



Universidade Federal de Pernambuco
Centro de Ciências Exatas e da Natureza
Programa de Pós-Graduação em Estatística

JOSÉ JAIRO DE SANTANA E SILVA

**ALGUNS MODELOS NA FAMÍLIA NADARAJAH-HAGHIGHI G COM
APLICAÇÕES A DADOS CLÍNICOS E DE FADIGA**

Recife

2019

JOSÉ JAIRO DE SANTANA E SILVA

**ALGUNS MODELOS NA FAMÍLIA NADARAJAH-HAGHIGHI G COM
APLICAÇÕES A DADOS CLÍNICOS E DE FADIGA**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística do Centro de Ciências Exatas e da Natureza da Universidade Federal de Pernambuco, como requisito parcial à obtenção do título de mestre em Estatística.

Área de Concentração: Ciência Exatas e da Terra

Orientadora: Maria do Carmo Soares Lima

Recife

2019

Catálogo na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

S586a Silva, José Jairo de Santana e
Alguns modelos na família Nadarajah-Haghighi G com aplicações a dados clínicos e de fadiga / José Jairo de Santana e Silva. – 2019.
88 f.: il., fig., tab.

Orientadora: Maria do Carmo Soares Lima.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2019.
Inclui referências e anexo.

1. Estatística. 2. Distribuição exponencial-exponencializada. I. Lima, Maria do Carmo Soares (orientador). II. Título.

310 CDD (23. ed.) UFPE- MEI 2019-122

JOSÉ JAIRO DE SANTANA E SILVA

**ALGUNS MODELOS NA FAMÍLIA NADARAJAH-HAGHIGHI G COM
APLICAÇÕES A DADOS CLÍNICOS E DE FADIGA**

Dissertação apresentada ao Programa de
Pós-Graduação em Estatística da
Universidade Federal de Pernambuco,
como requisito parcial para a obtenção do
título de Mestre em Estatística.

Aprovada em: 22 de julho de 2019.

BANCA EXAMINADORA

Prof.(^a) Maria do Carmo Soares de Lima
UFPE

Prof.(^o) Leandro Chaves Rêgo
UFC

Prof.(^o) Frank Sinatra Gomes da Silva
UFRPE

À minha família, por me apoiar em todo momento e por ser um porto seguro. Mãe, seu amor e cuidado me deram forças pra continuar a senhora devo tudo. Pai, és meu maior exemplo.

AGRADECIMENTOS

Agradeço primeiramente a Deus por me permitir conseguir chegar até aqui e concluir essa etapa. A caminhada foi difícil, porém pude crescer muito nesse tempo.

Aos meus pais, Diva e Adeilson, por todo suporte e apoio nesse desafio. Aos meus irmãos Jaqueline, Júnior e Patrícia, a minha Tia Fátima e a minha namorada Alícia, por todo apoio e entendimento, devo muito a vocês, obrigado.

A todos os meus amigos e colegas, Adenice, Anabeth, André, Anny, Cris, Eduardo, Fernando, Joas, Jordan, Luis e Zé por estarem comigo nos bons e nos não tão bons momentos, a ajuda de vocês foi muito importante para que eu pudesse chegar aqui, devo muito a todos vocês, sou muito grato a todos! Em especial, devo um agradecimento a César Diogo, Lucas Miranda e João Eudes por estarem comigo na luta desde o início. As meninas, Cristine, Larissa e Ranah por estarem lado a lado comigo. Também a Alice, por toda ajuda inicial para a pesquisa.

Aos meus amigos e professores de Caruaru, por entenderem um pouco da minha ausência e por toda a amizade.

A todos os meus professores que muito contribuíram na minha formação, Raydonal, Cribari, Gauss e Audrey. E um agradecimento especial para a minha professora e orientadora Maria do Carmo, por toda paciência e dedicação comigo, sua ajuda foi de extrema importância.

Ao pessoal da secretaria em especial a Valéria por toda ajuda e empenho, meu muito obrigado.

A Capes pelo apoio financeiro.

RESUMO

Pesquisas envolvendo dados de sobrevivência e confiabilidade são de importância significativa em diversos campos do conhecimento como, por exemplo, em engenharia, biomedicina, ciências sociais dentre outros. Com o avanço na análise de dados, se faz necessário o uso de distribuições de probabilidade que melhor se ajustem a esses bancos de dados. A proposta de novas distribuições de probabilidade é um tema de pesquisa considerado por diversos autores, que são motivados a propor distribuições que tenham mais flexibilidade em termos de formas da função taxa de falha, por exemplo, e que, portanto, se ajustem a uma gama maior de banco de dados. Neste trabalho, consideraremos apenas aplicações a dados clínicos e de fadiga de equipamentos. Aqui, entende-se dados clínicos como aqueles relacionados a experimentos realizados em humanos ou animais. Além disso, realizamos uma revisão bibliográfica para mostrar o progresso das pesquisas em novas classes de distribuições. A ênfase maior será dada a família T-X de distribuições, proposta por Alzatreeh *et al.* (2013) e em classes propostas a partir dessa família. Em seguida, explicitamos a classe de distribuições que utilizamos como base para esse estudo: a Nadarajah Haghighi-G proposta por Dias (2016). A seguir, propomos a distribuição Nadarajah-Haghighi Exponencial-Exponencializada (NH-EE), que possui bastante flexibilidade em sua função taxa de falha. Fornecemos expressões matemáticas explícitas para algumas de suas propriedades estruturais; realizamos estimação pontual para o modelo utilizando o método da máxima verossimilhança e, por fim, realizamos duas aplicações a dados reais mostrando que o modelo se adequa bem aos bancos de dados escolhidos, quando comparado com algumas distribuições já conhecidas na literatura. No capítulo seguinte, propomos a Nadarajah-Haghighi Rayleigh (NH-R), que possui mais flexibilidade em relação a sua distribuição-base. Nesse capítulo, fornecemos de forma explícita, algumas propriedades estruturais da nova distribuição, como momentos e função geradora de momentos. Realizamos um estudo de estimação pontual através do método da máxima verossimilhança em que introduzimos censura a direita com mecanismo aleatório, assim como no capítulo anterior. Finalmente, realizamos duas aplicações a dados reais. Para verificar a adequabilidade do modelo comparamos com extensões da distribuição Rayleigh e verificamos que o modelo proposto é o que melhor se ajusta aos bancos de dados considerados. Por fim, no último capítulo propomos a distribuição Nadarajah-Haghighi Gumbel (NH-Gu), com quatro parâmetros que tem suporte nos reais. Nesse capítulo, fornecemos as principais propriedades matemáticas do modelo, realizamos estimação pontual considerando o método da máxima verossimilhança introduzindo censura a direita com mecanismo aleatório e

ilustramos a aplicabilidade da nova distribuição realizando duas aplicações empíricas a dados reais. A leitura dos capítulos que tratam das distribuições NH-EE, NH-R e NH-Gu pode ser feita de forma independente

Palavras-chave: Distribuição exponencial-exponencializada. Distribuição Gumbel. Distribuição Rayleigh. Família T-X de distribuições. Método de máxima verossimilhança. Simulação.

ABSTRACT

Researches involving survival and reliability data are of significant importance in several research fields, such as engineering, biomedicine, social sciences, among others. With the advancement in data analysis, it is necessary the use of probability distributions that best fit these data sets. The proposal of new probability distributions is a research topic considered by several authors, who are motivated to propose distributions that are more flexible in terms of the failure rate function shapes, for example; therefore, fitting a larger range of data sets. In this work, we consider only applications to clinical and fatigue data of equipment. In this work, clinical data are those related to experiments performed on humans or animals. Besides that, we perform a bibliographic review to show the progress of the research in new classes of distributions. The greater emphasis will be given to the T-X family of distributions, proposed by Alzatreeh *et al.* (2013), and in classes proposed from this family. Next, we explain the class of distributions that we use as a basis in this study: the Nadarajah Haghighi-G, proposed by Dias (2016). In addition, we propose the Nadarajah-Haghighi Exponential-Exponentiated (NH-EE) distribution, which has enough flexibility in its failure rate function. We provide explicit mathematical expressions for some of its structural properties and performed a point estimation evaluation for the model using the maximum likelihood method and performed two applications with real data, showing that the model is well suited to the chosen data sets, when compared to some distributions well known in the literature. In the next chapter, we propose the Nadarajah-Haghighi Rayleigh (NH-R) distribution, which has more flexibility than its baseline distribution. In this chapter, we explicitly provide some of the structural properties of the new distribution, such as moments and moment generating function. We performed a point estimation study via the maximum likelihood method in which we introduced right censorship with random mechanism, as in the previous chapter. Finally, we conduct two applications utilizing real data sets. In order to verify the suitability of the model, we compare it with extensions of the Rayleigh distribution and verify that the proposed model best fits the data sets used. Finally, in the last chapter, we propose the distribution Nadarajah-Haghighi Gumbel (NH-Gu), with four parameters that have support in the whole set of real numbers. We provide the main mathematical properties of the model, perform point estimation considering the maximum likelihood method introducing right censorship with random mechanism and illustrate the applicability of the new distribution by performing two applications with real data. The reading of the Chapters related to the NH-E, NH-R, and NH-Gu distributions can be done independently.

Keywords: Exponential exponentiated distribution. Gumbel distribution. Maximum likelihood method. Rayleigh distribution. Simulation. T-X family of distributions.

LISTA DE FIGURAS

Figura 1 – Gráfico ilustrativo de algumas curvas TTT	21
Figura 2 – Gráficos para a densidade e a função taxa de falha para a distribuição NH-EE para diferentes valores de parâmetros.	41
Figura 3 – TTT plot	47
Figura 4 – TTT plot	49
Figura 5 – Gráficos para a densidade e a função taxa de falha da distribuição NH-R.	53
Figura 6 – TTT plot: Fibras expostas a pressão	63
Figura 7 – TTT plot	64
Figura 8 – Gráficos para a densidade e a função taxa de falha da distribuição NH- Gu.	68
Figura 9 – TTT plot	78
Figura 10 – TTT plot	79

LISTA DE TABELAS

Tabela 1	– Estatísticas descritivas - Primeira Aplicação	47
Tabela 2	– Aplicação tempo até o surgimento de uma DCV	48
Tabela 3	– Estatísticas descritivas - Segunda Aplicação	49
Tabela 4	– Aplicação - Tensão até a ruptura de um fio	49
Tabela 5	– Teste da razão de verossimilhanças	50
Tabela 6	– Algumas extensões da distribuição Rayleigh	52
Tabela 7	– Resultados da Simulação para $(\alpha, \gamma) = (1, 2)$	60
Tabela 8	– Resultados da Simulação para $(\alpha, \gamma) = (2, 3)$	60
Tabela 9	– Resultados da Simulação para $(\alpha, \gamma) = (2.5, 1.8)$	61
Tabela 10	– Resultados da Simulação para $(\alpha, \gamma) = (1.7, 2.8)$	61
Tabela 11	– Estatísticas descritivas - Primeira Aplicação	62
Tabela 12	– Aplicação 1	63
Tabela 13	– Aplicação 2	65
Tabela 14	– Teste da razão de verossimilhanças	65
Tabela 15	– Algumas extensões da distribuição Gumbel	67
Tabela 16	– Resultados da simulação com $\alpha = 1.5, \lambda = 2.3, \mu = 1, \beta = 0.5$	75
Tabela 17	– Resultados da simulação com $\alpha = 1, \lambda = 4, \mu = 1, \beta = 1$	75
Tabela 18	– Resultados da simulação com $\alpha = 2, \lambda = 3, \mu = 2.1, \beta = 2.5$	76
Tabela 19	– Resultados da simulação com $\alpha = 4, \lambda = 1.7, \mu = 3, \beta = 2.8$	76
Tabela 20	– Estatísticas descritivas - primeira aplicação	78
Tabela 21	– Guinea-Pig	79
Tabela 22	– Estatísticas descritivas - Segunda Aplicação	80
Tabela 23	– Fibras de Carbono expostas a pressão	80
Tabela 24	– Teste da razão de verossimilhanças	80

SUMÁRIO

1	INTRODUÇÃO	15
2	REVISÃO BIBLIOGRÁFICA	19
2.1	CONJUNTOS DE DADOS EM ANÁLISE DE SOBREVIVÊNCIA E CON- FIABILIDADE	19
2.1.1	Determinação da função taxa de falha empírica	20
2.2	ALGUMAS CLASSES DE DISTRIBUIÇÕES CONTÍNUAS	21
2.2.1	Família de distribuições T-X	22
2.3	INFERÊNCIA	26
2.4	SIMULAÇÃO	27
2.5	CRITÉRIOS PARA A SELEÇÃO DE MODELOS	28
2.5.1	Estatísticas de Teste	28
2.6	CONSIDERAÇÕES FINAIS	30
3	FAMÍLIA DE DISTRIBUIÇÕES NADARAJAH-HAGHIGHI-G	31
3.1	INTRODUÇÃO	31
3.2	PROPRIEDADES DA FAMÍLIA NH-G	32
3.2.1	Expansões em Série de Potência	33
3.2.2	Momentos	33
3.2.3	Função geradora de momentos	33
3.3	ESTIMAÇÃO POR MÁXIMA VEROSSIMILHANÇA	34
3.4	SIMULAÇÃO	35
3.5	APLICAÇÃO	36
3.6	CONSIDERAÇÕES FINAIS	37
4	DISTRIBUIÇÃO NADARAJAH-HAGHIGHI EXPONENCIAL EXPO- NENCIALIZADA: PROPRIEDADES E APLICAÇÕES	38
4.1	INTRODUÇÃO	38
4.2	DISTRIBUIÇÃO NADARAJAH-HAGHIGHI EXPONENCIAL-EXPONENCIALIZADA	40
4.3	EXPANSÕES UTILIZADAS	41
4.4	ALGUMAS PROPRIEDADES	42
4.4.1	Momentos	42
4.4.2	Função geradora de momentos	42
4.4.3	Entropia	43

4.4.4	Estatísticas de ordem	43
4.5	INFERÊNCIA E ESTIMAÇÃO	44
4.5.1	Estimação via máxima verossimilhança	44
4.6	APLICAÇÕES A DADOS REAIS	45
4.6.1	Conjunto de dados 1: tempo até a ocorrência de alguma doença cardiovascular em 68 pessoas.	46
4.6.2	Conjunto de dados 2: o comportamento da fadiga de tração de um fio de urdidura e sua influência no desempenho da tecelagem.	47
4.7	CONSIDERAÇÕES FINAIS	50
5	UMA NOVA EXTENSÃO DA DISTRIBUIÇÃO RAYLEIGH: SIMULAÇÃO E APLICAÇÕES	51
5.1	INTRODUÇÃO	51
5.2	DISTRIBUIÇÃO NADARAJAH-HAGHIGHI-RAYLEIGH	52
5.3	EXPANSÕES UTILIZADAS	54
5.4	ALGUMAS PROPRIEDADES	55
5.4.1	Momentos	55
5.4.2	Função Geradora de Momentos	55
5.4.3	Entropia	56
5.4.4	Estatísticas de Ordem	56
5.5	INFERÊNCIA E ESTIMAÇÃO	57
5.5.1	Estudo de simulação	58
5.6	APLICAÇÃO A DADOS REAIS	61
5.6.1	Conjunto de Dados 1: Fibras vidro expostas a pressão	62
5.6.2	Conjunto de dados 2: carcinoma de orofaringe	63
5.7	CONSIDERAÇÕES FINAIS	64
6	UMA NOVA EXTENSÃO DE UMA DISTRIBUIÇÃO DE VALOR EXTREMO: DISTRIBUIÇÃO NADARAJAH HAGHIGHI GUMBEL COM PROPRIEDADES E APLICAÇÕES	66
6.1	INTRODUÇÃO	66
6.2	EXPANSÕES UTILIZADAS	69
6.3	ALGUMAS PROPRIEDADES	69
6.3.1	Momentos	69
6.3.2	Função geradora de momentos	70

6.3.3	Entropia	70
6.3.4	Estatísticas de ordem	71
6.4	INFERÊNCIA E ESTIMAÇÃO	72
6.4.1	Estudo de simulação	73
6.5	APLICAÇÃO A DADOS REAIS	77
6.5.1	Conjunto de dados 1: porcos da índia infectados com vírus da tuberculose	77
6.5.2	Conjunto de dados 2: fibras de carbono expostas a pressão	78
6.6	CONCLUSÃO	80
7	CONSIDERAÇÕES FINAIS	81
7.1	TRABALHOS FUTUROS	81
7.2	CONTRIBUIÇÕES DAS PESQUISAS	81
	REFERÊNCIAS	82
	ANEXO A – MANIPULAÇÕES NH-R	88

1 INTRODUÇÃO

Estudos em análise de sobrevivência tomam conjuntos de dados que consideram tempo até a ocorrência de um evento de interesse, por exemplo. Esses tipos de dados são observados em diversas situações, como em dados clínicos (tempo até cura de pacientes, por exemplo) ou em dados de fadiga (tempo até a falha de lâmpadas, por exemplo), dentre outros.

Com o avanço nas análises de dados, se faz necessário o uso de modelos mais elaborados que possam ajustar melhor esses bancos de dados. Nesse sentido, a proposta de novas distribuições é tema de pesquisa de diversos autores. O objetivo, na maioria dos trabalhos dessa área, é a busca de distribuições que possuam mais flexibilidade no que tange ao ajuste dos bancos de dados de interesse.

Na literatura, há livros que detalham diferentes distribuições de probabilidade e suas propriedades, como é o exemplo da coleção Johnston *et al.* (1994), Kotz *et al.* (2004) e Johnson *et al.* (2005). Essa coleção contempla distribuições contínuas e discretas para os casos univariado e multivariado. Walck (2007) traz o assunto de distribuições de probabilidade com enfoque em aplicações. Com o avanço nas propostas de novas distribuições, já são encontrados livros apenas para extensões da distribuição Weibull, como é o caso do livro proposto por Murthy *et al.* (2004) e também para a classe das distribuições exponencializadas AL-Hussaini e Ahsanullah (2015).

Neste trabalho, apresentamos modelos que estendem as distribuições Exponencial Exponencializada, Rayleigh e Gumbel, usando como classe de distribuições a Nadarajah-Haghighi- G , proposta por Dias (2016) e que é baseada na família $T-X$ de distribuições, proposta por Alzaatreh *et al.* (2013).

É interessante mostrar que, apesar da ideia da família $T-X$ ter sido desenvolvida apenas em 2013, algumas famílias que foram propostas muito antes, também fazem parte dessa família, como será mais detalhado a seguir. Alguns exemplos são as famílias exponencializada- G (GOMPERTZ, 1825), beta - G (EUGENE *et al.*, 2002) e McDonald - G (ALEXANDER *et al.*, 2012).

Essa família vem sendo utilizada por muitos autores visando a geração de novas classes que estendem distribuições conhecidas na literatura. Aljarrah *et al.* (2014) mostraram como gerar classes de distribuições $T-X$ utilizando a função quantílica. No artigo ainda foi citado a sua relação com a geração de classes de distribuição em que é usada a acumulada.

Jamal *et al.* (2018) propuseram a classe *Exponencial Inversa Generalizada Tipo II- G* usando a família $T-X$ de distribuições. Nesse trabalho, foram encontradas as expressões para as

principais propriedades estruturais da família. Apresentações de modelos usando distribuições-base para mostrar a flexibilidade foram detalhadas; a classe se mostra flexível pois engloba varias formas para função densidade e a função taxa de falha. Para as aplicações, foram considerados conjuntos de dados assimétricos.

Mansour *et al.* (2019) propuseram o gerador *transmuted transmuted-G* usando a família T-X de distribuições, mostrando-se bastante flexível, uma vez que pode incorporar função taxa de falha nas formas crescente, decrescente, unimodal, forma de banheira e forma de N.

Além disso, foi proposto recentemente por (BRITO *et al.*, 2019), um método que também engloba a família T-X de distribuições. A proposta do autor foi de unificar diversas formas de geração de distribuições e classes de probabilidade conhecidas da literatura.

A seguir, descrevemos os objetivos dessa dissertação:

- Propor distribuições na família Nadarajah-Haghighi-G, escolhendo *distribuições-base* conhecidas na literatura e que já foram usadas para modelar certos tipos de conjuntos de dados;
- Desenvolver e estudar as principais propriedades matemáticas de cada distribuição proposta, tentando obtê-las em forma fechada e, quando não for possível, usaremos expressões em série de potência;
- Fazer uso do método de máxima verossimilhança, com o intuito de tentar obter estimadores em forma fechada para os parâmetros;
- Realizar um estudo de simulação com e sem censura objetivando avaliar as propriedades assintóticas dos estimadores dos respectivos modelos em estudo;
- Avaliar o desempenho das distribuições propostas a conjuntos de dados clínicos e de fadiga, comparando-as com distribuições já conhecidas na literatura. Entende-se aqui como *conjuntos de dados clínicos* aqueles que envolvem experimentos relacionados a doenças em humanos ou animais.

Essa dissertação está organizada da seguinte maneira. No capítulo 2, trouxemos uma revisão bibliográfica mostrando informações e resultados importantes em análise de sobrevivência. Também mostramos diferentes formas de gerar novas classes e distribuições que são conhecidos na literatura. Entre as formas de geração de classes focamos na classe de distribuições T-X e explicitamos algumas classes que foram geradas a partir desse gerador. Detalhamos como irá ocorrer o estudo de inferência dos parâmetros e também o uso do método Monte Carlo. A metodologia utilizada na aplicação de dados é mostrada nos critérios para a seleção de modelos, onde utilizaremos as estatísticas de testes definidas por Chen e Balakrishnan (1995), (A^*) e

(W^*) .

A classe de distribuições Nadarajah Haghighi- G , que tomamos como base dessa dissertação, é mostrada no capítulo 3, detalhando como a proposta de novas distribuições é feita a partir dessa família. Algumas propriedades e estimação que serão utilizadas nos demais capítulos são mostradas, assim como as vantagens dessa família sobre algumas outras já conhecidas na literatura.

No capítulo 4, propomos uma extensão da distribuição exponencial exponencializada, chamada Nadarajah-Haghighi-Exponencial Exponencializada, com quatro parâmetros. A taxa de falha da distribuição em questão tem forma crescente, decrescente e constante, sendo mais flexível que a distribuição-base. Para o novo modelo, desenvolvemos suas principais propriedades estruturais, como momentos, função geradora de momentos, estatísticas de ordem e entropia de Shannon. Realizamos inferência via máxima verossimilhança. Duas ilustrações empíricas são fornecidas tomando um conjunto de dados de sobrevivência e outro de fadiga, tendo por objetivo analisar a aplicabilidade do modelo proposto.

A distribuição Nadarajah-Haghighi-Rayleigh é proposta no capítulo 5, com dois parâmetros adicionais em relação a sua distribuição-base, a distribuição Rayleigh. Encontramos para a distribuição proposta formas unimodal e crescente para a função taxa de falha. No capítulo, fornecemos expressões explícitas para as funções densidade, acumulada e taxa de falha e também expressões para os momentos, função geradora de momentos, estatística de ordem e entropia de Shannon. No que tange a inferência dos parâmetros, realizamos inferência utilizando o método da máxima verossimilhança e, para verificação das propriedades assintóticas dos estimadores de máxima verossimilhança, introduzimos um estudo de simulação via Monte Carlo. Na ilustração empírica do modelo proposto, realizamos duas aplicações a dados reais para verificar se o modelo é competitivo. Na aplicação, consideramos dados clínicos e dados de fadiga.

No capítulo 6, propomos uma extensão de uma distribuição de valor extremo: a distribuição Gumbel. A nova distribuição possui quatro parâmetros, dois a mais que a distribuição-base, e possui formas unimodal e crescente, para a função taxa de falha. Para a distribuição proposta, fornecemos, de forma explícita, as principais propriedades estruturais: momentos, função geradora de momentos e estatísticas de ordem, por exemplo. Estimamos os parâmetros da distribuição via máxima verossimilhança e analisamos suas propriedades assintóticas introduzindo um estudo de simulação via Monte Carlo. Com o intuito de mostrar a aplicabilidade do modelo, realizamos duas aplicações a dados reais. Nas aplicações, consideramos um conjunto de dados clínicos e outro conjunto de dados de fadiga.

No capítulo 7, fizemos uma análise do que foi proposto no trabalho mostrando os pontos relevantes e pesquisas futuras que poderão ser realizadas. Também mostramos as contribuições que foram feitas em congressos nacionais e eventos locais.

2 REVISÃO BIBLIOGRÁFICA

Neste capítulo, faremos uma revisão bibliográfica contendo informações sobre o que é abordado nos capítulos seguintes. Mostraremos como são trabalhados dados de sobrevivência e fadiga. Em seguida, faremos uma breve revisão da família T-X de distribuições (ideia central dessa dissertação), bem como traremos alguns casos particulares. A metodologia para estimação dos parâmetros dos modelos e para as aplicações são mostradas em seguida.

2.1 CONJUNTOS DE DADOS EM ANÁLISE DE SOBREVIVÊNCIA E CONFIABILIDADE

Em análise de sobrevivência, a variável de interesse é o tempo até a ocorrência de um evento de interesse. Este tempo é chamado de tempo de falha; alguns exemplos de variáveis de interesse para esses tipos de dados é o tempo até a morte de um paciente, tempo até a falha de equipamentos, dentre outros. Como a variável tempo é sempre positiva, os conjuntos de dados possuem suporte nos reais positivos ($\mathbb{R}^+$). Sabendo que o suporte são os valores que a variável pode assumir para imagens positivas da densidade ou função de probabilidade.

Os dados de confiabilidade são frequentemente coletados com o objetivo de determinar os limites de uso seguro de um componente ou sistema. Pode-se acreditar que existe um nível de estresse que abaixo dele o material é seguro, mas acima desse nível existe uma probabilidade não desprezível de ruptura ou falha desse componente ou sistema.

Dados que possuem essas características podem ser encontrados em estudos clínicos e de confiabilidade. Exemplos de bancos de dados de confiabilidade e estudos clínicos podem ser encontrados em diversos livros e artigos como, por exemplo, em livros como Murthy *et al.* (2004) e Colosimo e Giolo (2006), ou em artigos como Bjerkedal *et al.* (1960) e Nichols e Padgett (2006). Há também bancos de dados implementados em pacotes do software estatístico R como, por exemplo o pacote *Survival* encontrado em <https://github.com/therneau/survival>

Uma outra característica desses dados é a presença de censura, que é a observação parcial da resposta, ou seja, para alguns elementos do estudo é desconhecido o tempo até a ocorrência do tempo do evento de interesse. Existem alguns tipos de censura, a direita e a esquerda, descritos a seguir:

- **TIPO I** - Aquela em que o estudo é encerrado após um período pré-determinado de tempo;
- **TIPO II**- Aquela em que o estudo é encerrado após ter ocorrido a falha em um número pré-estabelecido de indivíduos;
- **Aleatória**- Ocorre quando o indivíduo é retirado do estudo por causas alheias ao experi-

mento.

No entanto, consideramos aqui apenas censura à direita, em que o tempo de falha é superior ao tempo registrado, com mecanismo aleatório. Para mais detalhes, ver Colosimo e Giolo (2006).

