

UMA NOVA CLASSE DE DISTRIBUIÇÕES WEIBULL GENERALIZADA

Patrícia Silva Nascimento BARROS¹
Cícero Carlos Ramos de BRITO¹
Leandro Chaves RÉGO²
Wilson Rosa de OLIVEIRA¹
Frank GOMES-SILVA¹

- RESUMO: A distribuição Weibull é muito popular para modelar dados de sobrevivência. Muitas modificações da distribuição Weibull foram propostas nos últimos anos. Inspirado no método T-X (ALZAATREH; LEE; FAMOYE, 2013), propomos uma classe de distribuições Weibull generalizada a qual generaliza a classe Weibull-X. Um tratamento das propriedades matemáticas do novo gerador é realizado, sendo encontradas as expressões para a função de distribuição, função densidade, representação linear para a acumulada e densidade, momentos ordinários e centrais, função geradora de momentos, função característica, média e variância. Discutimos um caso particular do novo gerador e apresentamos, resumidamente, algumas de suas propriedades estruturais. Como uma alternativa à máxima verossimilhança, simulações de Bootstrap e Bootstrap corrigido foram realizadas a fim de comparar esses métodos de estimação. Ajustamos, a quatro conjuntos de dados reais, modelos paramétricos pelo método Bootstrap. Dessa forma, a classe de distribuições Weibull Generalizada, mesmo tendo mais parâmetros a serem estimados, mostrou-se capaz de se ajustar a dados de sobrevivência.
- PALAVRAS-CHAVE: Bootstrap; classe generalizada; método T-X; Weibull-X; máxima verossimilhança.

¹Universidade Federal Rural de Pernambuco - UFRPE, Departamento de Estatística e Informática, CEP: 52171-260, Recife, PE, Brasil. E-mail: *patriciaajp1408@hotmail.com*; *ciceroCarlosbrito@yahoo.com.br*; *wilson.rosa@gmail.com*; *franksinatrags@gmail.com*

²Universidade Federal de Pernambuco - UFPE, Departamento de Estatística, CEP: 50670-901, Recife, PE, Brasil. E-mail: *leandro@de.ufpe.br*

1 Introdução

A análise de sobrevivência é uma das áreas da estatística que mais cresceu nas últimas décadas (COLOSIMO; GIOLO, 2006). Esta área consiste de procedimentos estatísticos para analisar dados relacionados ao tempo até a ocorrência de algum evento de interesse (morte, cura, falha de um componente eletrônico, etc.) que, em geral, pode ser acomodado por uma distribuição de probabilidade. Nos últimos anos, têm sido desenvolvidas novas generalizações e/ou modificações de algumas distribuições utilizadas em análise de sobrevivência. Em particular, a distribuição Weibull, tendo as distribuições exponencial e Rayleigh como casos especiais, é uma distribuição muito utilizada para modelar dados em análise de sobrevivência (PHAM; LAI, 2007).

Mais recentemente, novas distribuições foram propostas com base em modificações da distribuição Weibull, sobretudo, para trabalhar com funções de taxa de risco não-monótonas.

Pham e Lai (2007) proporcionaram revisão de alguns desses modelos. Eles apresentaram a distribuição Weibull exponenciada (MUDHOLKAR; SRIVASTAVA; FRIEMER, 1995, 1996), a distribuição Weibull aditiva (XIE; LAI, 1995), a distribuição Weibull estendida (XIE; TANG; GOH, 2002), a distribuição Weibull modificada (LAI; XIE; MURTHY, 2003) e a distribuição Weibull estendida flexível (BEBBINGTON; LAI; ZITIKIS, 2007). Na direção de generalizações da Weibull a partir dos bem-estabelecidos geradores beta-G e gama-G podemos citar: a beta Weibull (FAMOYE; LEE; OLUMOLADE, 2005), a beta Weibull modificada (SILVA; ORTEGA; CORDEIRO, 2010), a gama Weibull Poisson (PERCONTINI et al., 2014) e a gama Weibull estendida (CORDEIRO et al., 2016).

A ideia de propor novas classes de distribuições tem despertado o interesse de estatísticos teóricos e aplicados, por gerarem distribuições mais flexíveis. Existe um interesse crescente no desenvolvimento de classes de distribuições generalizadas, as G-classes, por introdução de um ou mais parâmetros. A maioria dos métodos geradores de distribuição de probabilidade, desenvolvidos após a década de 80, são métodos de combinação por se basearem na ideia de combinar duas distribuições existentes, ou por adicionar parâmetros extras a alguma distribuição existente a fim gerar uma nova família de distribuições.

Na literatura, até os anos iniciais desta década, encontravam-se algumas formas de se transformar uma distribuição de probabilidade, são elas: a função geradora Marshall-Olkin (MasOlk-G) (MARSHALL; OLKIN, 1997), a função geradora de distribuições beta generalizada (beta-G) (EUGENE; LEE; FAMOYE, 2002; JONES, 2004), a função geradora gama (gama-G) (ZOGRAFOS; BALAKRISHNAN, 2009), a função geradora de distribuições Kumaraswamy (Kw-G) (CORDEIRO; CASTRO, 2011), função geradora de distribuições McDonald (Mc-G) (ALEXANDER et al., 2012), função geradora T-X (Weibull-X e gama-X) (ALZAATREH; LEE; FAMOYE, 2013).

Considere $r(t)$ a função densidade de probabilidade (fdp) de uma variável aleatória $T \in [a, b]$ ($-\infty \leq a < b < \infty$) e seja $W[G(x, \xi)]$ uma função da

acumulada $G(x, \xi)$ de uma variável aleatória X tal que $W[G(x, \xi)]$ satisfaz as seguintes condições:

$$\begin{cases} (i) & W[G(x, \xi)] \in [a, b], \\ (ii) & W[G(x, \xi)] \text{ é diferenciável e monótona não decrescente, e} \\ (iii) & W[G(x, \xi)] \rightarrow a \text{ quando } x \rightarrow -\infty \text{ e } W[G(x, \xi)] \rightarrow b \\ & \text{quando } x \rightarrow \infty, \end{cases} \quad (1)$$

em que ξ é o vetor paramétrico associado de dimensão $p \times 1$. Alzaatreh, Lee e Famoye (2013) definiram o método gerador T-X cuja função de distribuição acumulada (fda) é definida por

$$F(x) = \int_a^{W[G(x, \xi)]} r(t) dt, \quad (2)$$

em que $W[G(x, \xi)]$ satisfaz as condições dadas em (1). A distribuição em (2) admite o mesmo suporte advindo da distribuição-base. A fdp correspondente a (2) é expressa por

$$f(x) = \left\{ \frac{d}{dx} W[G(x, \xi)] \right\} r \{ W[G(x, \xi)] \}. \quad (3)$$

A distribuição Weibull é um dos modelos mais amplamente utilizados para modelar tempo de risco em teoria da confiabilidade e análise de sobrevivência. Entretanto, esta distribuição possui limitação quanto à forma acomodada por sua taxa de risco. A distribuição Weibull a dois parâmetros admite função de risco restrita às formas monótona crescente, constante e monótona decrescente, embora na literatura ela seja descrita como uma distribuição flexível. Em situações práticas, é comum funções de risco exibirem formas não-monótonas tais como banheira, unimodal, dentre outras. Portanto, existe um desejo pela procura de algumas generalizações ou modificações da distribuição Weibull que possam estabelecer mais flexibilidade na modelagem de tempos de vida.

Na literatura recente, quatro geradores baseados na distribuição Weibull foram propostos: beta Weibull-G por Cordeiro, Silva e Ortega (2012), Weibull-X por a Alzaatreh, Lee e Famoye (2013), Weibull-G por Bourguignon, Silva e Cordeiro (2014) e Weibull-X exponenciada por Alzagal, Famoye e Lee (2013).

Admitindo $r(t) = \lambda^\theta \theta t^{\theta-1} e^{-\lambda^\theta t^\theta}$ (t, θ, λ reais positivos) e fazendo $W[G(x, \xi)] = -\log[1 - G(x, \xi)]$ em (2), Alzaatreh, Lee e Famoye (2013) definiram a fda da classe Weibull-X por

$$\begin{aligned} F(x) &= \lambda \theta \int_0^{-\log[1-G(x, \xi)]} (\lambda t)^{\theta-1} e^{-(\lambda t)^\theta} dt \\ &= 1 - \exp[-\lambda^\theta \{-\log[1 - G(x, \xi)]\}^\theta]. \end{aligned} \quad (4)$$

A fdp correspondente a (4) é

$$f(x) = \lambda^\theta \theta \frac{g(x, \boldsymbol{\xi})}{1 - G(x, \boldsymbol{\xi})} e^{-\lambda^\theta \{-\log[1 - G(x, \boldsymbol{\xi})]\}^\theta} \{-\log[1 - G(x, \boldsymbol{\xi})]\}^{\theta-1}.$$

A taxa de risco é dada por

$$R(x) = \lambda^\theta \theta \{-\log[1 - G(x, \boldsymbol{\xi})]\}^{\theta-1} R_G(x)$$

em que $R_G(x) = \frac{g(x, \boldsymbol{\xi})}{1 - G(x, \boldsymbol{\xi})}$. Notamos que a quantidade $\lambda^\theta \theta \{-\log[1 - G(x, \boldsymbol{\xi})]\}^{\theta-1}$ funciona como um fator de correção para $R_G(x)$ que é a taxa de risco da distribuição-base. Alzaatreh e Ghosh (2015) apresentaram algumas propriedades e caracterizações da classe weibull-X. Neste artigo, propomos uma classe de distribuições chamada *nova Weibull-X* (NWX) a qual é muito flexível para modelagem de dados de sobrevivência podendo admitir taxa de risco na forma constante, crescente, decrescente, banheira, unimodal, J, J invertido, dentre outras.

