

## A Distribuição $\Phi^{\text{sup}}$ : propriedades e aplicações

### The $\Phi^{\text{sup}}$ distribution: properties and applications

DOI:10.34117/bjdv7n3-139

Recebimento dos originais: 08/02/2021

Aceitação para publicação: 08/03/2021

**Guilherme V. Muniz**

Departamento de Estatística  
Universidade Federal da Paraíba, Brazil  
Email: claudio@de.ufpb.br

**Claudio J. Tablada**

Departamento de Estatística  
Universidade Federal de Pernambuco, Brazil

#### RESUMO

Frequentemente, pesquisadores estão interessados em modelar eventos de diferente natureza recorrendo a propostas de modelos probabilísticos a fim de descrever os fenômenos com certa segurança, tais como níveis pluviométricos, inundações, velocidade dos ventos, entre outras aplicações. No escopo da teoria da probabilidade, a atenção é voltada para a Teoria dos Valores Extremos(TVE), em que eventos considerados raros são levados em consideração, e generalizações de famílias de distribuições englobadas no TVE são reconhecidas como distribuições de cauda mais pesada. Com uso do método  $T - X$  para gerar novas famílias de distribuições generalizadas, propõe-se a inserção de uma nova distribuição na literatura, utilizando uma distribuição de base advinda do TVE. Faz-se um estudo rigoroso desta afim de verificarmos suas aplicabilidades. Estrutura-se a sua obtenção, ilustra-se algumas de suas formas, apresenta-se o método de geração de valores do novo modelo, assim como é abordado a estimação dos parâmetros usando-se a teoria de máxima verossimilhança. Finalmente, ilustra-se a adequabilidade do novo modelo ao ajustá-lo a conjuntos de dados reais, compara-o com outros modelos existentes na literatura e constrói-se testes de hipótese a respeito da inserção do novo parâmetro, utilizando a metodologia para modelos não encaixados.

**Palavras-Chave:** Distribuições de probabilidade generalizadas. Distribuição Fréchet. Método T-X. Valores extremos.

#### ABSTRACT

Researchers are often interested in modeling events of different natures by resorting to probabilistic model proposals in order to describe phenomena with some certainty, such as rainfall levels, floods, wind velocity, among other applications. In the scope of probability theory, attention is focused on the Theory of Extreme Values (TVE), in which events considered rare are taken into account, and generalizations of families of distributions encompassed in the TVE are recognized as heavier-tailed distributions. Using the  $T - X$  method to generate new families of generalized distributions, it is proposed to introduce a new distribution into the literature, using a distribution derived from the TVE. A rigorous study of this distribution is made in order to verify its

applicability. Its obtaining is structured, some of its forms are illustrated, the method of generating values for the new model is presented, as well as the estimation of parameters using maximum likelihood theory. Finally, the suitability of the new model is illustrated by fitting it to real data sets, comparing it to other models in the literature, and constructing hypothesis tests regarding the insertion of the new parameter, using the methodology for ill-fitting models.

**Key words:** Generalized probability distributions. Fréchet distribution. T-X method. Extreme values.

## 1 INTRODUÇÃO

É comum recorrermos as distribuições de probabilidade, quando estamos interessados em descrever o comportamento de fenômenos reais. Por exemplo, pesquisadores podem estar interessados em entender o tempo de falha de um determinado equipamento, modelar a proporção de pessoas afetadas por uma determinada doença, captar o tempo de vida de um determinado medicamento, entre outros. No entanto, os modelos "clássicos" muitas vezes podem não apresentar a melhor adequação para determinados conjuntos dados reais Coles et al. (2001). Nesta conjuntura, há o constante interesse em obter novas distribuições advindas de um modelo probabilístico base.

Como resultado de práticas estatísticas, é comum obter-se um novo modelo propiciamente mais flexível, uma generalização mais precisamente, via inserção de um ou mais parâmetros a uma distribuição de base, uma  $G$ , ao fim que uma nova família de distribuições possa ser obtida Tablada & Cordeiro (2017). Estas famílias, obtidas ao fim da inserção, são referenciadas na literatura como famílias generalizadas de distribuições.

No escopo de métodos que generalizem as famílias de distribuições, recorre-se ao método T-X Alzaatreh et al. (2013), que leva em consideração a distribuição de 2 variáveis aleatórias quaisquer  $T$  (o transformado) e  $X$  (o transformador). Algumas classes para famílias advindas deste método já foram inseridas na literatura Ahmad et al. (2019).

De maneira mais particular, os fenômenos considerados raros, caracterizados por sua pouca frequência, mas de grande impacto em uma análise, centraliza a visão na teoria da probabilidade, para uma de suas particularidades, a teoria dos valores extremos, que vem a desempenhar um papel importante no surgimento de novos modelos Kotz & Nadarajah (2000). Como exemplo de sua usabilidade, os eventos de riscos, tais como cotações em bolsas de valores, velocidade máxima de ventos, precipitações pluviométricas, poluição no ar, entre outras. De maneira geral, as distribuições de valores

extremos, garantem margens de funcionamento para situações onde os eventos de mínima probabilidade apresentam consequências significativas em um todo.

Dentro da teoria dos valores extremos, propõe-se o estudo de um modelo generalizado da lei tipo II, Fréchet. Uma extensão da já conhecida, proposta por Maurice René Fréchet (1878-1973), donde esta é utilizada como uma possível distribuição limite para uma sequência de máximos de variáveis aleatórias (V.A) independentes e identicamente distribuídas (i.i.d.) Fréchet (1927). A ampla aplicabilidade do tipo II, se estende a mais de 50 tipos de dados reais conhecidos, tais como: velocidade dos ventos, aplicações hidrológicas, utilizações em processos estocásticos, econométricos, finanças de risco, em produções industriais de produtos, como algodão, metalurgia em geral, e outras Kotz & Nadarajah (2000).

Tamanho é a empregabilidade da Fréchet à análise de dados, fez diversas áreas des- tacarem, contribuindo de forma significativa no avanço destas Kotz & Nadarajah (2000). Nesta conjuntura, é crescente o interesse neste campo, por pesquisadores de áreas como: biologia, engenharias, economia, geografia, ciências atuariais, farmacêuticos e outros Har- low (2002). Diversas extensões da Fréchet já foram desenvolvidos pela ampla comuni- dade construída ao longo da histórica relação entre a estatística e os diversos campos da ciência enriquecendo a literatura, exemplos destas são as famílias : the exponentia- ted Fréchet (EFr) Cordeiro et al. (2013), beta Fréchet (BFr) Barreto-Souza et al. (2011), Marshall-Olkin Fréchet (MOF) Krishna et al. (2013), Kumaraswamy Fréchet (KwmF) Mead (2014), Transmuted Marshall-Olkin Fréchet (TMOF) Afify et al. (2015), Modified Fréchet (MoF) Tablada & Cordeiro (2017).

