o vetor de parâmetros e calculamos o viés e o erro quadrático médio (EQM) da seguinte maneira:

$$\text{Viés}_i = \frac{1}{10000} \sum_{j=1}^{10000} (\hat{k}_{ij} - k_i)$$

e

$$EQM_i = \frac{1}{10000} \sum_{j=1}^{10000} (\hat{k}_{ij} - k_i)^2,$$

em que k_i é o i -ésimo elemento do vetor de parâmetros $\mathbf{k} = (k_1, \dots, k_r)$ e \hat{k}_{ij} é a estimativa para k_i na j -ésima replicação.

Na Tabela 1 são apresentados os resultados para os valores do viés e do EQM, obtidos para os parâmetros, nas quatro amostras consideradas.

Tabela 1: Viés e EQM para as estimativas do modelo Normal $(\text{tg}(\frac{\pi}{2}\text{Go}), \log(1 - \text{GaI}))$.

n	Valor real				viés				EQM			
	θ	λ	α	β	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$
50	1,3	0,02	1,5	3,2	0,425	0,136	1,136	0,176	0,310	0,093	1,916	1,605
	0,5	0,1	1,7	4	0,555	0,235	1,022	0,556	0,436	0,214	1,830	3,783
	0,45	0,04	2,8	6,2	0,321	0,109	0,418	1,752	0,387	0,130	0,734	1,673
	0,7	0,2	1,4	3	0,406	0,136	1,366	0,649	0,421	0,041	2,702	2,544
	0,8	0,02	1,2	1,7	0,180	0,037	0,965	1,560	0,099	0,011	1,237	2,258
100	1,3	0,02	1,5	3,2	0,287	0,064	1,109	0,111	0,307	0,018	1,778	1,507
	0,5	0,1	1,7	4	0,438	0,119	0,556	0,296	0,276	0,096	1,783	2,265
	0,45	0,04	2,8	6,2	0,115	0,026	0,256	1,172	0,094	0,008	0,456	1,512
	0,7	0,2	1,4	3	0,292	0,126	1,234	0,634	0,247	0,023	2,201	2,416
	0,8	0,02	1,2	1,7	0,132	0,024	0,890	1,510	0,070	0,002	0,883	2,170
200	1,3	0,02	1,5	3,2	0,191	0,049	1,058	0,099	0,226	0,006	1,598	1,461
	0,5	0,1	1,7	4	0,283	0,043	0,330	0,197	0,124	0,018	0,358	1,884
	0,45	0,04	2,8	6,2	0,047	0,002	0,178	1,047	0,042	0,004	-0,306	1,135
	0,7	0,2	1,4	3	0,184	0,118	1,176	0,565	0,152	0,021	1,909	1,301
	0,8	0,02	1,2	1,7	0,072	0,020	0,864	1,478	0,065	0,001	0,879	2,088
500	1,3	0,02	1,5	3,2	0,071	0,046	0,981	0,042	0,218	0,006	1,287	1,307
	0,5	0,1	1,7	4	0,167	0,018	0,003	0,135	0,044	0,0005	0,050	1,278
	0,45	0,04	2,8	6,2	0,037	0,001	0,107	0,804	0,028	0,0008	0,303	1,037
	0,7	0,2	1,4	3	0,164	0,086	1,032	0,235	0,134	0,021	1,358	1,214
	0,8	0,02	1,2	1,7	0,001	0,017	0,024	1,373	0,064	0,0009	0,002	1,942

Os valores mostrados na Tabela 1, indicam, como seria de esperar, quanto maior o tamanho da amostra, menor o EQM.

5.4 Distribuição Normal $(\text{tg}(\frac{\pi}{2}\text{G}_1), \log(1 - \text{G}_2))$ aplicada a dados de quebra de fibras de carbono por stress (em Gba)

É consenso entre vários pesquisadores aplicados que quanto mais testada for uma massa de dados, mais indicada ela será para a validação de um novo modelo. A recíproca desse fato também é verdadeira. Neste sentido, apresentaremos, a seguir, uma aplicação a dados reais para a distribuição Normal $(\text{tg}(\frac{\pi}{2}\text{Go},$

$\log(1 - \text{GaI})$).