2 A nova classe

Considere uma distribuição-base $G(x, \boldsymbol{\xi})$ qualquer e o integrando $r(t)$ visto em (4). Definimos a fda da classe proposta pela seguinte expressão

$$\begin{aligned} F(x) &= \lambda \theta \int_{\delta[1 - G(x, \boldsymbol{\xi})]}^{\delta - \log[1 - G(x, \boldsymbol{\xi})]} (\lambda t)^{\theta-1} e^{-(\lambda t)^\theta} dt \\ &= e^{-\lambda^\theta [\delta(1 - G(x, \boldsymbol{\xi}))]^\theta} - e^{-\lambda^\theta \{\delta - \log[1 - G(x, \boldsymbol{\xi})]\}^\theta}, \quad \delta \geq 0, \quad x > 0. \end{aligned} \quad (5)$$

Note que a classe anterior admite a Weibull-X como subcaso quando fazemos $\delta = 0$. Esta classe não depende de função complicada e é razoavelmente tratável quanto mais simples for a distribuição-base $G(x, \boldsymbol{\xi})$. Diferentemente do que acontece na classe T-X (ALZAATREH; LEE; FAMOYE, 2013), utilizamos na acumulada geradora (5) limites de integração (tanto inferior, quanto superior) que constituem funções monótonas da distribuição-base $G(x, \boldsymbol{\xi})$. Em um certo sentido, nossa proposta generaliza o método gerador proposto por estes autores.

Se $G(x, \boldsymbol{\xi})$ for uma fda contínua, derivando-se $F(x)$ temos, para $x > 0$, a fdp associada dada por:

$$\begin{aligned} f(x) &= \theta \delta \lambda^\theta g(x) \{\delta[1 - G(x, \boldsymbol{\xi})]\}^{\theta-1} e^{-\lambda^\theta \{\delta[1 - G(x, \boldsymbol{\xi})]\}^\theta} \\ &\quad + \theta \lambda^\theta \frac{g(x)}{1 - G(x, \boldsymbol{\xi})} \{\delta - \log[1 - G(x, \boldsymbol{\xi})]\}^{\theta-1} e^{-\lambda^\theta \{\delta - \log[1 - G(x, \boldsymbol{\xi})]\}^\theta}. \end{aligned} \quad (6)$$

Numa tentativa de padronizar nomenclatura e levando-se em conta aquela existente na literatura, vamos estabelecer a partir de agora que $X \sim \text{NWX}(\delta, \theta, \lambda, \boldsymbol{\xi})$, quando a variável aleatória X admitir fdp dada em (6). Algumas vezes omitiremos a dependência do vetor $\boldsymbol{\xi}$ e simplesmente escreveremos $G(x, \boldsymbol{\xi}) = G(x)$.

2.1 Representação linear

As propriedades e os métodos de estimação dos parâmetros da classe exponenciada têm sido estudados por diversos autores, dentre eles destacam-se: Mudholkar, Srivastava e Freimer (1993) com a distribuição Weibull exponenciada, Gupta, Gupta e Gupta (1998) introduziram a classe geral de distribuições exponenciadas, Gupta e Kundu (1999) propuseram a distribuição exponencial generalizada, Nadarajah (2005) definiu a Pareto exponenciada, Nadarajah e Kotz (2006) propuseram a beta exponenciada e Nadarajah (2011) discutiu sobre as propriedades da distribuição exponencial exponenciada. A seguir será definida a distribuição exponenciada- G ($\exp -G$).

Seja $G(y)$ uma distribuição-base arbitrária. Diz-se que $Y \sim \exp-G(a)$, com $a > 0$, se a fda e a fdp de Y são, respectivamente

$$G_a(y) = G^a(y) \quad (7)$$

e

$$g_a(y) = a g(y) G^{a-1}(y), \quad (8)$$

em que $g(y) = \frac{dG(y)}{dy}$.

Para o caso particular $a = 1$, tem-se a própria distribuição-base. Muitos resultados obtidos da $\exp -G$ são utilizados em diversas classes generalizadas. Isto porque as expansões da fda e fdp dessas classes são combinações lineares das fda e fdp da $\exp -G$.

Podemos obter uma expansão da fda da seguinte maneira. Expandindo em série de Taylor as quantidades $M_1(x) = e^{-\lambda^\theta (\delta[1-G(x)])^\theta}$ e $M_2(x) = e^{-\lambda^\theta (\delta - \log[1-G(x)])^\theta}$, obtemos respectivamente

$$M_1(x) = \sum_{s=0}^{\infty} \frac{(-1)^s \lambda^{s\theta} \delta^{s\theta}}{s!} [1 - G(x)]^{s\theta} \quad \text{e} \quad M_2(x) = \sum_{k=0}^{\infty} a_k \left(G(x) - \frac{e-1}{e} \right)^k$$

em que $a_k = \frac{M_2^{(k)}\left(\frac{e-1}{e}\right)}{k!}$ e $M_2^{(k)}\left(\frac{e-1}{e}\right)$ é a k -ésima derivada de $M_2(x)$ avaliada no ponto $\frac{e-1}{e}$. Uma expansão para a fda (5) é obtida ao expandir os binômios que surgem nas quantidades $M_1(x)$ e $M_2(x)$. Portanto, podemos escrever a acumulada (5) como uma diferença de combinações lineares de exponenciadas da distribuição-base $G(x)$

$$F(x) = \sum_{s=0}^{\infty} \sum_{z=0}^{\infty} \binom{s\theta}{z} \frac{(-1)^{s+z}}{s!} \lambda^{s\theta} \delta^{s\theta} G_z(x) - \sum_{k=0}^{\infty} \sum_{j=0}^k \binom{k}{j} a_k (-1)^j \left(\frac{e-1}{e} \right)^j G_{k-j}(x). \quad (9)$$

Caso a distribuição $G(x)$ seja de variável aleatória contínua, temos $f(x) = \frac{dF(x)}{dx}$ e portanto

$$f(x) = \sum_{s=0}^{\infty} \sum_{z=0}^{\infty} \binom{s\theta}{z} \frac{(-1)^{s+z}}{s!} \lambda^{s\theta} \delta^{s\theta} g_z(x) - \sum_{k=0}^{\infty} \sum_{j=0}^k \binom{k}{j} a_k (-1)^j \left(\frac{e-1}{e}\right)^j g_{k-j}(x). \quad (10)$$

2.2 Expansões para os momentos

Os momentos são características de interesse muito importantes em qualquer análise estatística. Eles podem ser utilizados para estudar propriedades estatísticas tais como tendência, dispersão, assimetria e curtose. O m -ésimo momento de X , digamos μ_m , pode ser facilmente obtido a partir da Equação (10) e $Y_a \sim \text{exp-G}(a)$ como

$$\mu_m = E(X^m) = \sum_{s=0}^{\infty} \sum_{z=0}^{\infty} \binom{s\theta}{z} \frac{(-1)^{s+z}}{s!} (\lambda \delta)^{s\theta} E(Y_z^m) - \sum_{k=0}^{\infty} \sum_{j=0}^k \binom{k}{j} a_k (-1)^j \left(\frac{e-1}{e}\right)^j E(Y_{k-j}^m). \quad (11)$$

Em particular, a média de X pode ser calculada a partir da média de qualquer exponenciada fazendo $\mu = \mu_1$ conforme a Equação (11). A seguir, apresentamos os cálculos para expansão dos momentos centrais de ordem m . O m -ésimo momento de X , digamos μ'_m , segue em função dos momentos ordinários por meio da bem conhecida expressão

$$\mu'_m = \sum_{r=0}^m \binom{m}{r} (-1)^r \mu_{m-r} \mu^r. \quad (12)$$

Utilizando a Equação (12) e a expansão (10), podemos escrever os momentos centrais de X a partir dos momentos de ordem $m-r$ de exponenciadas

$$\mu'_m = E(X - \mu)^m = \sum_{r=0}^m \sum_{s=0}^{\infty} \sum_{z=0}^{\infty} \binom{s\theta}{z} \frac{(-1)^{s+z}}{s!} (\lambda \delta)^{s\theta} E(Y_z^{m-r}) - \sum_{r=0}^m \sum_{k=0}^{\infty} \sum_{j=0}^k \binom{k}{j} a_k (-1)^j \left(\frac{e-1}{e}\right)^j E(Y_{k-j}^{m-r}). \quad (13)$$

Em particular, podemos obter uma expansão para a variância fazendo $\sigma^2 = \mu'_2$. As Equações (11) e (13) constituem o principal resultado desta Seção.

2.3 Expansões para funções geradoras

Nesta seção derivamos uma expressão explícita para a função geratriz de momentos (fgm) $M_X(t)$ e função característica (fc) $\varphi_X(t)$. Temos, a partir da expansão dada em (10), o seguinte resultado para X

$$M_X(t) = \sum_{s=0}^{\infty} \sum_{z=0}^{\infty} \binom{s\theta}{z} \frac{(-1)^{s+z}}{s!} (\lambda \delta)^{s\theta} M_{Y_z}(t) - \sum_{k=0}^{\infty} \sum_{j=0}^k \binom{k}{j} a_k (-1)^j \left(\frac{e-1}{e}\right)^j M_{Y_{k-j}}(t), \quad (14)$$

em que $M_{Y_z}(t)$ e $M_{Y_{k-j}}(t)$ são as fgm's de Y_z e Y_{k-j} , respectivamente. De modo análogo ao realizado na obtenção da Equação (14), podemos obter uma expansão para $\varphi_X(t)$ da seguinte maneira

$$\varphi_X(t) = \sum_{s=0}^{\infty} \sum_{z=0}^{\infty} \binom{s\theta}{z} \frac{(-1)^{s+z}}{s!} (\lambda \delta)^{s\theta} \varphi_{Y_z}(t) - \sum_{k=0}^{\infty} \sum_{j=0}^k \binom{k}{j} a_k (-1)^j \left(\frac{e-1}{e}\right)^j \varphi_{Y_{k-j}}(t), \quad (15)$$

em que $\varphi_{Y_z}(t)$ e $\varphi_{Y_{k-j}}(t)$ são as fc's de Y_z e Y_{k-j} , respectivamente. As Equações (14) e (15) são o principal resultado desta seção e revelam que tanto a fgm de X quanto a fc de X podem ser expressas por meio das fgm's e fc's de exponenciadas da distribuição-base $G(x)$.