Seja T uma variável aleatória não negativa e contínua que representa o tempo até a ocorrência de algum evento. Denota-se por $f_T(t)$ sua função densidade de probabilidade (fdp) e $F_T(t)$ sua função de distribuição acumulada (fda). Com base nessas duas funções, podemos definir duas outras funções que são utilizadas em análise de sobrevivência e confiabilidade: as funções sobrevivência e taxa de falha.

A função de sobrevivência é o complementar da função de distribuição acumulada para o tempo até a falha. Essa função é dada por

$$S_T(t) = \mathbf{P}(T \geq t) = 1 - F_T(t),$$

em que temos como representação a probabilidade de que o evento em estudo sobreviva pelo menos até um tempo t , $t > 0$.

A função taxa de falha (ftf) tem a expressão dada a seguir

$$h_T(t) = \lim_{\delta t \rightarrow 0} \frac{\mathbf{P}(t \leq T \leq t + \delta t | T \geq t)}{\delta t} = \frac{f_T(t)}{S_T(t)}.$$

Essa função pode apresentar formas constante, crescente, decrescente, em forma de U (ou banheira), unimodal, banheira invertida e forma de N, a mais recentemente discutida na literatura. Quanto mais formas essa função apresentar, significa dizer que a distribuição proposta se adequará a uma gama maior de conjuntos de dados.

2.1.1 Determinação da função taxa de falha empírica

Como dito anteriormente, há varias formas que a ftf, de uma variável aleatória T , pode assumir. Uma metodologia que é bastante utilizada para a identificação do modelo, que pode melhor representar um banco dados, é proposto por Aarset (1987): o tempo total em teste (curva TTT). A metodologia toma a análise da forma da curva TTT em um gráfico. A curva TTT é obtida construindo um gráfico em que são plotados os pares $(r/n, G(r/n))$ que obtemos da seguinte expressão: $G(r/n) = [(\sum_{i=1}^r T_{i:n}) + (n-r)T_{r:n}] / (\sum_{i=1}^n T_{i:n})$, em que $r = 1, \dots, n$ e $T_{i:n}$, $i = 1 \dots, n$, é a estatística de ordem da amostra.

A Figura 1 nos mostra as diferentes formas que o gráfico TTT pode assumir. Quando a curva, obtida do gráfico TTT, é uma reta diagonal (Curva A), a função taxa de falha é constante; quando a curva é convexa (curva B) ou côncava (curva C) a função taxa de falha

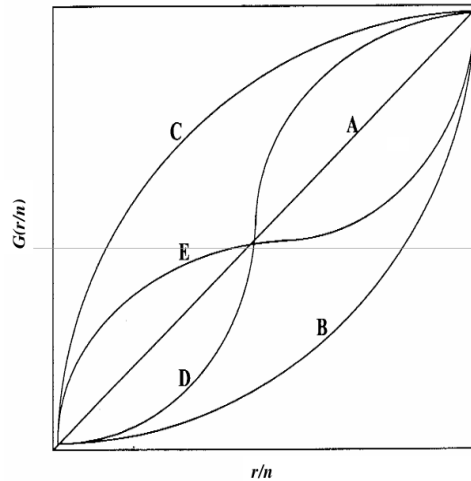


Figura 1 – Gráfico ilustrativo de algumas curvas TTT

é, respectivamente, monotonicamente decrescente ou crescente; quando a curva é convexa e depois côncava (curva D), a função taxa de falha tem forma de U, e no caso reverso (curva E) é unimodal.

2.2 ALGUMAS CLASSES DE DISTRIBUIÇÕES CONTÍNUAS

Na literatura, podemos encontrar diversas formas de geração de novas distribuições. Cada forma de gerar possui suas particularidades e motivações. A seguir mostraremos algumas formas que estão na literatura sobre a proposta de novas distribuições. Nesta seção, apresentamos como gerar famílias a partir do gerador T-X de distribuições, onde é escolhido uma função densidade de probabilidade. Em seguida, trazemos alguns casos particulares obtidos através do uso desse gerador.

Além da classe T-X, há algumas outras formas de geração de distribuições, como é o caso da adição de parâmetros, proposta inicialmente por Gompertz (1825) no cálculo de tabelas de mortalidade, com a manipulação da função de distribuição acumulada, $G(t) = (1 - \rho e^{\lambda t})^\alpha$ para $t > \lambda^{-1} \log \rho$, em que ρ, α e λ são números reais positivos. Ahuja e Nash (1967) introduziram a família Gompertz-Verhulst de distribuições, também estudando curvas de mortalidade. A distribuição Exponencial Exponencializada é um caso particular quando $\rho = 1$. Propriedades e métodos de estimação para a família das exponencializadas foram estudadas em diversos trabalhos como, por exemplo, Mudholkar *et al.* (1995), Mudholkar e Hutson (1996), Gupta e Kundu (1999), Gupta e Kundu (2001b).

nUma outra forma de gerar novas distribuições é pelo método da composição, que consiste em compor variáveis aleatórias discreta com contínua. Há diversas aplicações a dados

reais para distribuições com essa característica, uma vez que pode modelar o tempo até a falha de equipamentos em série, situação em que podemos ter valores pontuais ou contínuos. Um caso é a distribuição odd log-logistic Birnbaum–Saunders–Poisson proposta por Cordeiro *et al.* (2018).

Ainda sobre novas formas de propor distribuições, temos a proposta de mistura. A ideia é tomar mais de uma variável aleatória, iguais ou distintas, fazer a multiplicação de cada uma delas por um peso e adicioná-las. A condição aqui é que a soma total dos pesos seja igual a um, ou seja, satisfazendo as condições que podem ser vistas em Lindsay (1995). Na geração de uma nova distribuição por mistura, temos a definição $\sum \pi_j f(x; \xi_j)$, em que $\pi_j \geq 0$, $\sum_j \pi_j = 1$, π é definido como peso e $f(x; \xi_j)$ são as densidades das variáveis aleatórias que serão usadas para a mistura.

Um exemplo dessa forma de geração de novas distribuições pode ser vista em Erişoğlu *et al.* (2011), que apresenta a mistura entre duas variáveis aleatórias distintas considerando a seguinte expressão

$$f_{X,Y}(y; \psi) = \pi f_X(t; \theta_X) + (1 - \pi) f_Y(t; \theta_Y).$$

No artigo, os autores mostram algumas distribuições que foram geradas a partir dessa ideia como, por exemplo, a Gamma-Weibull. Zeghdoudi e Nedjar (2014) propôs a distribuição Gamma-Lindley. No artigo foi fornecido em forma fechada expressões para importantes propriedades matemáticas e mostrou-se a flexibilidade do modelo.

A seguir, mostraremos como é definido a família de distribuições pela T-X que gera a classe que trabalhamos nessa dissertação.

2.2.1 Família de distribuições T-X

Utilizado recentemente por diversos autores como mencionado anteriormente, o gerador T-X, definido por Alzaatreh *et al.* (2013), é de extrema relevância por englobar diversas famílias de distribuições. A fda desse gerador é dada por

$$F(x) = \int_a^{W[G(x; \boldsymbol{\eta})]} r(t) dt, \quad (2.1)$$

em que $G(x; \boldsymbol{\eta})$ é a fda da variável aleatória X , $r(t)$ é a fdp da variável aleatória T , definida no intervalo $[a, b]$ e $W[G(x; \boldsymbol{\eta})]$ é uma função de $G(x; \boldsymbol{\eta})$ que satisfaz as seguintes condições:

- $W[G(x; \boldsymbol{\eta})] \in [a, b]$;
- $W[G(x; \boldsymbol{\eta})]$ é diferenciável e monótona não decrescente;
- $W[G(x; \boldsymbol{\eta})] \rightarrow a$ quando $x \rightarrow -\infty$ e $W[G(x; \boldsymbol{\eta})] \rightarrow b$ quando $x \rightarrow \infty$.

A maneira como $W[G(x; \boldsymbol{\eta})]$ é definida nos dá uma nova família de distribuições. Para mais detalhes, ver Alzaatreh *et al.* (2013) e Dias (2016), por exemplo. Para o gerador proposto por Dias (2016), a $r(t)$ é a fdp da distribuição Nadarajah-Haghighi, e a função $W[G(x; \boldsymbol{\eta})]$ é da forma $-\log[1 - G(x; \boldsymbol{\eta})]$. Mais detalhes sobre o gerador NH-G são mostrados no capítulo 3.

É importante ressaltar que com a mesma densidade para $r(t)$ podemos ter diferentes classes de distribuições, pois a escolha da $W[G(x)]$ pode diferir; é o que acontece com as classes Weibull-G de Bourguignon *et al.* (2014), Weibull-X de Alzaatreh e Ghosh (2015) e a Nova Weibull-G proposta por Tahir *et al.* (2016). As classes geram distribuições diferentes, pois a $W[G(x)]$ não é a mesma, como veremos adiante.

Algumas famílias de distribuições que podem ser obtidas a partir do gerador T-X, são detalhadas a seguir

- Tomando como $r(t)$ a densidade da variável aleatória Weibull e variando o $W[g(x); \boldsymbol{\eta}]$ temos as seguintes classes propostas:

– Weibull-G

Proposta por Bourguignon *et al.* (2014), a classe de distribuições Weibull-G é obtida usando $W[G(x; \boldsymbol{\eta})] = \frac{G(x; \boldsymbol{\eta})}{1 - G(x; \boldsymbol{\eta})}$, em que $G(x)$ é a distribuição acumulada da variável aleatória X, e a $r(t)$ é a fdp da variável aleatória Weibull. A fda e fdp dessa classe de distribuições são dadas, respectivamente, por

$$F_{W-G}(x) = 1 - \exp \left\{ -a \left[\frac{G(x; \boldsymbol{\eta})}{1 - G(x; \boldsymbol{\eta})} \right]^b \right\} \quad (2.2)$$

e

$$f_{W-G}(x) = a b g(x; \boldsymbol{\eta}) \frac{g(x; \boldsymbol{\eta})^{b-1}}{[1 - G(x; \boldsymbol{\eta})]^{b+1}} \exp \left\{ -a \left[\frac{G(x; \boldsymbol{\eta})}{1 - G(x; \boldsymbol{\eta})} \right]^b \right\}, \quad (2.3)$$

em que $a > 0$ e $b > 0$ são, respectivamente, parâmetros de escala e forma e $\boldsymbol{\eta}$ é o vetor de parâmetros da distribuição-base G . Algumas distribuições foram propostas com base nessa classe de distribuições: Ramírez (2017) com a Weibull Nadarajah-Haghighi de 4 parâmetros, Oguntunde *et al.* (2015) com a Weibull exponencial de três parâmetros, dentre outras.

– Weibull-X

Utilizando também a densidade da Weibull, em $r(t)$ na Expressão 2.1, porém com $W[G(x)] = -\log[1 - F(x)]$, Alzaatreh e Ghosh (2015), propuseram a Weibull-X, onde foi verificado as propriedades da distribuição junto com aplicações usando

como distribuição-base a logística. Para a distribuição Weibull-logística foram fornecidas algumas propriedades bem como simulação e aplicações a dados reais. Essa nova família tem fda e fdp dadas, respectivamente por

$$F_{W-X}(x) = 1 - \exp \left\{ - \left(- \frac{\log(1 - G(x; \boldsymbol{\eta}))}{a} \right)^b \right\} \quad (2.4)$$

e

$$f_{W-X}(x) = \frac{bg(x; \boldsymbol{\eta})}{a(1 - G(x; \boldsymbol{\eta}))} \left(\frac{-\log(1 - G(x; \boldsymbol{\eta}))}{a} \right)^{b-1} \exp\{ -((\log(1 - G(x; \boldsymbol{\eta}))/a)^b) \}. \quad (2.5)$$

– Nova Weibull-G

Ainda com a $r(t)$ sendo a densidade da distribuição Weibull, Tahir *et al.* (2016) propuseram uma nova Weibull-G, em que $W[G(x)] = -\log[G(x)]$. No artigo, foi abordado um pouco das formas de gerar novas famílias através do gerador T-X. Para a nova família também foram fornecidas as propriedades matemáticas, formas para a estimação por máxima verossimilhança e, para ilustrar empiricamente duas distribuições que pertencem a essa classe, Weibull-log-logística e Weibull-Weibull, foram realizadas aplicação a dados reais. A fdp da nova Weibull-G é dada a seguir

$$f_{NW-G}(x) = ab \frac{g(x; \boldsymbol{\eta})}{G(x; \boldsymbol{\eta})} \{ -\log[G(x; \boldsymbol{\eta})] \}^{b-1} \exp\{ -a[-\log[G(x; \boldsymbol{\eta})]]^b \} \quad (2.6)$$

• Burr XII-G

Proposta por Santos (2016) a classe de distribuições Burr XII-G é obtida a partir do gerador T-X em que $W[G(x)] = G(x; \boldsymbol{\eta})/[1 - G(x; \boldsymbol{\eta})]$ e $r(t)$ é a densidade da variável aleatória Burr XII. A fda e fdp dessa família são dadas, respectivamente, por

$$F_{BXII-G}(x) = 1 - \left\{ 1 + \left[\frac{G(x; \boldsymbol{\eta})}{1 - G(x; \boldsymbol{\eta})} \right]^a \right\}^{-b} \quad (2.7)$$

e

$$f_{BXII-G}(x) = a b g(x; \boldsymbol{\eta}) \frac{G(x; \boldsymbol{\eta})^{a-1}}{1 - G(x; \boldsymbol{\eta})^{a+1}} \left\{ 1 + \left[\frac{G(x; \boldsymbol{\eta})}{1 - G(x; \boldsymbol{\eta})} \right]^a \right\}^{-b-1} \quad (2.8)$$

Essa classe de distribuições possui alguns casos especiais. Quando $a = 1$ em (2.7), a família recai na Lehmann tipo-II($b, \boldsymbol{\eta}$) (LEHMANN *et al.*, 1953). Se caso $b = 1$, obtemos a família odd log-logistic-G, Cruz *et al.* (2014).

• Kumaraswamy-G

Cordeiro e Castro (2011) introduziram a família de distribuições Kumaraswamy-G (Kw-G) considerando $r(t)$ como sendo a densidade da variável aleatória T seguindo uma variável

aleatória Kumaraswamy (KUMARASWAMY, 1980). A função $W[G(x)]$ é definida como $W[G(x; \boldsymbol{\eta})] = G(x; \boldsymbol{\eta})$. A família Kw-G tem fda e fdp dadas, respectivamente, por

$$F_{Kw-G}(x) = 1 - [1 - G(x; \boldsymbol{\eta})^a]^b \quad (2.9)$$

e

$$f_{Kw-G}(x) = a b g(x; \boldsymbol{\eta}) G(x; \boldsymbol{\eta})^{a-1} [1 - G(x; \boldsymbol{\eta})^a]^{b-1} \quad (2.10)$$

em que $a > 0$ e $b > 0$, são parâmetros de forma. Distribuições que foram propostas com base nessa classe de família são a Kw-generalized Rayleigh, distribuição proposta por Gomes *et al.* (2014) e a Kw-Weibull modificada, proposta por Cordeiro *et al.* (2014), por exemplo.

- Beta-G

Eugene *et al.* (2002) propôs a família de distribuições beta, essa família pode ser escrita a partir da família T-X tomando $H[G(x; \boldsymbol{\eta})] = G(x; \boldsymbol{\eta})$ e $r(t)$ como a densidade da distribuição beta. A fda e fdp da classe beta-G são dadas, respectivamente, por

$$F_{B-G}(x) = \frac{B_{G(x; \boldsymbol{\eta})}(a, b)}{B(a, b)} \quad (2.11)$$

e

$$f_{B-G}(x) = \frac{1}{B(a, b)} g(x; \boldsymbol{\eta}) G(x; \boldsymbol{\eta})^{a-1} [1 - G(x; \boldsymbol{\eta})]^b \quad (2.12)$$

em que $a > 0$, $b > 0$, $B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt$ é a função beta e $B_x(a, b) = \int_0^x t^{a-1} (1-t)^{b-1} dt$ é a função beta incompleta.

Uma distribuição dessa classe é a beta Weibull modificada, distribuição que possui bastante flexibilidade proposta por (SILVA *et al.*, 2010).

- Gamma-G

Proposta por (ZOGRAFOS; BALAKRISHNAN, 2009) e utilizando como $r(t)$ a densidade da distribuição Gamma e como $W[G(x; \boldsymbol{\eta})] = -\log[1 - G(x)]$ temos a classe Gamma-G, a fda e fda dessa classe são dadas, respectivamente, por

$$F(x) = \frac{1}{\Gamma(a)} \int_0^{-\log[1-G(x)]} t^{a-1} e^t dt \quad (2.13)$$

e

$$f(x) = \frac{g(x)}{\Gamma(a)} \{-\log[1 - G(x)]\}^{a-1} \quad (2.14)$$

A classe como exemplo as distribuições Ristić e Balakrishnan (2012) e Cordeiro *et al.* (2018).

2.3 INFERÊNCIA

Para a estimação dos parâmetros das distribuições propostas, foi utilizado o método da máxima verossimilhança. Consideramos a estimação utilizando a verossimilhança com censura à direita e mecanismo aleatório, afim de verificar as propriedades dos estimadores como a consistência, que nos diz que, ao aumentarmos o tamanho da amostra, o valor das estimativas tende ao valor verdadeiro. Outro aspecto seria observar o que acontece ao introduzirmos diferentes percentuais de censura, ou seja, observar se os valores das estimativas tem um vício maior, ao aumentarmos esses percentuais.

Sejam $X_1, \dots, X_n$ uma amostra aleatória de uma variável aleatória X com função densidade de probabilidade $f(x; \boldsymbol{\theta})$, em que $\boldsymbol{\theta}$ é o vetor de parâmetros da variável aleatória. A função de verossimilhança é definida da seguinte forma

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

Observe que a função de verossimilhança mostra que a contribuição de cada observação não censurada é a função de densidade, porém, se o dado é censurado, temos que considerar alguma penalização. Esta informação nos diz que o tempo de falha é maior que o tempo de censura observado. A penalização considerando o dado com censura, para a verossimilhança é a função de sobrevivência $S(T)$. Para mais detalhes ver Colosimo e Giolo (2006). A função de verossimilhança que considera esse caso é dada por

$$L(\boldsymbol{\theta}; x) \propto \prod_{i=1}^n [f(x; \boldsymbol{\theta})]^{\delta_i} [S(x; \boldsymbol{\theta})]^{1-\delta_i}, \quad (2.15)$$

em que δ_i é uma variável indicadora de tempo de falha ou censura. Sobre os tipos de censura, Lawless (2011) nos diz que há três tipos de censura e que são encontradas para casos de censura à direita e à esquerda.

Para encontrarmos o estimador de máxima verossimilhança, devemos encontrar o vetor de estimativas $\hat{\boldsymbol{\theta}}$ que maximiza $L(\boldsymbol{\theta})$. Como a função logaritmo é monótona, maximizar

$L(\boldsymbol{\theta})$ é equivalente a maximizar $l(\boldsymbol{\theta}) = \log[L(\boldsymbol{\theta})]$, e o resultado do sistema de equações

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0$$

nos dá essas estimativas. Porém, nem sempre os estimadores de máxima verossimilhança possuem forma fechada, então se faz necessário o uso de métodos numéricos para obtenção das estimativas, como os algoritmos de Newton Rapshson e BFGS, por exemplo. Para mais detalhes desses métodos ver, Press *et al.* (2007) e Nocedal e Wright (2006).

Quantidades que são avaliadas nessas estimativas são o viés, $B(\cdot)$, das estimativas e o Erro quadrático médio, $EQM(\cdot)$, que são definidos, respectivamente, por

$$\begin{aligned} B(\hat{\boldsymbol{\theta}}) &= E[\hat{\boldsymbol{\theta}}] - \boldsymbol{\theta} \text{ e} \\ EQM(\hat{\boldsymbol{\theta}}) &= Var[\hat{\boldsymbol{\theta}}] + B^2(\hat{\boldsymbol{\theta}}). \end{aligned}$$

Ao observarmos essas quantidades, podemos verificar se os estimadores de máxima verossimilhança para o modelo tem as propriedades de consistência e eficiência, ou seja, se ao aumentarmos o tamanho da amostra o seu viés e EQM tendem a zero.

2.4 SIMULAÇÃO

A simulação de Monte Carlo é usada para avaliar desempenhos de procedimentos inferenciais. A importância dos estudos de simulação se deve ao fato de replicarmos os resultados obtidos computacionalmente. Além disso, podemos, através de computação intensiva, obter valores que ficariam inviáveis por tomarem bastante tempo.

O método de Monte Carlo é um método numérico não determinístico (ou estatístico) usado para aproximar expressões matemáticas complexas e custosas de serem avaliadas com precisão. O método também fornece soluções aproximadas para uma grande variedade de problemas matemáticos que permitem a realização de experiências com amostras de números pseudo-aleatórios gerados em um computador.

Em estatística, é muito utilizado para a avaliação do desempenho de estimadores (pontuais e intervalares) e testes de hipóteses, usando como aporte teórico a Lei Fraca dos grandes números: a média amostral converge em probabilidade para a média populacional. Isso se dá pelo fato de que geramos de forma independente uma amostra finita de uma variável aleatória e para valores maiores de n a média dessa amostra tende a média da população, μ . Se $X_1, \dots, X_n$ são variáveis aleatórias independentes tais que $E(X_i) = \mu$ e $var(X_i) = \sigma^2 (< \infty)$ para

$i = 1, \dots, n$, então,

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \mu,$$

isto é, a média amostral converge em probabilidade para a média populacional. Para mais detalhes sobre a lei fraca ver James (1996).

Como os estimadores de máxima verossimilhança são assintoticamente não viesados, podemos fazer o estudo destes para pequenas e grandes amostras. Essa verificação é importante, pois podemos averiguar se esse viés é grande para pequenas amostras e propor estudos para a correção do mesmo de forma analítica (COX; SNELL, 1968) ou pelo método bootstrap (EFRON; TIBSHIRANI, 1993).

Com base em tudo que foi dito, a Simulação de Monte Carlo vem a ser uma ferramenta muito importante para verificação de algumas propriedades assintóticas apresentadas nesse trabalho.

2.5 CRITÉRIOS PARA A SELEÇÃO DE MODELOS

No ajuste das distribuições aos bancos de dados reais, algum critério deve ser selecionado para decidir qual modelo se ajusta melhor. Na busca de alguma medida que seja útil para a seleção de modelos, encontramos aquelas que usam a função de densidade, como AIC, BIC, CAIC e HQIC. Elas são utilizadas quando os modelos são encaixados ou autoregressivos. Elas são úteis quando comparamos dois modelos, porém elas não garantem a bondade do ajuste.

Com base nessa deficiência, buscamos medidas mais robustas para o critério de seleção de modelos. Na literatura, para a seleção do modelo que melhor representa o conjunto de dados, encontramos as estatísticas de teste. Nas estatísticas de teste são envolvidas as fdas escolhidas, diferente das medidas citadas anteriormente, que utilizam a fdp. A seguir, explicitamos as estatísticas que foram utilizadas para a análise dos nossos modelos.

2.5.1 Estatísticas de Teste

As estatísticas de teste são baseadas na $F_n(x)$, função de distribuição empírica (fde). Tome uma amostra de tamanho n . A fde é definida por

$$F_n(x) = \begin{cases} 0, & x < x_{(1)} \\ \frac{1}{n}, & x_{(i)} \leq x < x_{(i+1)}, \quad i = 1, \dots, n-1, \\ 1, & x_{(n)} \geq x \end{cases} \quad (2.16)$$

em que $x_{(i)}$ representa a i -ésima estatística de ordem. Note que $F_n(x)$ é uma função escada.

Embasado na diferença vertical entre a fde e a fda, D'Agostino (1986) mostra duas classes de estatísticas. A primeira classe de estatísticas é baseada no supremo, ou seja, o valor absoluto entre a diferença da fde e fda $|F_n(x) - F(x)|$. A estatística Kolmogorov-Smirnov (KS) usa essa diferença no seu cálculo.

Para as estatísticas da segunda classe, consideramos todas as diferenças entre a fde e fda. Essa classe é definida por

$$Q = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \psi(x) dF(x), \quad (2.17)$$

em que ψ é uma função de peso para a forma quadrática. Definidas por Chen e Balakrishnan (1995), as estatísticas de Cramér-von-Mises (W^*) e Anderson Darling (A^*) tomam $\psi(x) = 1$ e $\psi(x) = \{F(x)[1 - F(x)]\}^{-1}$. A expressão para essas estatísticas são dadas, respectivamente, por

$$W^* = \left\{ n \int_{-\infty}^{\infty} [F_n(x) - \hat{F}(x)]^2 d\hat{F}(x) \right\} \left(1 + \frac{0.5}{n} \right)$$

$$A^* = \left\{ n \int_{-\infty}^{\infty} \frac{[F_n(x) - \hat{F}(x)]^2}{\hat{F}(x) [1 - \hat{F}(x)]} d\hat{F}(x) \right\} \left(1 + \frac{0.75}{n} + \frac{2.25}{n^2} \right),$$

em que $\hat{F}(x)$ é a fda avaliada na estimativa de máxima verossimilhança (EMV) do vetor de parâmetros ϕ , denotado por $\hat{\phi}_n$. A hipótese nula, em questão, leva em consideração a adequação dos dados à fda proposta. Em ambos os casos, as estatísticas A^* e W^* nos evidenciam que quanto menor o valor, há mais evidências a favor da hipótese nula.

Na comparação de dois modelos encaixados, podemos utilizar o teste da razão de verossimilhanças (TRV) (BOLFARINE; SANDOVAL, 2001). Seja $\Theta = (\theta_1, \dots, \theta_p)^T$ o vetor de parâmetros de um modelo com dimensão p . Considere também um sub-modelo com vetor de parâmetros θ^0 , também de dimensão p , porém com k parâmetros conhecidos e $1 \leq k \leq p$. Esse teste tem por hipótese nula k igualdades que são definidas em θ^0 . Então,

$$H_0 : \theta_1 = \theta_1^0, \dots, \theta_k = \theta_k^0, \quad \theta_{k+1}, \dots, \theta_p,$$

em que $\theta_{k+1}, \dots, \theta_p$ são os parâmetros desconhecidos e $\theta_1^0, \dots, \theta_k^0$ são os parâmetros conhecidos. O TRV tem por definição a quantidade $-2\log(\lambda)$ em que

$$\lambda = \frac{\sup_{\Theta_0} L(\theta; x)}{\sup_{\Theta} L(\theta; x)}$$

e Q é a variável aleatória que toma os valores de λ . Então, $-2\log(Q)$ possui a distribuição assintótica Qui-quadrado com k graus de liberdade (χ_k^2). Tomando um nível de significância

$0 < \alpha < 1$, a hipótese nula é rejeitada quando $-2\log(\lambda) > q_{1-\alpha}$, em que $q_{1-\alpha}$ é o quantil da distribuição χ_k^2 .

Tomando por base essas quantidades apresentadas, podemos analisar qual modelo consegue se adequar melhor aos conjuntos de dados apresentados.

2.6 CONSIDERAÇÕES FINAIS

No presente capítulo, realizamos um estudo bibliográfico sobre o que está ocorrendo no contexto de novas distribuições e quais e como são tratadas. Damos maior ênfase a forma de geração proposta por Alzaatreh *et al.* (2013), mostrando classes que foram propostas e classes que podem ser obtidas a partir dessa ideia. Além disso, mostramos a estimação via máxima verossimilhança que é o método de inferência dos parâmetros que iremos utilizar. Para verificação das propriedades inferenciais do modelo, mostramos o método de simulação de Monte Carlo. E para as aplicações evidenciamos quais são os critérios para seleção do melhor modelo.