Neste artigo propomos uma nova extensão da Fréchet tri-paramétrica, chamada dis- tribuição Fréchet Sup, apresentamos o estudo de suas propriedades matemáticas, com a esperança de futuras aplicações reais desta. Com rigor, na seção 2.1 é apresentada sua obtenção, suas formas reduzidas como função acumulada, densidade e razão de risco. Formulamos os momentos da distribuição Sup [section 2.2], e o gerador de números alea- tórios da distribuição [section 2.3]. Na seção 3 uma simulação de monte Carlo é elaborada, exibindo critérios de adequação do ajuste. Na seção 4 são obtidos os estimadores via mé- todo da máxima verossimilhança. Após isso [section 5], é realizada uma aplicação em dados reais relativos a tempo em tráfego. Section 6 conclusão e considerações finais.

## 2 A DISTRIBUIÇÃO $\Phi^{\text{SUP}}$

A distribuição advém de um algoritmo que leva em consideração, duas variáveis aleatórias, seguindo uma mesma distribuição, logo mais será apresentado. Ao final da aplicação do algoritmo, para o conjunto de valores resultantes, o histograma esperado, descreverá uma curva. A esta curva, com suas possibilidades de parâmetros, chamaremos de Família  $G_n^{\text{SUP}}$ .

Seja  $G(x)$  a distribuição de 2 parâmetros Fréchet, estipulada por Maurice Fréchet (1927), e escolhida a dedo para este estudo por motivos de, ampla aplicabilidade junto a necessidade de introduzir modificações ou generalizações desta Kotz & Nadarajah (2000). Sua função acumulada e densidade, são dadas respectivamente por :

$$G(x; a, \beta) = e^{-\left(\frac{x}{a}\right)^\beta} \quad (1)$$

$$g(x; a, \beta) = \left(\frac{\beta}{x}\right) \left(\frac{a}{x}\right)^\beta e^{-\left(\frac{x}{a}\right)^\beta} \quad (2)$$

E sejam  $Z_1, Z_2, \dots, Z_n$  variáveis aleatórias independentes e identicamente distribuídas com função de distribuição acumulada  $G(x)$ , então  $G(x) = P(Z_j \leq x)$ ,  $j = 1, \dots, n$ .

Denotamos que  $Z = \max(Z_1, Z_2, \dots, Z_n)$ . Então, é sabido que  $Z$  tem função de distribuição acumulada dada por  $[G(x)]^n$ .

Agora, seja  $Y$  V.A independente de  $Z$  tal que  $Y$  tem também função de distribuição acumulada  $G(x)$ , analogamente  $P(Y \leq x) = G(x)$ ,  $\forall x \in \mathbb{R}$ . Finalmente, seja  $X = \min(Y, Z)$  então  $X \sim G_n^{\text{sup}}$

Por  $Y$  e  $Z$  serem funções da V.A.  $x$ , onde esta segue uma distribuição conhecida, expande-se a sua função probabilidade de mínimo

$$P(X \geq x) = P(Y \geq x, Z \geq x) = P(Y \geq x) \cdot P(Z \geq x) \quad (3)$$

$$= [1 - P(Y \leq x)][1 - P(Z \leq x)] = [1 - G(x)][1 - G^n(x)] \quad (4)$$

$$P(X \geq x) = 1 - G(x) - G^n(x) + G^{n+1}(x) \quad (5)$$

$$F_X(x) = P(X \leq x) = 1 - P(X \geq x) = \quad (6)$$

$$1 - [1 - G(x) - G^n(x) + G^{n+1}(x)] = G(x) + G^n(x) - G^{n+1}(x) \quad (7)$$

Relembra-se que a função Fréchet é uma função de dois parâmetros. E associada a metodologia anterior, com a clássica prática da estatística de melhorar um modelo fornecendo

maior flexibilidade, adicionando novos parâmetros Tablada & Cordeiro (2017). Então, nessa nova distribuição de 3 parâmetros repassasse a informação n, como nosso novo parâmetro  $0 < c < \infty$ . Ao fim, a generalizada  $G^{sup}$  para distribuição Fréchet é repassada como  $\Phi^{sup}$ .

## 2.1 FORMAS DA DENSIDADE E RAZÃO DE RISCO

Repassando a distribuição acumulada Fréchet em 7, obtêm-se a função acumulada da nova distribuição, expressada por

$$F(x, \alpha, \beta, c) = e^{-\left(\frac{\alpha}{x}\right)^\beta} + e^{-c\left(\frac{\alpha}{x}\right)^\beta} - e^{-(c+1)\left(\frac{\alpha}{x}\right)^\beta} \quad (8)$$

Derivando esta acumulada, obtêm-se a curva densidade :

$$f(x, \alpha, \beta, c) = \left(\frac{\beta}{x}\right)\left(\frac{\alpha}{x}\right)^\beta \left[ e^{-\left(\frac{\alpha}{x}\right)^\beta} + c e^{-c\left(\frac{\alpha}{x}\right)^\beta} - (c+1) e^{-(c+1)\left(\frac{\alpha}{x}\right)^\beta} \right] \quad (9)$$

É importante ressaltar, que além de lidarcom uma função continua de representatividade estatística, as propriedades ligadas aos parâmetros envolvidos definem um espaço de aplicação, como em  $\alpha, \beta, c > 0$ . Nesta conjuntura, os parâmetros  $\alpha$  e  $c$ , reais positivos, são associados a forma da distribuição. Conseqüentemente, o parâmetro  $\beta$ , cuja dimensão

esta em acordo com a dimensão da v.a., deve ser um real positivo, e reflete a escala da distribuição.

Prova : Um parâmetro  $\theta$  é dito ser de escala, se a distribuição da v.a  $\frac{x}{\theta}$  : não depende  $\theta$ . Em especifico para  $X \sim G^{sup}$