3 Estimação

3.1 Máxima verossimilhança

Várias abordagens para estimativas de parâmetros são propostas na literatura, sendo o método de máxima verossimilhança a mais utilizada. Os estimadores de máxima verossimilhança desfrutam de propriedades desejáveis e podem ser usados na construção de intervalos de confiança e testes de hipóteses. A aproximação normal para estes estimadores em distribuições de grandes amostras é facilmente manipulada, seja analítica ou numericamente.

Seja x_1, x_2, \dots, x_n uma amostra aleatória a partir de (6) e seja $\Theta = (\delta, \theta, \lambda, \xi)^T$ um vetor de parâmetros de dimensão p . O logaritmo da função de verossimilhança para Θ é dada por

$$l(\Theta) = n \log \theta + n \theta \log \lambda + \sum_{i=1}^n \log \left\{ \delta g(x, \xi) \{ \delta [1 - G(x, \xi)] \}^{\theta-1} e^{-\lambda^\theta [\delta(1-G(x, \xi))]} \right. \\ \left. + \frac{g(x, \xi)}{1-G(x, \xi)} \{ \delta - \log [1 - G(x, \xi)] \}^{\theta-1} e^{-\lambda^\theta \{ \delta - \log [1 - G(x, \xi)] \}^\theta} \right\}. \quad (16)$$

Assumimos que as seguintes condições de regularidade para o logaritmo da função de verossimilhança são válidas: i) o suporte de X associado à distribuição não depende de parâmetros desconhecidos; ii) o espaço paramétrico de X , digamos Ψ , é aberto e $l(\Theta)$ tem um máximo global em Ψ ; iii) para quase todos os x , existem as derivadas de quarta ordem do logaritmo da função de verossimilhança com respeito aos parâmetros do modelo e são contínuas em um subconjunto aberto de Ψ que contém o parâmetro verdadeiro; iv) a matriz de informação esperada é definida positiva e finita; v) os valores absolutos das derivadas de terceira ordem do logaritmo da função de verossimilhança em relação aos parâmetros são delimitadas por funções finitas das esperanças de X .

Seja a função score definida como $U(\Theta) = \frac{\partial l(\Theta)}{\partial \Theta}$. As componentes U_δ , U_θ , U_λ e U_ξ são dadas no Apêndice.

Os estimadores de máxima verossimilhança, $\hat{\Theta} = (\hat{\delta}, \hat{\theta}, \hat{\lambda}, \hat{\xi})^\top$ de $\Theta = (\delta, \theta, \lambda, \xi)^\top$ são obtidos resolvendo simultaneamente as equações não-lineares $U_\delta = 0$, $U_\theta = 0$, $U_\lambda = 0$ e $U_\xi = 0$. Estas equações não podem ser resolvidas analiticamente e softwares podem ser usados para resolvê-las numericamente. Aqui, adotamos o software R.

Para estimação intervalar e testes de hipóteses com respeito aos parâmetros do modelo, obtemos a matriz de informação de Fisher observada $J(\Theta)$ de dimensão 4×4 dada por

$$J(\Theta) = - \begin{pmatrix} U_{\delta\delta} & U_{\delta\theta} & U_{\delta\lambda} & U_{\delta\xi} \\ \cdot & U_{\theta\theta} & U_{\theta\lambda} & U_{\theta\xi} \\ \cdot & \cdot & U_{\lambda\lambda} & U_{\lambda\xi} \\ \cdot & \cdot & \cdot & U_{\xi\xi} \end{pmatrix}.$$

Os elementos de $J(\Theta)$ são calculados a partir de $U_{ij} = \frac{\partial^2 l(\Theta)}{\partial \theta_i \partial \theta_j}$, com $i, j \in \{\delta, \theta, \lambda, \xi\}$, e por questão de espaço não serão apresentados. A matriz $J(\Theta)$ é útil para obter intervalos de confiança aproximados para os parâmetros. Uma crítica que se faz as G-classes é que esses geradores, apesar de admitir dois a três parâmetros adicionais e portanto auxiliar na flexibilização do modelo, apresentam dificuldade na obtenção de suas estimativas por meio de recursos numéricos.

A estatística do *Teste da Razão de Verossimilhança* (TRV) é dada, sob $H_0 : \theta \in \Theta_0$ (vs $H_1 : \theta \notin \Theta_0$), por

$$\Lambda = 2 \log \left(\frac{\sup_{\theta \in \Theta} L(\theta)}{\sup_{\theta \in \Theta_0} L(\theta)} \right) = 2\ell(\hat{\theta}) - 2\ell(\theta_0) \stackrel{a}{\sim} \chi_{p-q}^2$$

em que $\ell(\hat{\theta})$ é o logaritmo da função de verossimilhança obtida no seu ponto máximo, assim denominado estimador de máxima verossimilhança, sob o modelo proposto com p parâmetros e $\ell(\theta_0)$ é o logaritmo da função de verossimilhança obtida no seu ponto máximo (estimador de máxima verossimilhança), sob o modelo reduzido (sub-modelo) com q parâmetros. Como Λ tem distribuição assintótica qui-quadrado, para se proceder o teste deve-se comparar a estatística Λ ao quantil desejado da distribuição χ^2 . Suponha que o modelo completo tem três parâmetros, enquanto

seu sub-modelo (modelo proposto sob H_0) tem um parâmetro. Se $\Lambda \geq \chi_2^2 = 5,99$ rejeita-se H_0 ao nível de confiança de 95%, isto é, o modelo proposto é melhor.

3.2 Bootstrap

Bootstrap é um método de reamostragem proposto por Efron (1979). Os métodos de Bootstrap são uma classe de métodos de Monte Carlo não paramétricos que estimam a distribuição da população por reamostragem. Existem muitas técnicas de reamostragem que visam estimar parâmetros de uma distribuição de interesse. Uma vantagem em utilizar a técnica de reamostragem Bootstrap é a generalidade com que pode ser aplicada, pois requer que menos suposições sejam feitas, além de geralmente fornecer respostas mais precisas.

Geralmente tomamos uma amostra aleatória X_1, \dots, X_n de uma distribuição F_X e construímos uma função $\hat{\theta} = g(X_1, \dots, X_n)$ chamada estatística a partir da qual se realizam inferências sobre um parâmetro θ da população. O método Bootstrap repete esse procedimento utilizando a amostra aleatória como base para se construir uma distribuição empírica F_n (WASSERMAN, 2004). O algoritmo a seguir ilustra todo o processo.

Algoritmo Bootstrap:

1. Simule $X_1^*, \dots, X_n^* \sim F_n$;
2. Calcule $\hat{\theta}^* = g(X_1^*, \dots, X_n^*)$;
3. Repita os passos (1) e (2) B vezes e compute $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*$;
4. Calcule

$$EP_{boot}(\hat{\theta}^*) = \sqrt{\frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b^* - \hat{\theta}_{(\cdot)}^*)^2}$$

em que $\hat{\theta}_{(\cdot)}^* = \frac{1}{B} \sum_{r=1}^B \hat{\theta}_r^*$.

Uma correção para as estimativas Bootstrap pode ser obtida definindo

$$\tilde{\theta} = \hat{\theta}^* - vies$$

e ao fazer $vies = \hat{\theta}_{(\cdot)}^* - \hat{\theta}^*$, chegamos à seguinte correção

$$\tilde{\theta} = 2\hat{\theta}^* - \hat{\theta}_{(\cdot)}^*.$$

Supondo que $\hat{\theta}_{(\cdot)}^*$ e $\hat{\theta}^*$ são independentes, o erro padrão de $\tilde{\theta}$ fica

$$EP(\tilde{\theta}) = \sqrt{4Var(\hat{\theta}^*) + Var(\hat{\theta}_{(\cdot)}^*)}$$

em que $Var(\hat{\theta}^*)$ é o quadrado do $EP_{boot}(\hat{\theta}^*)$. Para obtermos uma estimativa para $Var(\hat{\theta}_{(\cdot)}^*)$ repetimos o algoritmo de Bootstrap uma quantidade arbitrária C e tomamos ao final do processo a variância amostral de $\hat{\theta}_{(\cdot),1}^*, \dots, \hat{\theta}_{(\cdot),C}^*$.

Utilizamos essa técnica Bootstrap para obter as estimativas e respectivos erros-padrão de ajustes de densidades a conjuntos univariados de dados nas Seções 5 e 6.

4 Um caso particular

Nesta seção apresentamos um caso particular da classe proposta admitindo $\delta = 0$ e considerando como distribuição-base o modelo log-logístico (LL). A fdp e fda da distribuição LL são dadas, respectivamente, por $g(x) = \frac{(\beta/\alpha)(x/\alpha)^{\beta-1}}{1+(\frac{x}{\alpha})^{-\beta}}$ e $G(x) = \frac{1}{1+(\frac{x}{\alpha})^{-\beta}}$, em que $x > 0$, $\alpha > 0$ é um parâmetro de escala e $\beta > 0$ é um parâmetro de forma. De fato, ao considerar $\delta = 0$, obtemos a classe Weibull-X, e por este motivo vamos estudar nesta seção a distribuição *Weibull log-logística* (WLL). A fda da distribuição WLL é dada por

$$\begin{aligned} F_{\text{WLL}}(x) &= \int_0^{-\log\left\{1 - \frac{1}{1+(\frac{x}{\alpha})^{-\beta}}\right\}} \theta \lambda^\theta t^{\theta-1} e^{-(\lambda t)^\theta} dt \\ &= 1 - e^{-\left\{\lambda \log\left[1 + \left(\frac{x}{\alpha}\right)^\beta\right]\right\}^\theta}, \end{aligned} \quad (17)$$

em que $x > 0$, $\alpha > 0$, $\beta > 0$, $\theta > 0$ e $\lambda > 0$. Como a distribuição-base admitida é contínua, derivando-se $F_{\text{WLL}}(x)$ temos que a fdp associada é expressa por:

$$f_{\text{WLL}}(x) = \frac{\beta \theta \lambda \left(\frac{\alpha}{x}\right)^{\beta-1} \left\{\lambda \log\left[1 + \left(\frac{x}{\alpha}\right)^\beta\right]\right\}^{\theta-1} e^{-\left\{\lambda \log\left[1 + \left(\frac{x}{\alpha}\right)^\beta\right]\right\}^\theta}}{\alpha \left[1 + \left(\frac{x}{\alpha}\right)^\beta\right]}. \quad (18)$$

As distribuições log-logística (LL) e log-logística exponenciada (LLE) são os principais submodelos de WLL. As distribuições LL e LLE seguem de WLL quando fazemos, respectivamente, $\lambda = \theta = 1$ e $\theta = 1$.