Os dados são definidos a partir de Nichols e Padgett (2006) composto por 100 observações de quebra de fibras de carbono por stress (em Gba). Estes, são utilizados em aplicações de diversos trabalhos, dentre os quais podemos citar a distribuição Marshall-Olkin Fréchet transmutada (TMOFr), proposta por Afify *et al.* (2014b), a distribuição Gaussiana inversa exponenciada generalizada (EGIG), de Lemonte e Cordeiro (2010), a distribuição Weibull geométrica complementar transmutada (TCWG), de Afify *et al.* (2014a), a distribuição beta log-logística (BeLL), de Lemonte (2014), a distribuição Weibull beta modificada (MBeW) de Khan (2015) e a distribuição Weibull generalizada Weibull (WGW) de Yousof *et al.* (2018). Os dados estão na Tabela 2 a seguir.

Tabela 2: Dados definidos por Nichols e Padgett (2006) composto por 100 observações de quebra de fibras de carbono por stress (em Gba).

3,7	2,74	2,73	2,5	3,6	3,11	3,27	2,87	1,47	3,11	4,42	2,41	3,19	3,22
1,69	3,28	3,09	1,87	3,15	4,9	3,75	2,43	2,95	2,97	3,39	2,96	2,53	2,67
2,93	3,22	3,39	2,81	4,2	3,33	2,55	3,31	3,31	2,85	2,56	3,56	3,15	2,35
2,55	2,59	2,38	2,81	2,77	2,17	2,83	1,92	1,41	3,68	2,97	1,36	0,98	2,76
4,91	3,68	1,84	1,59	3,19	1,57	0,81	5,56	1,73	1,59	2	1,22	1,12	1,71
2,17	1,17	5,08	2,48	1,18	3,51	2,17	1,69	1,25	4,38	1,84	0,39	3,68	2,48
0,85	1,61	2,79	4,7	2,03	1,8	1,57	1,08	2,03	1,61	2,12	1,89	2,88	2,82
2,05	3,65												

Para estes dados, comparamos os ajustes das distribuições Normal ($\text{tg}(\frac{\pi}{2}\text{Go}), \log(1 - \text{GaI})$), Normal ($\text{tg}(\frac{\pi}{2}\text{N}), \log(1 - \text{GaI})$), Normal ($\text{tg}(\frac{\pi}{2}\text{GaI}), \log(1 - \text{W})$), Normal ($\text{tg}(\frac{\pi}{2}\text{GaI}), \log(1 - \text{C})$) e suas respectivas misturas GoGaI, NGaI, GaIWI e GaIC. Além dessas, também comparamos com as distribuições Go, GaI, W e N. Os cálculos foram realizados por meio do software R (versão 3.4.4). Na Tabela 3 é apresentado um resumo descritivo dos dados mostrados anteriormente.

Tabela 3: Estatísticas Descritivas

Média	Mediana	Moda	Variância	Assimetria	Curtose	Mínimo	Máximo
2,62	2,7	2,17	1,02	0,37	0,17	0,39	5,56

Observe que os dados possuem assimetria positiva e distribuição leptocúrtica.

Na Tabela 4 encontram-se as estimativas de máxima verossimilhança dos parâmetros, bem como seus respectivos erros-padrão.

Tabela 4: Estimativas e respectivos erros-padrão (entre parênteses)