$$P\left(\frac{X}{\alpha} \leq x\right) \text{ depende de } \alpha ?$$

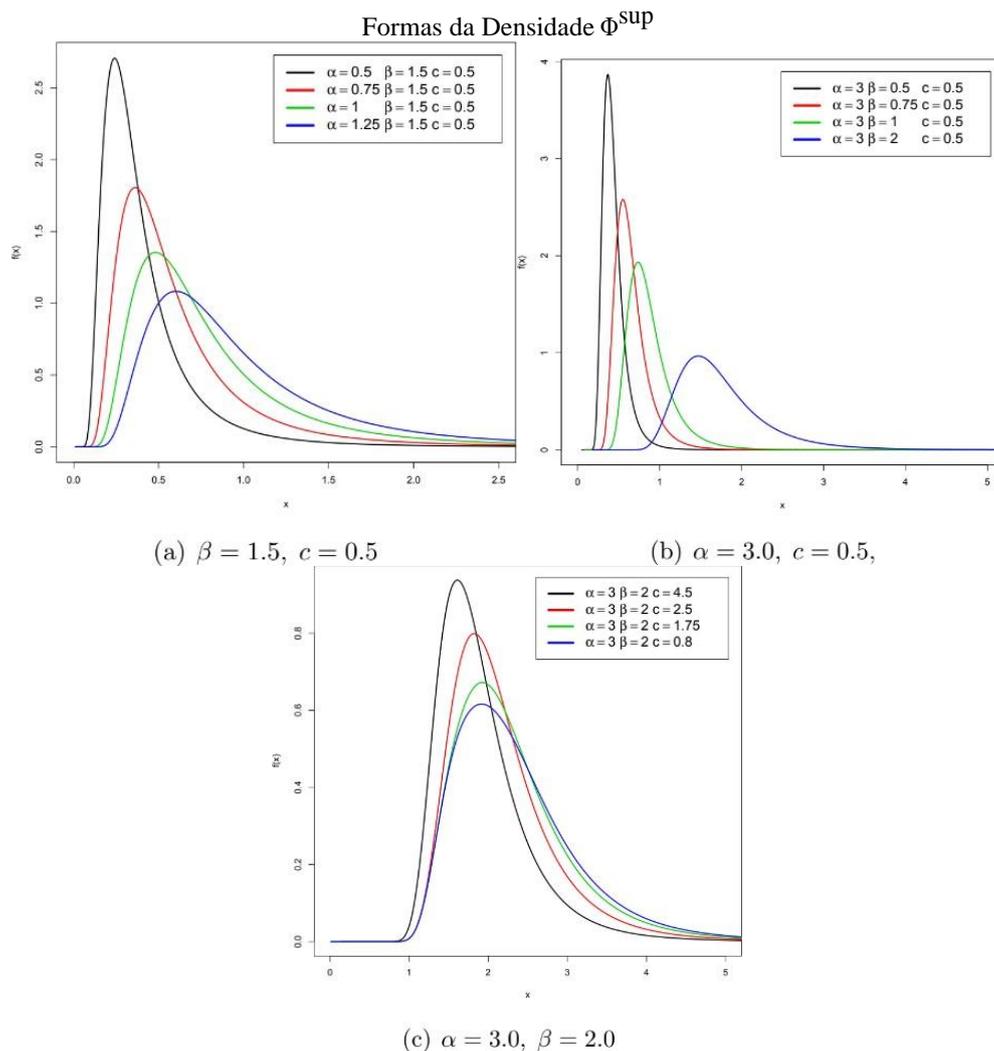
$$P\left(\frac{X}{\alpha} \leq x\right) = P(X \leq \alpha x) = F_X(\alpha x) = \Phi^{sup}(\alpha x) = F(\alpha x, \alpha, \beta, c) = e^{-x^{-\beta}} + e^{-c x^{-\beta}} - e^{-(c+1)x^{-\beta}} \quad (10)$$

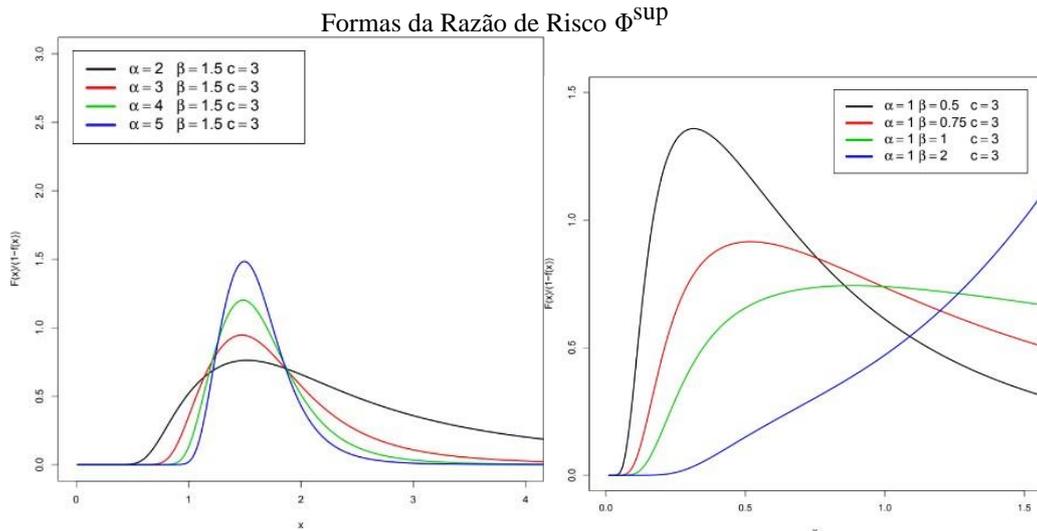
É comum em artigos da área, devido a aplicabilidade em análises de tempo de vida apresenta-se também a função razão de risco. Sabe-se que a linearidade da curva em

forma de banheira da função Razão de Risco é construída por  $h(x) = \frac{f(x)}{1-F(x)}$  Marshall et al. (2005). Logo, a função razão de risco da  $\Phi^{\text{sup}}$  é

$$h(x) = \frac{\left(\frac{\beta}{x}\right)\left(\frac{\alpha}{x}\right)^\beta \left[ e^{-(\frac{\alpha}{x})^\beta} + c e^{-c(\frac{\alpha}{x})^\beta} - (c+1) e^{-(c+1)(\frac{\alpha}{x})^\beta} \right]}{1 - e^{-(\frac{\alpha}{x})^\beta} - e^{-c(\frac{\alpha}{x})^\beta} + e^{-(c+1)(\frac{\alpha}{x})^\beta}} \quad (11)$$

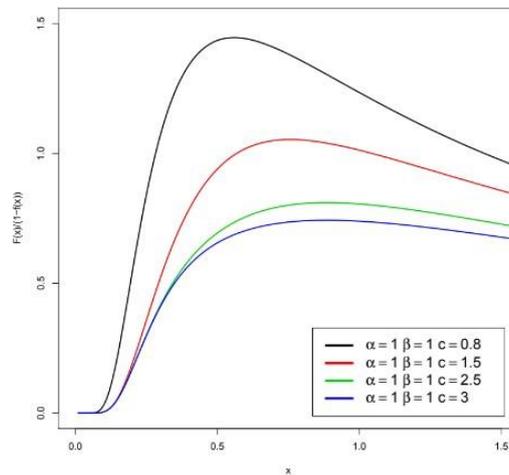
Conhecidas as formas algébricas da pdf e hzf, entender comportamento destas pode ser tarefa fácil. Tomando valores paramétricos a bel prazer, então, apela-se para o entendimento visual da função atentando para suas formas e escalas. Em Fig. 1 e Fig. 2, são exibidos plots da pdf e hzf, respectivamente, da  $\Phi^{\text{sup}}$  para alguns valores de parâmetros.





(a)  $\beta = 1.5, c = 3$

(b)  $\alpha = 1.0, c = 3,$



(c)  $\alpha = 1, \beta = 1$

## 2.2 GERADOR DE VALORES

A função geradora de valores pseudoaleatórios da distribuição em estudo, é de difícil obtenção, uma vez que as observações serão construídas a partir da composição de valores independentes entre si, das distribuições Fréchet e Fréchet Maximizada. Alternativamente, é proposto um algoritmo para gerar o número pseudoaleatório da  $\Phi^{\text{sup}}$ .