A seguir, apresentamos os cálculos do desenvolvimento da expansão da função de distribuição e da densidade da distribuição WLL.

Seguindo passo análogo àquele utilizado na obtenção da Equação (9) e depois de alguns desenvolvimentos algébricos, uma expansão para a fda da distribuição WLL pode ser expressa por:

$$\begin{aligned} F_{\text{WLL}}(x) &= 1 - \sum_{l, \nu, k=0}^{\infty} \sum_{j=0}^k \sum_{z=1}^{\infty} \binom{k}{j} \binom{j-k}{l} \binom{z}{\nu} \frac{a_k (-1)^{j+l+\nu}}{z} \\ &\quad \times \left(\frac{e-1}{e}\right)^j e^{-\left(\frac{x}{\alpha}\right)^{-\nu l \beta}}. \end{aligned}$$

A distribuição WLL é de variável contínua e portanto teremos $f_{\text{WLL}}(x) = \frac{dF_{\text{WLL}}(x)}{dx}$, ou seja,

$$f_{WLL}(x) = \sum_{l,\nu,k=0}^{\infty} \sum_{j=0}^k \sum_{z=1}^{\infty} \zeta_{l,\nu,k,j,z} h_{l,\nu,\alpha,\beta}(x) \quad (19)$$

em que

$$\zeta_{l,\nu,k,j,z} = \binom{k}{j} \binom{j-k}{l} \binom{z}{\nu} \frac{a_k (-1)^{j+l+\nu}}{z} \left(\frac{e-1}{e}\right)^j$$

e

$$h_{l,\nu,\alpha,\beta}(x) = l\nu\beta(\alpha^{-1}x)^{l\nu\beta} e^{-(\alpha^{-1}x)^{-l\nu\beta}}.$$

A Figura 1 contém a função densidade de probabilidade variando os valores de α , β , λ e θ . É possível observar que os formatos são variados, incluindo, exponencial, simétrica e assimétrica positiva dependendo dos valores dos parâmetros.

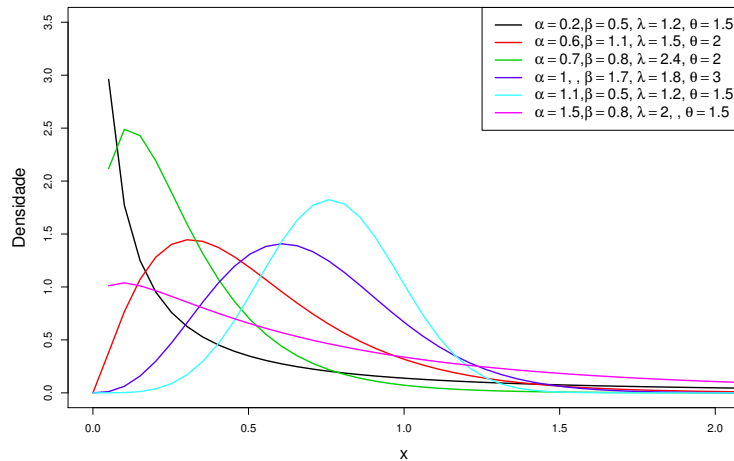


Figura 1 - Função densidade de probabilidade variando os valores de α , β , λ e θ .

A função de sobrevivência e a função de risco são definidas, respectivamente, por:

$$S_{WLL}(x) = 1 - F_{WLL}(x) = e^{-\left\{\lambda \log \left[1 + \left(\frac{x}{\alpha}\right)^\beta\right]\right\}^\theta}$$

e

$$R_{WLL}(x) = \frac{\beta\theta\lambda \left(\frac{x}{\alpha}\right)^{\beta-1} \left\{\lambda \log \left[1 + \left(\frac{x}{\alpha}\right)^\beta\right]\right\}^{\theta-1}}{\alpha \left[1 + \left(\frac{x}{\alpha}\right)^\beta\right]}.$$

A Figura 2 apresenta a função de risco variando os valores de α , β , λ e θ . A função de risco apresenta comportamentos constantes, crescentes e decrescentes.

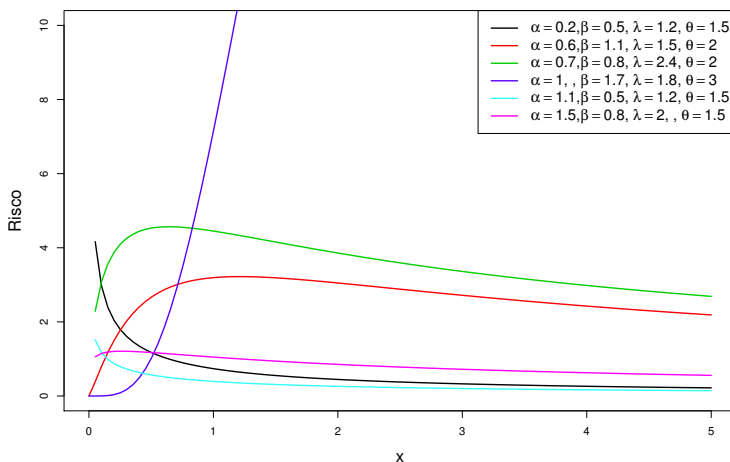


Figura 2 - Função de risco variando os valores de α , β , λ e θ .

4.1 Propriedades estruturais

Seja $X \sim WLL(\alpha, \beta, \lambda, \theta)$. Nesta seção apresentamos algumas propriedades estruturais para a distribuição WLL que são facilmente obtidas a partir da utilização direta da Equação (19). Na Tabela 1 encontram-se expressões explícitas para os momentos ordinários (M. O.), variância (Var), coeficientes de assimetria (C. A.), coeficiente de curtose (C. C.), fgm e fc para o modelo em questão.

Tabela 1 - Algumas propriedades estruturais para a distribuição WLL

Medidas	Expressões explícitas
M. O.	$E(X^r) = \sum_{l,\nu,k=0}^{\infty} \sum_{j=0}^k \sum_{z=1}^{\infty} \zeta_{l,\nu,k,j,z} \alpha^r \Gamma \left[1 + \frac{r}{l\nu\beta} \right]$
Var	$\sigma^2 = \sum_{l,\nu,k=0}^{\infty} \sum_{j=0}^k \sum_{z=1}^{\infty} \zeta_{l,\nu,k,j,z} \alpha^2 \left(\Gamma \left[1 + \frac{2}{l\nu\beta} \right] - \Gamma \left[1 + \frac{1}{l\nu\beta} \right]^2 \right)$
C. A.	$\gamma_1 = \sum_{l,\nu,k=0}^{\infty} \sum_{j=0}^k \sum_{z=1}^{\infty} \zeta_{l,\nu,k,j,z} \frac{\alpha^3 \Gamma \left[1 + \frac{3}{l\nu\beta} \right] - 3\mu\sigma^2 - \mu^3}{\sigma^3}$
C. C.	$\gamma_2 = \sum_{l,\nu,k=0}^{\infty} \sum_{j=0}^k \sum_{z=1}^{\infty} \zeta_{l,\nu,k,j,z} \frac{\alpha^4 \Gamma \left[1 + \frac{4}{l\nu\beta} \right] - 3\gamma_1 \mu \sigma^3 - 6\mu^2 \sigma^2 - \mu^4}{\sigma^4}$
fgm	$M_X(t) = \sum_{l,\nu,k=0}^{\infty} \sum_{j=0}^k \sum_{z=1}^{\infty} \zeta_{l,\nu,k,j,z} \left(1 + \sum_{m=0}^{\infty} \frac{(\alpha t)^{r+1}}{l\nu\beta m!} \Gamma \left[\frac{m+1}{l\nu\beta} \right] \right)$
fc	$\varphi_X(t) = \sum_{l,\nu,k=0}^{\infty} \sum_{j=0}^k \sum_{z=1}^{\infty} \zeta_{l,\nu,k,j,z} \left(1 + \sum_{m=0}^{\infty} \frac{(\alpha it)^{r+1}}{l\nu\beta m!} \Gamma \left[\frac{m+1}{l\nu\beta} \right] \right)$

O valor μ da média de X é facilmente encontrado ao admitirmos $r = 1$ na primeira linha da Tabela 1. Vale salientar que também podemos obter outras expressões explícitas para a maioria dessas medidas, utilizando para isso, as expansões dadas nas Seções 2.2 e 2.3. Note que essa obtenção, além de simples, é totalmente factível, uma vez que várias propriedades da distribuição log-logística exponenciada encontram-se documentadas na literatura, veja por exemplo Rosaiah, Kantan e Kumar (2006). As fórmulas que aparecem nesta seção e no restante do artigo são simples de serem calculadas a partir de softwares de linguagem algébrica tais como Maple, Matlab e Mathematica.

4.2 Função quantílica

Invertendo a fda (17), obtemos a função quantílica da distribuição WLL, digamos $x = Q(p)$, expressa por:

$$x = Q(p) = F_{\text{WLL}}^{-1}(p) = \alpha \left[e^{\frac{[-\log(1-p)]^{1/\theta}}{\lambda}} - 1 \right]^{1/\beta}. \quad (20)$$

Com a função quantílica (20), obtemos 450 números pseudo-aleatórios da distribuição $X \sim \text{WLL}(\alpha, \beta, \lambda, \theta)$ com $\alpha = 0, 3$, $\beta = 0, 5$, $\lambda = 1, 2$ e $\theta = 1, 5$. Os parâmetros estimados pelo método de máxima verossimilhança foram: $\alpha = 0, 139$, $\beta = 0, 448$, $\lambda = 0, 944$ e $\theta = 1, 786$. A Figura 3 mostra o histograma com os parâmetros verdadeiros e estimados da distribuição WLL. O valor do logaritmo da função de verossimilhança para os dados gerados foi de 0,202 e para os parâmetros verdadeiros 0,205, mostrando um ajuste razoável pelo método de máxima verossimilhança.

