

Alguns Resultados da Distribuição Marshall-Olkin Generalizada Exponenciada de Weibull

Kleber Napoleão N. O. Barros,

PPGBEA - UFRPE / DE - UEPB,

E-mail: kleber.barros@cct.uepb.br

Patrícia Silva N. Barros,

PPGBEA - UFRPE / DCE - UFPB,

E-mail: patricia.jp1408@hotmail.com

Kleber Régis Santoro

PPGBEA - UFRPE,

E-mail: krsantoro@yahoo.com.br

Resumo: *O presente trabalho generaliza a família de distribuições de Marshall-Olkin pela adição de parâmetros, tornando-a uma nova classe mais flexível, criando a nova distribuição de Marshall-Olkin Generalizada Exponenciada Weibull (MOGEW). É estudado o comportamento da função densidade de probabilidade MOGE e sua respectiva função de risco com resultados promissores. Se obtêm expansões da função de sobrevivência e da densidade de probabilidade e momentos de ordem s . Também são simulados um conjunto de dados de treino e um segundo de validação aos quais são ajustadas ajuste estimações de densidades da distribuição MOGEW.*

Palavras-chave: *Distribuição Marshall-Olkin Weibull Exponenciada, Distribuições Generalizadas, Análise de Sobrevivência.*

1 Introdução

A distribuição Weibull Exponenciada foi proposta em [8]. Algumas relações são obtidas em [5] para um caso particular, denominando-o de distribuição exponencial generalizada (EG) como alternativa às distribuições gama e de Weibull. Desde então, diversos trabalhos em análise de sobrevivência têm sido produzidos. Uma revisão detalhada da distribuição EG e outros resultados são apresentados em [9]. Seguindo a mesma linha, diversos outros trabalhos obtêm importantes generalizações de distribuições de tempos de vida pela adição de parâmetros.

O princípio de tais distribuições é simples. Dada uma distribuição com função de distribuição acumulada ¹ (fda ou fd) contínua $G(t)$, sua generalização ou exponenciação que pode ser escrita como

$$F(x) = G(x)^\alpha, \quad \alpha > 0. \quad (1)$$

Quando

$$G(x) = 1 - e^{-(\lambda x)^k}, \quad \text{para } x \geq 0, \quad (2)$$

¹A derivada da função de distribuição em relação à variável em questão - no caso contínuo - é denominada função densidade de probabilidade (fdp) e é simbolizada por

$$f(t) = \frac{dF(t)}{dt}.$$

tem-se a distribuição Weibull exponenciada. Adicionalmente, quando $k = 1$ se tem a distribuição exponencial exponenciada.

1.1 Distribuições Beta

As distribuições da família beta foram introduzidas por [9], com a inserção da função de distribuição acumulada da distribuição exponencial na integral da distribuição beta, formando-se a distribuição beta exponencial (BE). Seja uma distribuição com função de distribuição contínua $G(y)$. A distribuição beta-G será

$$F(x) = \frac{1}{B(\alpha, \beta)} \int_0^{G(x)} w^{\alpha-1}(1-w)^{\beta-1} dw,$$

com $a > 0$ e $b > 0$, em que

$$B(\alpha, \beta) = \int_0^1 w^{\alpha-1}(1-w)^{\beta-1} dw.$$

Esta classe de distribuições tem se desenvolvido bastante nos últimos anos [3]. Entretanto, as distribuições beta-G são funções da distribuição beta incompleta, dada por

$$B_p(\alpha, \beta) = \int_0^p w^{\alpha-1}(1-w)^{\beta-1} dw,$$

que não tem forma analítica fechada.

1.2 Distribuições Kumaraswamy

Para contornar o problema das distribuições beta, é proposta família de distribuições de Kumaraswamy [2], definindo-se as distribuições K-G por

$$F(x) = 1 - (1 - G(x)^\alpha)^\beta, \quad \alpha > 0, \beta > 0,$$

pois esta leva a uma função de distribuição analiticamente tratável para uma dada distribuição de base G .

1.3 Distribuições de Marshall-Olkin

Por outra linha, o seguinte método de generalização de distribuições foi introduzido por [7]:

$$F(x) = \frac{G(x)}{G(x) + \beta(1 - G(x))}, \quad \beta > 0. \tag{3}$$

Esta família de distribuições é, desde então, denominada família de distribuições de Marshall e Olkin. Procedimentos de estimação baseados no gráfico de probabilidade Weibull são implementados em [10] para obter as estimativas dos parâmetros da distribuição Marshall-Olkin Weibull.

O modelo Marshall-Olkin Lomax é apresentado em [4], mostrando que a mesma pode ser definida como uma mistura de distribuições exponenciais e obtiveram as funções de densidade, taxa de falha e log-verossimilhança.

