



Distribuição Kumaraswamy Burr XII: teoria e prática

Patrícia Ferreira Paranaíba^{*1} e Edwin Moises Marcos Ortega^{†2}

¹FAMAT - Universidade Federal de Uberlândia

²LCE - Universidade de São Paulo

Palavras-chave: Distribuição Burr XII. Kumaraswamy Burr XII. Máxima verossimilhança.

Resumo

A distribuição Burr XII (BXII) possui, como casos particulares, as distribuições normal, log-normal, gama, logística, valor extremo tipo I, entre outras. Diante disso, ela é considerada uma distribuição flexível no ajuste dos dados. A ideia de Cordeiro e Castro (2011) foi utilizada para o desenvolvimento de uma nova distribuição de probabilidade a partir da distribuição BXII, denominada distribuição Kumaraswamy Burr XII (KwBXII). As vantagens desse novo modelo reside na capacidade de acomodar várias formas da função risco, além disso, ele também se mostrou útil na discriminação de modelos. Para a estimação dos parâmetros, foi utilizado o método de máxima verossimilhança e para ilustrar a aplicação da nova distribuição foi utilizado um conjunto de dados reais.

Introdução

Diversos autores nos últimos anos têm concentrado seus esforços na generalização de família de distribuições de probabilidade, obtendo dessa forma, maior flexibilidade e, conseqüentemente, um maior ganho na modelagem de dados. Cordeiro e Castro (2011) apresentaram a classe de distribuições Kumaraswamy. A classe modela as diferentes formas das funções 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; estimar os parâmetros da distribuição Kumaraswamy Burr XII, por meio do método de máxima verossimilhança, utilizando dados de resistência de fibras de vidro; comparar o ajuste da distribuição Kumaraswamy Burr XII com os ajustes das distribuições Burr XII e log-logística.

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. A função de distribuição acumulada (fda) é definida como:

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

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:

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

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

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

*patriciaparanaiba@ufu.br

†edwin@usp.br.

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

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

$$f(x) = abg(x)G^{a-1}(x)\{1 - G^a(x)\}^{b-1}, \quad (4)$$

$$S(x) = [1 - G^a(x)]^b \quad \text{e} \quad h(x) = \frac{abg(x)G^{a-1}(x)}{1 - G^a(x)}, \quad (5)$$

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 (6) e (7), respectivamente.

$$\begin{aligned} f(x) &= abck s^{-c} x^{c-1} \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k-1} \left\{1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right\}^{a-1} \\ &\times \left[1 - \left\{1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right\}^a\right]^{b-1}, \quad x > 0, \end{aligned} \quad (6)$$

e

$$F(x) = 1 - \left[1 - \left\{1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right\}^a\right]^b. \quad (7)$$

A função risco da distribuição KwBXII é dada por:

$$\begin{aligned} h(x) &= abck s^{-c} x^{c-1} \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k-1} \left\{1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right\}^{a-1} \\ &\times \left[1 - \left\{1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right\}^a\right]^{-1}. \end{aligned} \quad (8)$$

Para avaliar o desempenho do modelo KwBXII, foi utilizado um conjunto de dados que está relacionado com as ciências dos materiais: o conjunto de dados reais refere-se a resistência de fibras de vidro de 1,5 cm, medidos no Laboratório *National Physical*, Inglaterra. O conjunto de dados ($n = 51$) foi utilizado anteriormente por Smith e Naylor (1987) e Cordeiro e Lemonte (2011).

Seja T uma variável aleatória com distribuição KwBXII com vetor de parâmetros $\theta = (a, b, s, k, c)^T$. Supondo que os dados consistam de n observações independentes e $t_i = \min(T_i, C_i)$ para $i = 1, \dots, n$, o logaritmo da função de verossimilhança, considerando a distribuição KumbXII, é dado por:

$$\begin{aligned} l(\theta) &= r \left[\log(a) + \log(b) + \log(c) + \log(k) - c \log(s) \right] + (c-1) \sum_{i \in F} \log(x_i) \\ &+ (a-1) \sum_{i \in F} \log(u_i) + \left(\frac{k+1}{k}\right) \sum_{i \in F} \log(1-u_i) + (b-1) \sum_{i \in F} \log(1-u_i^a) \\ &+ b \sum_{i \in C} \log(1-u_i^a), \end{aligned}$$

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

Consequentemente, o estimador de máxima verossimilhança (EMV) $\hat{\theta}$ de θ é obtido numericamente a partir das equações não lineares,

$$U_a(\theta) = U_b(\theta) = U_s(\theta) = U_k(\theta) = U_c(\theta) = 0.$$

O ajuste da distribuição KwBXII, foi comparado com o das distribuições Burr XII e log-logística (LL), por meio das estatísticas AIC (Critério de Informação de Akaike), BIC (Critério de Informação Bayesiano) e CAIC (Critério de Informação Akaike Consistente) e pelo teste da razão de verossimilhança (TRV). As análises foram implementadas no *software R* (R Development Core Team, 2011).