Distribuição	Estimativas				
Normal ($\text{tg}(\frac{\pi}{2}\text{Go}), \log(1 - \text{GaI})$)	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	
	1,142 (0,201)	0,022 (0,013)	3,108 (0,647)	6,726 (1,837)	
GoGaI	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	\hat{w}
	1,784 (0,259)	0,007 (0,005)	4,616 (1,021)	9,313 (2,314)	0,490 (0,061)
Normal ($\text{tg}(\frac{\pi}{2}\text{N}), \log(1 - \text{GaI})$)	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	
	3,047 (0,121)	1,062 (0,144)	3,062 (0,692)	6,670 (2,017)	
NGaI	\hat{a}	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	\hat{w}
	99,999 (0,121)	99,999 (0,144)	5,802 (0,691)	13,005 (0,045)	0,010 (0,007)
Normal ($\text{tg}(\frac{\pi}{2}\text{GaI}), \log(1 - \text{WI})$)	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\delta}$	
	4,075 (0,612)	1,617 (1,232)	4,3423 (0,684)	0,328 (0,019)	
GaIWI	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\delta}$	\hat{w}
	9,994 (1,283)	25,585 (0,010)	1,780 (0,879)	0,944 (1,476)	0,827 (0,097)
Normal ($\text{tg}(\frac{\pi}{2}\text{GaI}), \log(1 - \text{C})$)	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\epsilon}$	$\hat{\kappa}$	
	1,879 (0,370)	4,591 (1,184)	2,895 (0,120)	0,489 (0,118)	
GaIC	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\epsilon}$	$\hat{\kappa}$	\hat{w}
	3,355 (1,912)	7,104 (5,421)	-3,503 (0,080)	0,003 (0,087)	0,866 (0,049)
WW	\hat{a}_1	\hat{b}_1	\hat{a}_2	\hat{b}_2	\hat{w}
	2,454 (0,200)	2,854 (0,109)	8,541 (2,380)	3,134 (0,126)	0,783 (0,093)
Go	$\hat{\theta}$	$\hat{\lambda}$			
	0,791 (0,077)	0,076 (0,017)			
GaI	$\hat{\alpha}$	$\hat{\beta}$			
	4,448 (0,606)	9,519 (1,375)			
W	\hat{a}	\hat{b}			
	2,792 (0,214)	2,943 (0,111)			
N	$\hat{\sigma}$	$\hat{\mu}$			
	2,621 (0,100)	1,008 (0,071)			

Note que obtivemos valores pequenos apontando razoável precisão nas estimativas. Para todos os modelos, a menos do WW, a maximização ocorreu a partir do algoritmo L-BFGS-B, posto que este proporcionou melhores resultados e convergência para os mesmos. Por outro lado, para as misturas WW e NN a maximização se deu através do algoritmo EM, não havendo convergência para esta última.

A Tabela 5 apresenta os critérios AIC, CAIc, BIC, HQIC e os valores das versões modificadas das estatísticas de Anderson-Darling - A^* e Cramér-von Mises - W^* (mais detalhes em Chen e Balakrishnan, 1995).

Tabela 5: Critérios de informação e Estatísticas de teste

Distribuição	AIC	AICc	BIC	HQCI	A^*	W^*
Normal ($\text{tg}(\frac{\pi}{2}\text{Go}), \log(1 - \text{GaI})$)	286,398	300,819	296,819	290,615	0,197	0,027
GoGaI	292,055	310,081	305,081	297,327	0,324	0,048
Normal ($\text{tg}(\frac{\pi}{2}\text{N}), \log(1 - \text{GaI})$)	287,435	301,856	297,856	291,652	0,280	0,046
NGaI	323,049	341,076	336,076	328,321	2,070	0,337
Normal ($\text{tg}(\frac{\pi}{2}\text{GaI}), \log(1 - \text{WI})$)	292,803	307,224	303,224	297,020	0,379	0,044
GaIWI	302,633	320,659	315,659	307,904	0,803	0,138
Normal ($\text{tg}(\frac{\pi}{2}\text{GaI}), \log(1 - \text{C})$)	292,428	306,849	302,849	296,646	0,400	0,053
GaIC	298,114	316,14	311,14	303,385	0,658	0,130
WW	289,327	307,354	302,354	294,599	0,272	0,034
Go	302,25	309,46	307,46	304,359	1,767	0,229
GaI	321,474	328,684	326,684	323,583	2,871	0,522
W	287,059	294,269	292,269	289,167	0,420	0,063
N	289,541	296,751	294,751	291,649	0,471	0,058
C	324,431	331,641	329,641	326,539	1,676	0,173

Os critérios de informação podem ser usados como medidas relativas à qualidade do ajuste, de modo que os menores valores irão caracterizar os melhores modelos ajustados. Apesar de não serem muito indicados para comparação de modelos não-encaixados os critérios de informação do modelo Normal ($\text{tg}(\frac{\pi}{2}\text{Go}), \log(1 - \text{GaI})$) ficaram entre os mais competitivos em relação a todos os modelos ajustados. Por outro lado, as estatísticas A^* e W^* são normalmente usados para investigar a qualidade do ajuste de modelos probabilísticos, sejam eles encaixados ou não-encaixados. De acordo com as mesmas, para os modelos com valores mais baixos de A^* e W^* melhor será o ajuste aos dados.