3 FAMÍLIA DE DISTRIBUIÇÕES NADARAJAH-HAGHIGHI-G

RESUMO

No presente capítulo, falamos sobre a classe de distribuições Nadarajah-Haghighi- G (NH- G) que possui dois parâmetros adicionais: um parâmetro de escala e um de forma. Abordamos algumas propriedades gerais da família, tomando como base Dias (2016). No capítulo, fornecemos um pouco do que foi trabalhado pelo autor: inferência dos parâmetros via método da máxima verossimilhança, simulação de Monte Carlo para avaliação dos estimadores pontuais em distribuições propostas nessa classe, assim como aplicações a dados reais para mostrar a aplicabilidade das distribuições que são propostas nessa classe. A importância desse capítulo se faz necessário para os capítulos seguintes.

Palavras-chave: Distribuição Nadarajah-Haghighi; Máxima Verossimilhança; Simulação de Monte Carlo.

3.1 INTRODUÇÃO

Com o avanço das pesquisas com dados de sobrevivência, é esperado que se procure propor classes de distribuições que proporcionem maior flexibilidade e melhor ajuste aos dados em estudo. Uma das maneiras de se obter mais flexibilidade nos modelos é a adição de parâmetros na distribuição. Dias (2016) propôs a classe NH- G utilizando a geração de famílias proposta por Alzaatreh *et al.* (2013), conhecida por família T-X de distribuições.

Muitas famílias de distribuições são conhecidas na literatura. A família de distribuições Nadarajah Haghighi- G (NH- G) proposta por Dias (2016) é uma família obtida da distribuição proposta por Nadarajah e Haghighi (2011), que é uma alternativa para as distribuições Weibull e Gamma.

A distribuição NH, que é uma generalização da distribuição exponencial, é bastante utilizada em análise de sobrevivência, atuária, finanças, confiabilidade e biologia. Ela tem a vantagem de possuir mais formas que a sua distribuição-base

A família NH- G tem fda e fdp, dadas, respectivamente, por

$$\begin{aligned} F(x) &= \int_0^{-\log[1-G(x)]} \alpha \lambda (1 + \lambda t)^{\alpha-1} \exp[1 - (1 + \lambda t)^\alpha] dt \\ &= 1 - \exp\{1 - [1 - \lambda \log[1 - G(x)]]^\alpha\}, \end{aligned} \quad (3.1)$$

e

$$f(x) = \frac{\alpha \lambda g(x) \{1 - \lambda \log[1 - G(x)]\}^{\alpha-1} \exp\{1 - [1 - \lambda \log[1 - G(x)]]^\alpha\}}{1 - G(x)}, \quad x > 0; \alpha, \lambda > 0. \quad (3.2)$$

A família NH- G tem parâmetros adicionais α (de forma) e λ (de escala), ambos positivos e não nulos. Essa família tem como casos especiais a distribuição-base, quando $\alpha = \lambda = 1$ e a família Lehmann tipo II, quando $\alpha = 1$, Lehmann *et al.* (1953). Além disso note que as Equações (3.1) e (3.2) não envolvem funções robustas, entende-se com função robusta aquela que pode ocasionar problemas em soluções matemáticas, mesmo que implementadas em softwares, o que é um ponto positivo dessa família no ponto de vista analítico.

A função taxa de falha (ftf) é importante em análise de sobrevivência, pois suas curvas podem ter interpretações práticas, como a curva em forma de U, por exemplo, que pode exemplificar o tempo de vida de uma pessoa, em que na parte inicial a probabilidade de morte é alta, diminui com o tempo até determinada idade e depois começa a crescer. A ftf para essa classe de distribuições é, portanto, dada por

$$h(x) = \frac{\alpha \lambda g(x) \{1 - \lambda \log[1 - G(x)]\}^{\alpha-1}}{1 - G(x)}.$$

Para a geração de números aleatórios das distribuições dessa classe, consideramos o método da inversão, em que é utilizada a função quantílica da variável aleatória NH- G . Essa função é obtida invertendo a função de distribuição acumulada. A função quantílica dessa classe é dada por

$$X = G^{-1} \left\{ 1 - \exp \left[\frac{1 - [1 - \log(1 - u)]^{1/\alpha}}{\lambda} \right] \right\},$$

em que G^{-1} é a inversa da fda da distribuição de base G .

Nas próximas seções serão apresentadas propriedades matemáticas obtidas por Dias (2016) para a NH- G .

3.2 PROPRIEDADES DA FAMÍLIA NH- G

Algumas propriedades da família de distribuições NH- G , serão explicitadas a seguir, como expansão para a densidade e acumulada, expressões para momentos e função geradora de momentos.

3.2.1 Expansões em Série de Potência

Usando a série de potência para a função exponencial e a expansão binomial generalizada, obtemos a função densidade da NH-G que pode ser expressa como uma combinação linear infinita funções de densidade da exponencial-G (exp-G). Então, se $X \sim \text{NH-G}$, a fdp de X pode ser expressa como

$$f(x) = \sum_{m=0}^{\infty} w_m h_{m+1}(x; \xi), \quad (3.3)$$

em que

$$w_m = \sum_{i,j=0}^{\infty} \sum_{l=0}^{\infty} \frac{e \alpha \lambda^{j+1} (-1)^{i+l+m}}{(m+1) i!} \binom{l-1}{m} \binom{\alpha(i+1)-1}{j} \left[\binom{j}{l} + \sum_{k=0}^{\infty} p_k(j) \binom{k+j+1}{l} \right], \quad (3.4)$$

$p_0(c) = c/2$, $p_1(c) = c(3c+5)/24$, $p_2(c) = c(c^2+5c+6)/48$, $p_3(c) = c(15c^3+150c^2+485c+502)/5760$, etc, e $h_{m+1} = (m+1)g(x)G(x)^m$ é a função densidade da exp-G com parâmetro de potencia m .

Consequentemente, a fda de X é dada por

$$F(x) = \sum_{m=0}^{\infty} w_m H_{m+1}(x) \quad (3.5)$$

em que $H_{m+1}(x) = G(x)^{m+1}$ é a densidade da exp-G com parâmetro de potência $m+1$.

3.2.2 Momentos

Sejam $Y_m \sim \text{exp-G}(m, \xi)$ e $h_m(x)$ a fdp de Y_m . A expressão para o n -ésimo momento de X pode ser obtido de (3.3) e usando o teorema da convergência monótona. Obtemos então

$$E(X^n) = \sum_{m=0}^{\infty} w_m E(Y_{m+1}^n). \quad (3.6)$$

As expressões em forma fechada para os momentos das exponencializadas foram obtidas por Nadarajah e Kotz (2006) e, com base nessas propriedades, obtemos as propriedades da NH-G.

3.2.3 Função geradora de momentos

Seja $M_X(t) = E(e^{tX})$ a função geradora de momentos (fgm) de X . Usando a Equação (3.3) e o teorema da convergência monótona podemos escrever

$$M_X(t) = \sum_{m=0}^{\infty} w_m M_{m+1}(t), \quad (3.7)$$

em que $M_{m+1}(t)$ é a fgm de Y_{m+1} . Consequentemente, $M_X(t)$ pode ser determinada tomando por base a fgm da distribuição exp- G .

Alternativamente, podemos escrever $M_X(t)$ de (3.7) como segue

$$M_X(t) = \sum_{m=0}^{\infty} (m+1)w_m\rho(t, m), \quad (3.8)$$

em que

$$\rho(t, m) = \int_{-\infty}^{\infty} e^{tx} G(x)^m g(x) dx = \int_0^1 \exp[tQ_{G(u)}] u^m du \text{ e}$$

$Q_{G(\cdot)}$ denota a função quantílica da distribuição de base.

3.3 ESTIMAÇÃO POR MÁXIMA VEROSSIMILHANÇA

Nesta seção, detalhamos os estimadores de máxima verossimilhança dos parâmetros da família de distribuições de amostras completas. Seja $x_1, \dots, x_n$ observações obtidas da NH- G com parâmetros α, λ e ξ . Seja $\theta = (\alpha; \lambda; \xi)^\top$, sendo $p \times 1$ a dimensão do vetor de parâmetros. A função completa de log-verossimilhança para θ é dada por

$$\begin{aligned} l(\theta) = & n + n \log(\alpha) + n \log(\lambda) + \sum_{i=1}^n \log[g(x_i; \xi)] - \sum_{i=1}^n \log[1 - G(x_i; \xi)] \\ & + (\alpha - 1) \sum_{i=1}^n \log\{1 - \lambda[1 - G(x_i; \xi)]\} - \sum_{i=1}^n \log\{1 - \lambda \log[1 - G(x_i; \xi)]\}^\alpha. \end{aligned} \quad (3.9)$$

A função de log-verossimilhança pode ser maximizada diretamente usando a função *Optim* no R, com algum método iterativo, como o BFGS, e suas derivadas são calculadas numericamente. Podemos também obter o vetor escore onde cada elemento desse vetor é obtido derivando a Equação (3.9) em relação aos seus parâmetros. Esse vetor é necessário para utilização em algoritmos de otimização como Newton-Raphson ou para o uso no BFGS de derivadas analíticas. Os componentes da função escore $U_n(\theta) = (U_\alpha = \partial l_n / \partial \alpha; U_\lambda = \partial l_n / \partial \lambda; U_\xi l_n / \partial \xi)^\top$

são respectivamente:

$$\begin{aligned}
U_{\alpha} &= \frac{n}{\alpha} + \sum_{i=1}^n \log\{1 - \lambda \log[1 - G(x_i; \boldsymbol{\xi})]\} \\
&\quad - \sum_{i=1}^n \{1 - \lambda \log[1 - G(x_i; \boldsymbol{\xi})]\}^{\alpha} \log\{1 - \lambda \log[1 - G(x_i; \boldsymbol{\xi})]\}, \\
U_{\lambda} &= \frac{n}{\lambda} - (\alpha - 1) \sum_{i=1}^n \frac{\log[1 - G(x_i; \boldsymbol{\xi})]}{1 - \lambda \log[1 - G(x_i; \boldsymbol{\xi})]} \\
&\quad + \alpha \sum_{i=1}^n \{1 - \lambda \log[1 - G(x_i; \boldsymbol{\xi})]\}^{\alpha-1} \log[1 - G(x_i; \boldsymbol{\xi})] \text{ e} \\
U_{\boldsymbol{\xi}} &= \sum_{i=1}^n \frac{\dot{g}(x_i; \boldsymbol{\xi})}{g(x_i; \boldsymbol{\xi})} + \lambda(\alpha - 1) \sum_{i=1}^n \frac{\dot{G}(x_i; \boldsymbol{\xi})[1 - G(x_i; \boldsymbol{\xi})]^{-1}}{\{1 - \lambda \log[1 - G(x_i; \boldsymbol{\xi})]\}} \\
&\quad - \alpha\lambda \sum_{i=1}^n \frac{\{1 - \lambda \log[1 - G(x_i; \boldsymbol{\xi})]\}^{\alpha-1} \dot{G}(x_i; \boldsymbol{\xi})}{1 - G(x_i; \boldsymbol{\xi})} + \sum_{i=1}^n \frac{\dot{G}(x_i; \boldsymbol{\xi})}{1 - G(x_i; \boldsymbol{\xi})},
\end{aligned}$$

em que $\dot{g}(x_i; \boldsymbol{\xi}) = \partial g(x_i; \boldsymbol{\xi}) / \partial \boldsymbol{\xi}_k$ e $\dot{G}(x_i; \boldsymbol{\xi}) = \partial G(x_i; \boldsymbol{\xi}) / \partial \boldsymbol{\xi}_k$ para $k = 1, \dots, p$.

Podemos notar que, para a estimação intervalar dos parâmetros do modelo, precisamos encontrar a matriz de informação observada

$$J_n(\boldsymbol{\theta}) = \begin{pmatrix} U_{\alpha\alpha} & U_{\alpha\lambda} & | & U_{\alpha\boldsymbol{\xi}}^{\top} \\ U_{\lambda\alpha} & U_{\lambda\lambda} & | & U_{\lambda\boldsymbol{\xi}}^{\top} \\ \text{---} & \text{---} & | & \text{---} \\ U_{\boldsymbol{\xi}\alpha} & U_{\boldsymbol{\xi}\lambda} & | & U_{\boldsymbol{\xi}\boldsymbol{\xi}} \end{pmatrix}.$$

Os seus elementos podem ser computados numericamente em softwares como o *Maple* (GARVAN, 2001) ou *MATHEMATICA* (WOLFRAM, 1999). Esses softwares de computação simbólica tem uma boa habilidade em lidar com expressões analíticas de considerável tamanho e complexidade.

3.4 SIMULAÇÃO

Nesta seção, detalhamos como ocorreu a simulação feita por Dias (2016) para uma distribuição na classe NH-G. Foi considerado o modelo Nadarajah Haghighi-Exponencial Exponencializada (NH-EE), distribuição com quatro parâmetros α, λ, β e δ . Sua fdp é dada por

$$f(x) = \frac{\alpha\lambda\beta\delta \exp^{-\delta x} (1 - \exp^{-\delta x})^{\beta-1} \exp\{1 - [1 - \lambda \log[1 - (1 - \exp^{-\delta x})^{\beta}]]^{\alpha}\}}{[1 - (1 - \exp^{-\delta x})^{\beta}]\{1 - \lambda \log[1 - (1 - \exp^{-\delta x})^{\beta}]\}^{1-\alpha}}. \quad (3.10)$$

Os resultados foram obtidos através de 5000 réplicas de Monte Carlo utilizando a linguagem de programação R. Para cada réplica, foi gerada uma amostra de tamanho n da distribuição NH-EE($\alpha, \lambda, \beta, \delta$). Para a maximização da função de log-verossimilhança total, $l(\boldsymbol{\theta})$, foi utilizado o

BFGS, como método numérico. Para a geração de números aleatórios, foi utilizado o método da inversão. Os parâmetros verdadeiros considerados pelo autor foram $\alpha = 1.5$, $\lambda = 2$, $\beta = 0.5$ e $\delta = 2.5$. Os tamanhos de amostra foram de $n = 25, 50, 100, 200$ e 400 . Foram analisados como quantidades o viés de cada média de estimativa e o erro quadrático médio (EQM). Na análise, se verificou que ao aumentar o tamanho da amostra, o viés e o EQM tendiam a zero, como esperado. Para mais detalhes, ver Dias (2016).

3.5 APLICAÇÃO

Para mostrar a aplicabilidade empírica, Dias (2016) escolheu dois bancos de dados de fadiga. O primeiro conjunto de dados é sobre o tempo de ruptura de fibras de vidro de 1.5cm expostas a uma força e foi obtido por trabalhadores do Laboratório Nacional de Física do Reino Unido. A proposta de modelos que consigam se ajustar bem a esse banco de dados pode trazer melhoria para o setor de análise, podendo estimar melhor o estresse até ruptura desse material. O modelo considerado para esse conjunto de dados foi o NH-Gompertz (NH-Go), que possui densidade dada por

$$f(x) = \alpha \lambda \theta e^{\gamma x} \left[1 + \frac{\lambda \theta (e^{\gamma x} - 1)}{\gamma} \right]^{\alpha-1} \exp \left\{ 1 - \left[1 + \frac{\lambda \theta (e^{\gamma x} - 1)}{\gamma} \right]^{\alpha} \right\}.$$

Para concorrer com esse modelo, foram utilizadas algumas distribuições já propostas na literatura, que são as distribuições

- Beta Weibull (BW) Famoye *et al.* (2005);
- Kumaraswamy Weibull (KW) Cordeiro *et al.* (2010);
- Beta Gompertz (BGo) Jafari *et al.* (2014).

Na comparação, foram utilizados as estatísticas de testes modificada *Anderson Darlin* (A^*) e *Crammer Von misses* (W^*), Chen e Balakrishnan (1995). O modelo NH-Go teve as menores quantidades nessas duas estatísticas de testes, mostrando, assim, a superioridade dessa classe com essa distribuição para esses dados de fadiga. Para mais detalhes sobre essas quantidades e sobre a aplicação, ver Dias (2016).

A segunda aplicação a dados reais também é sobre dados de fadiga. O conjunto de dados diz respeito à fibras de carbono expostas a uma força até sua ruptura. Para mais detalhes sobre esses dados, ver Nichols e Padgett (2006). A distribuição escolhida para esse banco de dados foi a NH-Fréchet (NH-Fr), que tem fdp dada por

$$f(x) = \frac{\alpha \lambda \delta \sigma^{\delta} e^{-\left(\frac{\sigma}{x}\right)^{\delta}} x^{-(\delta+1)} \{1 - \lambda \log[1 - e^{-\left(\frac{\sigma}{x}\right)^{\delta}}]\}^{\alpha-1} \exp\{1 - [1 - \lambda \log[1 - e^{-\left(\frac{\sigma}{x}\right)^{\delta}}]]^{\alpha}\}}{1 - e^{-\left(\frac{\sigma}{x}\right)^{\delta}}}.$$

Para concorrer com esse modelo, o autor considerou as distribuições

- Beta Fréchet (BFr) (NADARAJAH; GUPTA, 2004);
- Kumaraswamy Fréchet (KFr) (MEAD, 2014);
- Fréchet (Fr) (FRECHET, 1927).

O resultado dessa aplicação revelou que a distribuição NH-Fr se ajustou melhor a esse banco de dados, mostrando assim sua boa aplicabilidade em dados de fadiga.

Dessa forma, analisando a boa aplicabilidade dessa nova classe de distribuições juntamente com o fato de sua simplicidade nas expressões da fda e fdp, julgamos necessário o estudo de novas distribuições pertencentes a essa família, bem como a verificação de adequabilidade a outros tipos de conjuntos de dados.

3.6 CONSIDERAÇÕES FINAIS

No presente capítulo detalhamos a classe Nadarajah-Haghighi G proposta por Dias (2016), no capítulo mostramos as principais propriedades que são utilizadas para a geração de novos modelos, fornecemos as propriedades que são encontradas na tese do autor, como a expressão para a geração da fda e fdp, função quantílica, expansões úteis que nos ajudam a calcular os momentos, função geradora de momentos. Para o processo de inferência via máxima verossimilhança, a expressão geral que pode ser utilizada para qualquer distribuição-base também é fornecida, assim como as componentes do vetor escore. Mostramos também como ocorreram os estudos de simulação e aplicação em modelos que são da classe, para assim motivar os possíveis leitores que se interessem por trabalhos futuros.

4 DISTRIBUIÇÃO NADARAJAH-HAGHIGHI EXPONENCIAL EXPONENCIALIZADA: PROPRIEDADES E APLICAÇÕES

RESUMO

Nesse capítulo, introduzimos uma extensão da distribuição exponencial exponencializada com base na família trabalhada por Dias (2016). Para essa nova distribuição, encontramos expressões para as funções taxa de falha, quantílica e de sobrevivência. Fornecemos expressões para as principais propriedades estruturais do modelo, como momentos, função geradora de momentos, entropia de Shannon e estatísticas de ordem, por exemplo. Realizamos inferência pontual utilizando o método da máxima verossimilhança. Para ilustração empírica, realizamos duas aplicações a dados reais considerando um conjunto de dados clínicos e outro de fadiga de equipamentos, tendo por objetivo mostrar a aplicabilidade do novo modelo.

Palavras-chave: Distribuição Nadarajah-Haghighi; Distribuição Exponencial Exponencializada; Família Nadarajah-Haghighi-G; Estimação via máxima verossimilhança; Método de Monte Carlo.

4.1 INTRODUÇÃO

A distribuição exponencial exponencializada (EE) é uma distribuição biparamétrica contínua com suporte em $\mathbb{R}^+$ e constitui uma extensão da distribuição exponencial (E). Esta última é caracterizada por ter uma função taxa de falha constante e é considerada como uma distribuição que possui expressões matemáticas mais simples, sendo utilizada, no entanto, em vários problemas teóricos e aplicados, como em análise de sobrevivência, para a observação de tempo de falha de materiais. É também uma das primeiras variáveis aleatórias que são trabalhadas em modelos de sobrevivência. No nosso caso, a base da distribuição proposta é obtida usando o gerador exponencial, obtendo, assim, a distribuição EE, que tem função densidade nas formas decrescente e unimodal.

Seja X uma variável aleatória que segue uma distribuição exponencial exponencializada com parâmetros a e b , $X \sim EE(a, b)$. Então sua função de distribuição acumulada (fda) é dada por

$$G(x; \beta) = [1 - \exp(-ax)]^b, \quad (4.1)$$

em que ambos os parâmetros são positivos e de forma. A função densidade de probabilidade (fdp) é dada por

$$g(x; \beta) = ab \exp(-ax) [1 - \exp(-ax)]^{b-1}. \quad (4.2)$$

Recentemente uma nova família de distribuições, chamada Nadarajah-Haghighi- G , proposta por Dias (2016) foi gerada a partir da variável aleatória Nadarajah-Haghighi (NH). Para diferentes distribuições de base $G(x)$, $x > 0$, o autor define a classe Nadarajah-Haghighi- G (NH- G), com fda e fdp, respectivamente, dadas por

$$F(x) = 1 - \exp \{1 - [1 - \lambda \log [1 - G(x)]]^\alpha\} \quad (4.3)$$

e

$$f(x) = \frac{\alpha \lambda g(x) \{1 - \lambda \log [1 - G(x)]\}^{\alpha-1} \exp \{1 - [1 - \lambda \log [1 - G(x)]]^\alpha\}}{1 - G(x)}. \quad (4.4)$$

Note que (4.4) não envolve nenhuma função especial, como o caso da gama incompleta e beta incompleta. Além disso, se $\alpha = 1$ em 4.3, obtemos a família Lehmann tipo II- G , e se $\alpha = \lambda = 1$ voltamos para a distribuição-base.

A função taxa de falha (ftf), digamos $h(\cdot)$, é uma importante função em análise de sobrevivência. Seja $X \sim \text{NH-}G(\lambda, \alpha, \boldsymbol{\theta})$ em que $\boldsymbol{\theta}$ é o vetor de parâmetros da distribuição-base. A função taxa de falha de X é dada por

$$h(x) = \frac{\alpha \lambda g(x) \{1 - \lambda \log [1 - G(x)]\}^{\alpha-1}}{1 - G(x)}. \quad (4.5)$$

Usando a ideia de Qian (2012), é possível fornecer intervalos paramétricos com as formas correspondentes da função taxa de falha. Primeiramente, considere o logaritmo da Equação (4.5). Assim, obtemos

$$\log[h(x)] = \log(\alpha) + \log(\lambda) + \log[g(x)] + (\alpha - 1) \log\{1 - \lambda \log[1 - G(x)]\} - \log[1 - G(x)].$$

Derivando ambos os lados da última equação, teremos:

$$\begin{aligned} \frac{h'(x)}{h(x)} &= \frac{g'(x)}{g(x)} + \frac{\lambda(\alpha - 1)g(x)}{[1 - G(x)]\{1 - \lambda \log[1 - G(x)]\}} + \frac{g(x)}{1 - G(x)} \\ \iff h'(x) &= h(x) \left\{ \frac{g'(x)}{g(x)} + \frac{\lambda(\alpha - 1)g(x)}{[1 - G(x)]\{1 - \lambda \log[1 - G(x)]\}} + \frac{g(x)}{1 - G(x)} \right\}, \\ \iff h'(x) &= h(x)s(x), \end{aligned}$$

em que $g'(x) = \frac{\partial g(x)}{\partial x}$. Então, o sinal de $h'(x)$ é o mesmo sinal de $s(x)$, já que $h(x) > 0$.

4.2 DISTRIBUIÇÃO NADARAJAH-HAGHIGHI EXPONENCIAL-EXPONENCIALIZADA

Substituindo as Equações (4.1) e (4.2) nas Equações (4.3) e (4.4) obteremos as expressões para a fda e fdp da NH-EE, como a seguir. Seja $\gamma = \lambda\beta$, então

$$F(x; \alpha, \lambda, a, b) = 1 - \exp \left\{ 1 - \left[1 - \lambda \log \left[1 - [1 - \exp(-ax)]^b \right] \right]^\alpha \right\} \quad (4.6)$$

e

$$f(x; \alpha, \lambda, a, b) = \frac{\alpha \lambda a b \exp(-ax) [1 - \exp(-ax)]^{b-1} \exp \left\{ 1 - \left[1 - \lambda \log \left[1 - [1 - \exp(-ax)]^b \right] \right]^\alpha \right\}}{\left\{ 1 - \lambda \log \left[1 - [1 - \exp(-ax)]^b \right] \right\}^{1-\alpha} \left\{ 1 - [1 - \exp(-ax)]^b \right\}} \quad (4.7)$$

respectivamente, em que α é parâmetro de forma e é positivo, λ é parâmetro de escala e a e b são ambos de forma e positivos. Evidentemente, a densidade dada em (4.7) não envolve funções mais robustas analiticamente. Note ainda que se $b = 1$ as Equações (4.6) e (4.7) correspondem as funções densidade e acumulada, respectivamente, de uma variável aleatória que segue distribuição NH. Além disso, se $\alpha = 1$, obtemos a distribuição Lehmann tipo II exponencial exponencializada (LII-EE), se $\alpha = b = 1$, temos a distribuição Lehmann tipo II exponencial (LII-E) e se $\alpha = \lambda = 1$, obteremos a distribuição EE. Ou seja, a distribuição NH-EE recai em quatro outras distribuições.

Além disso, se $X \sim \text{NH-EE}(\alpha, \lambda, a, b)$, sua função taxa de falha (ftf) é dada por

$$h(x; \alpha, \lambda, a, b) = \frac{\alpha \lambda a b \exp(-ax) [1 - \exp(-ax)]^{b-1}}{\left\{ 1 - \lambda \log \left[1 - [1 - \exp(-ax)]^b \right] \right\}^{1-\alpha} \left\{ 1 - [1 - \exp(-ax)]^b \right\}}. \quad (4.8)$$

Portanto, se a distribuição de base for a exponencial exponencializada, seguindo a ideia de Qian (2012), da Equação (4.8) obtemos

$$\log[h(x)] = \log \left(\frac{\alpha \lambda a b \exp(-ax) [1 - \exp(-ax)]^{b-1}}{\left\{ 1 - \lambda \log \left[1 - [1 - \exp(-ax)]^b \right] \right\}^{1-\alpha} \left\{ 1 - [1 - \exp(-ax)]^b \right\}} \right)$$

Derivando ambos os lados da última equação, com respeito a x , obtemos

$$\begin{aligned} \frac{h'(x)}{h(x)} &= - \frac{a(b - (\alpha - 1)\lambda K + 2K - 1) + \lambda(-2bK + e^{ax}(K - 1) + b) \log(1 - K) - e^{ax}(K - 1)}{(e^{ax} - 1)(K - 1)(\lambda \log(1 - K) - 1)} \\ \Leftrightarrow h'(x) &= h(x)s(x), \end{aligned}$$

em que $K = (1 - e^{-ax})^b$. Consequentemente, o sinal de $h'(x)$ é o mesmo sinal de $s(x)$ com $h(x) > 0$.