1. Gerar 2 observações independentes da distribuição Uniforme  $U(0, 1)$   $u$  e  $v$
2. Seja  $y$  uma observação da Fréchet, e  $z$  uma observação da Fréchet maximizada, isso é, sejam  $u \sim U(0, 1)$   $y = Q_{\alpha, \beta}(u_j) \sim G(x)$  "Seja  $u$  advindo de uma distribuição uniforme padronizada,  $y$  é a função cujo domínio é  $u$ , e imagem  $G(x)$ ".

e  $v \sim U(0, 1)$   $z = Q_{\alpha, \beta, c}(u_j) \sim G(x)^c$  "Seja  $v$  advindo de uma distribuição uniforme padronizada,  $z$  é a função cujo domínio é  $u$ , e a imagem  $G(x)^c$ "

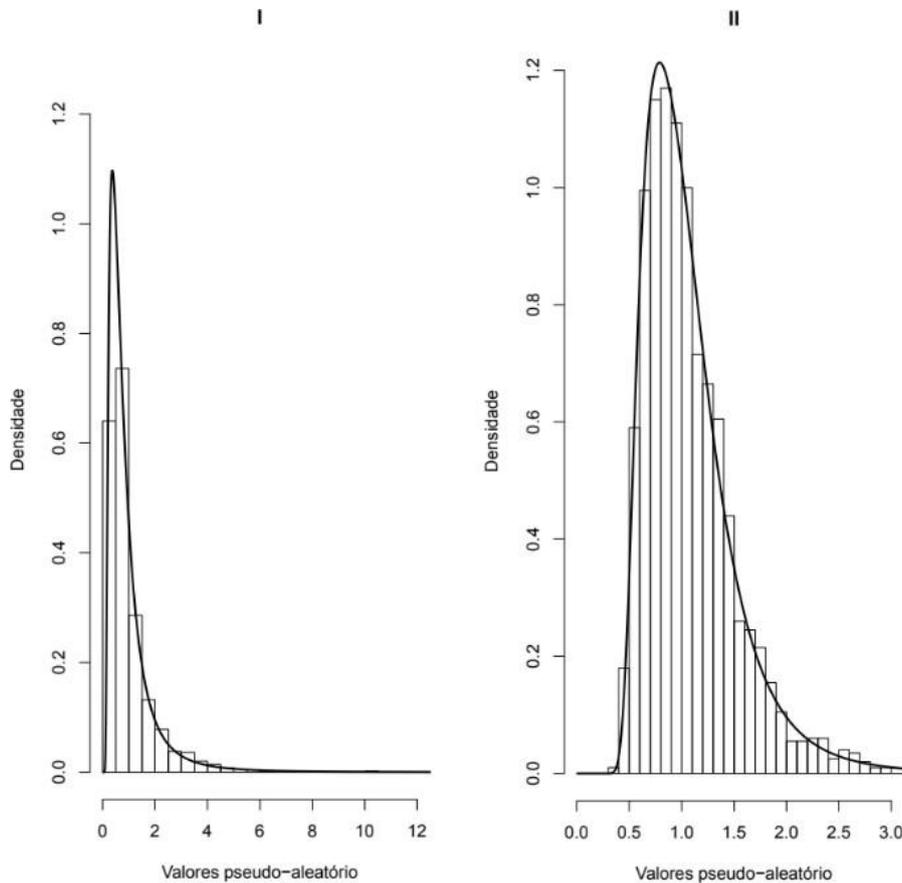
3. Ora,  $y = Q_{\alpha, \beta, c}(u_j) = \frac{\alpha}{\left[\log\left(\frac{1}{u}\right)\right]^{\frac{1}{\beta}}}$

Pois  $Q_{\alpha, \beta}(u_j)$  é a função quantilica da distribuição Frechet

4. Repassando  $z = Q_{\alpha, \beta, c}(u_j) = \frac{\alpha}{\left[\frac{1}{c} \log\left(\frac{1}{u}\right)\right]^{\frac{1}{\beta}}}$  já que  $Q_{\alpha, \beta, c}(u_j)$  é a quantilica da Frechet maximizada

5. Ao fim, ponto a ponto obtêm-se da composição  $x = \min(y, z)$  observações, donde  $x \sim \Phi^{\text{sup}}$

Repetindo a rotina de obtenção das observações da  $\Phi^{\text{sup}}$  um número razoável de vezes, é possível verificar a curva  $\Phi^{\text{sup}}$  sob os dados pseudoaleatórios como nas situações I com  $n = 400$ ,  $\alpha = 0.5$ ,  $\beta = 2.5$ ,  $c = 0.75$  e para II adotando-se  $n = 400$ ,  $\alpha = 1.5$ ,  $\beta = 1.5$ ,  $c = 0.75$



### 3 ESTIMAÇÃO DE MÁXIMA VEROSSIMILHANÇA

Na literatura, diversos métodos de estimação de parâmetros são apresentados. Mas em atualidade, a estimação via método da máxima verossimilhança é a mais empregada pela comunidade. O MLE é um método que fornece estimações desejáveis para a construção dos intervalos de confiança, estatística essa de grande utilidade. Nesta seção nos estimamos os parâmetros da distribuição  $\Phi^{\text{sup}}$ , onde  $x = (x_1, x_2, \dots, x_n)^T$  é uma amostra de tamanho  $n$  de  $X \sim \Phi^{\text{sup}}(\alpha, \beta, c)$  e  $\theta = (\alpha, \beta, c)^T$  o vetor de parâmetros. A função log-verossimilhança para a amostra  $x$ , é denotada por  $l_f(\theta; x)$  e dada

$$l_f(\theta; x) = c \log(\beta) + \beta n \log(\alpha) - (\beta + 1) n \sum_{i=1}^n \log(x_i) + \sum_{i=1}^n \log(\psi) \quad (12)$$

Onde  $\psi$  deve ser repassado por  $e^{-\left(\frac{\alpha}{x_i}\right)^\beta} + c e^{-c\left(\frac{\alpha}{x_i}\right)^\beta} - (c + 1) e^{-(c+1)\left(\frac{\alpha}{x_i}\right)^\beta}$

A maximização da função 12, obtida da diferenciação em relação ao vetor de parâmetros  $\theta$ , e em seguida resolvendo o sistema de equações resultante, pode ser calculada via R (optim e MaxLik functions) SAS(PROC NLMIXED), e Ox program(sub-routine

MaxBFGS). Seja nosso vetor de escores(derivadas)  $U = (U_\alpha, U_\beta, U_c)$  composto por

$$U_\alpha = \frac{\beta}{\alpha x} - n \beta x^{-\beta} \left(\frac{\alpha}{x}\right)^{\beta-1} \left[ \frac{e^{-\left(\frac{\alpha}{x}\right)^\beta} + c^2 e^{-c\left(\frac{\alpha}{x}\right)^\beta} - (c + 1)^2 e^{-(c+1)\left(\frac{\alpha}{x}\right)^\beta}}{e^{-\left(\frac{\alpha}{x}\right)^\beta} + c e^{-c\left(\frac{\alpha}{x}\right)^\beta} - (c + 1) e^{-(c+1)\left(\frac{\alpha}{x}\right)^\beta}} \right] \quad (13)$$

$$U_\beta = n(\beta^{-1} + \log(\alpha) - \log(x)) - \frac{\left(\frac{\alpha}{x}\right)^\beta \log\left(\frac{\alpha}{\beta}\right) \left[ e^{-\left(\frac{\alpha}{x}\right)^\beta} + c^2 e^{-c\left(\frac{\alpha}{\beta}\right)^\beta} - (c + 1)^2 e^{-(c+1)\left(\frac{\alpha}{\beta}\right)^\beta} \right]}{e^{-\left(\frac{\alpha}{x}\right)^\beta} + c e^{-c\left(\frac{\alpha}{\beta}\right)^\beta} - (c + 1) e^{-(c+1)\left(\frac{\alpha}{\beta}\right)^\beta}} \quad (14)$$

$$U_c = n \frac{\left[ e^{-c\left(\frac{\alpha}{x}\right)^\beta} - e^{-(c+1)\left(\frac{\alpha}{x}\right)^\beta} + \left(\frac{\alpha}{x}\right)^\beta \left[ (c + 1) e^{-(c+1)\left(\frac{\alpha}{x}\right)^\beta} - c e^{-c\left(\frac{\alpha}{x}\right)^\beta} \right] \right]}{e^{-\left(\frac{\alpha}{x}\right)^\beta} + c e^{-c\left(\frac{\alpha}{\beta}\right)^\beta} - (c + 1) e^{-(c+1)\left(\frac{\alpha}{\beta}\right)^\beta}} \quad (15)$$