5 Aplicação: Dados Simulados

Seja $\psi = (\alpha, \beta, \lambda, \theta)$ os parâmetros da distribuição WLL. Simulamos 450 valores da distribuição WLL com $\alpha = 0, 3$, $\beta = 0, 5$, $\lambda = 1, 2$, $\theta = 1, 5$ para testar a adequação da distribuição proposta. A seguir, ajustamos aos dados, a densidade com os parâmetros verdadeiros e sua estimativa de máxima verossimilhança (EMV) via algoritmo de estimação BFGS. Repetimos a simulação 1000 vezes utilizando máxima verossimilhança, reamostragem Bootstrap e Bootstrap corrigido. Na Figura 4 podemos observar o histograma e os respectivos ajustes aos dados. Verificamos que os ajustes dos EMVs, Bootstrap e Bootstrap corrigido guardam relação de semelhança segundo inspeção gráfica.

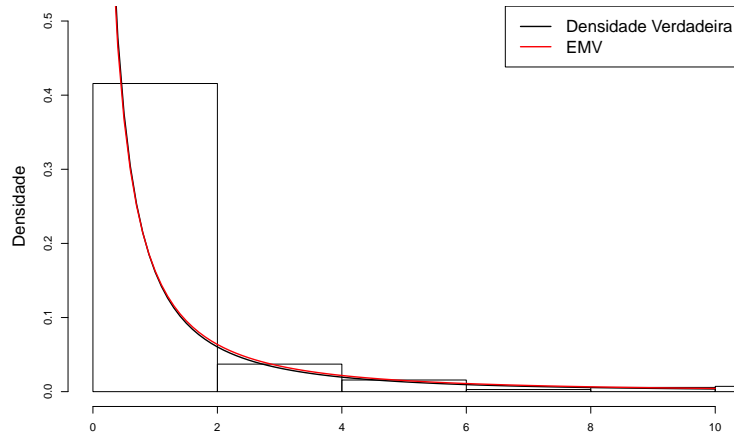


Figura 3 - Histograma com os valores gerados, ajustado a partir da densidade com parâmetros verdadeiros e densidade com estimativas de máxima verossimilhança (EMV) da distribuição WLL.

Na Tabela 2 apresentamos os valores dos parâmetros verdadeiros, as estimativas de máxima verossimilhança, Bootstrap e Bootstrap corrigido para os dados e os respectivos erros-padrão entre parênteses. O teste de Wald, para as EMVs, ao nível de significância de 5% para $H_0 : \alpha = 0$ contra $H_1 : \alpha \neq 0$, $H_0 : \lambda = 0$ contra $H_1 : \lambda \neq 0$ e $H_0 : \theta = 0$ contra $H_1 : \theta \neq 0$, apresenta valores da estatística de teste, digamos W , menores que 2, rejeitando assim H_0 e concluindo que as estimativas de α , λ e θ não são significativas. Para o parâmetro β , obtivemos o valor de $W = 2,68$ e concluímos que a EMV de β é significativa ao nível de 5%. As estimativas pelo método Bootstrap e Bootstrap corrigido são todas significativas, ao nível de 5% pelo teste de Wald.

Tabela 2 - Estimativas dos parâmetros com os respectivos erros-padrão

	α	β	λ	θ
Parâmetros	0,3	0,5	1,2	1,5
EMV	0,119 (0,353)	0,439 (0,164)	0,911 (0,606)	1,845 (1,105)
Bootstrap	0,130 (0,030)	0,449 (0,010)	0,927 (0,069)	1,794 (0,113)
Bootstrap Corrigido	0,158 (0,060)	0,462 (0,019)	0,983 (0,138)	1,684 (0,227)

Na Tabela 3 apresentamos os critérios de Akaike (AIC), Akaike corrigido (AICC), Bayesiano (BIC) e Hannan-Quinn (HQIC) e estatísticas de testes de aderência para os diferentes métodos de estimação empregados, bem como para

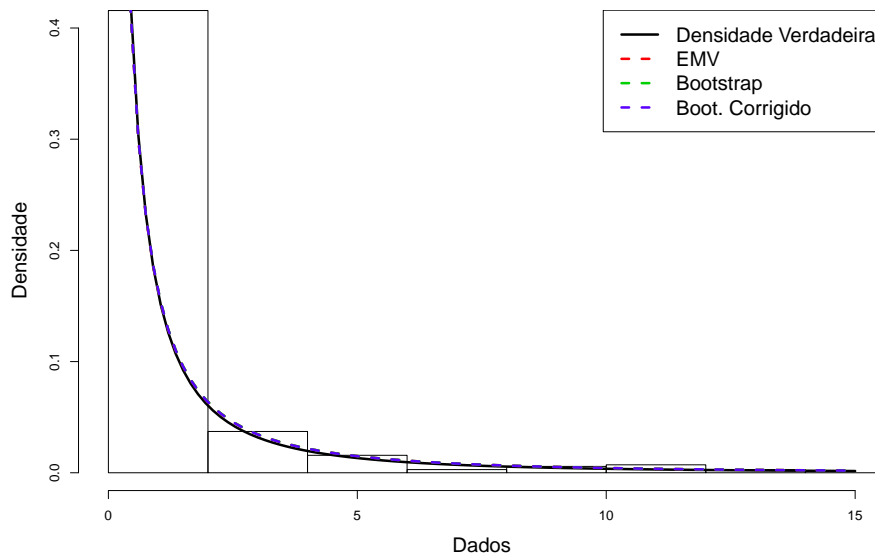


Figura 4 - Histograma para 450 valores gerados de uma distribuição WLL com $\alpha = 0,3$; $\beta = 0,5$; $\lambda = 1,2$; $\theta = 1,5$ e os ajustes para os dados.

os parâmetros verdadeiros. As estatísticas de teste utilizadas nesta tabela foram Cramér-von Mises (W^*), Anderson-Darling (A^*) e Kolmogorov-Smirnov (K-S), com p -valores entre parênteses para o teste K-S. Verificamos que os valores dos critérios e das estatísticas para os estimadores de máxima verossimilhança, Bootstrap e Bootstrap corrigido são semelhantes. Portanto, estes resultados apontam a metodologia Bootstrap como um importante método alternativo inferencial.

Tabela 3 - Critérios de informação e Estatísticas de teste para aderência das estimativas de máxima verossimilhança, Bootstrap e Bootstrap corrigido para os dados simulados

	Valores verdadeiros	EMV	Bootstrap	Boot. corrigido
<i>AIC</i>	587,422	585,794	585,804	585,884
<i>AICC</i>	587,843	586,215	586,225	586,303
<i>BIC</i>	597,842	596,214	596,224	596,303
<i>HQIC</i>	591,639	590,011	590,021	590,100
W^*	0,0117	0,0121	0,0122	0,0125
A^*	0,093	0,099	0,098	0,097
K-S (p -valor)	0,035 (0,787)	0,018 (0,999)	0,019 (0,999)	0,029 (0,921)

6 Aplicação: Dados reais

Nesta seção, usamos quatro conjuntos de dados reais para comparar os ajustes da WLL com dois de seus submodelos e em seguida com alguns modelos bem estabelecidos na literatura. Em cada caso, os parâmetros são estimados utilizando o método Bootstrap, por meio do pacote *boot*, através do software R. Primeiro, descrevemos os conjuntos de dados e fornecemos as EMVs (com correspondentes erros-padrão entre parênteses) dos respectivos parâmetros e os critérios AIC, AICC, BIC e HQIC. As estatísticas W^* , A^* e K-S (com p -valores entre parênteses) também são apresentadas. Para os critérios e as estatísticas, quanto menor seus valores, melhor o ajuste ao conjunto de dados. No final de cada subseção, apresentamos histogramas dos dados para estabelecermos uma comparação visual das densidades ajustadas. Em particular, na próxima subseção, utilizamos o TRV para testar o modelo proposto versus seus submodelos.

6.1 Modelos aninhados

Nesta subseção, ajustamos o modelo $X \sim \text{WLL}(\alpha, \beta, \lambda, \theta)$ para dois conjuntos de dados reais e comparamos com os ajustes dos submodelos LL e LLE.

(i) Dados: rolamentos

Os dados obtidos a partir de Lawless (2011) representam a resistência de rolamentos rígidos de esferas. Consistem no número de milhões de revoluções antes da falha para cada um dos 23 rolamentos de esferas no teste de vida. Os dados estão na Tabela 4:

Tabela 4 - Dados de resistência de rolamentos rígidos de esferas

17,88	28,92	33,00	41,52	42,12	45,60	48,40	51,84	51,84	51,96
55,56	67,80	68,64	68,64	68,88	84,12	93,12	98,64	105,12	105,84
128,04	173,40	54,12	127,92						

(ii) Dados: fibra nervosa

Os dados apresentados em Cox e Lewis (1966) e relatados em Jørgensen (2012) correspondem a um registro de 799 intervalos entre os pulsos ao longo de uma fibra nervosa. Por questão de espaço não os apresentaremos.

Na Tabela 5 estão os parâmetros estimados pelo método Bootstrap para os modelos estudados. Para os dois conjuntos de dados, verificamos pelo teste de Wald, ao nível de 5%, que as estimativas dos parâmetros são significativas para todos os modelos.