Gráficos da densidade e da função taxa de falha da distribuição proposta podem ser vistos na Figura 2. Note que temos as formas crescente, decrescente e unimodal para a fdp e para ftf temos as formas, constante, crescente e decrescente. Dessa forma, concluímos que a distribuição NH-EE é mais flexível que a distribuição EE.

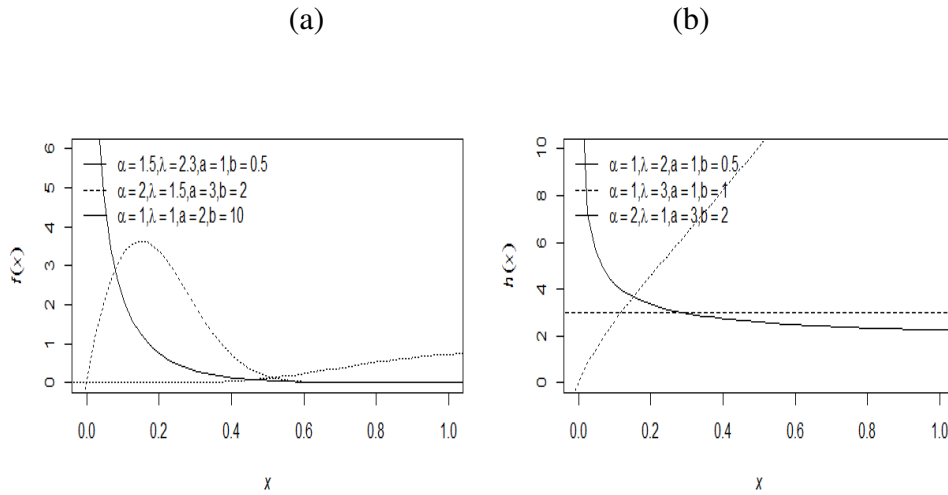


Figura 2 – Gráficos para a densidade e a função taxa de falha para a distribuição NH-EE para diferentes valores de parâmetros.

Denotamos por $X \sim \text{NH-EE}(\alpha, \lambda, a, b)$ uma variável aleatória que possui densidade (4.7). Para realizar o estudo de simulação geramos variáveis aleatórias $u \sim U(0, 1)$ e depois substituímos na função quantílica. Assim, geramos números da variável aleatória X , fazendo:

$$X = \frac{-1}{a} \log \left\{ 1 - \left\{ 1 - \exp \left[\frac{1 - [1 - \log(1 - u)]^{1/\alpha}}{\lambda} \right] \right\}^{1/b} \right\}. \quad (4.9)$$

Da Equação (4.9), tomando $u = 0.5$ obtemos a mediana da distribuição NH-EE, dada por:

$$q = \frac{-1 + (1.6931472)^{1/\alpha}}{\gamma}.$$

4.3 EXPANSÕES UTILIZADAS

Algumas expansões são úteis para essa distribuição. Dias (2016) mostrou que a NH-G tem propriedades derivadas das "Exponencializadas-G" (exp-G). Com isso, podemos escrever as Equações (4.6) e (4.7) como uma combinação de fda's e fdp's de exp-EE's, respectivamente, e, com isso, obter propriedades estruturais da distribuição NH-EE.

De Dias (2016), podemos escrever (4.7) como

$$f(x) = \sum_{m=0}^{\infty} w_m h_{m+1}(x), \quad (4.10)$$

em que $h_{m+1}(x) = (m+1)g(x)G(x)^m$ é a fdp da exp-EE, g é a fdp da exponencial exponencializada e G é a sua fda. Para mais detalhes sobre a Equação (4.10) e o seus coeficientes w'_m s, ver Dias (2016). Baseada em (4.10), várias propriedades estruturais da NH-EE, podem ser obtidas a partir das propriedades da exp-EE.

Integrando a Equação (4.10) e o usando o teorema da convergência monótona, a fda de X é expressa por

$$F(x) = \sum_{m=0}^{\infty} w_m H_{m+1}(x), \quad (4.11)$$

em que $H_{m+1}(x)$ é a acumulada de Y_{m+1} .

4.4 ALGUMAS PROPRIEDADES

Nessa seção, traremos algumas propriedades obtidas do modelo proposto.

4.4.1 Momentos

Seja $Y_m \sim \text{exp-EE}(m, \beta)$. O n -ésimo momento de X pode ser expresso, usando a expansão (4.10), por

$$E(X^n) = \sum_{m=0}^{\infty} w_m E(Y_{m+1}^n),$$

em que $E(Y_{m+1}^n)$ é o n -ésimo momento de Y [ver Nadarajah (2011)]. Média e variância de X , com $\beta = 1$, são dados por

$$E(X) = \sum_{m=0}^{\infty} w_m [\psi(m+1) - \psi(1)] \text{ e } V(X) = \sum_{m=0}^{\infty} w_m [\psi'(1) - \psi'(m+1)],$$

em que $\psi(\cdot)$ e suas derivadas são as funções digama e poligama, respectivamente.

4.4.2 Função geradora de momentos

Com base na Equação (4.10), denotamos a função geradora de momentos (fgm) de X , $M_X(t)$, por

$$M_X(t) = \sum_{m=0}^{\infty} w_m M_{m+1}(t),$$

em que $M_{m+1}(t)$ é a função geradora de momentos de Y_{m+1} . Consequentemente, $M_X(t)$ pode ser obtida a partir da função geradora de momentos da distribuição exp-E com parâmetro adicional

$m + 1$. Se $\beta = 1$, de Gupta e Kundu (1999), a fgm de X é dada por

$$M_X(t) = \Gamma(1-t) \sum_{m=0}^{\infty} w_m \frac{\Gamma(m+2)}{\Gamma(m+2-t)}.$$

Além disso, a fgm de X , com $\alpha = 1$ e $|t| < 1$, coincide com a fgm de uma distribuição exponencial com parâmetro γ . Todas as outras propriedades para $\alpha = 1$ já são conhecidas na literatura.

4.4.3 Entropia

Para medir o grau de variação ou de incerteza da variável aleatória X , usaremos uma ferramenta da Teoria da informação, conhecida como entropia de Shannon. A entropia de Shannon de uma variável aleatória X é definida por

$$\begin{aligned} E\{-\log[f(x)]\} &= -\log(\alpha) - \log(\lambda) - \log(a) - \log(b) - aE(X) - (b-1)E[1 - \exp(-aX)] \\ &\quad - 1 + \alpha E\left\{\log\left[1 - \lambda \log\left[1 - [1 - \exp(-aX)]^b\right]\right]\right\} \\ &\quad - (\alpha-1)E\left\{\log\left\{1 - \lambda \log\left[1 - [1 - \exp(-aX)]^b\right]\right\}\right\} \\ &\quad + E\left\{\log\left\{1 - [1 - \exp(-aX)]^b\right\}\right\}. \end{aligned}$$

Se $\alpha = b = 1$, temos, após algumas manipulações algébricas,

$$E\{-\log[f(x)]\} = -\log(\lambda) - \log(a) - 1 + E[\log(1 + a\lambda X)]$$

4.4.4 Estatísticas de ordem

Estatísticas de ordem são muito importantes, tanto para o uso em problemas práticos como em problemas teóricos. Podemos exemplificar seu uso, relacionando a problemas que visam encontrar valores discrepantes, problemas de engenharia e de controle de qualidade, assim como em problemas de estimação. Suponha que $X_1, \dots, X_n$ seja uma amostra aleatória de variáveis aleatórias que seguem a distribuição NH-EE. Denote por $X_{1:n}, \dots, X_{n:n}$ as estatísticas de ordem. A fdp da i -ésima estatística de ordem de uma amostra aleatória $X_1, \dots, X_n$ da distribuição NH-EE é dada por

$$f_{i:n}(x) = C f(x) F^{i-1}(x) [1 - F(x)]^n - i,$$

em que $C = n! / [(i-1)!(n-i)!]$, $F(x)$ e $f(x)$ são a acumulada (4.6) e a densidade (4.7) da NH-EE. Então,

$$f_{i:n}(x) = K \sum_{j=0}^{n-1} (-1)^j \binom{n-1}{j} f(x) F^{i+j-1}(x).$$

A função densidade de probabilidade $X_{i:n}$ pode ser expressada de (4.10) e (4.11) como

$$f_{i:n}(x) = K \sum_{j=0}^{n-1} (-1)^j \binom{n-1}{j} \left\{ \sum_{r=0}^{\infty} w_r (r+1) [1 - \exp(-\beta x)]^r [\beta \exp(-\beta x)] \right\} \\ \times \left\{ \sum_{m=0}^{\infty} w_m [1 - \exp(-\beta x)]^{m+1} \right\}^{i+j-1}.$$

Usando séries de potência para um inteiro positivo, temos

$$\left[\sum_{r=0}^{\infty} w_m [1 - \exp(-\beta x)]^{m+1} \right]^{i+j-1} = \sum_{m=0}^{\infty} f_{i+j-1,m} [1 - \exp(-\beta x)]^{i+j+m-1},$$

em que $f_{i+j-1,0} = (w_0)^{i+j-1}$ e (para $m \geq 1$)

$$f_{i+j-1,m} = (mw_0)^{-1} \sum_{p=1}^m [p(i+j) - m] w_p f_{i+j-1,m-p}.$$

Consequentemente,

$$f_{i:n}(x) = \sum_{j=0}^{n-1} \sum_{r,m=0}^{\infty} c_{j,r,m} h_{i+j+r+m}(x), \quad (4.12)$$

em que

$$c_{j,r,m} = \frac{(-1)^j n!}{(i-1)!(n-1-j)!j!} \frac{(r+1)w_r f_{i+j-1,m}}{(i+j+r+m)}.$$

A Equação (4.12) é o principal resultado dessa seção e nos mostra que a densidade das estatísticas de ordem é escrita como uma tripla combinação linear de densidades da distribuição exp-EE. Portanto, podemos observar que algumas propriedades matemáticas (estatísticas de ordem, fgm, entre outros) podem ser obtidas através das respectivas propriedades da exp-EE.

4.5 INFERÊNCIA E ESTIMAÇÃO

Aqui, discutimos o uso do método de máxima verossimilhança para inferência e estimação pontual dos parâmetros do modelo proposto.

4.5.1 Estimação via máxima verossimilhança

Para a estimação dos parâmetros da distribuição NH-EE, utilizamos o método da máxima verossimilhança. Seja $x_1, \dots, x_n$ uma amostra aleatória de tamanho n da distribuição

NH-EE(α, λ, a, b). O logaritmo da função de verossimilhança para o vetor de parâmetros $\boldsymbol{\theta} = (\alpha, \lambda, a, b)^\top$, é expressado por

$$\begin{aligned} l(\boldsymbol{\theta}) &= n + n \log(\alpha) + n \log(\lambda) + n \log(a) + n \log(b) - a \sum_{i=1}^n x_i + (b-1) \sum_{i=1}^n \log[1 - \exp(-ax_i)] \\ &+ (\alpha-1) \sum_{i=1}^n \log \left\{ 1 - \lambda \log \left[1 - [1 - \exp(-ax_i)]^b \right] \right\} - \sum_{i=1}^n \log \left\{ 1 - [1 - \exp(-ax_i)]^b \right\} \\ &- \sum_{i=1}^n \left\{ 1 - \lambda \log \left[1 - [1 - \exp(-ax_i)]^b \right] \right\}^\alpha. \end{aligned} \quad (4.13)$$

A maximização de (4.13) é realizada utilizando rotinas estabelecidas como *nlm* ou *optimize* em um pacote estatístico do R. Derivando (4.13) em relação aos parâmetros, igualando a zero e resolvendo as equações simultaneamente obtemos os estimadores de máxima verossimilhança (EMV), $\hat{\boldsymbol{\theta}}$, de $\boldsymbol{\theta}$.

Essas equações não possuem forma fechada, então é necessário o uso de técnicas iterativas, como BFGS ou Newton-Raphson, e com o apoio de um software estatístico obtemos a solução dessas equações não lineares, que são as estimativas de máxima verossimilhança.

O vetor score tem seus componentes dados por

$$U_\lambda = - \frac{\alpha \lambda \log(1-K) (1 - \lambda \log(1-K))^\alpha - \alpha \lambda \log(1-K) + 1}{\lambda (\lambda \log(1-K) - 1)},$$

$$U_\alpha = \frac{1 - \alpha \log(1 - \lambda \log(1-K)) ((1 - \lambda \log(1-K))^\alpha - 1)}{\alpha},$$

e

$$\begin{aligned} U_a &= \frac{a \alpha b \lambda x K - a \alpha b \lambda x K (1 - \lambda \log(1-K))^\alpha - a b \lambda x K}{a (e^{ax} - 1) (K - 1) (\lambda \log(1-K) - 1)} \\ &- \frac{\lambda (-2abxK + K + abx + e^{ax}(ax-1)(K-1) - 1) \log(1-K)}{a (e^{ax} - 1) (K - 1) (\lambda \log(1-K) - 1)} \\ &+ \frac{K - 2abxK + abx + e^{ax}(ax-1)(K-1) - 1}{a (e^{ax} - 1) (K - 1) (\lambda \log(1-K) - 1)}, \end{aligned}$$

e

$$U_b = \frac{1}{b} - \frac{\log(1 - e^{-ax}) (\lambda K (\alpha ((1 - \lambda \log(1-K))^\alpha - 1) + 1) + (\lambda - 2\lambda K) \log(1-K) + 2K - 1)}{(K-1) (\lambda \log(1-K) - 1)},$$

em que $K = (1 - e^{-ax})^b$

4.6 APLICAÇÕES A DADOS REAIS

Realizamos aqui duas aplicações a dados reais, visando mostrar a flexibilidade do modelo proposto. Comparamos o modelo NH-EE com outros modelos conhecidos na literatura utilizando as estatísticas de teste. De fato, comparamos os ajustes do modelo NH-EE com oito outros modelos; a saber:

1. Exponencial (E),

2. Exponencial Exponencializada,(EE) (GUPTA; KUNDU, 2001b)
3. Nadarajah-Haghighi Exponencializada (ENH) (LEMONTE, 2013)
4. Burr XII Exponencializada, (Exp-BurrXII) (AL-HUSSAINI; HUSSEIN, 2011)
5. Weibull Exponencializada, (Exp-W) (MUDHOLKAR; SRIVASTAVA, 1993)
6. Gama Exponencializada, (Exp-Gama) (RISTIĆ; BALAKRISHNAN, 2012)
7. Gama exponencial (Gama-E) (SURIYA; JAN, 2015)
8. Gama-gama, (Gama-G), ver gerador (ZOGRAFOS; BALAKRISHNAN, 2009) .

Os EMV's dos parâmetros do modelo são calculados para avaliação. Essas estimativas foram obtidas a partir do método de iteração SANN. As estatísticas de teste Cramér-von Mises(W^*) e Anderson-Darling(A^*), definidas por Chen e Balakrishnan (1995), foram obtidas para comparar os modelos ajustados. Para essa avaliação, quanto menor os valores dessas estatísticas, melhor o ajuste aos dados. Todos os cálculos foram realizados usando o software estatístico R junto com o pacote *Adequacy Model*, Marinho (2016). Ambos conjuntos de dados foram maximizados utilizando o método numérico SANN.

4.6.1 Conjunto de dados 1: tempo até a ocorrência de alguma doença cardiovascular em 68 pessoas.

Doenças cardiovasculares (DCV) são consideradas como uma classe de doenças que afetam os vasos sanguíneos ou o coração. Entre as DCV's temos o acidente cardiovascular cerebral (AVC) e doenças arteriais coronárias (DAC). Essas doenças estão associadas à tabagismo, etilismo, sedentarismo e maus hábitos alimentares, hipertensão arterial, etc. Um estudo mais detalhado, sobre essas doenças, é encontrado em Mendis *et al.* (2011).

Estudos envolvendo tempo até a ocorrência dessas doenças, assim como proposta de distribuições que modelem bem esse tipo de estudo, se fazem necessários para ideias de prevenção a longo prazo. Um estudo que visa simular essas ocorrências pode ser encontrado em Lee e Wang (2003).

Suponha que os dados demográficos, pessoais, clínicos e laboratoriais sejam coletados a partir de uma entrevista e exame físico de 200 participantes em um estudo de DCV. Esses participantes, com idade entre 50 e 79 anos e livres de DCV no momento do exame de base, são seguidos por 10 anos. Durante o período de seguimento, 96 dos 200 participantes desenvolvem ou morrem de DCV. Consideramos um subconjunto dos dados com 68 participantes. O banco de dados foi retirado de Lee e Wang (2003).

O tempo de evento, T , de interesse é o tempo livre de DCV, que é definido como o tempo (em anos) desde o exame de base até a primeira vez que um participante foi diagnosticado como portador de DCV ou confirmado como morte por DCV.

Abaixo, na figura 3, plotamos o gráfico TTT, Aarset (1987), que possui forma predominantemente côncava indicando que distribuições que possuem ftf crescente conseguem se ajustar melhor a esse banco de dados. A distribuição NH-EE, por possuir forma crescente na ftf, pode modelar bem esses dados.

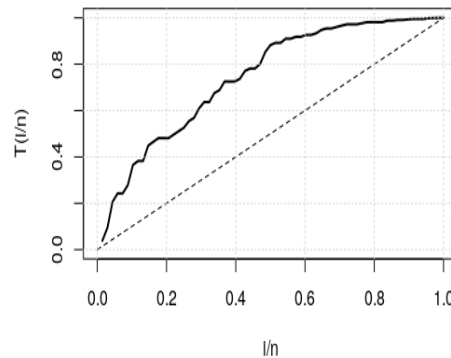


Figura 3 – TTT plot

As estatísticas descritivas para esse conjunto de dados é mostrada na Tabela 1. É possível notar que esta amostra tem uma pequena variância e assimetria e curtose negativas.

Tabela 1 – Estatísticas descritivas - Primeira Aplicação

Aplicação	Media	Mediana	Variância	Assimetria	Curtose	Min.	Max.
dados de coração	5.231	6.050	2.318791	-0.4316233	-1.125723	0.200	8.400

A tabela 2 nos fornece os valores dos EMV's dos parâmetros, bem como as quantidades para Erro Padrão (entre parênteses) a partir dos ajustes dos modelos citados e também as medidas W^* e A^* . Os resultados indicam que o modelo proposto possui os menores valores dessas estatísticas entre os modelos e, portanto, poderia ser escolhido como o melhor modelo para esses dados.

4.6.2 Conjunto de dados 2: o comportamento da fadiga de tração de um fio de urdidura e sua influência no desempenho da tecelagem.

Nas operações modernas de manufatura, requisitos rigorosos são impostos a fim de verificar as propriedades dos materiais que são utilizados. Com base nisso, se faz necessário que esses materiais sejam submetidos a testes de aquecimento intenso ou tensões mecânicas relativamente altas, por exemplo. Portanto, existe uma necessidade na proposta de métodos que permitam uma avaliação prévia na capacidade de suportar esses rigores impostos no processo de manufatura. Dentre esses processos, há o processo

Tabela 2 – Aplicação tempo até o surgimento de uma DCV

Model	α	λ	a	b	A^*	W^*
NH-EE(α, λ, a)	9.4410 (3.7363)	1.1957 (0.5780)	0.0511 (0.0182)	2.0761 (0.2812)	1.5729	0.2628
E(a)			0.1911 (0.0231)		2.7517	0.4598
EE(a, b)			3.1838 (0.5961)	0.3525 (0.0408)	2.9588	0.4960
NH(α, λ)	0.0355 (0.0141)	4.1871 (1.4975)			2.1564	0.3594
ENH(α, λ, b)	3.8725 (0.8087)	2.2953 (0.3706)		0.0499 (0.0123)	2.2042	0.3678
Exp-BurXII(α, λ, b)	10.1071 (4.3989)	0.8923 (0.1940)		1.7577 (0.5025)	5.0739	0.8795
Exp-W(α, λ, b)	7.9335 (0.2476)	9.8753 (0.3192)		0.1663 (0.0223)	2.5951	0.4113
Exp-Gamma(α, λ, b)	17.3798 (7.2648)	1.9771 (0.7279)		0.1531 (0.0672)	2.3209	0.3857
Gamma-Exp(α, λ)	3.1819 (0.5196)	0.6088 (0.1077)			2.7391	0.4578
Gamma-G(α, λ, b)	11.9571 (0.5960)	1.3267 (0.1166)		0.2264 (0.0370)	2.2959	0.3814

de tecelagem têxtil. Picciotto (1970), em sua dissertação de mestrado, propõe técnicas para modelar o quanto um fio pode suportar após tensões impostas.

A urdidura é o conjunto de fios colocados em primeiro lugar, paralelos uns aos outros, no sentido do comprimento do tear. É a base do tecido, definindo a largura e o comprimento da peça. No estudo, foi medido a tensão em ciclos até a ruptura desse fio, a fim de saber sua resistência.

Os dados são de Picciotto (1970), em sua dissertação de mestrado, e surgiram em teste no ciclo em que o fio falhou, o estudo foi feito com 100 fios. Os dados foram também utilizados por Shanker *et al.* (2015), que buscaram modelar dados de confiabilidade.

A proposta de modelos que consigam ajustar bem esses bancos de dados podem ser úteis para informação sobre o processo de tecelagem e, também, em outros bancos de dados de fadiga.

O gráfico TTT, Aarset (1987), do conjunto de dados sobre a fadiga de um fio de urdidura, é mostrado na Figura 4. Analisando o gráfico, podemos notar que temos uma curva predominantemente côncava, indicando, assim, que distribuições que possuam ftf crescente podem modelar bem esse banco de dados. A distribuição NH-EE possui essa forma na ftf.

Na Tabela 3 mostramos as estatísticas descritivas do banco de dados. Notamos que esta amostra tem uma variância alta e assimetria e curtose positivas, características opostas ao primeiro conjunto de dados aqui considerado. A tabela 4 nos fornece os valores dos EMV's dos parâmetros assim como as quantidades para Erro Padrão (entre parênteses) a partir dos ajustes dos modelos citados.

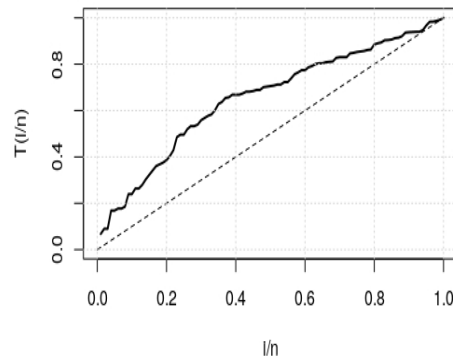


Figura 4 – TTT plot

Tabela 3 – Estatísticas descritivas - Segunda Aplicação

Aplicação	Media	Mediana	Variância	Assimetria	Curtose	Min.	Max.
Tensão no fio	222.0	195.5	144.6181	1.339683	2.7435	15.0	829.0

Também é fornecido as medidas W^* e A^* que nos indicarão qual modelo irá melhor ajustar os dados. Os resultados indicam que o modelo proposto possui os menores valores dessas estatísticas entre os modelos e, portanto, poderia ser escolhido como o melhor modelo para ajustar esses dados. A tabela 5 lista a

Tabela 4 – Aplicação - Tensão até a ruptura de um fio

Model	α	λ	a	b	A^*	W^*
NH-EE(α, λ, a, b)	1.3953 (0.2927)	1.4521 (0.5043)	0.0033 (0.0004)	1.8114 (0.2363)	0.5344	0.0936
E(a)				0.0044 (0.0004)	0.6842	0.1257
EE(a, b)			2.3798 (0.3629)	0.0072 (0.0007)	0.7901	0.1461
NH(α, λ)	0.0029 (0.0004)	1.3445 (0.1544)			0.5732	0.1028
ENH(α, λ, b)	0.6707 (0.1041)	4.0927 (1.1834)		0.0204 (0.0094)	1.4377	0.2655
Exp-BurXII(α, λ, b)	15.6295 (2.9309)	3.3249 (0.7410)		0.1818 (0.0405)	3.7624	0.6871
Exp-W(α, λ, b)	14.5100 (6.0888)	0.4532 (0.0478)		14.2137 (3.8518)	198.8453	32.1663
Exp-Gama(α, λ, b)	3.8877 (1.3179)	0.0128 (0.0028)		0.5016 (0.1908)	0.6050	0.1094
Gamma-Exp(α, λ)	2.2658 (0.2924)	0.01024 (0.0014)			0.6712	0.1232
Gamma-G(α, λ, b)	1.6679 (0.5835)	0.0096 (0.0013)		1.3159 (0.4838)	0.6861	0.1261

estatística de razão de verossimilhanças para comparar os modelos encaixados que foram ajustados em

ambas as aplicações. Note que rejeitamos a hipótese nula ($H_0 : \alpha = \lambda = 1$) nas duas aplicações indicando indícios em favor da nova distribuição.

Tabela 5 – Teste da razão de verossimilhanças

Aplicação	Estatística W	p-value
Doença Cardiovascular	26.1622	0.000002
Tensão no fio	16.3584	0.000280

4.7 CONSIDERAÇÕES FINAIS

Propomos a distribuição Nadarajah-Haghighi Exponential Exponencializada (NH-EE) para estender a distribuição exponencial exponencializada, realizando um estudo matemático da nova distribuição, incluindo expressões explícitas para os momentos, função geradora de momentos, mediana e função quantílica. As aplicações realizadas mostram a eficiência do modelo proposto, indicando que a nova distribuição fornece melhor ajuste do que as outras distribuições consideradas, concluindo, assim, que o modelo se mostra bastante competitivo no ajuste de dados clínicos e de fadiga.

5 UMA NOVA EXTENSÃO DA DISTRIBUIÇÃO RAYLEIGH: SIMULAÇÃO E APLICAÇÕES

RESUMO

Uma nova distribuição de três parâmetros, Nadarajah-Haghighi-Rayleigh, é proposta usando a família desenvolvida em Dias (2016). Obtemos algumas propriedades estruturais do novo modelo, incluindo expressões explícitas para os momentos e função geradora. O método de máxima verossimilhança é usado para estimar os parâmetros do modelo. Um estudo de simulação é realizado para verificar as propriedades assintóticas dos parâmetros e a utilidade da distribuição proposta é ilustrada por meio de dois conjuntos de dados reais.

Palavras-chave: Distribuição Nadarajah-Haghighi; Distribuição Rayleigh; Método da máxima verossimilhança; Simulação de Monte Carlo.

5.1 INTRODUÇÃO

A distribuição de Rayleigh (R) é uma distribuição de probabilidade contínua com suporte em $\mathbb{R}^+$. Essa distribuição é usada constantemente quando a magnitude geral de um vetor está relacionada a seus componentes direcionais. Além disso, mais recentemente, extensões do modelo Rayleigh foram usadas para ajustar dados de fadiga. Uma variável aleatória $X \sim R(\sigma)$ tem função de distribuição acumulada (fda) dada por

$$G(x; \sigma) = 1 - \exp \left[-\frac{x^2}{2\sigma^2} \right], \quad (5.1)$$

em que $\sigma > 0$ é parâmetro de escala. Sua função densidade de probabilidade (pdf) é dada por

$$g(x; \sigma) = \frac{x}{\sigma^2} \exp \left[-\frac{x^2}{2\sigma^2} \right]. \quad (5.2)$$

Recentemente, novas classes de distribuições para dados de sobrevivência são propostas para fornecer grande flexibilidade na modelagem de dados assimétricos. A Tabela 6 mostra algumas extensões da distribuição R.