O MLE  $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{c})^T$  pode ser estimado fixando  $U_{\alpha} = U_{\beta} = U_c = 0$  e resolvendo as equações simultaneamente. Essas equações não foram resolvidas de forma analítica, e softwares estatísticos podem ser usados para resolvê-las numericamente, por meio de técnicas iterativas como o algoritmo de Newton-Raphson. Para esta nova distribuição, todas as segundas derivadas existem.

#### 4 APLICAÇÃO

É de grande interesse averiguar se o modelo Fréchet Sup é adequado para descrever fenômenos reais. Dessa forma, ajusta-se o novo modelo a dois conjuntos de dados. O primeiro conjunto de dados reportado por Jørgensen (1982), e consiste em 128 observações de dados de tráfego, isso é, o tempo entre um veículo e outro, em determinado ponto de uma rodovia. O segundo conjunto, consiste em 63 observações que representam a força mensurada em GPa das fibras de carbono testadas sob a tensão a um calibre de 10mm, e inicialmente reportado por Bader & Priest (1982), e reutilizado por Kundu et al. (2008). Tais amostras consistem dos seguintes dados:

Traffic data - 2.8, 3.4, 1.4, 14.5, 1.9, 2.8, 2.3, 15.3, 1.8, 9.5, 2.5, 9.4, 1.1, 88.6, 1.6, 1.9, 1.5, 33.7, 2.6, 12.9, 16.2, 1.9, 20.3, 36.8, 40.1, 70.5, 2, 8, 2.1, 3.2, 1.7, 56.5, 23.7, 2.4, 21.4, 5.1, 7.9, 20.1, 14.9, 5.6, 51.7, 87.1, 1.2, 2.7, 1, 1.5, 1.3, 24.7, 72.6, 119.8, 1.2, 6.9, 3.9, 1.6, 3, 1.8, 44.8, 5, 3.9, 125.3, 22.8, 1.9, 15.9, 6, 20.6, 12.9, 3.9, 13, 6.9, 2.5, 12.3, 5.7, 11.3, 2.5, 1.6, 7.6, 2.3, 6.1, 2.1, 34.7, 15.4, 4.6, 55.7, 2.2, 6, 1.8, 1.9, 1.8, 42, 9.3, 91.7, 2.4, 30.6, 1.2, 8.8, 6.6, 49.8, 58.1, 1.9, 2.9, 0.5, 1.2, 31, 11.9, 0.8, 1.2, 0.8, 4.7, 8.3, 7.3, 8.8, 1.8, 3.1, 0.8, 34.1, 3, 2.6, 3.7, 41.3, 29.7, 17.6, 1.9, 13.8, 40.2, 10.1, 11.9, 11, 0.2.

Strength Data - 1.901, 2.132, 2.203, 2.228, 2.257, 2.350, 2.361, 2.396, 2.397, 2.445, 2.454, 2.474, 2.518, 2.522, 2.525, 2.532, 2.575, 2.614, 2.616, 2.618, 2.624, 2.659, 2.675, 2.738, 2.740, 2.856, 2.917, 2.928, 2.937, 2.937, 2.977, 2.996, 3.030, 3.125, 3.139, 3.145, 3.220, 3.223, 3.235, 3.243, 3.264, 3.272, 3.294, 3.332, 3.346, 3.377, 3.408, 3.435, 3.493, 3.501, 3.537, 3.554, 3.562, 3.628, 3.852, 3.871, 3.886, 3.971, 4.024, 4.027, 4.225, 4.395, 5.020

Ajusta-se esses conjuntos de dados às distribuições Fréchet, Weibull, Fréchet exponencializada, Weibull exponencializada, Fréchet modificada, Weibull modificada, Log-Logística e Fréchet Sup. Para o cálculo dos estimadores de máxima verossimilhança utiliza-se o software R (versão 3.5.3, pacote "AdequacyModel"). Para a maximização da função log-verossimilhança (12) usa-se o método BFGS com derivadas numéricas. Valo-

res iniciais para a estimação via BFGS dos parâmetros podem ser obtidos, comparando o histograma referente aos dados, com as formas da densidade da distribuição para diferentes valores de parâmetros, utilizando pacotes computacionais, como Mathematica ou Maple. Com o propósito de comparação, usa-se o critério de Cramér-Von Mises( $W^*$ ), e o critério Anderson-Darling( $A^*$ ) Chen & Balakrishnan (1995). Em geral, menores valores para estas estatísticas, indicam distribuições com melhor ajustamento aos dados.

As estatísticas MLE são dadas com seus erros padrão em parênteses nas tabelas 1 e 3. Os valores de  $W^*$  e  $A^*$  são apresentados nas tabelas 2 e 4. Não são considerados mais critérios de comparação, pois o modelo Fréchet não é um submodelo da distribuição Fréchet Sup. Em consequência, esse fato inviabiliza usar o critério da informação de Akaike(AIC) e o critério de informação Bayesiana (BIC) (usados somente para o caso de modelos encaixados), e faz das estatísticas bondade de ajuste  $A^*$  e  $W^*$  as mais apropriadas.

Essas tabelas revelam que a distribuição Fréchet Sup possui os menores valores para as estatísticas  $W^*$  e  $A^*$ , entre os modelos ajustados. Logo, é plausível afirmar que o novo modelo fornece o melhor ajustamento para os dados apresentados.

As figuras 1 e 2 apresentam os correspondentes histogramas dos dados com as curvas das densidades das distribuições Fréchet, Fréchet modificada, Log-logística e Fréchet Sup. Via Q-Q Plots, verifica-se a qualidade do ajustamento aos dados, como mostram as figuras seguintes, para as distribuições com melhor adesão aos dados. Desses gráficos, pode-se também concluir que a distribuição Fréchet Sup apresenta um bom ajustamento

aos dados, quando comparado com as outras extensões da distribuição Fréchet.