Por estas estatísticas, a Distribuição Normal ($\text{tg}(\frac{\pi}{2}\text{Go}), \log(1 - \text{GaI})$) venceu a maioria dos modelos ajustados, inclusive a mistura WW que é um modelo bem competitivo para dados bimodais. Vale salientar que os modelos Normal($\text{tg}(\frac{\pi}{2}\text{N}), \log(1 - \text{GaI})$), Normal ($\text{tg}(\frac{\pi}{2}\text{GaI}), \log(1 - \text{WI})$) e Normal ($\text{tg}(\frac{\pi}{2}\text{GaI}), \log(1 - \text{C})$) tiveram um desempenho superior (para todas as estatísticas), quando comparados às suas versões mistas NGaI, GaIWI e GaC, respectivamente.

O gráfico dos ajustes para o modelo Normal ($\text{tg}(\frac{\pi}{2}\text{Go}), \log(1 - \text{GaI})$) e os três melhores modelos, segundo as estatísticas A^* e W^* , são apresentados na Figura 8.

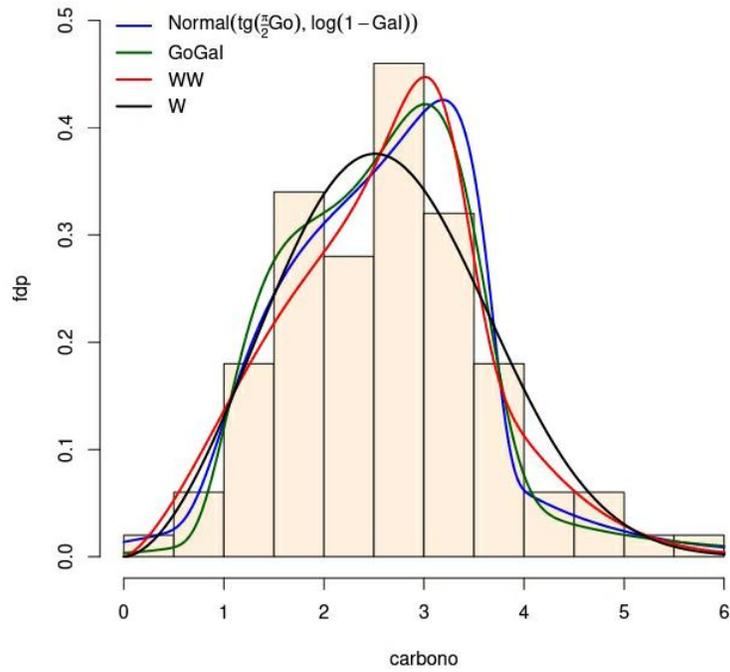


Figura 8: Histograma

Observe que a Normal ($\text{tg}(\frac{\pi}{2}\text{Go}), \log(1 - \text{Gal})$) está bem próxima a mistura GoGal, no entanto, ressaltamos que análises gráficas são meramente utilizadas com o intuito da visualização. O que de fato importa são as análises numéricas, e estas indicam que o nosso modelo é o melhor.

6 Considerações Finais

Tomando por base o método gerador de classes de distribuições de probabilidade apresentado em Brito *et al.* (2019) e Brito (2014) foi introduzida uma nova classe, denominada Classe Normal $(\text{tg}(\frac{\pi}{2}G_1), \log(1 - G_2))$. Foi feito o estudo de identificabilidade, assim como as expansões para a fda e para a fdp da nova classe. Foram mostradas algumas propriedades de caracterização, que sevem para qualquer modelo gerado pela classe, tais como, função de risco, momentos e momentos centrais de ordem m , função geradora de momentos, função característica, o coeficiente geral e os desvios, médio e quantílico. Foram calculadas as derivadas da função de log-verossimilhança da classe e feito o estudo do suporte.