Assim como os autores descritos na Tabela 6, procurando introduzir mais flexibilidade para a distribuição Rayleigh, utilizamos a classe de distribuições Nadarajah Haghighi-G, proposta por Dias (2016). Neste contexto, para qualquer distribuição-base com fda $G(x)$, $x > 0$, o autor definiu a classe Nadarajah-Haghighi-G (NH-G) com fda $F(x)$ e pdf $f(x)$, respectivamente, dadas por

$$F(x) = 1 - \exp \left\{ 1 - [1 - \lambda \log [1 - G(x)]]^\alpha \right\} \quad (5.3)$$

Tabela 6 – Algumas extensões da distribuição Rayleigh

Modelo	Autor
Slashed Rayleigh exponencializada	Salinas <i>et al.</i> (2015)
Kumaraswamy Rayleigh exponencializada	Rashwan (2016)
Marshall-Olkin Rayleigh generalizada estendida	MirMostafae <i>et al.</i> (2017)
Transmuted Rayleigh	Merovci (2014)
Rayleigh Generalizada	Raqab e Madi (2011)
Transmuted Rayleigh modificada inversa	Khan e King (2015)

e

$$f(x) = \frac{\alpha \lambda g(x) \{1 - \lambda \log [1 - G(x)]\}^{\alpha-1} \exp \{1 - [1 - \lambda \log [1 - G(x)]]^\alpha\}}{1 - G(x)}, \quad (5.4)$$

em que $g(x)$ e $G(x)$ são, respectivamente, fdp e fda da distribuição-base. Note que a classe proposta por Dias (2016) não possui funções complicadas. Isso é uma vantagem dessa classe. É possível notar, também, que se $\alpha = \lambda = 1$, temos como caso especial a distribuição-base e se $\alpha = 1$ obtemos o modelo Lehmann type II-G.

5.2 DISTRIBUIÇÃO NADARAJAH-HAGHIGHI-RAYLEIGH

Aqui, nós propomos um novo modelo chamado Nadarajah-Haghighi-Rayleigh (NH-R), distribuição que estende o modelo R. Seja $\gamma = \lambda/\sigma^2$. Então o modelo NH-R tem fda e fdp dadas, respectivamente, por

$$F(x) = 1 - \exp \left\{ 1 - \left[1 + \frac{\gamma x^2}{2} \right]^\alpha \right\} \quad (5.5)$$

e

$$f(x) = \alpha \gamma x \left\{ 1 + \frac{\gamma x^2}{2} \right\}^{\alpha-1} \exp \left\{ 1 - \left[1 + \frac{\gamma x^2}{2} \right]^\alpha \right\}. \quad (5.6)$$

A função taxa de falha especifica a taxa instantânea de falha ou morte no tempo t condicional à sobrevivência do elemento até o tempo t , e é definida por

$$h(t) = \lim_{\Delta \rightarrow \infty} \frac{P(t \leq T \leq T + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)},$$

em que $S(t) = 1 - F(t)$ é chamada de função de sobrevivência. A função taxa de falha que corresponde a (5.6) é dada por

$$h(x) = \alpha \gamma x \left\{ 1 + \frac{\gamma x^2}{2} \right\}^{\alpha-1}. \quad (5.7)$$

Qian (2012), a partir da função taxa de falha, conseguiu mostrar analiticamente o comportamento das curvas da ftf.

Portanto, se a distribuição-base for a Rayleigh, seguindo a ideia de Qian (2012), podemos determinar os intervalos para os quais temos as formas da ftf, para o caso proposto. Seja $z = \frac{\gamma x^2}{2} \implies x = \sqrt{\frac{2z}{\gamma}}, z > 0$.

$$\begin{aligned} \log[h(x)] &= \log\left[h\left(\sqrt{\frac{2z}{\gamma}}\right)\right] = \log[r(z)] \\ &= \log(\alpha) + \log(\gamma) + \frac{1}{2} \log\left[\left(\frac{2z}{\gamma}\right)\right] + (\alpha - 1) \log(1 + z) \\ \log[r(z)] &= \log(\alpha) + \frac{1}{2} \log(\gamma) + \frac{1}{2} [\log(z) + \log(\alpha)] + (\alpha - 1) \log(1 + z) \end{aligned}$$

Derivando ambos os lados da última equação em relação a z , obtemos

$$\begin{aligned} \frac{r'(z)}{r(z)} &= \frac{1}{2z} + \frac{(\alpha - 1)}{1 + z} \Leftrightarrow r'(z) = r(z) \left[\frac{1}{2z} + \frac{(\alpha - 1)}{1 + z} \right] \Leftrightarrow r'(z) = r(z) \left[\frac{z(2\alpha - 1) + 1}{2z(1 + z)} \right] \\ \Leftrightarrow r'(z) &= r(z)s(x). \end{aligned}$$

Consequentemente, o sinal de $r'(z)$ é o mesmo sinal de $s(z)$ com $r(z) > 0$. Note que o parâmetro σ não muda as formas da ftf, como era de esperar. Com base nisso podemos observar que

- Se $0 < \alpha < \frac{1}{2} \implies$ ftf unimodal;
- Se $\alpha \geq \frac{1}{2} \implies$ ftf crescente.

Os gráficos das funções de densidade e taxa de falha para alguns valores de parâmetros são fornecidos na Figura 5. Esta figura indica que o modelo proposto é mais flexível que a distribuição de Rayleigh.

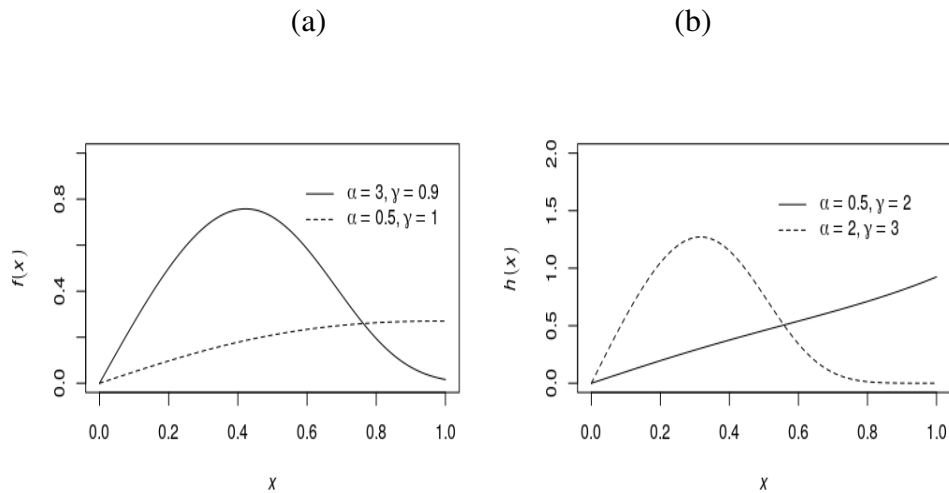


Figura 5 – Gráficos para a densidade e a função taxa de falha da distribuição NH-R.

Quando o $\alpha = 1$ (Lehmann tipo II-R) a Equação (5.6) é dada por

$$f(x) = \gamma x \exp\left(\frac{-\gamma x^2}{2}\right).$$

Outro aspecto interessante desta distribuição é a ligação com a família exponencial. Podemos reescrever a última equação como

$$f(x) = x \exp \left\{ \frac{-\gamma x^2}{2} + \log(\gamma) \right\},$$

o que implica que o modelo NH-R (com $\alpha = 1$) pertence à família exponencial, com

$$h(x) = x, \eta(\gamma) = -\gamma/2, T(x) = x^2, e b(\gamma) = \log(\gamma).$$

Além disso o novo modelo é facilmente simulado da seguinte forma: se U é uma variável aleatória uniforme $(0, 1)$, então

$$X = Q_R \left(1 - \exp \left[\frac{1 - [1 - \log(1 - u)]^{1/\alpha}}{\lambda} \right] \right) = \sigma \left(-2 \frac{[1 - \log(1 - u)]^{1/\alpha}}{\lambda} \right)^{1/2},$$

em que $Q_R(\cdot)$ é a função quantílica de R .

E, após algumas manipulações tomando $\gamma = \lambda / \sigma^2$, temos que

$$Q_R(u) = \left(-2 \frac{[1 - \log(1 - u)]^{1/\alpha}}{\gamma} \right)^{1/2} \quad (5.8)$$

Depois de obtida a função quantil, podemos obter a mediana da NH-R fazendo $u = 0.5$, onde temos

$$X = \left(-2 \left[\frac{1 - (1.6931472)^{1/\alpha}}{\gamma} \right] \right)^{1/2}.$$

5.3 EXPANSÕES UTILIZADAS

Algumas expansões são úteis para as Equações (5.3) e (5.4) que podem ser derivadas usando o conceito das distribuições exponencializadas. Para uma fda de uma distribuição-base arbitrária $G(\cdot)$, é dito que uma variável aleatória T tem distribuição exponencializada- G com parâmetro adicional de potência $a > 0$, digamos $T \sim \exp-G(a)$, se sua fdp e fda forem

$$H_a(t) = G^a(t) \text{ e } h_a(t) = a g(t) G^{a-1}(t),$$

respectivamente. As propriedades das distribuições exponencializadas foram estudadas por muitos autores, ver Mudholkar e Srivastava (1993) e Mudholkar *et al.* (1995) para Weibull exponencializada, Gupta e Kundu (2001b) e Nadarajah e Gupta (2007) para distribuição gama exponencializada, por exemplo.

De Dias (2016), podemos escrever a Equação (5.6) como

$$f(x) = \sum_{m=0}^{\infty} w_m h_{m+1}(x), \quad (5.9)$$

em que $h_{m+1}(x) = (m+1)g(x)G(x)^m$ e substituindo em $G(x)$ e $g(x)$ a fda e fdp da distribuição Rayleigh, temos a Rayleigh exponencializada, exp-R, e seu parâmetro de exponencialização é $m+1$. Para mais detalhes sobre (5.9) e seus coeficientes, ver Dias (2016). Com base na Equação (5.9) várias propriedades estruturais da distribuição NH-R podem ser obtidas conhecendo as propriedades da distribuição exp-R.

Integrando a Equação (5.9) e utilizando o teorema da convergência monótona, a fda de X é expressa por

$$F(x) = \sum_{m=0}^{\infty} w_m H_{m+1}(x), \quad (5.10)$$

em que $H_{m+1}(x) = \left\{1 - \exp\left[-\frac{x^2}{2\sigma^2}\right]\right\}^{m+1}$ é a acumulada de Y_{m+1} .

5.4 ALGUMAS PROPRIEDADES

Nessa seção, desenvolvemos expressões fechadas, algumas em séries de potência, usando os resultados da seção anterior para momentos, função geradora de momentos e entropia para uma variável aleatória $X \sim \text{NH-R}(\alpha, \lambda, \sigma)$

5.4.1 Momentos

Seja Y_m uma variável aleatória com função densidade $h_m(x)$, tal que $Y_m \sim \text{exp-R}(m, \sigma)$. O n -ésimo momento de X pode ser expresso de (5.9) e usando os resultados de Mahmoud e Ghazal (2017), temos

$$E(X^n) = \sum_{m=0}^{\infty} w_m E(Y_{m+1}^n) \quad (5.11)$$

em que

$$E(Y_{m+1}^n) = 2^{\frac{n}{2}} \sigma^n (m+1) \sum_{i=0}^{M(m+1)} \frac{(-1)^i \binom{m}{i} \Gamma(\frac{n}{2} + 1)}{(i+1)^{\frac{n}{2}+1}} \quad \text{e,} \quad M(m+1) = \begin{cases} m, & m \in \mathbb{Z}^+, \\ \infty, & \text{caso contrário.} \end{cases}$$

5.4.2 Função Geradora de Momentos

Seja $M_X(t)$ a função geradora de momentos (fgm), de X . Então, usando a expansão (5.9) e o teorema da convergência monótona, podemos escrever

$$M_X(t) = \sum_{m=0}^{\infty} w_m M_{m+1}(t), \quad (5.12)$$

em que $M_{m+1}(t)$ é a fgm de Y_m . Consequentemente, $M_X(t)$ é obtido a partir da função geradora da distribuição exp-R.

Alternativamente, podemos calcular $M_X(t)$ de (5.12) como

$$M_X(t) = \sum_{m=0}^{\infty} (m+1)w_m \rho(t, m),$$

e

$$\rho(t, m) = \int_{-\infty}^{\infty} e^{tx} G(x)^m g(x) dx = \int_0^1 \exp[tQ_G(u)] u^m du,$$

em que $G(x)$, $g(x)$ e $Q_G(u)$ são respectivamente, acumulada, densidade e a quantílica da distribuição Rayleigh. Para mais detalhes sobre essa equação e seus coeficientes, ver Dias (2016)

Depois de algumas manipulações algébricas envolvendo expansões em séries de potência para um número real e um inteiro, segue que

$$M_X(t) = \sum_{m=0}^{\infty} (m+1)w_m \sum_{j=0}^{\infty} p_j(m)t^j, \quad (5.13)$$

em que $p_j(m) = \frac{f_j}{m+1} + g_j(m)$. Mais detalhes sobre (5.13) e os coeficientes envolvidos, ver Anexo A.

Se $\alpha = 1$, a fgm de X é dada, de forma fechada, por

$$M_X(t) = \frac{t\sigma\sqrt{\pi}e^{\frac{(t\sigma)^2}{2\lambda}} \left[\mathcal{E}\left(\frac{t\sigma}{\sqrt{2\lambda}}\right) + 1 \right] + \frac{\sqrt{2\lambda}}{\sigma}}{\sqrt{2\lambda}},$$

em que $\mathcal{E}(\cdot)$ é a função erro.

5.4.3 Entropia

A entropia de Shannon para uma variável aleatória X é definida por $E\{-\log[f(X)]\}$. Da equação (5.6), obtemos

$$\begin{aligned} E\{-\log[f(X)]\} &= -\log(\alpha) - \log(\gamma) - E[\log(X)] - (\alpha - 1)E\left\{\log\left[1 + \frac{\gamma X^2}{2}\right]\right\} \\ &\quad - 1 + E\left\{\left[1 + \frac{\gamma X^2}{2}\right]^\alpha\right\} \end{aligned}$$

Para $\alpha = 1$, obtemos

$$\begin{aligned} E\{-\log[f(X)]\} &= -\log(\alpha) - \log(\gamma) - E[\log(X)] + \frac{\gamma}{2}E(X^2) \\ &= -\log(\alpha) - \log(\gamma) + \frac{\xi}{4} + \log\left(\frac{\gamma}{2}\right) + \frac{1}{2} \\ &= -\log(\alpha) - \log(2) + \frac{\xi}{4} + \frac{1}{2}, \end{aligned}$$

em que ξ é a constante de Euler e tem valor aproximado de 0,577216.

5.4.4 Estatísticas de Ordem

Estatísticas de ordem é uma ferramenta importante para inferência e estatística não paramétrica. Suponha que $X_1, \dots, X_n$ seja uma amostra de variáveis aleatórias que seguem a distribuição NH-R.

Denote por $X_{1:n}, \dots, X_{n:n}$ as estatísticas de ordem. A fdp da i -ésima estatística de ordem de uma amostra aleatória $X_1, \dots, X_n$ da distribuição NH-R é dada por

$$f_{i:n}(x) = K f(x) F^{i-1}(x) [1 - F(x)]^{n-i},$$

em que $K = n! / [(i-1)!(n-i)!]$, $F(x)$ e $f(x)$ são a acumulada (5.5) e a densidade (5.6) da NH-R, respectivamente. Então,

$$f_{i:n}(x) = K \sum_{j=0}^{n-1} (-1)^j \binom{n-1}{j} f(x) F^{i+j-1}(x).$$

A função densidade de probabilidade $X_{i:n}$ pode ser expressada de (5.9) e (5.10) como

$$\begin{aligned} f_{i:n}(x) &= K \sum_{j=0}^{n-1} (-1)^j \binom{n-i}{j} \left\{ \sum_{r=0}^{\infty} w_r (r+1) \left\{ 1 - \exp \left[-\frac{x^2}{2\sigma^2} \right] \right\}^r \left\{ \frac{x}{\sigma^2} \exp \left[-\frac{x^2}{2\sigma^2} \right] \right\} \right\} \\ &\times \left\{ \sum_{m=0}^{\infty} w_m \left[1 - \exp \left(-\frac{x^2}{2\sigma^2} \right) \right]^{m+1} \right\}^{i+j-1}. \end{aligned} \quad (5.14)$$

Usando séries de potência para um inteiro positivo, temos

$$\left\{ \sum_{m=0}^{\infty} w_m \left[1 - \exp \left(-\frac{x^2}{2\sigma^2} \right) \right]^{m+1} \right\}^{i+j-1} = \sum_{m=0}^{\infty} f_{i+j-1,m} \left[1 - \exp \left(-\frac{x^2}{2\sigma^2} \right) \right]^{1+j+m-1}$$

em que $f_{i+j-1,0} = w_0^{i+j-1}$ e (para $m \geq 1$)

$$f_{i+j-1,m} = (mw_0)^{-1} \sum_{p=1}^m [p(i+j) - m] w_p f_{i+j-1,m-p}.$$

Consequentemente,

$$f_{i:n}(x) = \sum_{j=0}^{n-1} \sum_{r,m=0}^{\infty} c_{j,r,m} h_{i+j+r+m}(x), \quad (5.15)$$

em que

$$c_{j,r,n} = \frac{(-1)^j n!}{(i-1)!(n-1-j)!j!} \frac{(r+1)w_r f_{i+j-1,m}}{(i+j+r+m)}.$$

O principal resultado dessa seção é a Equação (5.15), que nos mostra que a densidade das estatísticas de ordem é escrita como uma combinação linear tripla de densidades da distribuição exp-R. Portanto, podemos observar que algumas propriedades matemáticas (estatísticas de ordem, momentos, função geradora de momentos) podem ser obtidas através das propriedades da exp-R

5.5 INFERÊNCIA E ESTIMAÇÃO

Nessa seção, discutimos estimação via máxima verossimilhança para a inferência dos parâmetros da distribuição proposta.

Seja $x_1, \dots, x_n$ uma amostra aleatória de tamanho n da variável aleatória X que segue uma distribuição NH-R, $X \sim \text{NH-R}(\alpha, \gamma)$. A função de log-verossimilhança para o modelo é dada por

$$\begin{aligned} l(\boldsymbol{\theta}) &= n + n \log(\alpha) + n \log(\gamma) + \sum_{i=1}^n [\log(x_i)] + (\alpha - 1) \sum_{i=1}^n \log \left\{ 1 + \frac{\gamma x_i^2}{2} \right\} \\ &\quad - \sum_{i=1}^n \left(1 + \frac{\gamma x_i^2}{2} \right)^\alpha. \end{aligned} \quad (5.16)$$

em que $\boldsymbol{\theta}$ é o vetor de parâmetros, $\boldsymbol{\theta} = (\alpha, \gamma)$.

A maximização de (5.16) é realizada utilizando o software estatístico R. Derivando (5.16) em relação aos parâmetros, igualando a zero, e denotando por U as primeiras derivadas, temos $U_\alpha(\boldsymbol{\theta}) = 0$, $U_\gamma(\boldsymbol{\theta}) = 0$, e resolvendo-os simultaneamente obtemos os estimadores de máxima verossimilhança (EMV), $\hat{\boldsymbol{\theta}}$, de $\boldsymbol{\theta}$.

Geralmente os estimadores de máxima verossimilhança não possuem forma fechada. Para a solução desse sistema de equações não lineares se faz necessário o uso de técnicas de otimização não linear, como por exemplo, o algoritmo Newton-Raphson ou quase-Newton Nocedal e Wright (2006); Press *et al.* (2007). O vetor escore, correspondente a Equação (5.16) é dado por

$$U_\alpha = \frac{n}{\alpha} + \sum_{i=1}^n \log \left(1 + \frac{\gamma x_i^2}{2} \right) - \sum_{i=1}^n \left(1 + \frac{\gamma x_i^2}{2} \right)^\alpha \log \left(1 + \frac{\gamma x_i^2}{2} \right),$$

e

$$U_\gamma = \frac{n}{\gamma} + (\alpha - 1) \sum_{i=1}^n \frac{x_i^2}{2 + \gamma x_i^2} - \alpha \frac{x_i^2}{2} \sum_{i=1}^n \left(1 + \frac{\gamma x_i^2}{2} \right)^{\alpha-1}$$

5.5.1 Estudo de simulação

Nesta seção, é apresentado um estudo de simulação para verificar se os estimadores de máxima verossimilhança satisfazem as suas propriedades assintóticas. Para verificação dessas propriedades, é necessário observar se ao aumentar o tamanho da amostra, a média das estimativas tendem ao valor verdadeiro. Para isso, escolhemos os valores de tamanho de amostra de $n = \{80, 200, 500\}$. Além disso, inserimos no estudo censura a direita com mecanismo aleatório. Determinamos as porcentagens de 0%, 10% e 20% de censura, afim de verificar se ao introduzirmos censura na simulação, para os mesmos tamanhos de amostra o viés aumenta. Isso deve ocorrer, pois, ao termos dados censurados, alguma informação é perdida.

A função de verossimilhança para dados censurados, considerando censura aleatória a direita é mostrada a seguir

$$L(\boldsymbol{\theta}) \propto \prod_{i=1}^n [f(t_i; \boldsymbol{\theta})]^{\delta_i} [S(t_i; \boldsymbol{\theta})]^{1-\delta_i}, \quad (5.17)$$

em que $f(t_i; \boldsymbol{\theta})$, $S(t_i; \boldsymbol{\theta})$ e δ_i são, respectivamente, função densidade, função de sobrevivência e variável indicadora que representa tempo de falha ou censura na forma

$$\delta_i = \begin{cases} 1, & \text{se é tempo de falha} \\ 0, & \text{se é tempo de censura} \end{cases}. \quad (5.18)$$

Para mais detalhes sobre as Equações (5.17) e (5.18) ver Colosimo e Giolo (2006).

Os estimadores de máxima verossimilhança são os valores de $\boldsymbol{\theta}$ que maximizam $L(\boldsymbol{\theta})$. Maximizar $L(\boldsymbol{\theta})$ é equivalente a maximizar o logaritmo de $L(\boldsymbol{\theta})$, ou seja, $\log[L(\boldsymbol{\theta})]$ e, após isso, resolver o sistema de equações:

$$U(\boldsymbol{\theta}) = \frac{\partial \log(L(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} = 0. \quad (5.19)$$

Porém, nem sempre esse sistema de equações possui forma fechada, então se faz necessário o uso de técnicas iterativas como, por exemplo, o método BFGS (NOCEDAL; WRIGHT, 2006).

Para a geração de números aleatórios da variável NH-R, utilizamos o método da inversão em que geramos ocorrências independentes da distribuição $U \sim U(0, 1)$ e substituímos o números gerados na expressão 5.8, assim gerando números aleatórios para a NH-R.

O método de simulação de Monte Carlo serve para a verificação da lei fraca dos grandes números, que nos diz que

$$\frac{\sum_{i=1}^n X_i}{n} \xrightarrow{P} \mu, \quad (5.20)$$

em que " $\xrightarrow{P}$ " indica convergência em probabilidade e μ é a média populacional. A Equação (5.20) nos diz que quando o tamanho da amostra aumenta, a média das estimativas tende ao valor verdadeiro.

Para a simulação feita, consideramos o seguinte algoritmo

- Geramos de forma independente, números aleatórios da distribuição $U \sim U(0, 1)$;
- Utilizando o método da inversão, geramos números aleatórios para a distribuição NH-R;
- Para determinarmos os elementos que possuem ou não censura, geramos ocorrências da distribuição Bernoulli em que a a probabilidade de sucesso é $p = 1 - C$, em que C é o percentual de censura, dessa forma temos uma coleção de Bernoulli;
- Maximizamos a Expressão (5.17);
- Replicamos a maximização mil vezes;
- Obtemos as médias das estimativas, viés e EQM.

As simulações foram realizadas com os valores $(\alpha, \gamma) = (1, 2)$, para o primeiro cenário, para o segundo cenário, os valores dos parâmetros foram $(\alpha, \gamma) = (2, 3)$, no terceiro cenário os valores foram $(\alpha, \gamma) = (2.5, 1.8)$ e para o quarto cenário os valores $(\alpha, \gamma) = (1.7, 2.8)$. As quantidades observadas para o estudo foram a média das estimativas do vetor $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\gamma})$, os valores dos vieses e seus respectivos Erros Quadráticos Médios (EQM's). Os resultados são mostrados nas tabelas 7, 8, 9 e 10.

Podemos observar que nesse estudo de simulação, as propriedades assintóticas são satisfeitas para os valores de parâmetros escolhidos, pois ao aumentarmos o tamanho da amostra o viés das estimativas diminui, assim como o valor do EQM. Também podemos notar que ao aumentarmos o percentual de censura com os mesmos tamanhos de amostra o valor do viés aumenta.