Figura 1: Traffic Data

Adesão ao Traffic Data

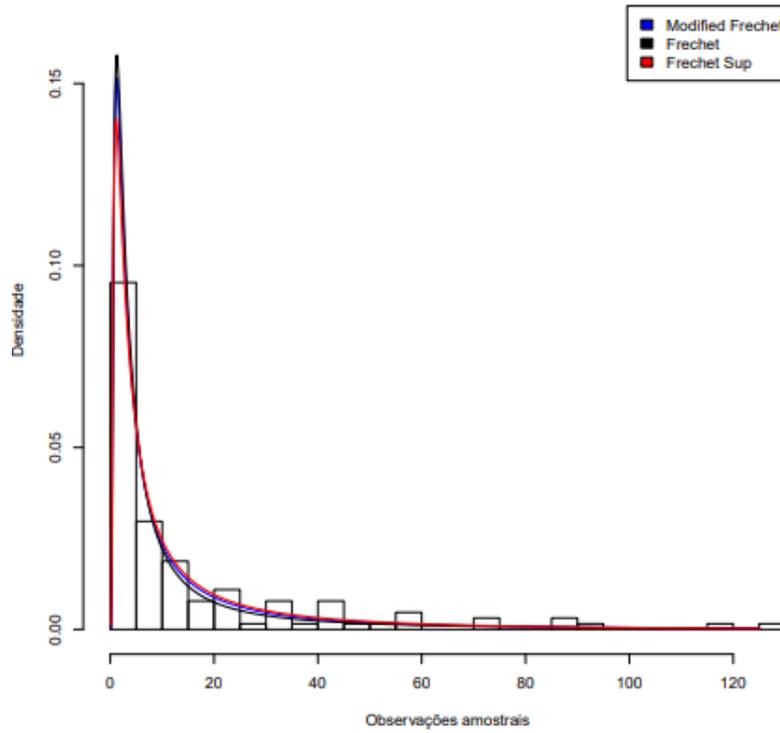


Figura 2: Strength Data

Adesão ao Strength Data

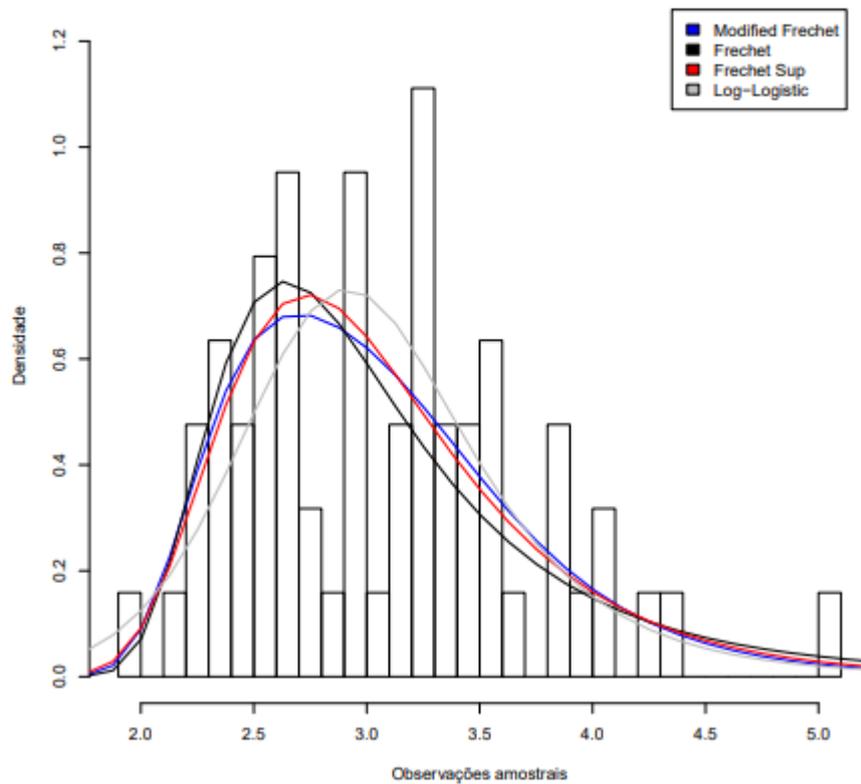


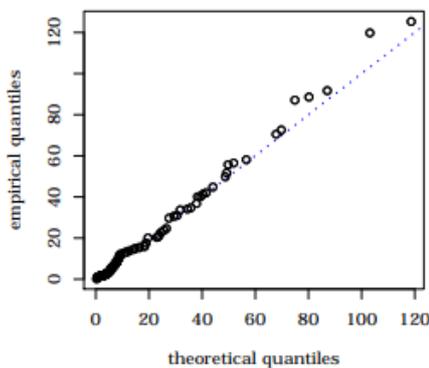
Tabela 1: Estimação do parâmetros - MLE - Traffic Data

MLE			
DISTRIBUIÇÃO	$\hat{\alpha}$	$\hat{\beta}$	$\hat{c}$
Fréchet	0.8182 (0.42635)	3.3133 (0.0587)	-
Weibull	0.7462 (0.0488)	12.8485 (1.6156)	-
Log-Logistic	6.0580 (0.7711)	1.2355 (0.08878)	-
Exp Fréchet	3.8938 (3.5121)	0.4329 (0.1641)	23.3179 (38.4556)
Exp Weibull	22.422 (34.643)	0.2486 (0.0303)	20.7411 (11.2547)
Modified Frechet	4.0280 (0.0051)	0.6887 (11.4324)	0.0162 (0.0048)
Frechet Sup	0.7787 (0.0587)	3.4479(0.4263) (0.0587)	9.8948(4.7707)

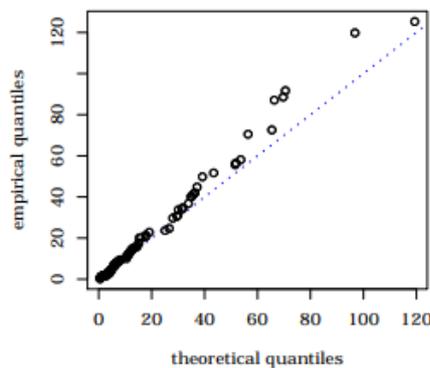
Tabela 2: Estatísticas de Bondade de Ajuste - Traffic Data

DISTRIBUIÇÃO	Estatísticas	
	W*	A*
Fréchet	0.1588584	1.043192
Weibull	0.5259943	3.205901
Exp Fréchet	0.2098713	1.265178
Exp Weibull	0.2405113	1.447344
Modified Frechet	0.1432725	0.8815578
Log-Logistic	0.3257547	1.976889
Frechet Sup	<b>0.1209889</b>	<b>0.7876428</b>

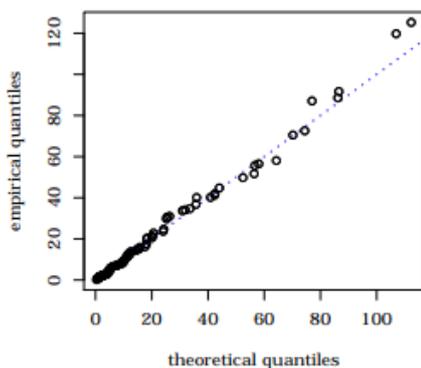
Q-Q Modified Frechet e Traffic Data



Q-Q Frechet e Traffic Data



Q-Q Frechet Sup e Traffic Data



Q-Q Exp Fréchet e Traffic Data

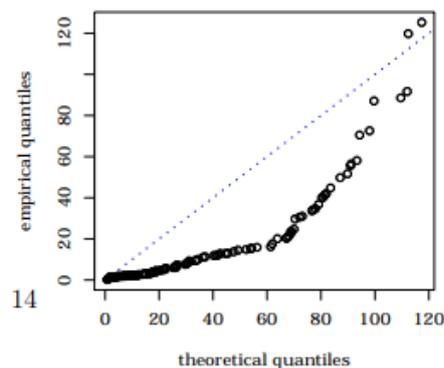


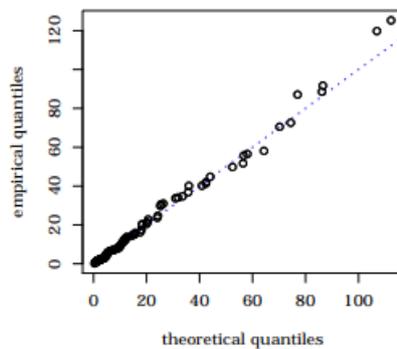
Tabela 3: Estimação do parâmetros - MLE - Strength Data