A tabela 5 mostra os valores dos critérios e estatísticas de teste de aderência para os modelos utilizados. Tanto para os dados de rolamentos, quanto para os dados de fibra nervosa, podemos perceber que todos os critérios indicam que o modelo WLL se ajusta melhor a esses conjuntos de dados em comparação aos demais modelos. Verificamos também que o modelo WLL obteve os menores valores para

Tabela 5 - EMVs, critérios de informação e estatísticas de aderência

Rolamentos	WLL	LL	LLE
α	2,262 (0,358)	7,169 (0,423)	3,255 (0,033)
β	1,385 (0,152)	0,665 (0,027)	1,721 (0,015)
λ	0,208 (0,013)	1 -	0,201 (0,005)
θ	4,006 (0,740)	1 -	1 -
<i>AIC</i>	254,576	312,937	305,472
<i>AICC</i>	255,266	313,137	305,879
<i>BIC</i>	263,149	317,224	311,901
<i>HQIC</i>	257,948	314,623	308,001
<i>W*</i>	0,041	0,038	0,042
<i>A*</i>	0,236	0,244	0,292
K-S (p-valor)	0,269 (0,0502)	0,675 ($< 10^{-10}$)	0,490 ($< 10^{-6}$)
Fibra nervosa	WLL	LL	LLE
α	0,412 (0,018)	0,143 (0,002)	0,288 (0,002)
β	0,707 (0,038)	1,611 (0,001)	1,368 (0,002)
λ	2,046 (0,012)	1 -	2,041 (0,011)
θ	1,925 (0,109)	1 -	1 -
	WLL	LL	LLE
<i>AIC</i>	-839,989	-787,232	-824,117
<i>AICC</i>	-839,300	-787,032	-823,710
<i>BIC</i>	-831,417	-782,945	-817,687
<i>HQIC</i>	-836,618	-785,546	-821,588
<i>W*</i>	0,152	0,619	0,250
<i>A*</i>	1,258	4,593	2,043
K-S (p-valor)	0,036 (0,247)	0,044 (0,093)	0,039 (0,159)

as estatísticas A^* , W^* e K-S, com uma única exceção nos dados de rolamentos para o modelo LL que apresentou $W^* = 0,038$, menor do que os demais.

Na Tabela 6 apresentamos os TRVs para os dois conjuntos de dados. Para os dados de rolamentos e fibra nervosa, rejeitamos todas as hipóteses nulas em favor da nova distribuição. Nas Figuras 5 e 6 podemos observar os histogramas com as estimativas dos modelos WLL, LL e LLE para os dados de rolamentos e fibra nervosa, respectivamente. Verificamos que em ambos os casos modelo WLL

se ajustou melhor aos dados em questão do que os seus submodelos.

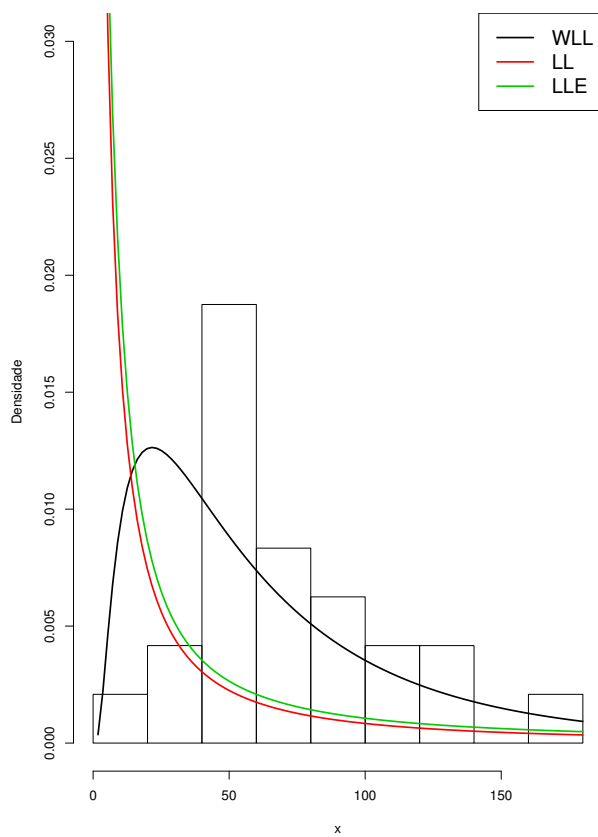


Figura 5 - Histograma com as estimativas dos modelos WLL, W e ELL para os dados de rolamentos.

Tabela 6 - TRVs

Rolamentos	Hipóteses	Λ
WLL vs LL	$H_0 : \lambda = \theta = 1$ vs $H_1 : H_0$ não é verdade	62,717 ($\geq \chi_2^2 = 5,99$)
WLL vs LLE	$H_0 : \theta = 1$ vs $H_1 : H_0$ não é verdade	52,896 ($\geq \chi_1^2 = 3,84$)
Fibra nervosa	Hipóteses	Λ
WLL vs LL	$H_0 : \lambda = \theta = 1$ vs $H_1 : H_0$ não é verdade	56,510 ($\geq \chi_2^2 = 5,99$)
WLL vs LLE	$H_0 : \theta = 1$ vs $H_1 : H_0$ não é verdade	17,237 ($\geq \chi_1^2 = 3,84$)

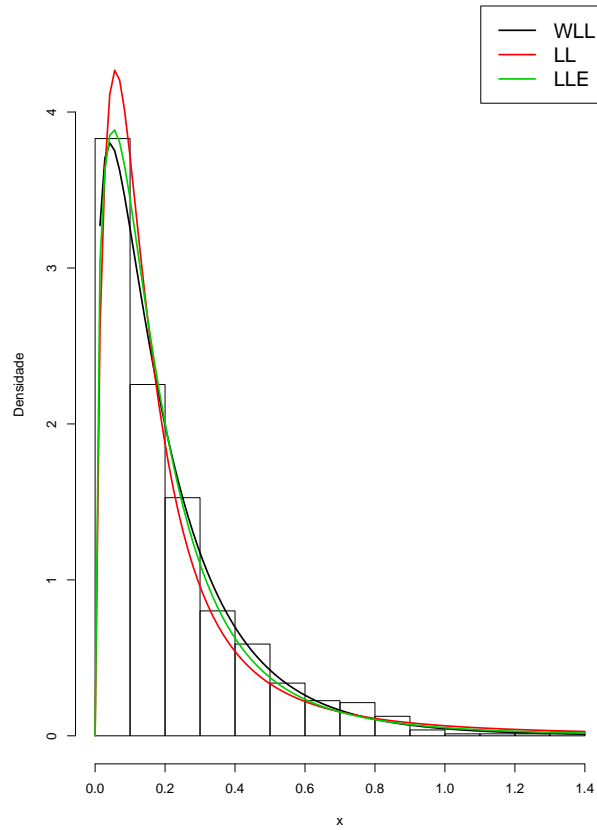


Figura 6 - Histograma com as estimativas dos modelos WLL, W e ELL para os dados de fibra nervosa.

6.2 Modelos não aninhados

Ajustamos $X \sim \text{WLL}(\alpha, \beta, \lambda, \theta)$ para um conjunto de dados reais e comparamos com os modelos Beta Exponencial Generalizada (BGE), Fréchet (F) e Beta Fréchet (BF) cujas densidades são expressas, respectivamente, por:

$$f_{\text{BEG}}(x) = \frac{1}{B(a, b)} e^{-\lambda x} [1 - e^{-\lambda x}]^{\alpha a - 1} \{[1 - e^{-\lambda x}]^{\alpha}\}^{b-1},$$

$$f_{\text{F}}(x) = \lambda \sigma^{\lambda} x^{-(\lambda+1)} e^{-\left(\frac{\sigma}{x}\right)^{\lambda}}$$

e

$$f_{\text{BF}}(x) = \frac{1}{B(a, b)} \lambda \sigma^{\lambda} x^{-(\lambda+1)} e^{-a\left(\frac{\sigma}{x}\right)^{\lambda}} \left\{1 - e^{-\left(\frac{\sigma}{x}\right)^{\lambda}}\right\}^{b-1}.$$

(i) *Dados: carbono*

Os dados são definidos a partir de Nichols e Padgett (2006) e composto por 100 observações de quebra de fibras de carbono por stress (em Gba). Esses dados foram ajustados por Barreto-Souza, Cordeiro e Simas (2011) e Barreto-Souza, Santos e Cordeiro (2010), respectivamente, às distribuições BF e BGE. Eles se encontram na Tabela 7:

Tabela 7 - Dados definidos por Nichols e Padgett (2006) composto por 100 observações de quebra de fibras de carbono por stress (em Gba)

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

(ii) *Dados: fibra de vidro*

Os dados obtidos a partir de Smith e Naylor (1987) representam a resistência de fibras de vidro de 1,5 cm, medidos no National Physical Laboratory, Inglaterra. Infelizmente, as unidades de medida não são dadas no artigo. Barreto-Souza, Cordeiro e Simas (2011) e Barreto-Souza, Santos e Cordeiro (2010) ajustaram esse mesmo conjunto de dados aos modelos BF e BGE, respectivamente. Na Tabela 8 apresentamos os dados que consistem de 63 observações.

Tabela 8 - Dados que representam a resistência de fibras de vidro de 1,5 cm, medidos no National Physical Laboratory, Inglaterra

0,55	0,93	1,25	1,36	1,49	1,52	1,58	1,61	1,64	1,68	1,73	1,81	2,00
1,27	1,39	1,49	1,53	1,59	1,61	1,66	1,68	1,76	1,82	2,01	0,77	1,11
1,50	1,54	1,60	1,62	1,66	1,69	1,76	1,84	2,24	0,81	1,13	1,29	1,48
1,61	1,62	1,66	1,70	1,77	1,84	0,84	1,24	1,30	1,48	1,51	1,55	1,61
1,70	1,78	1,89	0,74	1,28	1,50	1,63	1,04	1,42	1,55	1,67		

Na Tabela 9 encontramos os parâmetros estimados pelo método de Bootstrap para os modelos utilizados. Verificamos, para os dois conjuntos de dados, pelo teste de Wald (ao nível de 5%), que todas essas estimativas são significativas.