Tabela 7 – Resultados da Simulação para $(\alpha, \gamma) = (1, 2)$

Censura	n	$(\hat{\alpha}, \hat{\gamma})$	Viés($\hat{\alpha}, \hat{\gamma}$)	EQM($\hat{\alpha}, \hat{\gamma}$)
Censura = 0%	80	(1.0060, 2.0194)	(0.0060, 0.0194)	(0.0050, 0.0537)
	200	(1.0030, 2.0061)	(0.0030, 0.0061)	(0.0021, 0.0220)
	500	(1.0012, 2.0003)	(0.0012, 0.0003)	(0.0008, 0.0080)
Censura = 10%	80	(0.9882, 1.7921)	(-0.0117, -0.2078)	(0.0054, 0.0855)
	200	(0.9838, 1.7754)	(-0.0161, -0.2245)	(0.0025, 0.0677)
	500	(0.9834, 1.7997)	(-0.0165, -0.2002)	(0.0013, 0.0468)
Censura=20%	80	(0.9588, 1.5141)	(-0.0411, -0.4858)	(0.0076, 0.2660)
	200	(0.9601, 1.5445)	(-0.0398, -0.4554)	(0.0040, 0.2205)
	500	(0.9612, 1.5689)	(-0.0387, -0.4310)	(0.0025, 0.1909)

Tabela 8 – Resultados da Simulação para $(\alpha, \gamma) = (2, 3)$

Censura	n	$(\hat{\alpha}, \hat{\gamma})$	Viés($\hat{\alpha}, \hat{\gamma}$)	EQM($\hat{\alpha}, \hat{\gamma}$)
Censura = 0%	80	(2.0136, 3.0234)	(0.0136, 0.0234)	(0.0194, 0.0759)
	200	(2.0051, 3.0080)	(0.0051, 0.0080)	(0.0083, 0.0313)
	500	(1.99994, 3.0002)	(0.00005, 0.0002)	(0.0031, 0.0116)
Censura = 10%	80	(2.0289, 2.9127)	(0.0289, -0.0872)	(0.0227, 0.0794)
	200	(2.0201, 2.8949)	(0.0201, -0.1050)	(0.0092, 0.0400)
	500	(2.0115, 2.9026)	(0.0115, -0.0973)	(0.0036, 0.0206)
Censura=20%	80	(2.0511, 2.7565)	(0.0511, -0.2434)	(0.0277, 0.1242)
	200	(2.0381, 2.7681)	(0.0381, -0.2318)	(0.0118, 0.0810)
	500	(2.0277, 2.7742)	(0.0277, -0.2257)	(0.0048, 0.0613)

Tabela 9 – Resultados da Simulação para $(\alpha, \gamma) = (2.5, 1.8)$

Censura	n	$(\hat{\alpha}, \hat{\gamma})$	Viés($\hat{\alpha}, \hat{\gamma}$)	EQM($\hat{\alpha}, \hat{\gamma}$)
Censura = 0%	80	(2.5161, 1.8141)	(0.0161, 0.0141)	(0.0300, 0.0249)
	200	(2.5063, 1.8043)	(0.0063, 0.0043)	(0.0124, 0.0101)
	500	(2.5005, 1.7994)	(0.0005, -0.0005)	(0.0047, 0.0039)
Censura = 10%	80	(2.5350, 1.7818)	(0.0350, -0.0181)	(0.0353, 0.0252)
	200	(2.5237, 1.7718)	(0.0237, -0.0281)	(0.0142, 0.0110)
	500	(2.5147, 1.7710)	(0.0147, -0.0289)	(0.0056, 0.0048)
Censura=20%	80	(2.5617, 1.7317)	(0.0617, -0.0682)	(0.0428, 0.0292)
	200	(2.5450, 1.7318)	(0.0450, -0.0681)	(0.0180, 0.0149)
	500	(2.5317, 1.7309)	(0.0317, -0.0690)	(0.0069, 0.0087)

Tabela 10 – Resultados da Simulação para $(\alpha, \gamma) = (1.7, 2.8)$

Censura	n	$(\hat{\alpha}, \hat{\gamma})$	Viés($\hat{\alpha}, \hat{\gamma}$)	EQM($\hat{\alpha}, \hat{\gamma}$)
Censura = 0%	80	(1.7116, 2.8228)	(0.0116, 0.0228)	(0.0140, 0.0723)
	200	(1.7044, 2.8078)	(0.0044, 0.0078)	(0.0058, 0.0297)
	500	(1.7003, 2.7997)	(0.00034, -0.0002)	(0.0023, 0.0110)
Censura = 10%	80	(1.7212, 2.6718)	(0.0212, -0.1281)	(0.0161, 0.0817)
	200	(1.7133, 2.6542)	(0.0133, -0.1457)	(0.0064, 0.0475)
	500	(1.7067, 2.6670)	(0.0067, -0.1329)	(0.0026, 0.0279)
Censura=20%	80	(1.7330, 2.4680)	(0.0330, -0.3319)	(0.0191, 0.1659)
	200	(1.7232, 2.4861)	(0.0232, -0.3138)	(0.0079, 0.1219)
	500	(1.7148, 2.4996)	(0.0148, -0.3003)	(0.0031, 0.0991)

5.6 APLICAÇÃO A DADOS REAIS

Apresentamos duas aplicações para comprovar empiricamente a flexibilidade da distribuição NH-R. Comparamos os resultados dos ajustes do modelo NH-R e outros modelos utilizando estatísticas de testes apropriadas. De fato, comparamos os ajustes do NH-R com sete outros modelos de sobrevivência:

1. Distribuição Rayleigh (R) (RAYLEIGH, 1880);
2. Distribuição Exponentiated Rayleigh (Exp-R) (MAHMOUD; GHAZAL, 2017);
3. Distribuição Kumaraswamy Rayleigh (Kw-R), ver gerador em Cordeiro e Castro (2011);
4. Distribuição Birnbaum Saunders (BS) (BIRNBAUM; SAUNDERS, 1969);
5. Distribuição Gamma-Birnbaum Saunders (GBS) (CORDEIRO *et al.*, 2016);
6. Distribuição Exponentiated Nadarajah-Haghighi (ENH) (LEMONTE, 2013).

Os EMV's dos parâmetros do modelo foram calculados e usamos estatísticas de teste visando compará-los quanto à adequacidade aos dados. Essas estimativas foram avaliadas usando o método BFGS. As estatísticas de teste, Cramér-von Mises (W^*) e Anderson-Darling (A^*), definidas por Chen e

Balakrishnan (1995), são computadas para comparação dos modelos ajustados. Em geral, quanto menor os valores dessas estatísticas, melhor o ajuste aos dados. Todos os cálculos foram realizados usando o software estatístico R utilizando o pacote *Adequacy Model* proposto por Marinho (2016). Para os chutes iniciais, usamos um método heurístico com o pacote GenSA, também do R.

5.6.1 Conjunto de Dados 1: Fibras vidro expostas a pressão

A fibra de vidro é encontrada em diversos materiais no nosso dia a dia como, por exemplo, em malhas, placas e até mesmo em diversos segmentos industriais. Além disso, seu uso é de extrema importância na fabricação de aviões. Um dos motivos do seu uso extenso é o fato de que este material é resistente à ação de corrosão e agentes químicos, por exemplo. Podemos ver isso em Wang e Restrepo (2001) que verificou que mantas de polímero reforçadas com fibras de vidro conseguem fazer com que evitem flambagem longitudinal em colunas. Ainda sobre a importância, Beaudoin (1990) nos mostra que a fibra de vidro, se acrescentada na mistura do concreto, traz mais resistência.

No entanto, apesar de suas boas características, o desempenho das fibras de vidro podem apresentar queda devido, dentre outros fatores, ao diâmetro e comprimento. Assim, é comum pensar que se diminuirmos o diâmetro das fibras, aumentaremos a sua resistência mecânica.

Nesse sentido, faz-se necessário o estudo visando testar a resistência das fibras de vidro exposta a alguma força externa como, por exemplo, o estudo discutido em Smith e Naylor (1987).

Como primeira aplicação, consideramos o conjunto de dados que consiste de $n = 63$ observações de fibras de vidro de 1,5 cm, originalmente obtidas por trabalhadores do Laboratório Nacional de Física do Reino Unido. A unidade de medida para o estresse não foi informada no artigo.

O gráfico TTT, (AARSET, 1987), para esse banco de dados, é apresentado na Figura 6, e nos apresenta uma curva predominantemente côncava, indicando que a função taxa de falha empírica é crescente. Logo, o modelo proposto é indicado para o ajuste dos dados, que consegue modelar a forma crescente.

As estatísticas descritivas para esse conjunto de dados é mostrada na Tabela 11. Note que esse conjunto de dados apresenta variância pequena, assimetria negativa e curtose positiva. A tabela 12 nos fornece os valores dos EMV's dos parâmetros assim como as quantidades para Erro Padrão (entre parênteses) a partir dos ajustes dos modelos citados assim como as medidas W^* e A^* . Os resultados indicam que o modelo proposto possui os menores valores dessas estatísticas entre os modelos e, portanto, poderia ser escolhido como o melhor modelo para esses dados.

Tabela 11 – Estatísticas descritivas - Primeira Aplicação

Aplicação	Media	Mediana	Variância	Assimetria	Curtose	Min.	Max.
Fibras de vidro	1.507	0.3241	0.3241	-0.8785	0.8001	0.550	2.240

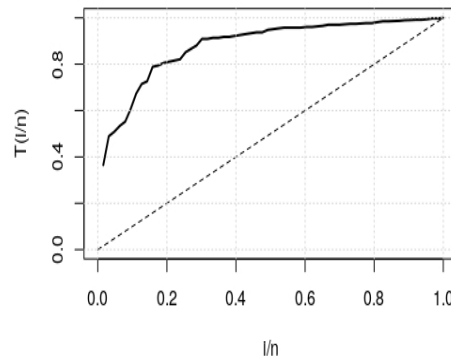


Figura 6 – TTT plot: Fibras expostas a pressão

Tabela 12 – Aplicação 1

Model	α	γ	σ	W^*	A^*
NH-R(α, γ)	7.9143 (2.8365)	0.0803 (0.0304)		0.3014	1.6509
R(σ)			1.0894 (0.0683)	0.4654	2.5538
Exp-R(α, σ)	5.4860 (1.1848)		0.7164 (0.0391)	0.5594	3.0722
Kw-R(α, γ, σ)	0.7166 (0.0186)	99990.38 (27277.07)	3506.206 (31.5246)	0.5244	2.8772
BS(α, γ)	0.2621 (0.0233)	1.4566 (0.0476)		0.7180	3.9152
GBS(α, γ, σ)	1.0148 (0.1255)	0.0408 (0.0119)	19.6343 (3.0482)	0.5616	3.0810
ENH(α, γ, σ)	0.0825 (2.0061)	11.2842 (0.0372)	9.7800 (4.7105)	0.4991	2.739

5.6.2 Conjunto de dados 2: carcinoma de orofaringe

Uma das doenças que ainda mais letais do mundo é o câncer. Diversos estudos buscam informações sobre o tempo até a morte de indivíduos com câncer de diferentes tipos. Podemos ver isso nos estudos de Prentice (1973) analisando o tempo até a morte de indivíduos com câncer de pulmão ou em Ibrahim *et al.* (2014) sobre o tempo até a morte de pacientes com melanoma cutâneo. Entre as diferentes variações de câncer, há um que se associa ao tabagismo e etilismo, que é o câncer de Orofaringe. Avaliar o perfil epidemiológico e a taxa de sobrevivência de pacientes com câncer de boca pode ser um bom indicador para a prevenção da doença.

Devido a alta incidência de ocorrências e também por sua agressividade, a busca de informações sobre o câncer de orofaringe se fazem necessárias. Um estudo que apresentou ações que modifiquem positivamente os indicadores epidemiológicos do câncer de orofaringe foi feito por Torres-Pereira *et al.*

(2012). Uma caracterização sobre esse tipo de enfermidade pode ser visto em Dedivitis *et al.* (2004). Ainda sobre o estudo de câncer de boca, Moro *et al.* (2018) fez uma análise de sobrevida em pacientes.

Para a segunda aplicação a dados reais, consideramos um conjunto de dados encontrado em <http://www.umass.edu/statdata/statdata/stat-survival.html>. O banco de dados é sobre o tempo até a morte de 195 pessoas com carcinoma de faringe.

Nessa aplicação, o gráfico TTT Aarset (1987) apresenta uma forma predominantemente côncava, Figura 7, indicando que uma distribuição com taxa de falha crescente pode modelar bem esses dados. A distribuição NH-R possui ftf com forma crescente e, portanto, pode ser uma boa opção para modelar esses dados.

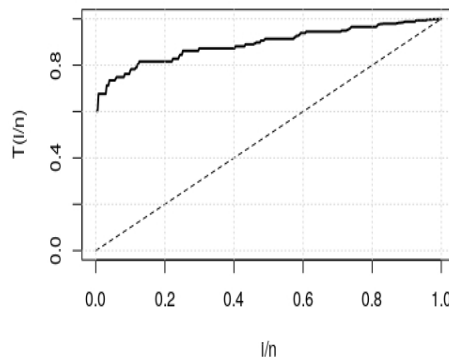


Figura 7 – TTT plot

As quantidades para seleção do melhor modelo foram as estatísticas A^* e W^* . Na Tabela 13, colocamos as estimativas e o erro padrão entre em parênteses, assim como as quantidades das estatísticas.

A Tabela 14 lista a estatística da razão de verossimilhança (RV) para comparar modelos encaixados que foram ajustados em ambas as aplicações. Note que rejeitamos a hipótese nula ($H_0 : \alpha = \lambda = 1$) nos dois cenários em favor da nova distribuição.

5.7 CONSIDERAÇÕES FINAIS

Neste capítulo, propomos a distribuição Nadarajah-Haghighi Rayleigh (NH-R) para estender a distribuição Rayleigh, fornecendo um tratamento matemático da nova distribuição, incluindo expansões para função de densidade. Além disso, derivamos expressões explícitas para os momentos e para a função geradora. A estimação dos parâmetros do modelo é abordada pelo método da máxima verossimilhança. Aplicações da distribuição NH-R a dois conjuntos de dados reais mostram que a nova distribuição fornece consistentemente melhores ajustes do que outras distribuições competitivas, e um estudo de

Tabela 13 – Aplicação 2

Model	α	λ	σ	W^*	A^*
NH-R(α, λ, σ)	7.4895 4.1943 ⁻⁴	1.0896 $\times 10^{-5}$ 3.7405 $\times 10^{-7}$		0.3661	1.9660
R(σ)			95.7791 (2.4534)	0.6657	3.7468
Exp-R(α, σ)	62.9731 (2.0991)		6.5110 (0.8882)	0.4711	2.5819
Kw-R(α, λ, σ)	1.6171 (0.0140)	99.9936 $\times 10^3$ (6875.4535)	3503.1228 (52.9526)	0.9166	5.2532
BS(α, λ)	0.26749 (0.02615)	108.2223 (3.6425)		0.3872	2.0880
GBS(α, λ, σ)	0.2298 0.0045	80.9344 2.1095 $\times 10^{-6}$	3.3466 0.03043768	0.4046	2.1892
ENH(α, λ, σ)	15.4918 (1.4959)	0.0098 (0.0005)	1.7103 (0.0667)	0.3941	2.1286

Tabela 14 – Teste da razão de verossimilhanças

Aplicação	Estatística W	p-value
Fibras de Vidro	32.9984	6.8310 $\times 10^{-8}$
Carcinoma	53.6940	2.1902 $\times 10^{-12}$

simulação considerando diferentes percentuais de censura foi realizado afim de analisar se as propriedades assintóticas dos parâmetros do modelo são satisfeitas, verificamos que essas propriedades foram satisfeitas. Esperamos que o modelo proposto possam atrair aplicações mais amplas na literatura das distribuições de vida em fadiga.

6 UMA NOVA EXTENSÃO DE UMA DISTRIBUIÇÃO DE VALOR EXTREMO: DISTRIBUIÇÃO NADARAJAH HAGHIGHI GUMBEL COM PROPRIEDADES E APLICAÇÕES

RESUMO

Neste capítulo é introduzido uma nova extensão da distribuição de valor extremo, a distribuição de Gumbel, com quatro parâmetros. Para esse novo modelo encontramos expressões em forma fechada para as funções densidade, acumulada, taxa de falha, sobrevivência e quantílica. Fornecemos as principais propriedades matemáticas para a nova distribuição como momentos, função geradora de momentos, entropia de Shannon e estatísticas de ordem. Para a estimação dos parâmetros consideramos o método da máxima verossimilhança verificamos as propriedades assintóticas do modelo considerando a verossimilhança com censura à direita e mecanismo aleatório. Por fim, para ilustração empírica, realizamos duas aplicações com dados reais, um conjunto de dados clínicos e outro conjunto de dados de fadiga de equipamentos, verificando assim a boa competitividade no novo modelo.

Palavras-chave: Distribuição Nadarajah-Haghighi; Distribuição Gumbel; Método da máxima verossimilhança; Fadiga de equipamentos; Simulação.

6.1 INTRODUÇÃO

Considerada uma distribuição muito importante para a modelagem de valores extremos, a distribuição Gumbel, proposta por Gumbel (1958), é uma distribuição amplamente aplicada em engenharia, análise de frequência de inundações entre outros campos. Ela é, também, conhecida como distribuição do valor extremo e surge quando tomamos o logaritmo natural de uma variável aleatória com distribuição Weibull.

Seja uma variável aleatória X com distribuição Gumbel. Então, sua fdp e fda são,

$$g(x) = \beta^{-1} \exp\left(-\frac{x-\mu}{\beta}\right) \exp\left[-\exp\left(-\frac{x-\mu}{\beta}\right)\right] \quad (6.1)$$

e

$$G(x) = \exp\left[-\exp\left(-\frac{x-\mu}{\beta}\right)\right], \quad (6.2)$$

respectivamente, em que $\mu \in \mathbb{R}$ é parâmetro de locação e $\beta > 0$ é parâmetro de escala. Escrevemos, então, $X \sim Gu(\mu, \beta)$

Na busca por mais flexibilidade para a modelagem de dados, é encontrada na literatura algumas extensões da distribuição de Gumbel, como por exemplo, as citadas na Tabela 15:

Tabela 15 – Algumas extensões da distribuição Gumbel

Modelo	Autor
1. Gamma Gumbel	ver gerador Zografos e Balakrishnan (2009)
2. Beta Gumbel (Beta-Gu) distribution	Nadarajah e Kotz (2004)
3. Gumbel Exponencializada	Nadarajah (2006)
4. Kumaraswamy Gumbel	Cordeiro <i>et al.</i> (2012)
5. Gumbel Generalizada Exponencializada	Andrade <i>et al.</i> (2015)
6. Slash Gumbel	Gómez <i>et al.</i> (2019)

Seguindo a ideia dos autores das distribuições propostas acima e usando a família de distribuições Nadarajah Haghighi- G , em que G é a fda de é uma variável aleatória Gumbel, propomos a distribuição com 4 parâmetros denotada por Nadarajah Haghighi-Gumbel (NH-Gu). A fda e fdp dessa nova distribuição são dadas, respectivamente, por

$$F(z) = 1 - \exp \left\{ 1 - [1 - \lambda \log [1 - \exp [-\exp(-z)]]]^\alpha \right\} \quad (6.3)$$

e

$$\begin{aligned} f(z) &= \alpha \lambda \beta^{-1} \exp \left\{ -z - \exp(-z) + 1 - [1 - \lambda \log [1 - \exp [-\exp(-z)]]]^\alpha \right\} \\ &\times \{1 - \lambda \log [1 - \exp [-\exp(-z)]]\}^{\alpha-1} \{1 - \exp [-\exp(-z)]\}^{-1} \end{aligned} \quad (6.4)$$

em que $z = (x - \mu)/\beta$. Claramente, se $\alpha = \lambda = 1$, temos a distribuição Gumbel e se $\alpha = 1$ temos a distribuição Lehmann Type II-Gu (LII-Gu).

A função taxa de falha especifica a taxa instantânea de falha, , ou morte, no tempo t condicional à sobrevivência do elemento até o tempo t , e é definida por

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq T + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}, \quad (6.5)$$

em que $S(t) = 1 - F(t)$.

Para a distribuição NH-Gu, temos por ftf a seguinte expressão

$$h(z) = \frac{\alpha \lambda \beta^{-1} \{1 - \lambda \log [1 - \exp [-\exp(-z)]]\}^{\alpha-1} \exp[-z - \exp(-z)]}{\{1 - \exp[-\exp(-z)]\}} \quad (6.6)$$

em que $z = (x - \mu)/\beta$.

Portanto, se a distribuição de base for a Gumbel, seguindo a ideia de Qian (2012), da última equação obtemos

$$\begin{aligned} \log[h(x)] &= \log(\alpha) + \log(\lambda) - \log(\beta) + (\alpha - 1) \log\{1 - \lambda \log[1 - \exp[-\exp(-z)]]\} \\ &- \log[1 - \exp[-\exp(-z)]] - z - \exp(-z). \end{aligned}$$

Derivando ambos os lados da última equação, obtemos

$$\begin{aligned}
 \frac{h'(x)}{h(x)} &= -1 + \exp(-z) + \frac{\lambda(\alpha - 1)\beta^{-1} \exp[-z - \exp(-z)]}{\{1 - \exp[-\exp(-z)]\}\{1 - \lambda \log[1 - \exp[-\exp(-z)]]\}} \\
 &+ \frac{\beta^{-1} \exp[-z - \exp(-z)]}{1 - \exp[-\exp(-z)]} \Leftrightarrow \\
 \Leftrightarrow h'(x) &= h(x) \left\{ -1 + \exp(-z) + \frac{\lambda(\alpha - 1)\beta^{-1} \exp[-z - \exp(-z)]}{\{1 - \exp[-\exp(-z)]\}\{1 - \lambda \log[1 - \exp[-\exp(-z)]]\}} \right. \\
 &+ \left. \frac{\beta^{-1} \exp[-z - \exp(-z)]}{1 - \exp[-\exp(-z)]} \right\} \Leftrightarrow \\
 \Leftrightarrow h'(x) &= h(x)s(x)
 \end{aligned}$$

Logo, o sinal de $h'(x)$ depende apenas no sinal de $s(x)$, já que $h(x) > 0$

Abaixo, na Figura 8, segue os gráficos para a densidade e da função taxa de falha, fixando os valores dos parâmetros da distribuição de base em $(\beta, \mu) = (1, 0)$ e variando os valores de α e λ . Note que temos a função taxa de falha em forma crescente e unimodal.

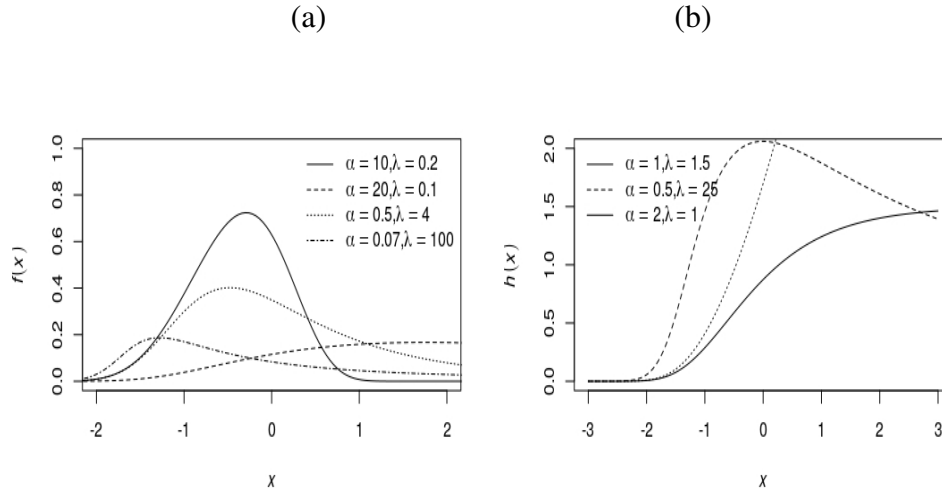


Figura 8 – Gráficos para a densidade e a função taxa de falha da distribuição NH-Gu.

Para a geração de números aleatórios da variável aleatória (v.a) $X \sim \text{NH-Gu}(\alpha, \lambda, \mu, \beta)$, usando o método da inversão, precisamos, primeiramente, gerar números aleatórios para a variável aleatória $U \sim U(0, 1)$. A função quantílica, q , da NH-Gu, $Q_{\text{NH-Gu}}(u) = F^{-1}(u)$, pode ser expressa em termos da q da Gu, $(Q_{\text{Gu}}(\cdot))$, que é dada por

$$Q_{\text{Gu}}(u) = \mu - \beta \log[-\log(u)].$$

Invertendo a Equação (6.3), obtemos a fq de X , que pode ser escrita como

$$\begin{aligned} F^{-1}(u) &= Q_{NH-Gu}(u) = Q_{Gu} \left\{ 1 - \exp \left[\frac{1 - [1 - \log(1 - u)]^{1/\alpha}}{\lambda} \right] \right\} \\ &= \mu - \beta \log \left\{ -\log \left\{ 1 - \exp \left[\frac{1 - [1 - \log(1 - u)]^{1/\alpha}}{\lambda} \right] \right\} \right\}, \end{aligned} \quad (6.7)$$

para $0 < u < 1$.

Da Equação (6.7) e tomando $u = 0.5$, obtemos a expressão para a mediana da variável aleatória, que é dada a seguir

$$F^{-1}(0.5) = \mu - \beta \log \left\{ -\log \left\{ 1 - \exp \left[\frac{1 - [1.6931]^{1/\alpha}}{\lambda} \right] \right\} \right\}$$

6.2 EXPANSÕES UTILIZADAS

Para a obtermos algumas propriedades matemáticas dessa distribuição, utilizamos algumas expansões em séries. Dias (2016) mostrou que a NH-G possui propriedades derivadas das "Exponencializada-G" (exp-G). Com isso podemos escrever as expressões (6.3) e (6.4) como uma combinação de exp-G e, assim, obter as propriedades estruturais da distribuição NH-Gu.

As propriedades das distribuições exponencializadas foram estudadas por diversos autores nos últimos anos, ver Nadarajah e Kotz (2006), Nadarajah (2006). Com base em Dias (2016), podemos escrever a expressão (6.2) como

$$f(x) = \sum_{m=0}^{\infty} w_m h_{m+1}(x), \quad (6.8)$$

em que $h_{m+1}(x) = (m+1)g(x)G(x)^m$ é a fdp da exp-Gu($m+1, \mu, \beta$), de parâmetro de exponencialização $m+1$. Para mais detalhes sobre (6.8) e sobre seus coeficientes w'_m , veja (DIAS, 2016). Com base na Equação (6.8) várias propriedades estruturais da distribuição NH-Gu podem ser obtidas conhecendo-se as respectivas propriedades da distribuição exp-Gu.

6.3 ALGUMAS PROPRIEDADES

Nessa seção, mostramos algumas propriedades obtidas do modelo proposto.

6.3.1 Momentos

Como Dias (2016) mostrou, a classe de distribuições NH-G goza de propriedades das exponencializadas, ou seja, podemos encontrar as propriedades da NH-Gu, com base nas propriedades da exp-Gu.

Seja Y_m uma variável aleatória seguindo distribuição exp-Gu($m+1, \mu, \beta$), com densidade $h_{m+1}(x)$. O n -ésimo momento de X pode ser expresso de (6.8) por

$$E(X^n) = \int_{-\infty}^{\infty} x^n \sum_{m=0}^{\infty} w_m h_{m+1} dx. \quad (6.9)$$

De Andrade *et. al* (2015) o n -ésimo momento de Y_m é dado por

$$E(X^n) = \sum_{m=0}^{\infty} \sum_{i=0}^n (m+1) \binom{n}{i} (-\beta)^i \mu^{n-i} w_m \left(\frac{\partial}{\partial a} \right)^i [(m+1)^{-a} \Gamma(a)]|_{a=1} \quad (6.10)$$

em que $\Gamma(\cdot)$ denota a função gama. Assim, substituindo (6.10) em (6.9) nós obtemos o n -ésimo momento de X . Para mais detalhes sobre a Equação (6.10), veja Andrade *et al.* (2015).

6.3.2 Função geradora de momentos

Com base na Equação (6.2), denotamos a função geradora de momentos (fgm) de X , $M_X(t)$, por

$$M_X(t) = \sum_{m=0}^{\infty} w_m M_{m+1}(t),$$

em que $M_{m+1}(t)$ é a função geradora de momentos da exp-Gu($m+1, \mu, \beta$).

A fgm de X é obtida de Andrade *et al.* (2015) como

$$M(t) = \exp(t\mu) \Gamma(1-t\beta) \sum_{m=0}^{\infty} (m+1)^{t\beta} w_m.$$

6.3.3 Entropia

A entropia de Shannon de uma variável aleatória X é definida por $E\{-\log[f(X)]\}$. Utilizando a equação (6.4), obtemos

$$\begin{aligned} E\{-\log[f(Z)]\} &= -\log(\alpha) - \log(\lambda) + \log(\beta) + E(Z) + E[\exp(-Z)] - 1 \\ &+ E\{[1 - \lambda \log[1 - \exp[-\exp(-Z)]]]^\alpha\} \\ &- (\alpha - 1)E\{\log[1 - \lambda \log[1 - \exp[-\exp(-Z)]]]\} \\ &+ E\{\log[1 - \exp[-\exp(-Z)]]\}, \end{aligned} \quad (6.11)$$

em que $Z = (X - \mu)/\beta$.

