Distribuição Kumaraswamy Burr XII: uma abordagem bayesiana

Patrícia Ferreira Paranaíba [1], Edwin Moises Marcos Ortega [2], Marcelino Alves Rosa de Pascoa [3]

Introdução

Diversos autores nos últimos anos têm concentrado seus esforços na generalização de família de distribuições de probabilidade, obtendo dessa forma, maior flexibilidade e consequentemente, um maior ganho na modelagem de dados. Cordeiro e Castro (2011) apresentaram uma classe de generalização baseada na distribuição Kumaraswamy. A classe modela as seguintes formas de risco: crescente, decrescente, unimodal e forma de U.

Nesse contexto, os objetivos deste trabalho foram: propor uma nova distribuição, denominada distribuição Kumaraswamy Burr XII, que é uma junção da distribuição Kumaraswamy com a distribuição Burr XII e utilizar a metodologia bayesiana para estimar os parâmetros do modelo proposto.

1 Material e métodos

A distribuição Burr XII (BXII), considerada em Zimmer; Keats e Wang (1998), é uma distribuição comumente encontrada para modelar dados de sobrevivência. Sua função de distribuição acumulada (fda) é definida como:

$$G(x; s, k, c) = 1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}, \; x \geq 0, \tag{1}$$

em que $k > 0$ e $c > 0$ são os parâmetros de forma e $s > 0$ é o parâmetro de escala.

A classe de distribuições Kumaraswamy, proposta por Cordeiro e Castro (2011), baseia-se na generalização da distribuição Kumaraswamy proposta por Kumaraswamy (1980) para variáveis limitadas inferiormente e superiormente.

Seja "Kw" uma abreviação para o nome da distribuição, a função de distribuição acumulada da distribuição Kumaraswamy é dada por:

$$G_{\mathrm{Kw}}(x; a, b) = 1 - (1 - x^a)^b, \quad 0 < x < 1 \quad \text{e} \quad a, b > 0. \tag{2}$$

Substituindo a variável aleatória x por $G(x)$, e $G_{\mathrm{Kw}}(x; a, b)$ por $F(x)$ na equação (2), obtém-se a classe de distribuições Kumaraswamy, cuja função de distribuição acumulada é dada por:

$$F(x) = 1 - \{1 - G^a(x)\}^b, \quad 0 < G(x) < 1 \quad \text{e} \quad a, b > 0, \tag{3}$$

em que $a > 0$ e $b > 0$ são os novos parâmetros de forma.

[1]UFU. e-mail: *patriciaparanaiba@ufu.br*
[2]ESALQ-USP. e-mail: *edwin@usp.br*
[3]UFMT. e-mail: *marcelino.pascoa@gmail.com*

A função densidade de probabilidade, função de sobrevivência e função risco são dadas por:

$$f(x) = a\, b\, g(x)\, G^{a-1}(x)\, \{1 - G^a(x)\}^{b-1},$$

$$S(x) = [1 - G^a(x)]^b \qquad \text{e} \qquad h(x) = \frac{abg(x)G^{a-1}(x)}{1 - G^a(x)},$$

respectivamente.

A nova distribuição, denominada distribuição Kumaraswamy Burr XII (KwBXII) é uma composição baseada na classe de distribuições Kumaraswamy e na distribuição Burr XII. Substituindo $G(x)$ na equação (3) pela fda da Burr XII, dada na equação (1), obtém-se, então, a distribuição KwBXII com cinco parâmetros a, b, s, k e c, cujas funções de densidade de probabilidade e distribuição acumulada são definidas pelas equações (4) e (1), respectivamente.

$$\begin{aligned} f(x) &= a\, b\, c\, k\, s^{-c} x^{c-1} \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k-1} \left\{1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right\}^{a-1} \\ &\times \left[1 - \left\{1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right\}^a\right]^{b-1}, \;\; x > 0, \end{aligned}$$

e

$$F(x) = 1 - \left[1 - \left\{1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right\}^a\right]^b.$$

A função risco da distribuição KwBXII é dada por:

$$\begin{aligned} h(x) &= a\, b\, c\, k\, s^{-c} x^{c-1} \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k-1} \left\{1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right\}^{a-1} \\ &\times \left[1 - \left\{1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right\}^a\right]^{-1}. \end{aligned}$$

Para avaliar o desempenho do modelo KwBXII, foi utilizado um conjunto de dados ($n = 101$) que representam a resistência à ruptura de vertentes kevlar 49/epoxy que foram submetidos a pressão constante no nível de 90% de estresse até que todos falhassem, para a obtenção de dados completos com tempos exatos da falha. O conjunto de dados foi utilizado por Andrews e Herzberg (1985) e mais recentemente por Cooray e Ananda (2008).