Resultados e discussões

Na Tabela 1, podem ser vistas as EMVs (e os correspondentes erros-padrão que estão entre parênteses) dos parâmetros e os valores das estatísticas dos modelos KwBXII, BXII e LL. Os resultados indicam que o modelo KwBXII tem os menores valores de AIC, CAIC e BIC quando comparado com os modelos BXII e LL, portanto, sendo o mais adequado para o ajuste dos dados de fibras de vidro. Para ilustrar a utilização da distribuição KwBXII para discriminação de modelos, comparou-se a distribuição KwBXII com alguns de seus sub-modelos por meio do TRV. As hipóteses testadas, bem como a estatística do teste são apresentados na Tabela 2. Os resultados sugerem que o modelo KwBXII produz um ajuste mais adequado quando comparado com as distribuições BXII e LL.

Além disso, as funções de densidade de probabilidade e de distribuição acumulada estimadas para os modelos KwBXII, BXII e LL, são representadas graficamente na Figura 1(a) e 1(b), respectivamente. Dessa figura, observa-se que para os dados de fibras de vidro a distribuição KwBXII apresenta um ajuste mais adequado se comparado aos outros modelos probabilísticos.

Tabela 1: EMVs dos parâmetros dos modelos KwBXII, BXII e LL, correspondentes erros-padrão (entre parênteses) e valores das estatísticas AIC, CAIC e BIC, para os dados de fibras de vidro.

Modelo	a	b	s	k	c	AIC	CAIC	BIC
KwBXII	0,1322 (0,0333)	0,8944 (0,4094)	1,6440 (0,0474)	0,3358 (0,2073)	32,3362 (1,4639)	25,2	26,5	34,9
BXII	1 (-)	1 (-)	2,3424 (0,9581)	10,4426 (1,6855)	5,6369 (0,9406)	33,2	33,7	39,0
LL	0,6865 (0,0215)	7,5388 (0,9255)				40,8	41,1	44,7

Tabela 2: TRV para os dados de fibras de vidro.

Fibras	Hipóteses	Estatística w	valor- p
KwBXII vs BXII	$H_0 : a = b = 1$ vs $H_1 : H_0$ é falsa	12,0	0,00248
KwBXII vs LL	$H_0 : a = b = k = 1$ vs $H_1 : H_0$ é falsa	21,6	<0,0001

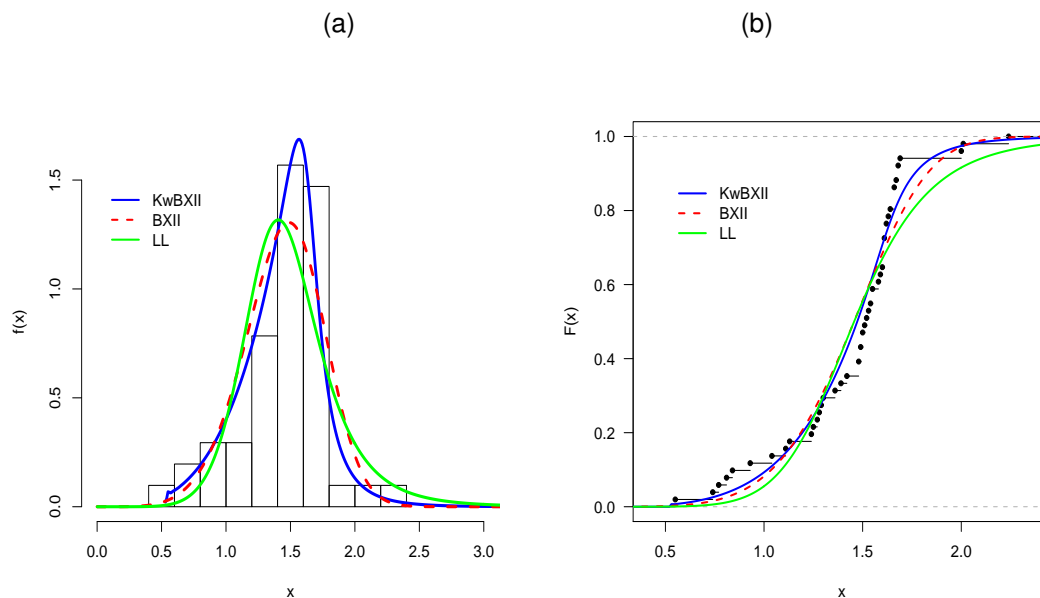


Figura 1: (a) Densidades estimadas para os modelos KwBXII, BXII e LL para os dados de fibras de vidro. (b) Funções de distribuição acumulada estimadas e função de distribuição acumulada empírica para os dados de fibras de vidro.

Conclusão

A distribuição KwBXII apresentou um ajuste mais adequado para modelar os dados de resistência de fibras de vidro, segundo o teste da razão de verossimilhança e as estatísticas AIC, CAIC e BIC quando comparada com as distribuições BXII e LL.

Referências

CORDEIRO, G.M.; CASTRO, M. A new family of generalized distributions. **Journal of Statistical Computation and Simulation**, London, v.81, p.883-898, 2011.

CORDEIRO, G.M.; LEMONTE, A.J. The β -Birnbaum Saunders distribution: An improved distribution for fatigue life modeling. **Computational Statistics and Data Analysis**, London, v.55, p.55, 2011.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Disponível em: <<http://www.R-project.org>>. Acesso em: 20 jan. 2011.

SMITH, R.L; NAYLOR, J.C. A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. **Applied Statistics**, New Jersey, v.36, p.358-369, 1987.

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.