Se $\alpha = 1$ na Equação (6.11), temos a seguinte expressão

$$E\{-\log[f(Z)]\} = -\log(\lambda) + \log(\beta) + E(Z) + E[\exp(-Z)] + (1 - \lambda)E\{\log[1 - \exp[-\exp(-Z)]]\}.$$

6.3.4 Estatísticas de ordem

No estudo de valores discrepantes dos dados, o uso de estatísticas de ordem se faz necessário. As estatísticas de ordem podem ser encontradas em problemas de engenharia e controle de qualidade. Suponha que $X_1, \dots, X_n$ seja uma amostra de variáveis aleatórias que seguem a distribuição NH-Gu. Denote por $X_{1:n}, \dots, X_{n:n}$ as estatísticas de ordem. A fdp da i -ésima estatística de ordem de uma amostra aleatória $X_1, \dots, X_n$ da distribuição NH-Gu é dada por

$$f_{i:n}(x) = K f(x) F^{i-1}(x) [1 - F(x)]^{n-i},$$

em que $K = n! / [(i-1)!(n-i)!]$, $F(x)$ e $f(x)$ são a acumulada (6.3) e a densidade (6.4), respectivamente, da NH-Gu. Então,

$$f_{i:n}(x) = K \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} f(x) F^{i+j-1}(x).$$

Integrando a Equação (6.2) e utilizando o teorema da convergência monótona, a fda de X é expressa por

$$F(x) = \sum_{m=0}^{\infty} w_m H_{m+1}(x), \quad (6.12)$$

em que $H_{m+1}(x) = \left\{ \exp \left[-\exp \left(-\frac{x-\mu}{\beta} \right) \right] \right\}^{m+1}$ é a acumulada de Y_{m+1} . A função densidade de probabilidade $X_{i:n}$ pode ser expressada de (6.2) e (6.12) como

$$\begin{aligned} f_{i:n}(x) &= K \sum_{j=0}^{n-1} (-1)^j \binom{n-i}{j} \left\{ \sum_{r=0}^{\infty} w_r (r+1) \left\{ \exp \left[-\exp \left(-\frac{x-\mu}{\beta} \right) \right] \right\}^r \left\{ \beta^{-1} \exp \left(-\frac{x-\mu}{\beta} \right) \right. \right. \\ &\quad \times \left. \left. \exp \left[-\exp \left(-\frac{x-\mu}{\beta} \right) \right] \right\} \right\} \left\{ \sum_{m=0}^{\infty} w_m \left\{ \exp \left[-\exp \left(-\frac{x-\mu}{\beta} \right) \right] \right\}^{m+1} \right\}^{i+j-1}. \end{aligned}$$

Usando séries de potência para um inteiro positivo, temos

$$\left\{ \sum_{m=0}^{\infty} w_m \left\{ \exp \left[-\exp \left(-\frac{x-\mu}{\beta} \right) \right] \right\}^{m+1} \right\}^{i+j-1} = \sum_{m=0}^{\infty} f_{i+j-1,m} \left\{ \exp \left[-\exp \left(-\frac{x-\mu}{\beta} \right) \right] \right\}^{i+j+m-1}.$$

em que $f_{i+j-1,0} = w_0^{i+j-1}$ e (para $m \geq 1$)

$$f_{i+j-1,m} = (mw_0)^{-1} \sum_{p=1}^m [p(i+j) - m] w_p f_{i+j-1,m-p}.$$

Consequentemente

$$f_{i:n}(x) = \sum_{j=0}^{n-1} \sum_{r,m=0}^{\infty} c_{j,r,m} h_{i+j+r+m}(x), \quad (6.13)$$

em que

$$c_{j,r,n} = \frac{(-1)^j n!}{(i-1)!(n-1-j)!j!} \frac{(r+1)w_r f_{i+j-1,m}}{(i+j+r+m)}.$$

A Equação (6.13) é o principal resultado dessa seção e nos mostra que a densidade das estatísticas de ordem é escrita como uma combinação linear tripla de densidades da distribuição exp-Gu. Portanto, podemos observar que algumas propriedades matemáticas (estatísticas de ordem, momentos, função geradora de momentos) podem ser obtidas através das propriedades da exp-Gu

6.4 INFERÊNCIA E ESTIMAÇÃO

Aqui, consideramos a estimação dos parâmetros desconhecidos da distribuição NH-Gu pelo método da máxima verossimilhança. Seja $x_1, \dots, x_n$ uma amostra aleatória de tamanho n da distribuição NH-Gu $(\alpha, \lambda, \mu, \beta)$. A função de log-verossimilhança para o vetor de parâmetros $\boldsymbol{\theta} = (\alpha, \lambda, \mu, \beta)^T$ é expressa por

$$\begin{aligned} l(\boldsymbol{\theta}) &= n + n \log(\alpha) + n \log(\lambda) - n \log(\beta) - \sum_{i=1}^n \left(\frac{x_i - \mu}{\beta} \right) \\ &\quad - \sum_{i=1}^n \exp \left[-\frac{(x_i - \mu)}{\beta} \right] - \sum_{i=1}^n \log \left\{ 1 - \exp \left[-\exp \left(-\frac{(x_i - \mu)}{\beta} \right) \right] \right\} \\ &\quad + (\alpha - 1) \sum_{i=1}^n \log \left\{ 1 - \lambda \log \left[1 - \exp \left(-\exp \left(-\frac{x_i - \mu}{\beta} \right) \right) \right] \right\} \\ &\quad - \sum_{i=1}^n \left\{ 1 - \lambda \log \left[1 - \exp \left(-\exp \left(-\frac{x_i - \mu}{\beta} \right) \right) \right] \right\}^\alpha. \end{aligned} \quad (6.14)$$

A maximização de (6.14) é realizada utilizando a ferramenta estatística R. Derivando (6.14) em relação aos parâmetros e igualando a zero, e chamando de U , temos $U_\alpha(\boldsymbol{\theta}) = 0$, $U_\lambda(\boldsymbol{\theta}) = 0$, $U_\mu(\boldsymbol{\theta}) = 0$, $U_\beta(\boldsymbol{\theta}) = 0$, e resolvendo-os simultaneamente obtemos os estimadores de máxima verossimilhança (EMV), $\hat{\boldsymbol{\theta}}$, de $\boldsymbol{\theta}$.

Como, geralmente, esse sistema é não linear, torna-se necessário usar um algoritmo de otimização. Por exemplo, o algoritmo Newton-Raphson ou quase-Newton Nocedal e Wright (2006); Press *et al.* (2007)

O vetor escore tem como componentes as seguintes equações

$$\begin{aligned} U_\alpha(\boldsymbol{\theta}) &= \frac{1}{\alpha} - \log \{ 1 - \lambda \log [1 - \exp(B)] \} \{ 1 - \lambda \log [1 - \exp(B)] \}^\alpha \\ &\quad + \log \{ 1 - \lambda \log [1 - \exp(B)] \}, \end{aligned}$$

$$\begin{aligned} U_\lambda(\boldsymbol{\theta}) &= \frac{1}{\lambda} + \alpha \log [1 - \exp(B)] \{ 1 - \lambda \log [1 - \exp(B)] \}^{\alpha-1} \\ &\quad - \frac{(\alpha - 1) \log [1 - \exp(B)]}{1 - \lambda \log [1 - \exp(B)]}, \end{aligned}$$

$$\begin{aligned}
U_{\beta}(\boldsymbol{\theta}) &= -\frac{1}{\beta} + \frac{\alpha\lambda(x-\mu)\exp[-\frac{x-\mu}{\beta}+B]\{1-\lambda\log[1-\exp(B)]\}^{\alpha-1}}{\beta^2[1-\exp(B)]} \\
&\quad - \frac{(\alpha-1)\lambda(x-\mu)e^{-\frac{x-\mu}{\beta}+B}}{\beta^2[1-\exp(B)][1-\lambda\log(1-\exp(B))]} \\
&\quad + \frac{(x-\mu)B}{\beta^2} - \frac{(x-\mu)e^{-\frac{x-\mu}{\beta}+B}}{\beta^2[1-\exp(B)]} + \frac{x-\mu}{\beta^2} e
\end{aligned}$$

$$\begin{aligned}
U_{\mu}(\boldsymbol{\theta}) &= \frac{1}{\beta} + \frac{\alpha\lambda B + B(1-\lambda\log(1-\exp(B)))^{\alpha-1}}{\beta(1-\exp(B))} \\
&\quad - \frac{(\alpha-1)\lambda - \exp(-\frac{x-\mu}{\beta}+B)}{\beta[1-\exp(B)][1-\lambda\log(1-\frac{x-\mu}{\beta}+B)]} - \frac{B}{\beta} \\
&\quad - \frac{-\exp(-\frac{x-\mu}{\beta}+B)}{\beta[1-\exp(B)]}.
\end{aligned}$$

em que $B = -\exp(-\frac{x-\mu}{\beta})$

6.4.1 Estudo de simulação

Neste capítulo, é apresentado um estudo de simulação para verificar se os estimadores de máxima verossimilhança satisfazem as suas propriedades assintóticas. A verossimilhança usada, inclui censura a direita com mecanismo aleatório, para verificarmos o que acontece se introduzirmos a censura.

Para verificação dessas propriedades, é necessário observar se ao aumentar o tamanho da amostra, a média das estimativas tendem ao valor verdadeiro assim como o viés e o EQM tendem a zero. Para a censura, é necessário observar se ao aumentarmos o seu percentual, o valor do viés aumenta em relação ao percentual de menor censura.

Os tamanhos de amostra escolhidos são de $\{80, 200, 500\}$ e os percentuais de censura são de $\{0\%, 10\%, 20\%\}$. A verossimilhança utilizada para essa maximização é um pouco diferente da mostrada em 6.14. Nesta consideramos censura a direita com mecanismo aleatório; a função de verossimilhança é mostrada a seguir

$$L(\boldsymbol{\theta}) \propto \prod_{i=1}^n [f(t_i; \boldsymbol{\theta})]^{\delta_i} [S(t_i; \boldsymbol{\theta})]^{1-\delta_i}, \quad (6.15)$$

em que $f(t_i; \boldsymbol{\theta})$, $S(t_i; \boldsymbol{\theta})$ e δ_i são, respectivamente, função densidade, função de sobrevivência e variável indicadora que representa tempo de falha ou censura

$$\delta_i = \begin{cases} 1, & \text{se é tempo de falha} \\ 0, & \text{se é tempo de censura} \end{cases} \quad (6.16)$$

. Para mais detalhes sobre as Equações (6.15) e (6.16) ver Colosimo e Giolo (2006).

Os estimadores de máxima verossimilhança são os valores de θ que maximizam $L(\theta)$; maximizar $L(\theta)$ é equivalente a maximizar o logaritmo $L(\theta)$, ou seja, $\log(L(\theta))$ e após isso resolvendo os sistema de equações:

$$U(\theta) = \frac{\partial \log(L(\theta))}{\partial \theta} = 0. \quad (6.17)$$

O método de simulação de Monte Carlo serve para a verificação da lei fraca dos grandes números, que nos diz que

$$\frac{\sum_{i=1}^n x_i}{n} \xrightarrow{P} \mu, \quad (6.18)$$

ou seja, quando o tamanho da amostra aumenta a média das estimativas tende ao valor verdadeiro.

Para a simulação feita, consideramos o seguinte algoritmo

- Geramos de forma independente, números aleatórios da distribuição $U \sim U(0, 1)$;
- Utilizando o método da inversão, geramos números aleatórios para a distribuição NH-Gu;
- Para determinarmos os elementos que possuem ou não censura, geramos ocorrências da distribuição Bernouli em que a a probabilidade de sucesso é $p = 1 - C$, em que C é o percentual de censura, dessa forma temos uma coleção de Bernoulli;
- Maximizamos a Expressão (6.15);
- Replicamos a maximização mil vezes;
- Obtemos as médias das estimativas, viés e EQM.

As quantidades observadas para o estudo foram a média das estimativas do vetor $\hat{\theta} = (\hat{\alpha}, \hat{\lambda}, \hat{\beta}, \hat{\mu})$ e seus respectivos Erros Quadráticos Médios (EQM). Consideramos quatro cenários: o primeiro com $(\alpha, \lambda, \beta, \mu) = (1.5, 2.3, 1, 0.5)$ e o segundo com $(\alpha, \lambda, \beta, \mu) = (1.0, 4.0, 1.0, 1.0)$, o terceiro com $(\alpha, \lambda, \beta, \mu) = (2, 3, 2.1, 2.5)$ e o quarto com $(\alpha, \lambda, \beta, \mu) = (4, 1.7, 3, 2.8)$. Os resultados são mostrados nas Tabelas 16, 17, 18 e 19.

Tabela 16 – Resultados da simulação com $\alpha = 1.5, \lambda = 2.3, \mu = 1, \beta = 0.5$

Censura	n	$\text{Est}(\hat{\alpha}, \hat{\lambda}, \hat{\mu}, \hat{\beta})$	$\text{EQM}(\hat{\alpha}, \hat{\lambda}, \hat{\mu}, \hat{\beta})$	
C=0%	80	(1.5098, 2.3187, 1.0002, 0.4995)	(0.0098, 0.0187, 0.0002, -0.0004)	(0.0108, 0.0526, 0.0009, 0.0008)
	200	(1.5040, 2.3062, 1.0006, 0.4993)	(0.0040, 0.0062, 0.0006, -0.0006)	(0.0044, 0.0214, 0.0003, 0.0003)
	500	(1.5004, 2.2997, 1.0004, 0.4996)	(0.0004, -0.0002, 0.0004, -0.0003)	(0.0016, 0.0079, 0.0001, 0.0001)
C=10%	80	(1.4036, 2.0882, 1.0276, 0.4933)	(-0.0963, -0.2117, 0.0276, -0.0066)	0.0177, 0.0835, 0.0016, 0.0011)
	200	(1.4120, 2.1048, 1.0250, 0.4936)	(-0.0879, -0.1951, 0.0250, -0.0063)	(0.0116, 0.0559, 0.0010, 0.0003)
	500	(1.4088, 2.0989, 1.0250, 0.4939)	(-0.0911, -0.2010, 0.0250, -0.0060)	(0.0097, 0.0466, 0.0007, 0.0001)
C=20%	80	(1.3172, 1.9094, 1.0515, 0.4871)	(-0.1827, -0.3905, 0.0515, -0.0128)	(0.0416, 0.1880, 0.0035, 0.0010)
	200	(1.3122, 1.8989, 1.0518, 0.4870)	(-0.1878, -0.4010, 0.0518, -0.0129)	(0.0386, 0.1753, 0.0030, 0.0005)
	500	(1.3093, 1.8938, 1.0516, 0.4874)	(-0.1906, -0.4061, 0.0516, -0.0125)	(0.0376, 0.1703, 0.0028, 0.0003)

Tabela 17 – Resultados da simulação com $\alpha = 1, \lambda = 4, \mu = 1, \beta = 1$

Censura	n	$\text{Est}(\hat{\alpha}, \hat{\lambda}, \hat{\mu}, \hat{\beta})$	$\text{Viés}(\hat{\alpha}, \hat{\lambda}, \hat{\mu}, \hat{\beta})$	$\text{EQM}(\hat{\alpha}, \hat{\lambda}, \hat{\mu}, \hat{\beta})$
C=0%	80	(1.0064, 4.0376, 1.0009, 0.9990)	(0.0064, 0.0376, 0.0009, -0.0009)	(0.0048, 0.2141, 0.0041, 0.0032)
	200	(1.0026, 4.0113, 1.0015, 0.9989)	(0.0026, 0.0113, 0.0015, -0.0010)	(0.0019, 0.0868, 0.0016, 0.0013)
	500	(1.0002, 3.9993, 1.0009, 0.9996)	(0.0002, -0.0006, 0.0009, -0.0003)	(0.0007, 0.0318, 0.0006, 0.0005)
C=10%	80	(0.9367, 3.5830, 1.0554, 0.9878)	(-0.0632, -0.4169, 0.0554, -0.0121)	0.00819, 0.3416, 0.0073, 0.0036)
	200	(0.9316, 3.5498, 1.0576, 0.9873)	(-0.0684, -0.4501, 0.0576, -0.0126)	0.0064, 0.2706, 0.0050, 0.0016)
	500	(0.9391, 3.5994, 1.0496, 0.9894)	(-0.0608, -0.4005, 0.0496, -0.0105)	(0.0043, 0.1862, 0.0031, 0.0007)
C=20%	80	(0.8425, 3.0277, 1.1324, 0.9687)	(-0.1574, -0.9722, 0.1324, -0.0312)	(0.0282, 1.0651, 0.0220, 0.0048)
	200	(0.8534, 3.0885, 1.1200, 0.9732)	(0.14651, -0.9114, 0.1200, -0.0267)	(0.0229, 0.8822, 0.01623, 0.0023)
	500	(0.8618, 3.1355, 1.1119, 0.9750)	(-0.1381, -0.8644, 0.1119, -0.0249)	(0.0196, 0.7668, 0.0132, 0.0013)

Tabela 18 – Resultados da simulação com $\alpha = 2, \lambda = 3, \mu = 2.1, \beta = 2.5$

Censura	n	$\text{Est}(\hat{\alpha}, \hat{\lambda}, \hat{\mu}, \hat{\beta})$	$\text{Viés}(\hat{\alpha}, \hat{\lambda}, \hat{\mu}, \hat{\beta})$	$\text{EQM}(\hat{\alpha}, \hat{\lambda}, \hat{\mu}, \hat{\beta})$
Censura = 0%	80	(2.0242, 3.0426, 2.0935, 2.4985)	(0.0242, 0.0426, -0.0064, -0.0014)	(0.0188, 0.0742, 0.0125, 0.0134)
	200	(2.0100, 3.0170, 2.0974, 2.4990)	(0.0100, 0.0170, -0.0025, -0.0009)	(0.0075, 0.0290, 0.0049, 0.0054)
	500	(2.0028, 3.0064, 2.0992, 2.5005)	(0.0028, 0.0064, -0.0007, 0.0005)	(0.0029, 0.0111, 0.0019, 0.0020)
Censura = 10%	80	(1.8344, 2.6721, 2.2449, 2.4071)	(-0.1655, -0.3278, 0.1449, -0.0928)	(0.0430, 0.1642, 0.0338, 0.0225)
	200	(1.8355, 2.6733, 2.2382, 2.4132)	(-0.1644, -0.3266, 0.1382, -0.0867)	(0.0333, 0.1289, 0.0241, 0.0131)
	500	(1.8580, 2.7187, 2.2154, 2.4291)	(-0.1419, -0.2812, 0.1154, -0.0708)	(0.0228, 0.0880, 0.0154, 0.0072)
Censura=20%	80	(1.6184, 2.2778, 2.4177, 2.3045)	(-0.3815, -0.7221, 0.3177, -0.1954)	(0.1581, 0.5623, 0.1140, 0.0531)
	200	(1.7039, 2.4307, 2.3440, 2.3501)	(-0.2960, -0.5692, 0.2440, -0.1498)	(0.0931, 0.3425, 0.0647, 0.0284)
	500	(1.7247, 2.4656, 2.3248, 2.3625)	(-0.2752, -0.5343, 0.2248, -0.1374)	(0.0781, 0.2930, 0.0525, 0.0210)

Tabela 19 – Resultados da simulação com $\alpha = 4, \lambda = 1.7, \mu = 3, \beta = 2.8$

Censura	n	$\text{Est}(\hat{\alpha}, \hat{\lambda}, \hat{\mu}, \hat{\beta})$	$\text{Viés}(\hat{\alpha}, \hat{\lambda}, \hat{\mu}, \hat{\beta})$	$\text{EQM}(\hat{\alpha}, \hat{\lambda}, \hat{\mu}, \hat{\beta})$
Censura = 0%	80	(4.0482, 1.7225, 2.9923, 2.8003)	(0.0482, 0.0225, -0.0076, 0.0003)	(0.0757, 0.0185, 0.0108, 0.0129)
	200	(4.0195, 1.7091, 2.9973, 2.8001)	(0.0195, 0.0091, -0.0026, 0.0001)	(0.0295, 0.0072, 0.0043, 0.0050)
	500	(4.0078, 1.7030, 2.9990, 2.8008)	(0.0078, 0.0030, -0.0009, 0.0008)	(0.0115, 0.0028, 0.0017, 0.0020)
Censura = 10%	80	(3.6687, 1.5364, 3.1392, 2.6876)	(-0.3312, -0.1635, 0.1392, -0.1123)	(0.1713, 0.0413, 0.0303, 0.0255)
	200	(3.6710, 1.5380, 3.1336, 2.6936)	(-0.3289, -0.1619, 0.1336, -0.1063)	(0.1326, 0.0321, 0.0222, 0.0165)
	500	(3.7169, 1.5608, 3.1119, 2.7114)	(-0.2830, -0.1391, 0.1119, -0.0885)	(0.0900, 0.0218, 0.0143, 0.0098)
Censura=20%	80	(3.2377, 1.3333, 3.3086, 2.5638)	(-0.7622, -0.3666, 0.3086, -0.2361)	(0.6305, 0.14561, 0.1062, 0.0688)
	200	(3.4086, 1.4121, 3.2366, 2.6164)	(-0.5913, -0.2878, 0.2366, -0.1835)	(0.3715, 0.0881, 0.0603, 0.0389)
	500	(3.4468, 1.4316, 3.2194, 2.6322)	(-0.5531, -0.2683, 0.2194, -0.1677)	(0.3145, 0.0741, 0.0499, 0.0301)

Podemos observar nesse estudo de simulação que as propriedades assintóticas são satisfeitas, pois ao aumentarmos o tamanho da amostra o viés das estimativas diminui, assim como o valor do EQM. É possível observar, também, que ao aumentarmos o percentual de censura, o viés aumenta em relação ao valor que possui menos censura, como era de se esperar.

6.5 APLICAÇÃO A DADOS REAIS

Para ilustrar a aplicação da metodologia estudada, utilizou-se dois conjuntos de dados. Ambos dados sem censura. O objetivo da aplicação a dados reais é verificar empiricamente que para os bancos de dados escolhidos o modelo paramétrico proposto, NH-Gu, é superior a alguns modelos já conhecidos na literatura. O critério de comparação é feito com base no pacote *Adequacy Model* proposto por Marinho (2016), que nos fornece as estatísticas de teste, Anderson Darling, A^* e Cramer-Von-Misses, W^* . Todas essas calculadas através do método de otimização *Simulated Annealing* (SANN). Os ajustes são comparados com a distribuição-base e com algumas outras extensões dessa distribuição-base, que são

1. Gumbel (Gu) (GUMBEL, 1958),
2. Gamma Gumbel (Gamma-Gu) , ver Zografos e Balakrishnan (2009)
3. Beta Gumbel (Beta-Gu) (NADARAJAH; KOTZ, 2004),
4. Gumbel Exponencializada (Exp-Gu) (NADARAJAH, 2006),
5. Gumbel Generalizada Exponencializada (EG-Gu)(ANDRADE *et al.*, 2015).

6.5.1 Conjunto de dados 1: porcos da Índia infectados com vírus da tuberculose

Diversos estudos são feitos com cobaias afim de analisar a viabilidade de tratamentos de doenças em humanos, como é o caso de estudos sobre a tuberculose, por exemplo. Devido a sua alta susceptibilidade à tuberculose, o estudo com porcos da Índia pode ser encontrado na literatura, em diversos livros e artigos. Sobre esse tema, Andrade *et al.* (2006) nos traz um pouco da história de como foram os primeiros estudos com cobaias. Bjerkedal *et al.* (1960) mostra um experimento com cobaias infectadas com o vírus da tuberculose. Nesse estudo, foi avaliado o tempo até a morte em diversos cenários.

A tuberculose foi uma doença que matou muitas pessoas durante as décadas passadas. Apesar do decréscimo nos casos, ainda há uma quantidade considerável de indivíduos infectados no Brasil, segundo dados do Ministério da Saúde. A melhor forma é a prevenção da doença. Uma das melhores formas de prevenção em crianças é tomar a vacina BCG. Esses dados e outras formas de prevenção podem ser encontrados no portal do Ministério da Saúde (<http://portalms.saude.gov.br/saude-de-a-z/tuberculose#prevencao>).

A proposta de modelos que possam ajustar bancos de dados desse tipo são, portanto, úteis. Para a primeira aplicação a dados reais, consideramos o conjunto de dados de Bjerkedal *et al.* (1960),

dividindo as observações por 1000. Estes dados (medidos em dias/1000) é sobre o tempo até a morte de 72 porcos da índia com o vírus *tubercle bacilli*.

O gráfico TTT, Aarset (1987), apresentado na Figura 9 nos mostra uma curva côncava, porém com uma pequena parte final abaixo da reta diagonal indicando que a função de taxa de falha empírica é unimodal. A distribuição NH-Gu possui essa forma na ftf, então se mostra um bom modelo para ajustar esses dados.

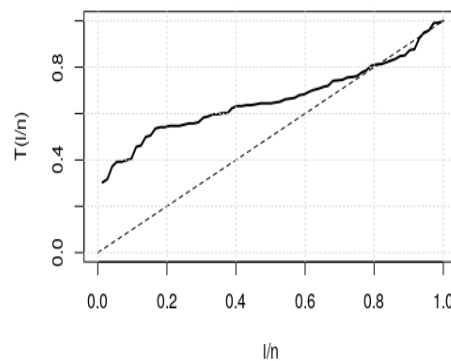


Figura 9 – TTT plot

Algumas estatísticas descritivas desses dados são fornecidas na Tabela 20. As quantidades observadas nos mostram uma variância bem pequena e assimetria e curtose positivas.

Tabela 20 – Estatísticas descritivas - primeira aplicação

Aplicação	Media	Mediana	Variância	Assimetria	Curtose	Min.	Max.
Porcos da índia	0.141	0.102	0.011	2.463	6.074	0.043	0.598

A Tabela 21 nos fornece os valores dos EMV's dos parâmetros e as quantidades para Erro Padrão (entre parênteses) a partir dos ajustes dos modelos citados assim como as medidas W^* e A^* . Os resultados indicam que o modelo proposto possui os menores valores dessas estatísticas entre os modelos e, portanto, poderia ser escolhido como o melhor modelo para esses dados.

6.5.2 Conjunto de dados 2: fibras de carbono expostas a pressão

Por conter um baixo peso e alta rigidez, materiais com fibras de carbono são utilizados em diversos setores que buscam materiais com alto desempenho. Placas de fibra de carbono podem ser encontradas em aeronaves, meios de transporte em geral e para reforço estrutural em obras da construção civil, por exemplo. Dentre seus benefícios, uma das maiores vantagens do uso da fibra de carbono é não sofrer corrosão, que é uma vantagem em relação ao aço. Podemos ver trabalhos envolvendo fibras de

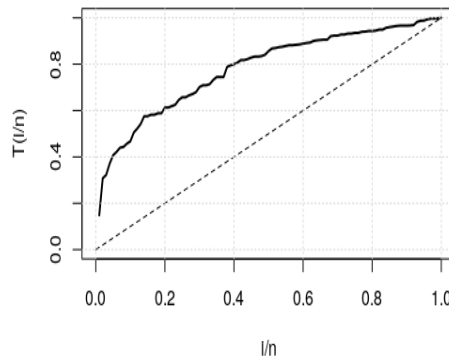
Tabela 21 – Guinea-Pig

Model	α	λ	μ	β	W^*	A^*
NH-Gu($\alpha, \lambda, \mu, \beta$)	0.2492 (0.0224)	685.3412 (0.1693)	0.3251 (0.0328)	0.1338 (0.0173)	0.0503	0.3173
Gu(μ, β)			0.1032 (0.0066)	0.0549 (0.0056)	0.4737	2.8104
Gamma-Gu(α, μ, β)	1.7140 (0.9592)		1.8497 (0.9866)	19.9790 (2.8913)	0.5141	3.0408
Exp-Gu(α, μ, β)	999.9808 (0.1667)		0.0063 (0.0016)	12.0443 (1.6563)	0.3187	1.8859
EG-Gu($\alpha, \lambda, \mu, \beta$)	106.4632 (0.0039)	147.7822 (0.0122)	0.3980 (0.0339)	0.0033 (0.0022)	0.1718	1.0014
Beta-Gu($\alpha, \lambda, \mu, \beta$)	354.4642 (18.0132)	818.5109 (14.3995)	0.6576 (0.1160)	2.8703 (0.2420)	0.1877	1.1014

carbono em Souza *et al.* (2016), que estudou o comportamento das fibras de carbono elevadas a diferentes temperaturas.