Além disso, foi apresentada a distribuição normal $(\text{tg}(\frac{\pi}{2}Go), \log(1 - GaI))$ gerada pela classe e para ilustrar sua utilidade, foram feitas aplicações para um conjunto de dados simulados e para um conjunto de dados reais. O modelo ajustado foi comparado com outras distribuições competitivas, assim como com a mistura convexa de duas Weibulls, quanto às Estatísticas AIC, AICc, BIC, HQIC e os testes de aderência de Anderson-Darling (A^*) e Cramér-von Mises (W^*). Os resultados gerais indicam que a Normal $(\text{tg}(\frac{\pi}{2}Go), \log(1 - GaI))$ supera as outras distribuições em comparação.

Como sugestão para trabalhos futuros, destacam-se a obtenção de outros modelos e subclasses a partir da Classe Normal $(\text{tg}(\frac{\pi}{2}G_1), \log(1 - G_2))$; realizar o estudo dos parâmetros, analisando se são de escala ou de forma; aplicar o modelo proposto a outros conjuntos de dados e desenvolver as expressões para a função de risco. Ademais, espera-se que a nova classe e seus modelos derivados possam ter aplicações em outras áreas do conhecimento.

7 Referências

AARTS, R. M. **Lauricella functions**. From MathWorld, A Wolfram Web Resource, created by Eric W. Weisstein. 2000. (<http://mathworld.wolfram.com/LauricellaFunctions.html>).

ABREU, Y. K. L. Emprego de funções de densidade de probabilidade na modelagem da distribuição diamétrica de clones de eucalyptus spp. no Polo Gesseiro do Araripe. **Dissertação** (Mestrado) - Programa de Pós-Graduação em Biometria e Estatística Aplicada, Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife, Janeiro, 2018.

AFIFY, Ahmed Z.; NOFAL, Zohdy M.; BUTT, Nadeem Shafique. Transmuted complementary Weibull geometric distribution. **Pakistan Journal of Statistics and Operation Research**, p. 435-454, 2014a.

AFIFY, Ahmed Z. et al. The transmuted Marshall-Olkin Fréchet distribution: properties and applications. **International Journal of Statistics and Probability**, 2014b.

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, New York, v. 19, p. 716-723, 1974.

ALEXANDER, C., CORDEIRO, G. M., ORTEGA, E. M., SARABIA, J. M. Generalized beta-generated distributions. **Computational Statistics and Data Analysis**, v. 56, n. 6, p. 1880-1897, 2012.

ALVES, L. C.; LEITE, I. C.; MACHADO, C. J. Fatores associados à incapacidade funcional dos idosos no Brasil: análise multinível. **Revista de Saúde pública**, v. 44, p. 468-478, 2010.

ALZAATREH, A.; LEE, C.; FAMOYE, F. A new method for generating families of continuous distributions. **Metron**, v. 71, n. 1, p. 63-79, 2013.

ANDERSON, T.; DARLING, D. Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. **Annals of Mathematical Statistics**, v. 23, p. 193-212, 1952.

BARROS, P. S. N. Classes de distribuições weibull generalizada: teoria e aplicações. **Tese** (Doutorado) — Programa de Pós-Graduação em Biometria e Estatística Aplicada, Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife, Maio 2015.

BASTOS, J.; ROCHA, C. Análise de sobrevivência: conceitos básicos. **Arquivos de Medicina**, v. 20, n. 5-6, p. 185-187, 2006.

BERNARDI, J. V. E.; FOWLER, H. G.; LANDIM, P. M. B. Um estudo de impacto ambiental utilizando análises estatísticas espacial e multivariada. **Holos Environment**, v. 1, n. 2, p. 162-172, 2001.

BITTENCOURT, H. R.; VIALI, L. Contribuições para o ensino da distribuição normal ou curva de Gauss em cursos de graduação. **SIPEM** (Seminário Internacional de Pesquisa em Educação Matemática), 2006.