MLE			
DISTRIBUIÇÃO	$\hat{\alpha}$	$\hat{\beta}$	$\hat{c}$
Fréchet	2.7214 (0.0668)	5.4337 (0.5078)	-
Weibull	5.0494 (0.4557)	3.3147 (0.0877)	-
Log-Logistic	2.9932 (0.0770)	8.6456 (0.8902)	-
Exp Fréchet	7.0091 (8.4608)	2.3663 (1.0269)	4.2913 (1.6052)
Exp Weibull	38.2091 (72.3628)	0.2346 (0.0446)	25.0686 (18.2253)
Modified Fréchet	50.4579 (219.5719)	1.3828 (1.5478)	1.4580 (0.5646)
Fréchet Sup	4.9553 (0.5760)	2.7576 (0.0805)	5.6529(3.3678)

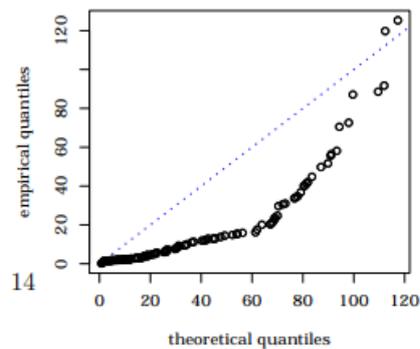
Tabela 4: Estatísticas de Bondade de Ajuste - Strength Data

DISTRIBUIÇÃO	Estatísticas	
	W*	A*
Fréchet	0.1149678	0.6419976
Weibull	0.1284381	0.8920839
Exp Fréchet	0.06144788	0.3297812
Exp Weibull	0.2375775	1.429076
Modified Fréchet	0.07696315	0.4017374
Log-Logistic	0.08889368	0.5017535
Fréchet Sup	<b>0.05538169</b>	<b>0.3011118</b>

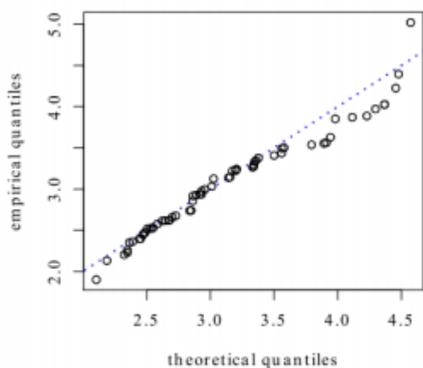
Q-Q Fréchet Sup e Traffic Data



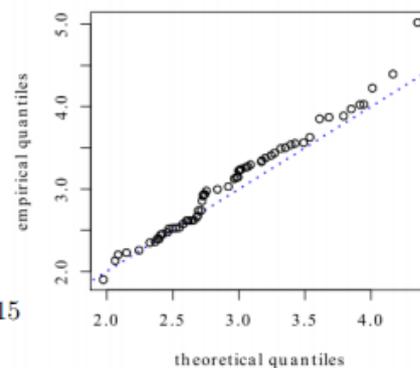
Q-Q Exp Fréchet e Traffic Data



Q-Q Fréchet Sup e Strength Data



Q-Q Log-Logistic e Strength Data



#### 4.1 INFERÊNCIA COM O MODELO NÃO ENCAIXÁVEL

Em muitas situações, é de interesse analisar quão significativa é a introdução de um parâmetro em uma extensão de uma distribuição na modelagem de um conjunto de dados. Por exemplo, podemos testar a hipótese nula de que a distribuição de base da o melhor ajustamento a uma amostra versus a alternativa, de que à distribuição generalizada é a mais adequada para a modelagem. Quando os modelos são encaixados, há procedimentos clássicos, como o teste da razão de verossimilhança (LR), que podem ser usados. Entretanto, estes não são adequados para testes de hipóteses de modelos não encaixados, como acontece com os modelos Fréchet e Fréchet Sup. Na literatura, há propostas de procedimentos para testes de hipóteses com modelos não encaixados. Aqui, considera-se a aproximação baseada na generalização do teste LR introduzido por Vuong(1989). Para o procedimento, consideramos uma amostra aleatória  $x = (x_1, x_2, \dots, x_n)$  de uma distribuição desconhecida  $f_0$ . No nosso caso, temos a densidade da Fréchet  $\phi(x)$  como distribuição de base, e a densidade da Fréchet Sup  $\phi^{sup}(x)$  como a distribuição generalizada. Define-se

$$D_{\phi}^{sup} = \mathbb{E}_{f_0} \{ \log [\phi^{sup}(0, \theta_*)] - \log [\phi(0, v_*)] \}$$

Em que  $\mathbb{E}_{f_0}(\cdot)$  denota a esperança com respeito à verdadeira densidade  $f_0(x)$  e  $\theta_* = \underset{\theta}{\operatorname{argmax}} \mathbb{E}_{f_0} \{ l_{\phi}^{sup}(\theta) \}$ ,  $v_* = \underset{v}{\operatorname{argmax}} \mathbb{E}_f \{ l_{\phi}(v) \}$ , são os pseudo-valores verdadeiros de  $\theta$  e  $v$ , respectivamente (ULLOA; PESARAN,2006). Baseado no critério da informação de Kullback-Leibler (KULLBACK,1997), um método de escolha pode ser definido usando a quantidade  $D_{\phi}^{sup}$ . Configurando as seguintes hipóteses e suas alegações.

- A.  $H_0 : D_{\phi}^{sup} = 0$ , indicando que os modelos Fréchet e Fréchet Sup são equivalentes, versus
- B.  $H_{\phi sup} : D_{\phi}^{sup} > 0$ , indicando que o modelo Fréchet Sup é significativamente melhor em relação ao modelo de base Fréchet, ou
- C.  $H_{\phi sup} : D_{\phi}^{sup} < 0$ , indicando que o modelo Fréchet Sup é significativamente pior em relação ao modelo de base Fréchet.