Apesar dos estudos mostrando o grande benefício do uso de fibras de carbono, se faz necessário estudos que procurem verificar o quão essas fibras são resistentes. Sobre o estudo da resistência das fibras de carbono, Nichols e Padgett (2006) analisam o limite de força exercida até a ruptura de fibras de carbono, em GPa. Esse conjunto de dados, constitui o estudo da nossa segunda aplicação. As estatísticas descritivas do banco de dados é mostrada na Tabela 22 e nos mostrando uma variância pequena e assimetria e curtose positivas, porém pequenas

O gráfico TTT, Aarset (1987), apresentada na Figura 10 nos mostra uma curva predominantemente côncava indicando que a função de taxa de falha empírica é crescente. A distribuição NH-Gu possui essa forma na ftf, mostrando-se, então, um bom modelo para ajustar esses dados.

**Figura 10 – TTT plot**

As quantidades mostradas na tabela 23 são as estimativas e seus respectivos erros padrão,

Tabela 22 – Estatísticas descritivas - Segunda Aplicação

Aplicação	Media	Mediana	Variância	Assimetria	Curtose	Min.	Max.
Carbono	2.621	2.700	1.027	0.362	0.043	0.390	5.560

assim como as estatísticas de teste, W^* e A^* . Os resultados obtidos comprovam que o modelo proposto se mostra superior, quando comparado aos outros, para esse banco de dados.

Tabela 23 – Fibras de Carbono expostas a pressão

Model	α	λ	μ	β	W^*	A^*
NH-Gu($\alpha, \lambda, \mu, \beta$)	0.5149 (0.0774)	348.3888 (14.1494)	8.6359 (1.0455)	3.6317 (0.5137)	0.0728	0.4135
Gu(μ, β)			2.1327 (0.0962)	0.9090 (0.0681)	0.1837	0.9370
Gamma-Gu(α, μ, β)	0.9777 (0.4105)		5.5495 (1.8246)	2.1045 (0.3638)	0.1467	0.7508
Exp-Gu(α, μ, β)	0.1326 (0.0418)		33.6213 (8.4464)	8.1469 (1.7669)	0.1275	0.6547
EG-Gu($\alpha, \lambda, \mu, \beta$)	10.5246 (0.0295)	0.7205 (0.0907)	4.8403 (0.0100)	0.7072 (0.0329)	0.1799	0.9250
Beta-Gu($\alpha, \lambda, \mu, \beta$)	125.4398 (<0.00001)	679.1719 (<0.0001)	16.7752 (0.6310)	22.8061 (1.0084)	0.7007	3.9775

A tabela 24 lista a estatística RV para comparar modelos ajustados em ambas as aplicações. Note que rejeitamos a hipótese nula ($H_0 : \alpha = \lambda = 1$) nas duas aplicações indicando indícios em favor da nova distribuição, considerando um nível de significância de 10%

Tabela 24 – Teste da razão de verossimilhanças

Aplicação	Estatística W	p-value
Porcos da índia	24.117	5.795087×10^{-06}
Fibras de Carbono	5.4952	0.06408147

6.6 CONCLUSÃO

Neste capítulo, propomos uma extensão do modelo de Gumbel (Gu) chamado de distribuição Nadarajah-Haghighi-Gumbel, (NH-Gu), com dois parâmetros extras. Investigamos algumas propriedades estruturais da nova distribuição, incluindo expansões para a densidade e função de distribuição acumulada, expressões explícitas para os momentos ordinários, funções geradora e quantílica. Os parâmetros do modelo são estimados via máxima verossimilhança. Finalmente, a flexibilidade do modelo proposto é ilustrado por meio de duas aplicações para conjuntos de dados reais. Uma simulação com censura é feita verificando as propriedades assintóticas dos estimadores de máxima verossimilhança.

7 CONSIDERAÇÕES FINAIS

No presente trabalho, propomos três novas distribuições de probabilidade NH-EE, NH-R e NH-Gu, com base na família NH-G, proposta por Dias (2016). Encontramos expressões explícitas para as propriedades estruturais dos modelos, como momentos, função geradora de momentos, estatísticas de ordem e função quantílica. Estimamos os parâmetros dos modelos propostos utilizando o método da máxima verossimilhança, achando expressões para o vetor escore. Através do método de Monte Carlo, realizamos um estudo de simulação e verificamos que os estimadores de máxima verossimilhança satisfazem as propriedades assintóticas para os modelos NH-R e NH-Gu. Por fim, realizamos aplicações com dados clínicos e de fadiga para os três novos modelos, verificando, assim, que para esses tipos de bancos de dados as distribuições se mostram competitivas.

7.1 TRABALHOS FUTUROS

Dando continuidade a pesquisa apresentada, pretendemos estender o pacote estatístico *AdequacyModel*, no R, visando realizar aplicações a dados com censura.

Para a NH-EE pretendemos realizar um estudo de simulação com censura, semelhante ao realizado para as outras distribuições propostas.

Outra pesquisa, que visamos dar continuidade, é a proposta de modelos de regressão para essas distribuições, pois sentimos a necessidade contribuir na literatura com modelos que possam trabalhar com covariáveis.

Já é bastante observado que os estimadores de máxima verossimilhança para pequenas amostras costumam ser viesados. Dado essa deficiência, uma outra pesquisa que temos interesse em realizar é a correção do viés desse estimador de máxima verossimilhança, para os modelos propostos. Pretendemos realizar a correção de forma analítica, obtendo expressões para o cálculo da correção do viés, assim como correções via reamostragem bootstrap.

7.2 CONTRIBUIÇÕES DAS PESQUISAS

- Apresentação de pôster no Simpósio Nacional de Probabilidade e Estatística: 23º SINAPE THE NADARAJAH-HAGHIHI EXPONENTIAL DISTRIBUTION: PROPERTIES AND APPLICATIONS. 2018. (Apresentação de Trabalho/Simpósio).
- Apresentação oral na Jornada Científica Júnior- UFPE SILVA, J. J. S.; LIMA, M. C. S. . DISTRIBUIÇÃO NADARAJAH-HAGHIGHI GUMBEL: PROPRIEDADES E APLICAÇÕES. 2018. (Apresentação de Trabalho/Colóquio).

REFERÊNCIAS

- AARSET, M. V. How to identify a bathtub hazard rate. **IEEE Transactions on Reliability**, IEEE, v. 36, n. 1, p. 106–108, 1987.
- AHUJA, J.; NASH, S. W. The generalized gompertz-verhulst family of distributions. **Sankhyā: The Indian Journal of Statistics, Series A**, JSTOR, p. 141–156, 1967.
- AL-HUSSAINI, E. K.; AHSANULLAH, M. **Exponentiated distributions**. [S.l.]: Springer, 2015. v. 21.
- AL-HUSSAINI, E. K.; HUSSEIN, M. Estimation using censored data from exponentiated burr type xii population. **American Open Journal of Statistics**, v. 1, p. 33–45, 2011.
- ALEXANDER, C. *et al.* Generalized beta-generated distributions. **Computational Statistics & Data Analysis**, Elsevier, v. 56, n. 6, p. 1880–1897, 2012.
- ALJARRAH, M. A. *et al.* On generating T-X family of distributions using quantile functions. **Journal of Statistical Distributions and Applications**, Springer, v. 1, n. 1, p. 2, 2014.
- ALZAATREH, A.; GHOSH, I. On the Weibull-x family of distributions. **Journal of Statistical Theory and Applications**, Atlantis Press, v. 14, n. 2, p. 169–183, 2015.
- ALZAATREH, A. *et al.* A new method for generating families of continuous distributions. **METRON**, v. 71, 08 2013.
- ANDRADE, A. *et al.* **Animais de laboratório: criação e experimentação**. [S.l.]: SciELO-Editora FIOCRUZ, 2006.
- ANDRADE, T. *et al.* The exponentiated generalized gumbel distribution. **Revista Colombiana de Estadística**, Universidad Nacional de Colombia., v. 38, n. 1, p. 123–143, 2015.
- BEAUDOIN, J. J. **Handbook of Fiber-Reinforced Concrete. Principles, Properties, Developments and Applications**. [S.l.: s.n.], 1990.
- BIRNBAUM, Z. W.; SAUNDERS, S. C. A new family of life distributions. **Journal of applied probability**, Cambridge University Press, v. 6, n. 2, p. 319–327, 1969.
- BJERKEDAL, T. *et al.* Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli. **American Journal of Hygiene**, v. 72, n. 1, p. 130–48, 1960.
- BOLFARINE, H.; SANDOVAL, M. C. **Introdução à inferência estatística**. [S.l.]: SBM, 2001. v. 2.
- BOURGUIGNON, M. *et al.* The weibull-g family of probability distributions. **Journal of Data Science**, v. 12, n. 1, p. 53–68, 2014.
- BRITO, C. R. D. *et al.* Method for generating distributions and classes of probability distributions: the univariate case. **Hacettepe Journal of Mathematics and Statistics**, v. 48, n. 3, p. 897–930, 2019.
- CHEN, G.; BALAKRISHNAN, N. A general purpose approximate goodness-of-fit test. **Journal of Quality Technology**, Taylor & Francis, v. 27, n. 2, p. 154–161, 1995.
- COLOSIMO, E.; GIOLO, S. **Análise de sobrevivência aplicada**. 1ª edição. **São Paulo: Editora Edgard Blücher**, 2006.
- CORDEIRO, G. *et al.* A new extended Birnbaum–Saunders model: Properties, regression and applications. **Stats**, Multidisciplinary Digital Publishing Institute, v. 1, n. 1, p. 32–47, 2018.
- CORDEIRO, G. M.; CASTRO, M. de. A new family of generalized distributions. **Journal of statistical computation and simulation**, Taylor & Francis, v. 81, n. 7, p. 883–898, 2011.

- CORDEIRO, G. M. *et al.* An extended Birnbaum–Saunders distribution: Theory, estimation, and applications. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 45, n. 8, p. 2268–2297, 2016.
- CORDEIRO, G. M. *et al.* The kumaraswamy gumbel distribution. **Statistical Methods & Applications**, Springer, v. 21, n. 2, p. 139–168, 2012.
- CORDEIRO, G. M. *et al.* The kumaraswamy weibull distribution with application to failure data. **Journal of the Franklin Institute**, Elsevier, v. 347, n. 8, p. 1399–1429, 2010.
- CORDEIRO, G. M. *et al.* The kumaraswamy modified weibull distribution: theory and applications. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 84, n. 7, p. 1387–1411, 2014.
- COX, D. R.; SNELL, E. J. A general definition of residuals. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 30, n. 2, p. 248–265, 1968.
- CRUZ, J. D. *et al.* A new family of distributions: Theory and applications in survival analysis. submitted. **Statistics**, 2014.
- D’AGOSTINO, R. B. **Goodness-of-fit-techniques**. [S.l.]: CRC press, 1986. v. 68.
- DEDIVITIS, R. A. *et al.* Características clínico-epidemiológicas no carcinoma espinocelular de boca e orofaringe. **Rev Bras Otorrinolaringol**, SciELO Brasil, v. 70, n. 1, p. 35–40, 2004.
- DIAS, C. R. B. **New continuous distributions applied to lifetime data and survival analysis**. Tese (Doutorado) — Doctoral thesis, Universidade Federal de Pernambuco, 2016.
- EFRON, B.; TIBSHIRANI, R. J. An introduction to the bootstrap, volume 57 of. **Monographs on Statistics and applied probability**, p. 17, 1993.
- ERIŞOĞLU, Ü. *et al.* A mixture model of two different distributions approach to the analysis of heterogeneous survival data. **International Journal of Computational and Mathematical Sciences**, v. 5, n. 2, 2011.
- EUGENE, N. *et al.* Beta-normal distribution and its applications. **Communications in Statistics-Theory and methods**, Taylor & Francis, v. 31, n. 4, p. 497–512, 2002.
- FAMOYE, F. *et al.* The beta-weibull distribution. **Journal of Statistical Theory and Applications**, v. 4, n. 2, p. 121–136, 2005.
- FRECHET, M. Sur la loi de probabilité de l’écart maximum. **Ann. de la Soc. Polonaise de Math.**, 1927.
- GARVAN, F. **The maple book**. [S.l.]: Chapman and Hall/CRC, 2001.
- GOMES, A. E. *et al.* A new lifetime model: the kumaraswamy generalized rayleigh distribution. **Journal of statistical computation and simulation**, Taylor & Francis, v. 84, n. 2, p. 290–309, 2014.
- GÓMEZ, Y. M. *et al.* Gumbel distribution with heavy tails and applications to environmental data. **Mathematics and Computers in Simulation**, Elsevier, v. 157, p. 115–129, 2019.
- GOMPERTZ, B. Xxiv. on the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. in a letter to francis baily, esq. frs &c. **Philosophical transactions of the Royal Society of London**, The Royal Society London, n. 115, p. 513–583, 1825.
- GUMBEL, E. J. Statistical theory of floods and droughts. **Journal of the Institution of Water Engineers and Scientists**, n. 12, p. 157–184, 1958.
- GUPTA, R. D.; KUNDU, D. Theory & methods: Generalized exponential distributions. **Australian & New Zealand Journal of Statistics**, Wiley Online Library, v. 41, n. 2, p. 173–188, 1999.

GUPTA, R. D.; KUNDU, D. Exponentiated exponential family: an alternative to gamma and weibull distributions. **Biometrical Journal: Journal of Mathematical Methods in Biosciences**, Wiley Online Library, v. 43, n. 1, p. 117–130, 2001b.

IBRAHIM, J. G. *et al.* Bayesian survival analysis. **Wiley StatsRef: Statistics Reference Online**, Wiley Online Library, 2014.

JAFARI, A. A. *et al.* The beta-gompertz distribution. **Revista Colombiana de Estadística**, Universidad Nacional de Colombia., v. 37, n. 1, p. 141–158, 2014.

JAMAL, F. *et al.* Type II General Inverse Exponential family of distributions. Working paper or preprint. 2018. Disponível em: <<https://hal.archives-ouvertes.fr/hal-01914244>>.

JAMES, B. R. **Probabilidade: um curso em nível intermediário**. [S.l.: s.n.], 1996.

JOHNSON, N. L. *et al.* **Univariate discrete distributions**. [S.l.]: John Wiley & Sons, 2005. v. 444.

JOHNSTON, N. *et al.* **Continuous Univariate Distributions (Volume 2)**. [S.l.]: New York: Wiley, 1994.

KHAN, M. S.; KING, R. Transmuted modified inverse rayleigh distribution. **Austrian Journal of Statistics**, v. 44, n. 3, p. 17–29, 2015.

KOTZ, S. *et al.* **Continuous multivariate distributions, Volume 1: Models and applications**. [S.l.]: John Wiley & Sons, 2004. v. 1.

KUMARASWAMY, P. A generalized probability density function for double-bounded random processes. **Journal of Hydrology**, Elsevier, v. 46, n. 1-2, p. 79–88, 1980.

LAWLESS, J. F. **Statistical models and methods for lifetime data**. [S.l.]: John Wiley & Sons, 2011. v. 362.

LEE, E. T.; WANG, J. **Statistical methods for survival data analysis**. [S.l.]: John Wiley & Sons, 2003. v. 476.

LEHMANN, E. L. *et al.* The power of rank tests. **The Annals of Mathematical Statistics**, Institute of Mathematical Statistics, v. 24, n. 1, p. 23–43, 1953.

LEMONTE, A. J. A new exponential-type distribution with constant, decreasing, increasing, upside-down bathtub and bathtub-shaped failure rate function. **Computational Statistics & Data Analysis**, Elsevier, v. 62, p. 149–170, 2013.

LINDSAY, B. G. Mixture models: theory, geometry and applications. In: JSTOR. **NSF-CBMS regional conference series in probability and statistics**. [S.l.], 1995. p. i–163.

MAHMOUD, M.; GHAZAL, M. Estimations from the exponentiated rayleigh distribution based on generalized type-ii hybrid censored data. **Journal of the Egyptian Mathematical Society**, Elsevier, v. 25, n. 1, p. 71–78, 2017.

MANSOUR, M. *et al.* The transmuted transmuted-g family: Properties and applications. **Journal of Nonlinear Science and Applications**, v. 12, 01 2019.

MARINHO, P. R. D. Some new families of continuous distributions. Universidade Federal de Pernambuco, 2016.

MEAD, M. E. A. A note on Kumaraswamy Fréchet distribution. **Australia**, v. 8, p. 294–300, 2014.

MENDIS, S. *et al.* **Global atlas on cardiovascular disease prevention and control**. [S.l.]: Geneva: World Health Organization, 2011.

- MEROVCI, F. Transmuted generalized rayleigh distribution. **Journal of Statistics Applications & Probability**, Natural Sciences Publishing Corp, v. 3, n. 1, p. 9, 2014.
- MIRMOSTAFEE, S. *et al.* The marshall–olkin extended generalized rayleigh distribution: Properties and applications. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 46, n. 2, p. 653–671, 2017.
- MORO, J. d. S. *et al.* Oral and oropharyngeal cancer: epidemiology and survival analysis. **Einstein (São Paulo)**, SciELO Brasil, v. 16, n. 2, 2018.
- MUDHOLKAR, G. S.; HUTSON, A. D. The exponentiated weibull family: some properties and a flood data application. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 25, n. 12, p. 3059–3083, 1996.
- MUDHOLKAR, G. S.; SRIVASTAVA, D. K. Exponentiated weibull family for analyzing bathtub failure-rate data. **IEEE transactions on reliability**, IEEE, v. 42, n. 2, p. 299–302, 1993.
- MUDHOLKAR, G. S. *et al.* The exponentiated weibull family: A reanalysis of the bus-motor-failure data. **Technometrics**, Taylor & Francis, Ltd., American Statistical Association, American Society for Quality, v. 37, n. 4, p. 436–445, 1995. ISSN 00401706. Disponível em: <<http://www.jstor.org/stable/1269735>>.
- MURTHY, D. P. *et al.* **Weibull models**. [S.l.]: John Wiley & Sons, 2004. v. 505.
- NADARAJAH, S. The exponentiated gumbel distribution with climate application. **Environmetrics: The official journal of the International Environmetrics Society**, Wiley Online Library, v. 17, n. 1, p. 13–23, 2006.
- NADARAJAH, S. **The exponentiated exponential distribution: a survey**. [S.l.]: Springer, 2011.
- NADARAJAH, S.; GUPTA, A. The beta fréchet distribution. **Far East Journal of Theoretical Statistics**, v. 14, 01 2004.
- NADARAJAH, S.; GUPTA, A. K. The exponentiated gamma distribution with application to drought data. **Calcutta Statistical Association Bulletin**, SAGE Publications Sage India: New Delhi, India, v. 59, n. 1-2, p. 29–54, 2007.
- NADARAJAH, S.; HAGHIGHI, F. An extension of the exponential distribution. **Statistics**, Taylor & Francis, v. 45, n. 6, p. 543–558, 2011.
- NADARAJAH, S.; KOTZ, S. The beta gumbel distribution. **Mathematical Problems in engineering**, Hindawi, v. 2004, n. 4, p. 323–332, 2004.
- NADARAJAH, S.; KOTZ, S. The exponentiated type distributions. **Acta Applicandae Mathematica**, Springer, v. 92, n. 2, p. 97–111, 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.
- NOCEDAL, J.; WRIGHT, S. **Numerical optimization**. [S.l.]: Springer Science & Business Media, 2006.
- OGUNTUNDE, P. *et al.* The weibull-exponential distribution: Its properties and applications. **Journal of Applied Sciences**, Asian Network for Scientific Information, v. 15, n. 11, p. 1305–1311, 2015.
- PICCIOTTO, R. **Tensile fatigue characteristics of a sized polyester/Viscose Yarn and their effect on weaving performance**. Tese (Doutorado) — North Carolina State University at Raleigh, 1970.
- PRENTICE, R. L. Exponential survivals with censoring and explanatory variables. **Biometrika**, Oxford University Press, v. 60, n. 2, p. 279–288, 1973.

PRESS, W. H. *et al.* **Numerical recipes 3rd edition: The art of scientific computing**. [S.l.]: Cambridge university press, 2007.

Qian, L. The fisher information matrix for a three-parameter exponentiated weibull distribution under type ii censoring. **Statistical Methodology**, Elsevier, v. 9, n. 3, p. 320–329, 2012.

RAMÍREZ, F. A. P. New generalized nadarajah-haghighi distributions in survival analysis. Universidade Federal de Pernambuco, 2017.

RAQAB, M. Z.; MADI, M. T. Inference for the generalized rayleigh distribution based on progressively censored data. **Journal of Statistical Planning and Inference**, Elsevier, v. 141, n. 10, p. 3313–3322, 2011.

RASHWAN, N. I. A note on kumaraswamy exponentiated rayleigh distribution. **Journal of Statistical Theory and Applications**, Atlantis Press, v. 15, n. 3, p. 286–295, 2016.

RAYLEIGH, L. On the resultant of a large number of vibrations of the same pitch and of arbitrary phase. **Phil. Mag., Series 5**, v. 43, n. 261, p. 259–272, 1880.

RISTIĆ, M. M.; BALAKRISHNAN, N. The gamma-exponentiated exponential distribution. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 82, n. 8, p. 1191–1206, 2012.

SALINAS, H. S. *et al.* Slashed exponentiated rayleigh distribution. **Revista Colombiana de Estadística**, Universidad Nacional de Colombia., v. 38, n. 2, p. 453–466, 2015.

SANTOS, R. G. D. **Uma nova classe de distribuições e um novo modelo para tempo de vida: propriedades e aplicações**. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2016.

SHANKER, R. *et al.* On modeling of lifetimes data using exponential and lindley distributions. **Biometrics & Biostatistics International Journal**, v. 2, n. 5, p. 1–9, 2015.

SILVA, G. O. *et al.* The beta modified weibull distribution. **Lifetime data analysis**, Springer, v. 16, n. 3, p. 409–430, 2010.

SMITH, R. L.; NAYLOR, J. A comparison of maximum likelihood and bayesian estimators for the three-parameter weibull distribution. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, Wiley Online Library, v. 36, n. 3, p. 358–369, 1987.

SOUZA, C. Sales Reis de *et al.* Estudo do comportamento térmico de laminados carbono/epóxi submetidos a múltiplos ciclos térmicos. **Polímeros: Ciência e Tecnologia**, Associação Brasileira de Polímeros, v. 26, 2016.

SURIYA, J.; JAN, T. Tx family of gamma-exponential distribution and its structural properties. In: **Proceedings of International Conference on Advances in Computers, Communication and Electronic Engineering, University of Kashmir**. [S.l.: s.n.], 2015. p. 207–213.

TAHIR, M. *et al.* A new weibull-g family of distributions. **Hacettepe Journal of Mathematics and Statistics**, v. 45, n. 2, p. 629–647, 2016.

TORRES-PEREIRA, C. C. *et al.* Abordagem do câncer da boca: uma estratégia para os níveis primário e secundário de atenção em saúde. **Cadernos de Saúde Pública**, SciELO Public Health, v. 28, p. s30–s39, 2012.

WALCK, C. Statistical distributions for experimentalists. **Particle Physics Group**, 2007.

WANG, Y. C.; RESTREPO, J. I. Investigation of concentrically loaded reinforced concrete columns confined with glass fiber-reinforced polymer jackets. **Structural Journal**, v. 98, n. 3, p. 377–385, 2001.

WOLFRAM, S. **The mathematica book**. [S.l.]: Cambridge university press, 1999.

ZEGHDOUDI, H.; NEDJAR, S. Gamma-lindley distribution and its application. **Journal of Applied Probability and Statistics**, v. 11, 11 2014.

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

ANEXO A – MANIPULAÇÕES NH-R

Aqui, mostramos um pouco das manipulações algébricas usadas na Equação (5.13). Temos

$$\begin{aligned}\rho(t, m) &= \int_0^1 \exp[tQ_R(u)]u^m du \\ &= \int_0^1 \exp[t\sigma\sqrt{-2\log(1-u)}]u^m du\end{aligned}$$

e, usando expansões para as funções exponencial e logaritmo, obtemos:

$$\rho(t, m) = \int_0^1 \sum_{j=0}^{\infty} \frac{(t\sigma)^j}{j!} 2^{j/2} \left(\sum_{i=1}^{\infty} b_i u^i \right)^{j/2} u^m du,$$

em que $b_i = \frac{(-1)^{2i}}{i}$.

Agora, usando expansão em série de potência para um número real e trocando a integral pelo somatório, obtemos:

$$\rho(t, m) = \sum_{l,j=0}^{\infty} \frac{2^{j/2}(t\sigma)^j}{j!} v_i(j/2) \int_0^1 \left(\sum_{i=1}^{\infty} b_i u^i \right)^l u^m du,$$

em que

$$v_i(j/2) = \sum_{n=1}^{\infty} (-1)^{n-1} \binom{n}{i} (j/2)_n / n! \text{ e } (j/2)_n = \frac{j}{2} \left(\frac{j}{2} - 1 \right) \dots \left(\frac{j}{2} - n + 1 \right).$$

Usamos ao longo do artigo uma equação de Seção 0,314 de Gradshteyn e Ryzhik para uma série de potências para um inteiro positivo. Assim, a última equação é dada por

$$\rho(t, m) = \sum_{j=0}^{\infty} \frac{2^{j/2}(t\sigma)^j}{j!} \frac{v_i(j/2)}{m+1} + \sum_{j=0}^{\infty} \sum_{l,i=1}^{\infty} \frac{2^{j/2}(t\sigma)^j}{j!} \frac{v_i(j/2)}{i+m+1} a_{l,i},$$

em que

$$a_{l,i} = (ib_0)^{-1} \sum_{k=1}^i [k(l+1-i)] b_k a_{l,i-k}$$

Então, depois de alguma manipulação algébrica, obtemos

$$M_X(t) = \sum_{m=0}^{\infty} (m+1) w_m \sum_{j=0}^{\infty} p_j(m) t^j,$$

em que $p_j(m) = \frac{f_j}{m+1} + g_j(m)$,

$$f_j = \frac{2^{j/2}(t\sigma)^j v_i(j/2)}{j!} \text{ e } g_j = \sum_{l,i=1}^{\infty} \frac{2^{j/2}(t\sigma)^j}{j!} \frac{v_i(j/2)}{i+m+1} a_{l,i}.$$