Para uma análise bayesiana, assume-se a seguinte densidade conjunta *a priori*, dada por:

$$\pi(a, b, s, k, c) \propto \pi(a) \times \pi(b) \times \pi(s) \times \pi(k) \times \pi(c),$$

em que $a \sim \Gamma(a_1, b_1)$, $b \sim \Gamma(a_2, b_2)$, $s \sim \Gamma(a_3, b_3)$, $k \sim \Gamma(a_4, b_4)$ e $c \sim \Gamma(a_5, b_5)$, sendo $\Gamma(a_i, b_i)$ a distribuição gama com média a_i/b_i, variância a_i/b_i^2 e função densidade de probabilidade dada por:

$$f(\upsilon; a_i, b_i) = \frac{b_i^{a_i} \upsilon^{a_i - 1} \exp(-\upsilon b_i)}{\Gamma(a_i)},$$

em que $\upsilon > 0$, $a_i > 0$ e $b_i > 0$. Todos os hiperparâmetros são especificados. Assumindo independência entre os parâmetros a, b, s, k e c, a distribuição conjunta *a posteriori* para a, b, s, k e c é dada por:

$$\begin{aligned}\pi(a,b,s,k,c|y) &\propto \left(a\,b\,c\,k\,s^{-c}\right)^r \prod_{i\in F} x_i^{c-1} u_i^{a-1}(1+u_i)^{\frac{k+1}{k}}\,(1-u_i^a)^{b-1} \prod_{i\in C}(1-u_i^a)^b \\ &\times\ \pi(a,b,s,k,c). \qquad (4)\end{aligned}$$

O algoritmo Metropolis-Hasting permite simular observações de distribuições a partir das densidades condicionais *a posteriori.* Assim, para o uso de tal algoritmo consideram-se as seguintes condicionais *a posteriori* completas para os parâmetros a, b, s, k e c:

$$\pi(a|y,b,s,k,c) \propto a^r \prod_{i\in F} u_i^a\,(1-u_i^a)^{b-1} \prod_{i\in C}(1-u_i^a)^b \times \pi(a),$$

$$\pi(b|y,a,s,k,c) \propto b^r \prod_{i\in F}(1-u_i^a)^b \prod_{i\in C}(1-u_i^a)^b \times \pi(b),$$

$$\pi(s|y,a,b,k,c) \propto s^{-(c\,r)} \prod_{i\in F} u_i^{a-1}(1+u_i)^{\frac{k+1}{k}}\,(1-u_i^a)^{b-1} \prod_{i\in C}(1-u_i^a)^b \times \pi(s),$$

$$\pi(k|y,a,b,s,c) \propto k^r \prod_{i\in F} u_i^{a-1}(1+u_i)\,(1-u_i^a)^{b-1} \prod_{i\in C}(1-u_i^a)^b \times \pi(k),$$

e

$$\pi(c|y,a,b,s,k) \propto s^{-(c\,r)} \prod_{i\in F} x_i^c\,u_i^{a-1}(1+u_i)^{\frac{k+1}{k}}\,(1-u_i^a)^{b-1} \prod_{i\in C}(1-u_i^a)^b \times \pi(c).$$

em que $u_i = 1 - \left[1+\left(\frac{x_i}{s}\right)^c\right]^{-k}$, r é o número total de falhas e F e C representam o número de observações que falharam ou foram censuradas, respectivamente.

Para as análises descritivas para os parâmetros do modelo KwBXII foi utilizado o *software* estatístico R, pacote BOA (*Bayesian Output Analysis*).

Foram adotadas as seguintes distribuições *a priori* pouco informativas, para os três conjuntos de dados: $a \sim \Gamma(0,01;0,01)$, $b \sim \Gamma(0,01;0,01)$, $s \sim \Gamma(0,01;0,01)$, $k \sim \Gamma(0,01;0,01)$ e $c \sim \Gamma(0,01;0,01)$,

Dessa forma, foram geradas duas cadeias paralelas e independentes com tamanho 50.000 para cada parâmetro usando o algoritmo de Metropolis-Hasting, desconsiderando as primeiras 5.000 iterações para eliminar o efeito dos valores iniciais. Além disso, considerou-se espaçamento de tamanho 10, para evitar problemas de correlação, obtendo-se uma amostra de tamanho 4.500. Para monitorar a convergência das amostras foi usado o fator de redução da escala $\hat{R}$, desenvolvido por Gelman e Rubin (1992). A Tabela 1 apresenta os sumários *a posteriori* para os parâmetros do modelo KwBXII, além de mostrar os valores do fator de redução da escala para os dados de estresse .