Tabela 9 - EMVs, critérios de informação e estatísticas de aderência

Carbono	WLL	BGE	F	BF
α	2,758 (0,188)	2,439 (0,233)	- -	- -
β	1,226 (0,087)	- -	- -	- -
λ	1,380 (0,086)	- -	1,768 (0,004)	1,307 (0,423)
θ	3,172 (0,159)	- -	- -	- -
a	- -	2,344 (0,654)	- -	5,978 (3,749)
b	- -	2,454 (0,169)	- -	0,992 (0,114)
σ	- -	0,582 (0,063)	1,886 (0,007)	4,479 (1,709)
<i>AIC</i>	291,026	295,853	350,292	314,820
<i>AICC</i>	291,716	296,253	350,492	315,220
<i>BIC</i>	299,598	304,425	354,578	323,393
<i>HQIC</i>	294,398	299,224	351,977	318,192
<i>W*</i>	0,085	0,169	0,756	0,382
<i>A*</i>	0,458	0,866	4,306	2,104
K-S (p-valor)	0,070 (0,711)	0,098 (0,291)	0,177 (0,004)	0,136 (0,049)
Fibra de vidro	WLL	BGE	F	BF
α	2,296 (0,266)	4,139 (0,449)	- -	- -
β	2,064 (0,119)	- -	- -	- -
λ	2,516 (0,374)	3,985 (1,171)	2,889 (0,060)	0,409 (0,020)
θ	3,304 (0,139)	- -	- -	- -
a	- -	4,287 (0,144)	- -	5,148 (2,913)
b	- -	1,305 (0,144)	- -	2,116 (0,146)
σ	- -	- -	1,259 (0,008)	2,585 (0,298)
<i>AIC</i>	40,335	58,733	97,714	71,131
<i>AICC</i>	41,025	59,133	97,914	71,531
<i>BIC</i>	48,908	67,305	102,000	79,703
<i>HQIC</i>	43,707	62,104	99,399	74,502
<i>W*</i>	0,282	0,613	1,225	0,804
<i>A*</i>	1,545	3,363	6,483	4,354
K-S (p-valor)	0,162 (0,066)	0,216 (0,005)	0,248 (0,001)	0,256 (0,0004)

Nas Figuras 7 e 8, apresentamos um histograma com os ajustes para os modelos WLL, BGE, F e BF. Por inspeção visual, notamos que o modelo WLL se ajustou melhor aos conjuntos de dados utilizados.

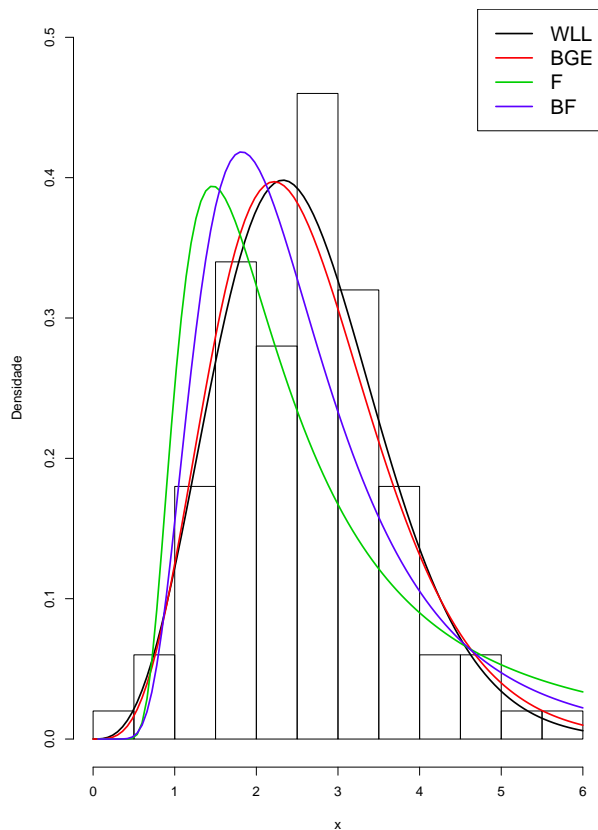


Figura 7 - Histograma com as estimativas dos modelos WLL, BGE, F e BF para os dados de carbono.

Os critérios e valores das estatísticas dos testes de aderência para os modelos utilizados são apresentados na Tabela 9. Para os dois conjuntos de dados, observamos que o modelo que obteve os menores valores foi o WLL. Verificamos, para os dados de carbono, bem como os dados de fibra de vidro, que este modelo apresenta os menores valores das estatísticas A^* , W^* e K-S. Estes resultados apontam que a distribuição WLL admite um melhor ajuste para estes conjuntos de dados, dentre os modelos estudados.

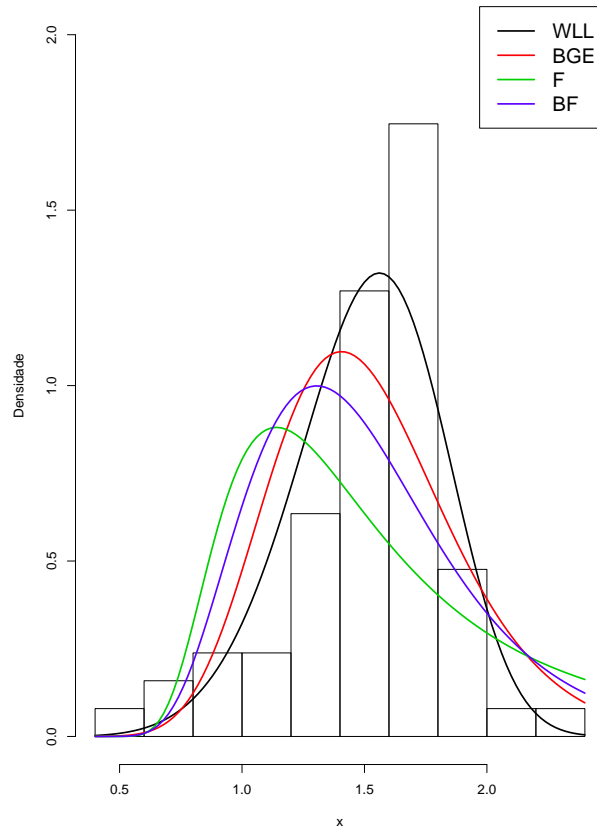


Figura 8 - Histograma com as estimativas dos modelos WLL, BGE, F e BF para os dados de fibra de vidro.

Conclusões

Definimos e estudamos uma nova classe de distribuição de probabilidades, denominada *nova Weibull-X* (NWX), a qual estende a classe Weibull-X. Para cada distribuição-base $G(x, \xi)$, definimos a correspondente NWX com três parâmetros adicionais, a fim de obter mais flexibilidade. Obtemos expressões explícitas de algumas características da nova classe tais como momentos ordinários e centrais, função geratriz de momentos e função característica. Indicamos e estudamos algumas propriedades estruturais de um caso básico deste gerador, admitindo a distribuição básica como sendo o modelo log-logístico. Em seguida, estudos de simulação foram conduzidos, com o intuito de avaliar estimativas obtidas a partir

dos métodos de máxima verossimilhança, Bootstrap e Bootstrap corrigido. Por fim, apresentamos quatro aplicações a dados reais, conduzida a partir do método Bootstrap, para demonstrar a importância deste novo gerador.

BARROS, P. S. N.; BRITO, C. C. R.; REGO, L. C.; OLIVEIRA, W. R.; GOMES-SILVA, F. A new class of generalized Weibull distributions. *Rev. Bras. Biom.*, Lavras, v.35, n.2, p.319-347, 2017.

■ **ABSTRACT:** *The Weibull distribution is a very popular distribution for modelling lifetime data. Many modifications of the Weibull distribution have been proposed in recent years. Inspired by the T-X method (ALZAATREH; LEE; FAMOYE, 2013), we propose a class of generalized Weibull distributions which generalizes the Weibull-X class. We investigate some of its mathematical properties including an expansion for the density function and explicit expressions for the ordinary and central moments, generating function and characteristic function. We discuss a particular case of the new generator and present briefly some of its structural properties. As an alternative to maximum likelihood, we performed Bootstrap simulations and corrected Bootstrap in order to compare these methods of estimation. The Bootstrap method is employed to estimate the model parameters. We fit one special model of the new class to four real data sets to demonstrate the flexibility of the proposed class.*

■ **KEYWORDS:** *Bootstrap; generalized class; T-X method; Weibull-X; maximum likelihood.*

Referências

ALEXANDER, C. et al. Generalized beta-generated distributions. *Computational Statistics and Data Analysis*, v. 56, p. 1880–1897, 2012.

ALZAATREH, A.; GHOSH, I. On the weibull-x family of distributions. *Journal of Statistical Theory and Applications*, v. 14, p. 169–183, 2015.

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

ALZAGHAL, A.; FAMOYE, F.; LEE, C. Exponentiated t-x family of distributions with some applications. *International Journal of Probability and Statistics*, v. 2, p. 31–49, 2013.

BARRETO-SOUZA, W.; CORDEIRO, G. M.; SIMAS, A. B. Some results for beta fréchet distribution. *Communications in Statistics - Theory and Methods*, Taylor & Francis, v. 40, n. 5, p. 798–811, 2011.

BARRETO-SOUZA, W.; SANTOS, A. H.; CORDEIRO, G. M. The beta generalized exponential distribution. *Journal of Statistical Computation and Simulation*, Taylor & Francis, v. 80, n. 2, p. 159–172, 2010.

- BEBBINGTON, M.; LAI, C. D.; ZITIKIS, R. A flexible weibull extension. *Reliability Engineering and System Safety*, v. 92, p. 719–726, 2007.
- BOURGUIGNON, M.; SILVA, R. B.; CORDEIRO, G. M. The weibull-g family of probability distributions. *Journal of Data Science*, v. 12, n. 1, p. 53–68, 2014.
- COLOSIMO, E. A.; GIOLO, S. R. Análise de sobrevivência aplicada. In: *ABE-Projeto Fisher*. Sao Paulo: Edgard Blücher, 2006.
- CORDEIRO, G. M.; CASTRO, M. A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, v. 81, p. 883–893, 2011.
- CORDEIRO, G. M. et al. The gamma extended weibull distribution. *Journal of Statistical Distributions and Applications*, v. 3:7, 2016.
- CORDEIRO, G. M.; SILVA, G. O.; ORTEGA, E. M. M. The beta extended weibull distribution. *Journal of Probability and Statistical Science*, v. 10, p. 15–40, 2012.
- COX, D.; LEWIS, P. The statistical analysis of series of events. John Wiley and Sons, 1966.
- EFRON, B. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, v. 7, p. 1–26, 1979.
- EUGENE, N.; LEE, C.; FAMOYE, F. Beta-normal distribution and its applications. *Communication in Statistics - Theory and Methods*, v. 31, p. 497–512, 2002.
- FAMOYE, F.; LEE, C.; OLUMOLADE, O. The beta-weibull distribution. *Journal of Statistical Theory and Applications*, v. 4, p. 121–136, 2005.
- GUPTA, R. C.; GUPTA, P. L.; GUPTA, R. D. Modeling failure time data by lehmann alternatives. *Communications in Statistics - Theory and Methods*, New York, v. 27, p. 887–904, 1998.
- GUPTA, R. D.; KUNDU, D. Generalized exponential distributions. *Australian e New Zealand Journal of Statistics*, Oxford, v. 41, n. 2, p. 173–188, 1999.
- JONES, M. C. Families of distributions arising from the distributions of order statistics. *Test*, v. 13, p. 1–43, 2004.
- JØRGENSEN, B. *Statistical properties of the generalized inverse Gaussian distribution*. [S.l.]: Springer Science & Business Media, 2012. v. 9.
- LAI, C. D.; XIE, M.; MURTHY, D. N. P. A modified weibull distribution. *IEEE Transactions Reliability*, v. 52, p. 33–37, 2003.
- LAWLESS, J. F. *Statistical models and methods for lifetime data*. [S.l.]: John Wiley & Sons, 2011. v. 362.