BOTELHO, V. A. V. A. Ajuste da distribuição gama aos dados de precipitação pluvial do município de

Lavras estado de Minas Gerais. **Dissertação** (Mestrado) - Programa de Pós-Graduação em Agronomia, Universidade Federal de Lavras. Lavras, Fevereiro 1998.

BOTREL, M. C. G.; VIEIRA, F. A. Análise da distribuição diamétrica de *Eucalyptus grandis* Hill ex Maiden. **Revista Científica Eletrônica de Engenharia Florestal**, v. 3, p. 6, 2005.

BOZDONGAN, H. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. **Psychometrika**. v.52, n.3, 345-370, Sep. 1987.

BRITO, C. C. R. de. Método gerador de distribuições e classes de distribuições probabilísticas. Recife: Universidade Federal Rural de Pernambuco, 2014. 241f. **Tese** (Doutorado em Biometria e Estatística Aplicada).

BRITO, C. C. R. ; CHAVES REGO, L. ; OLIVEIRA JUNIOR, W. R. ; GOMES-SILVA, F. Method for generating distributions and classes of probability distributions: the univariate case. **Hacettepe Journal of Mathematics and Statistics**, v. 48, p. 897-930, 2019.

BURNHAM, K. P.; ANDERSON, D. R. Model selection and multimodel inference: a practical information-theoretic approach. [S.l.]: **Springer**, 2002.

CASELLA, G.; BERGER, R. L. **Statistical inference**. Pacific Grove, CA: Duxbury, 2. ed., 2002.

CHEN, G.; BALAKRISHNAN, N. A general purpose approximate goodness-of-fit test. **Journal of Quality Technology**. 27, 154-161, 1995.

CORDEIRO, G. M.; NADARAJAH, S. Closed-form expressions for moments of a class of beta generalized distributions. **Brazilian journal of probability and statistics**, v. 25, n. 1, p. 14-33, 2011.

CORDEIRO, G. M.; ORTEGA; E. M. M.; DA CUNHA, D. C. C. The exponentiated generalized class of distributions. **Journal of data science**, v. 11, n. 1, p. 1-27, 2013.

COX, D. (1962), 'Further results on tests of separate families of hypotheses', **Journal of the Royal Statistical Society B** 24, 406-424.

DARLING, D. The kolmogorov-smirnov, cramer-von mises tests. **Annals of Mathematical Statistics**, v. 28, p. 823-838, 1957.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 39, n. 1, p. 1-22, 1977.

DRACHLER, M. D. L., ANDERSSON, M. C. S., LEITE, J. C. D. C., MARSHALL, T., AERTS, D. R. G. D. C., FREITAS, P. F., GIUGLIANNI, E. R. J. Desigualdade social e outros determinantes da altura em crianças: uma análise multinível. **Cadernos de Saúde Pública**, v. 19, n. 6, p. 1815-1825, 2003.

EUGENE, N.; LEE, C.; FAMOYE, F. Beta-normal distribution and its applications.