Nota-se, que os valores de $\hat{R}$ estão próximos de 1, indicando a convergência das cadeias para os dados de estresse. Além disso, a Figura 1 apresenta as densidades marginais *a posteriori* aproximadas pelo histograma, considerando as 4.500 observações amostrais geradas para os dados de estresse. As densidades marginais apresentadas sugerem que a distribuição *a posteriori* apresenta uma tendência à simetria para os parâmetros.

Tabela 1: Sumário *a posteriori* dos parâmetros do modelo KwBXII para os dados de estresse.

Parâmetro	Média	Desvio Padrão	HPD (95%)	$\widehat{R}$
a	0,0990	0,0255	(0,0510; 0,1495)	1,0014
b	0,8850	0,0171	(0,8523; 0,9188)	1,0020
s	1,6845	0,0301	(1,6238; 1,7398)	1,0019
k	0,3945	0,0403	(0,3180; 0,4735)	1,0017
c	6,8195	0,0504	(6,7196; 6,9141)	1,0025

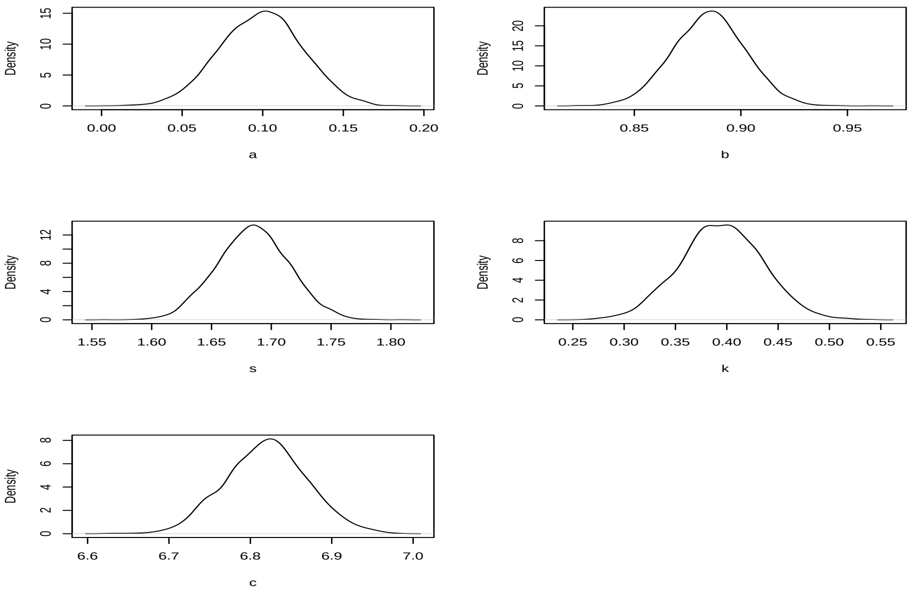

Figura 1: Estimativa por histograma das densidades marginais *a posteriori* para os parâmetros do modelo KwBXII para os dados de estresse.

2 Conclusão

Foi utilizada de forma satisfatória a metodologia bayesiana para as estimativas dos parâmetros da ditribuição proposta Kumaraswamy Burr XII.

Referencias Bibliográficas

ANDREWS, D.F; HERZBERG, A.M. **Data: A Collection of Problems from Many Fields for the Student and Research Worker**, New York: Springer Series in Statistics, 1985. 442p.

COORAY, K.; ANANDA, M.M.A. A generalization of the half-normal distribution wit applications to lifetime data. **Communication in Statistical - Theory and Methods**, New York, v.37, p.1323-1337, 2008.

CORDEIRO, G.M.; CASTRO, M. A new family of generalized distributions. *Journal of*

Statistical Computation and Simulation, London, v.81, p.883-898, 2011.

GELMAN, A.; RUBIN, D.B. Inference from iterative simulation using multiple sequences. **Statistical Science**, Hayward, v.7, p.457-472, 1992.

KUMARASWAMY, P. A generalized probability density function for double-bounded random processes. **Journal of Hydrology**, New York, v.46, p.79-88, 1980.

R CORE TEAM. *R*: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012. ISBN 3-900051-07-0, URL http://www.R-project.org/.

ZIMMER, W. J.; KEATS, J. B.; WANG, F. K. The Burr XII distribution in reliability analysis. *Journal of Quality Technology*, Milwaukee, v.30, p.389-394, 1998.