- MARSHALL, A. N.; OLKIN, I. A new method for adding a parameter to a family of distributions with applications to the exponential and weibull families. *Biometrika*, v. 84, p. 641–652, 1997.
- MUDHOLKAR, G.; SRIVASTAVA, D.; FREIMER, M. Exponentiated weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability*, v. 42, p. 229–302, 1993.
- MUDHOLKAR, G. S.; SRIVASTAVA, D. K.; FRIEMER, M. The exponential weibull family: A re-analysis of the bus-motor failure data. *Technometrics*, v. 37, p. 436–445, 1995.
- MUDHOLKAR, G. S.; SRIVASTAVA, D. K.; FRIEMER, M. A generalization of the weibull distribution with application to the analysis of survival data. *Journal of the American Statistical Association*, v. 91, p. 1575–1583, 1996.
- NADARAJAH, S. Exponentiated pareto distributions. *Statistics*, London, v. 39, p. 255–260, 2005.
- NADARAJAH, S. The exponentiated exponential distribution: a survey. *AStA Advances in Statistical Analysis*, New York, v. 95, p. 219–251, 2011.
- NADARAJAH, S.; KOTZ, B. S. The beta exponential distribution. *Reliability Engineering System Safety*, v. 91, p. 689–697, 2006.
- NICHOLS, M. D.; PADGETT, W. A bootstrap control chart for weibull percentiles. *Quality and reliability engineering international*, Wiley Online Library, v. 22, n. 2, p. 141–151, 2006.
- PERCONTINI, A. et al. A distribuição gama weibull poisson aplicada a dados de sobrevivência. *Tendências em Matemática Aplicada e Computacional*, v. 15, p. 165–176, 2014.
- PHAM, H.; LAI, C. D. On recent generalizations of the weibull distribution. *IEEE Transactions on Reliability*, v. 56, n. 3, p. 454–458, 2007.
- ROSAIAH, K.; KANTAN, R. R. L.; KUMAR, C. S. Reliability test plans for exponentiated log-logistic distribution. *Economic Quality Control*, v. 21, p. 279–289, 2006.
- SILVA, G. O.; ORTEGA, E. M. M.; CORDEIRO, G. M. The beta modified weibull distribution. *Lifetime Data Analysis*, v. 16, p. 409–430, 2010.
- SMITH, R. L.; NAYLOR, J. A comparison of maximum likelihood and bayesian estimators for the three-parameter weibull distribution. *Applied Statistics*, JSTOR, p. 358–369, 1987.
- WASSERMAN, L. *All of statistics: a concise course in statistical inference*. New York: Springer, 2004.

XIE, M.; LAI, C. D. Reliability analysis using an additive weibull model with bathtubshaped failure rate function. *Reliability Engineering and System Safety*, v. 52, p. 87–93, 1995.

XIE, M.; TANG, Y.; GOH, T. N. A modified weibull extension with bathtub failure rate function. *Reliability Engineering and System Safety*, v. 76, p. 279–285, 2002.

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

Recebido em 07.04.2016.

Aprovado após revisão em 23.09.2016.

Apêndice - Componentes da função escore

As componentes da função escore são:

$$U_{\delta} = \sum_{i=1}^n \left\{ e^{-(\lambda\delta)^{\theta} \bar{G}^{\theta}(x_i, \boldsymbol{\xi})} \theta \delta^{\theta-1} \bar{G}^{\theta}(x_i, \boldsymbol{\xi}) [1 - \delta^{\theta} \lambda^{\theta} \bar{G}^{\theta}(x_i, \boldsymbol{\xi})] \right. \\ \left. + e^{-\lambda^{\theta} H^{\theta}(x_i, \boldsymbol{\xi})} H^{\theta-2}(x_i, \boldsymbol{\xi}) [(\theta - 1) - \theta \lambda^{\theta} H^{\theta}(x_i, \boldsymbol{\xi})] \right\} \\ \times \left\{ e^{-(\lambda\delta)^{\theta} \bar{G}^{\theta}(x_i, \boldsymbol{\xi})} \delta^{\theta} \bar{G}^{\theta}(x_i, \boldsymbol{\xi}) + e^{-\lambda^{\theta} H^{\theta}(x_i, \boldsymbol{\xi})} H^{\theta-1}(x_i, \boldsymbol{\xi}) \right\}^{-1},$$

$$U_{\theta} = \frac{n}{\theta} + n \log \lambda + \sum_{i=1}^n \left\{ e^{-(\lambda\delta)^{\theta} \bar{G}^{\theta}(x_i, \boldsymbol{\xi})} \delta^{\theta} \bar{G}^{\theta}(x_i, \boldsymbol{\xi}) \{ \log[\delta \bar{G}^{\theta}(x_i, \boldsymbol{\xi})] \right. \\ \times [1 - (\lambda\delta)^{\theta} \bar{G}^{\theta}(x_i, \boldsymbol{\xi})] - (\lambda\delta)^{\theta} \bar{G}^{\theta}(x_i, \boldsymbol{\xi}) \log \lambda \} + e^{-\lambda^{\theta} H^{\theta}(x_i, \boldsymbol{\xi})} H^{\theta-1} \\ \times \bar{G}^{\theta}(x_i, \boldsymbol{\xi}) \{ \log[H(x_i, \boldsymbol{\xi})][1 - \lambda^{\theta} H^{\theta}(x_i, \boldsymbol{\xi})] - \lambda^{\theta} \log \lambda H^{\theta}(x_i, \boldsymbol{\xi}) \} \} \\ \times \left\{ e^{-(\lambda\delta)^{\theta} \bar{G}^{\theta}(x_i, \boldsymbol{\xi})} \delta^{\theta} \bar{G}^{\theta}(x_i, \boldsymbol{\xi}) + e^{-\lambda^{\theta} H^{\theta}(x_i, \boldsymbol{\xi})} H^{\theta-1}(x_i, \boldsymbol{\xi}) \right\}^{-1},$$

$$U_{\lambda} = \lambda^{-1} + \sum_{i=1}^n \frac{\theta \{ e^{-(\lambda\delta)^{\theta} \bar{G}^{\theta}(x_i, \boldsymbol{\xi})} \delta^{2\theta} \bar{G}^{2\theta}(x_i, \boldsymbol{\xi}) - e^{-\lambda^{\theta} H^{\theta}(x_i, \boldsymbol{\xi})} H^{2\theta-1}(x_i, \boldsymbol{\xi}) \}}{\lambda^{1-\theta} \{ e^{-(\lambda\delta)^{\theta} \bar{G}^{\theta}(x_i, \boldsymbol{\xi})} \delta^{\theta} \bar{G}^{\theta}(x_i, \boldsymbol{\xi}) + e^{-\lambda^{\theta} H^{\theta}(x_i, \boldsymbol{\xi})} H^{\theta-1}(x_i, \boldsymbol{\xi}) \}}$$

e

$$U_{\boldsymbol{\xi}_k} = \sum_{i=1}^n \left\{ e^{-(\lambda\delta)^{\theta} \bar{G}^{\theta}(x_i, \boldsymbol{\xi})} \delta^{\theta} \bar{G}^{\theta}(x_i, \boldsymbol{\xi}) \{ g'_k(x_i, \boldsymbol{\xi}) - g(x_i, \boldsymbol{\xi}) \bar{G}^{-1}(x_i, \boldsymbol{\xi}) G'_k(x_i, \boldsymbol{\xi}) \right. \\ \times [(\theta - 1) + \delta^{\theta} \theta \lambda^2 \bar{G}^{\theta}(x_i, \boldsymbol{\xi})] \} + e^{-\lambda^{\theta} H^{\theta}(x_i, \boldsymbol{\xi})} H^{\theta-1}(x_i, \boldsymbol{\xi}) \{ g'_k(x_i, \boldsymbol{\xi}) \\ + g(x_i, \boldsymbol{\xi}) \bar{G}^{-1}(x_i, \boldsymbol{\xi}) G'_k(x_i, \boldsymbol{\xi}) [1 + (\theta - 1) H^{-1}(x_i, \boldsymbol{\xi}) \\ - \theta \lambda^{\theta} H^{\theta-1}(x_i, \boldsymbol{\xi})] \} \} \times \left\{ e^{-(\lambda\delta)^{\theta} \bar{G}^{\theta}(x_i, \boldsymbol{\xi})} \delta^{\theta} g(x_i, \boldsymbol{\xi}) \bar{G}^{\theta}(x_i, \boldsymbol{\xi}) \right. \\ \left. + e^{-\lambda^{\theta} H^{\theta}(x_i, \boldsymbol{\xi})} g(x_i, \boldsymbol{\xi}) H^{\theta-1}(x_i, \boldsymbol{\xi}) \right\}^{-1}$$

em que $\bar{G}(x_i, \boldsymbol{\xi}) = 1 - G(x_i, \boldsymbol{\xi})$, $H(x_i, \boldsymbol{\xi}) = \delta - \log \bar{G}(x_i, \boldsymbol{\xi})$ e $t'_k(\cdot, \boldsymbol{\xi}) = \partial t(\cdot, \boldsymbol{\xi}) / \partial \boldsymbol{\xi}_k$.