- Communications in Statistics-Theory and methods**, v. 31, n. 4, p. 497-512, 2002.
- EVERITT, B., HAND D. **Finite Mixture Distributions**. London: Chapman Hall, (1981).
- EXTON, H. **Handbook of Hypergeometric Integrals: Theory, Applications, Tables, Computer Programs**. Halsted Press, New York. 1978.
- FAMOYE, F.; LEE, C.; OLUMOLADE, O. The beta-weibull distribution. **Journal of Statistical Theory and Applications**, v. 4, p. 121–136, 2005.
- FELGUEIRAS, M. M. Modelação estatística com misturas e pseudo-misturas. **Tese (Doutorado)** — Universidade de Lisboa, Faculdade de Ciências, Departamento de Estatística e Investigação Operacional, Lisboa, 2009.
- FRUHWIRTH, S. **Finite Mixture and Markov Switching Models**. New York: Springer, 2006.
- GOMES, A. E., DA-SILVA, C. Q., CORDEIRO, G. M., ORTEGA, E. M. The beta burr iii model for lifetime data. **Brazilian Journal of Probability and Statistics**, v. 27, p. 502–543, 2013.
- GOMPERTZ, B. On the nature of the function expressive of the law of human mortality, and on a new method of determining the value of life contingencies. **Philos. Trans. R. Soc. London**, London, v.115, n.1825, p.513-585, 1825.
- GRADSHTEYN, I., RYZHIK, I. **Table of Integrals, Series, and Products**. Academic Press, 2007, New York.
- GREENWOOD, J. A., LANDWEHR, J. M., MATALAS, N. C., WALLIS, J. R. Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. **Water resources research**, v. 15, n. 5, p. 1049-1054, 1979.
- HANNAN, E. J.; QUINN, B. G. The determination of the order of an autoregression. **Journal of the Royal Statistical Society, Series B (Methodological)**, JSTOR, p. 190–195, 1979.
- JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **All of statistics: a concise course in statistical inference**. [S.I.]: Springer, 2013.
- JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. **Distribuições univariadas contínuas**, volume 2 . John wiley sons, 1995.
- JÚNIOR, J. J. C. R. Aplicação da Análise de Fourier Elíptica no estudo da forma do corpo dos peixes marinhos demersais do Nordeste do Brasil. **Dissertação (Mestrado)** - Programa de Pós-Graduação em Biometria e Estatística Aplicada, Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife, Fevereiro, 2019.
- KITE, G.W. Frequency and risk analysis in hydrology, Water Resources Publications, **Fort Collins** v. 3, 395 p., 1978.

- LAFRAIA, J. R. B. Manual De Confiabilidade, Manutenibilidade: E DISPONIBILIDADE. **Qualitymark Editora Ltda**, 2001.
- LAWLESS, J. Statistical models and methods for lifetime data. **New York: Wiley**, 1982.
- LEE, C.; FAMOYE, F.; ALZAATREH, A. Y. Methods for generating families of univariate continuous distributions in the recent decades. **Wiley Interdisciplinary Reviews: Computational Statistics, Wiley Online Library**, v. 5, n. 3, p. 219–238, 2013.
- LEMONTE, Artur J. et al. The beta log-logistic distribution. **Brazilian Journal of Probability and Statistics**, v. 28, n. 3, p. 313-332, 2014.
- LEMONTE, Artur J.; CORDEIRO, Gauss M. The exponentiated generalized inverse Gaussian distribution. **Statistics probability letters**, v. 81, n. 4, p. 506-517, 2011.
- LIANG, H.; ZOU, G. Improved aic selection strategy for survival analysis. *Computational statistics data analysis*, **Elsevier**, v. 52, n. 5, p. 2538–2548, 2008.
- LUSTOSA, P. R. B.; GARCIA, S.; BARROS, N.R. **Aplicabilidade do método de simulação de Monte Carlo na previsão dos custos de produção de companhias industriais: o caso da Companhia Vale do Rio Doce**. 2010.
- MCLACHLAN, G.; PEEL, D. **Finite mixture models**. [S.l.]: John Wiley & Sons, 2004.
- MENDES, P. N. Ajuste dos modelos Gompertz e logístico na descrição das curvas de crescimento de tomates: uma abordagem Bayesiana. **Tese (Doutorado) - Programa de Pós-Graduação em Estatística e Experimentação Agropecuária**, Universidade Federal de Lavras, Lavras, Fevereiro, 2011.
- NADARAJAH, S.; KOTZ, b. The beta exponential distribution. **Reliability Engineering System Safety**, v. 91, p. 689–697, 2006.
- NICHOLS, M. D.; PADGETT, W. A bootstrap control chart for weibull percentiles. *Quality and reliability engineering international*, **Wiley Online Library**, v. 22, n. 2, p. 141–151, 2006.
- PEARSON, K. Contributions to the mathematical theory of evolution. **Philosophical Transactions of the Royal Society of London A** **185**, 71-110, 1894.
- PINHEIRO, J. C.; BATES, D. M. Mixed-effects models in S and S-PLUS. New York, **New York: Springer-Verlag**, p. 528, 2000.
- POPOVIC, B. V.; CORDEIRO, G. M.; ORTEGA, E. M.; PASCOA, M. A. R. A new extended mixture normal distribution. **Mathematical Communications**, v. 22, p. 53-73, 2017.
- RISTIĆ, M. M.; NADARAJAH, S. A new lifetime distribution. **Journal of Statistical Computation and Simulation**, v. 84, n. 1, p. 135-150, 2014.
- SABINO, C. V. S.; LAGE, L. V.; ALMEIDA, K. C. D. B. Uso de métodos estatísticos robustos na