Para testar essas hipóteses, consideramos a estatística GLR, definida por

$$GLR_n = \frac{LR_n(\hat{\theta}_n, \hat{v}_n)}{\sqrt{n\omega_n^2}}$$

Em que  $LR_n(\hat{\theta}_n, \hat{v}_n) = l^{sup}(\hat{\theta}_n) - l_{\varphi}(\hat{v}_n)$ , onde  $\hat{\theta}_n$  e  $\hat{v}_n$  são as MLE's dos modelo FréchetSup e Fréchet, respectivamente, e

$$\omega_n^2 = \frac{1}{n} \sum_{i=1}^n \left\{ \log [\phi^{sup}(x_i, \hat{\theta}_n) - \log [\phi(x_i, \hat{v}_n)]]^2 - \left\{ \frac{LR_n(\hat{\theta}_n, \hat{v}_n)}{n} \right\}^2 \right.$$

A estatística GLR é um estimador consistente de quantidade desconhecida  $D_{\phi}^{sup}$ . Sob certas condições de regularidade, são obtidos os seguintes resultados assintóticos da distribuição de  $GLR_n$  segundo o teorema 5.1 dado por Vuong (1989) :

- Sob  $H_0$   $GLR_n$  converge assintoticamente para  $N(0, 1)$ ;
- $H_{\varphi sup}$   $GLR_n \rightarrow +\infty$  quase certamente;
- $H_{\varphi}$   $GLR_n \rightarrow -\infty$  quase certamente;

Esses resultados proporcionam critérios simples para selecionar entre os modelos Fréchet e Fréchet Sup, ao fazer uso dos quantis da distribuição normal padrão. Com efeito, ao nível de significância de  $(100 \times \gamma)\%$ . sendo  $\frac{z\gamma}{2}$  o quantil da distribuição normal padrão.

- se  $GLR_n > \frac{z\alpha}{2}$ , rejeita-se a hipótese nula de que os modelos são equivalentes, favorecendo o modelo generalizado Fréchet Sup como melhor distribuição para ajustar aos dados;
- se  $GLR_n < \frac{z\alpha}{2}$ , rejeitamos a hipótese nula de que os modelos são equivalentes entre si, favorecendo o modelo de base Fréchet como melhor distribuição para ajustar os dados;
- se  $|GLR_n| < \frac{z\alpha}{2}$ , então não podemos distinguir entre os modelos Fréchet e Fréchet Sup para os dados considerados.

Na tabela 5 são expostas as Estatísticas GLR calculadas e o p-valor para os testes de hipóteses realizados mediante os conjuntos de dados mencionados anteriormente. Nesta, revela-se, a um nível de significância de 1%, a rejeição de hipótese de que os modelos são equivalentes, em favor de que o modelo Fréchet Sup é superior ao modelo Fréchet para modelar os dados relativos a Trafego em uma rodovia. Para o conjunto de dados Strengh Data, rejeitamos a hipótese nula, ao nível de significância de 5%, de que os modelos são equivalentes, novamente em favor do modelo Fréchet Sup. Diante destes

testes, temos evidência de que a inserção do novo parâmetro na distribuição de base foi significativa para modelar os dados.

Tabela 5: teste de hipóteses

Hipóteses	Aplicação	Estatística GLR	p-valor
$D_{\phi}^{sup} = 0$ vs $D_{\phi}^{sup} > 0$	Traffic Data	3.4543	$2.757 \times 10^{-4}$
	Strenght Data	2.2459	$1.23 \times 10^{-2}$

## 5 CONCLUSÃO

Neste artigo, o autor introduzir um novo modelo de três parâmetros, chamado de Fréchet Sup, sendo esta uma adição a família T-X, e uma extensão para a lei tipo II Fréchet. Usando o método T-X é exposto sua obtenção, e ilustra-se suas formas, como densidade e função taxa de falha. Também é desenvolvida uma função que gera valores pseudo-aleatorias do novo modelo fornecendo alguns gráficos para entender se estas estão em acordo com a distribuição. Formula-se a obtenção dos estimadores de máxima verosimilhança do novo modelo para amostras quaisquer. Utilizando de uma base de dados referente a tráfego em uma rodovia, e outra sobre a tensão de fibras de carbono a um calibre de 10mm, compara-se o desempenho do novo modelo com outras extensões da Fréchet como EF, MF, e a própria Fréchet, usando as estatísticas clássicas de goodness-of-fit e gráficos Q-Q. Os resultados confirmam que a nova distribuição é muito apropriada para aplicações em dados reais. Dessa forma, tem-se que o modelo desenvolvido nesse trabalho traz contribuições importantes na área da probabilidade e agrupamento para dados tipo-intervalo.

## AGRADECIMENTOS

O autor agradece pela leitura sobre o novo modelo triparametrico, e ressalta a carência quanto ao suporte financeiro da ciência no Brasil. O autor também agradece a seu orientador (Cláudio Javier) pelo ensino e oportunidade, assim como a sua companheira (Karolayne Leonardo) pelo financiamento a publicação, sem estes, o presente artigo não seria possível. Esta pesquisa teve um suporte parcial da CAPES, Brasil.

## REFERÊNCIAS

- Afify, A. Z., Hamedani, G., Ghosh, I., & Mead, M. (2015). The transmuted Marshall-Olkin Fréchet distribution: Properties and applications. *International Journal of Statistics and Probability*, 4, 132-148.
- Ahmad, Z., Hamedani, G., & Butt, N. S. (2019). Recent developments in distribution theory: A brief survey and some new generalized classes of distributions. *Pakistan Journal of Statistics and Operation Research*, 15 (1), 87–110.
- Alzaatreh, A., Lee, C., & Famoye, F. (2013). A new method for generating families of continuous distributions. *METRON*, 71, 63–79.
- Bader, M., & Priest, A. (1982). Statistical aspects of fibre and bundle strength in hybrid composites. *Progress in science and engineering of composites*, 1129–1136.
- Barreto-Souza, W., Cordeiro, G. M., & Simas, A. B. (2011). Some results for beta fréchet distribution. *Communications in Statistics?Theory and Methods*, 40 (5), 798–811.
- Chen, G., & Balakrishnan, N. (1995). A general purpose approximate goodness-of-fit test. *Journal of Quality Technology*, 27, 154–161.
- Coles, S., Bawa, J., Trenner, L., & Dorazio, P. (2001). *An introduction to statistical modeling of extreme values (Vol. 208)*. Springer.
- Cordeiro, G., Ortega, E. M., & da Cunha, D. (2013). The exponentiated generalized class of distributions. *Journal of Data Science*, 11, 1–27.
- Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Math. Polon.*, 6, 93–116.
- Harlow, D. G. (2002). Applications of the frechet distribution function. *International Journal of Materials and Product Technology*, 17 (5-6), 482–495.
- Jørgensen, B. (1982). Examples. In *Statistical properties of the generalized inverse gaussian distribution* (pp. 114–176). Springer.
- Kotz, S., & Nadarajah, S. (2000). *Extreme value distributions: theory and applications*. World Scientific.
- Krishna, E., Jose, K., Alice, T., & Ristić, M. M. (2013). The marshall-olkin fréchet distribution. *Communications in Statistics-Theory and Methods*, 42 (22), 4091–4107.
- Kundu, D., Kannan, N., & Balakrishnan, N. (2008). On the hazard function of Birnbaum-Saunders distribution and associated inference. *Computational Statistics & Data Analysis*, 52 (5), 2692–2702.
- Marshall, A. W., Olkin, I., et al. (2005). A new method for adding a parameter to a family of distributions with application to the exponential and weibull families. *Biometrika*, 92 (2), 505–505.

Mead, M. E. A. (2014). A note on kumaraswamy fréchet distribution. *Australia*, 8 , 294–300.

Tablada, C. J., & Cordeiro, G. M. (2017). The modified fréchet distribution and its properties. *Communications in Statistics-Theory and Methods*, 46 (21), 10617–10639.