análise ambiental. **Engenharia Sanitaria e Ambiental**, v. 19, n. SPE, p. 87-94, 2014.

SANTOS, E.; CUNHA, M. Interpretação crítica dos resultados estatísticos de uma meta-análise: Estratégias metodológicas. **Millenium-Journal of Education, Technologies, and Health**, n. 44, p. 85-98-85-98, 2016.

SANTOS, F. S. Modelagem Matemático-Computacional de Ventos na Região Nordeste do Brasil. **Dissertação** (Mestrado) - Programa de Pós-Graduação em Biometria e Estatística Aplicada, Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife, Novembro, 2019.

SELLITTO, M. Formulação estratégica da manutenção industrial com base na confiabilidade dos equipamentos. **Produção**, v.15, n.1, p.44-59, 2005.

SILVA, J. W. L. Modelagem da biomassa e da quantidade de carbono de clones de Eucalyptus da Chapada do Araripe-PE. **Dissertação** (Mestrado) - Programa de Pós-Graduação em Biometria e Estatística Aplicada, Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife, Fevereiro, 2016.

SILVEIRA, F. V. J. et al. Normal-G Class of Probability Distributions: Properties and Applications. **Symmetry**, v. 11, n. 11, p. 1407, 2019.

SOUZA, G. T. Nova Generalização para a Classe Beta-G de Distribuições de Probabilidade. **Dissertação** (Mestrado) — Programa de Pós-Graduação em Biometria e Estatística Aplicada, Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, Recife, Março 2016.

SOUZA, R. N. Dinâmica e distribuição diamétrica de espécies arbóreas em fragmento florestal. **Dissertação** (Mestrado em Ciências Florestais). Universidade Federal Rural de Pernambuco. Recife, 2017. 72f.

THAIR, M. H., CORDEIRO, G. M., ALZAATREH, A., MANSOOR, M., ZUBAIR, M. The Logistic-X family of distributions and its applications. **Communications in Statistics-Theory and Methods**, v. 45, n. 24, p. 7326-7349, 2016.

TAHIR, M. H., ZUBAIR, M., MANSOOR, M., CORDEIRO, G. M., ALIZADEH, M., HAMEDANI, G. A new Weibull-G family of distributions. **Hacettepe Journal of Mathematics and Statistics**, v. 45, n. 2, p. 629-647, 2016.

Teicher, H. Identifiability of mixtures. **The Annals of Mathematical Statistics** 32, 244—248, 1961.

THOM, H. CS. A note on the gamma distribution. **Monthly weather review**, v. 86, n. 4, p. 117-122, 1958.

VELOSO, G. G.; MALIK, A. M. Análise do desempenho econômico-financeiro de empresas de saúde. **RAE eletrônica**, v. 9, n. 1, p. 0-0, 2010.

WASSERMAN, L. All of statistics: a concise course in statistical inference. [S.l.]: Springer, 2004.

WEIBULL, W. Wide applicability. **Journal of applied mechanics**, 1951.

WITKOVSKY, V. (2001). Calculando a distribuição de uma combinação linear de variáveis gama invertidas. **Kybernetika**. 37 (1): 79–90.

YOUSOF, Haitham M. et al. A new Weibull class of distributions: theory, characterizations and applications. **Journal of Statistical Research of Iran JSRI**, v. 15, n. 1, p. 45-82, 2018.

ZOGRAFOS, K.; BALAKRISHNAN, N. On families of beta-and generalized gammagenerated distributions and associated inference. **Statistical Methodology, Elsevier**, v. 6, n. 4, p. 344-362, 2009.