Entre os objetivos deste trabalho se destaca a generalização a família de distribuições de Marshall-Olkin [7] pela introdução de novos parâmetros, obtendo uma maior flexibilidade ², sendo portanto capaz de modelar comportamentos diversos de dados positivos. Particularmente, são obtidas algumas quantidades para esta nova generalização com a distribuição de Weibull.

²Isto é, uma maior quantidade de formatos possíveis aos quais uma curva possa se ajustar.

1.4 Modelo Proposto

No presente trabalho, introduzimos alguns parâmetro à família Marshall-Olkin, ficando com a forma mais geral

$$F(x) = \left\{ \frac{G(x)^\alpha}{G(x)^\alpha + \beta[1 - G(x)^\gamma]^\delta} \right\}^\theta, \quad \alpha > 0, \beta > 0, \gamma > 0, \delta > 0, \theta > 0, \quad (4)$$

Naturalmente, esta nova classe engloba a classe de distribuições de Marshall-Olkin (3) quando $\alpha = \gamma = \delta = \theta = 1$, e a família de distribuições exponenciadas (1) quando $\beta = \gamma = \delta = \theta = 1$, o que é bastante conveniente a fim de comparação das propriedades.

1.5 Funções utilizadas em Análise de Sobrevidência

Uma função de interesse em Análise de Sobrevidência é a denominada função de Sobrevidência. Ela é definida como a probabilidade de uma observação (peça mecânica, pessoa, etc.) vir a falhar num tempo superior a t , sendo dada por

$$S(t) = P(T > t) = 1 - P(T \leq t).$$

A função de sobrevivência é o complemento da função de distribuição, isto é, $S(t) = 1 - F(t)$, e podemos interpretá-la como a probabilidade de uma observação não vir a falhar num tempo inferior a t .

Outra função muito importante neste campo é a função de risco ou função taxa de falha, definida por

$$h(t) = \lim_{h \rightarrow 0} \frac{P(t \leq T \leq t + h | T \geq t)}{h} = \frac{f(t)}{S(t)}, \quad (5)$$

em que $f(t) = F'(t)$ é a função de densidade de probabilidade.

A função de risco é interpretada como a probabilidade de um indivíduo falhar no instante t , desde que a falha não ocorreu antes de t . Representa, pois, o risco eminente do indivíduo falhar em t .

1.6 Bootstrap

Os métodos tradicionais de estimação de parâmetros, tais como Newton-Raphson, Gradiente e BFGS, frequentemente apresentam problemas nas estimativas da variância dos estimadores. O método bootstrap é uma técnica robusta para se obter estimativas e erros-padrões de parâmetros quaisquer [6]. No mundo real, tomamos uma amostra aleatória X_1, \dots, X_n de uma distribuição F e construímos uma função T_n chamada estatística da qual se realizam inferências. O método bootstrap essencialmente repete esse procedimento utilizando a amostra aleatória como base para se construir uma distribuição empírica \hat{F}_n . O esquema a seguir ilustra o método:

$$\begin{aligned} \text{Mundo Real: } F &\implies X_1, \dots, X_n \implies T_n = g(X_1, \dots, X_n) \\ \text{Mundo Bootstrap: } \hat{F}_n &\implies X_1^*, \dots, X_n^* \implies T_n^* = g(X_1^*, \dots, X_n^*) \end{aligned}$$

2 Desenvolvimento

Seja $X \sim MOGE(\alpha, \beta, \gamma, \delta, \theta)$ com função de distribuição dada por (4). A respectiva função de densidade é dada por

$$f(x) = \beta \theta g(x) G^{-1-\alpha}(x) [1 - G^\gamma(x)]^{\delta-1} \times [\alpha + (\gamma\delta - \alpha)G^\gamma(x)] \left\{ \frac{G^\alpha(x)}{G^\alpha(x) + \beta[1 - G^\gamma(x)]^\delta} \right\}^{\theta+1} \quad (6)$$

Utilizando a distribuição de Weibull (2) é possível a densidade da distribuição de Marshall-Olkin Generalizada Exponenciada de Weibull (*MOGEW*). Na Figura 1 é representada a densidade *MOGEW* se variando alguns parâmetros. É possível se observar que, mesmo se variando apenas um ou dois parâmetros, os formatos obtidos são bastante variados, incluindo forma leptocúrtica (curva vermelha tracejada), simétrica (preta cheia) assimétricas positivas (curvas cheias verde e azul) e bimodal (preta tracejada), o que nos dá um indício da flexibilidade do modelo.

A função de sobrevivência, por outro lado, é

$$S(x) = 1 - \left\{ \frac{G(x)^\alpha}{G(x)^\alpha + \beta [1 - G(x)^\gamma]^\delta} \right\}^\theta$$

Assim, a função de risco é, segundo a equação (5), dada por

$$h(x) = \frac{\beta \theta g(x) G(x)^{-1-\alpha} [1 - G(x)^\gamma]^{\delta-1} [\alpha + (-\alpha + \gamma\delta)G(x)^\gamma] \left\{ \frac{G(x)^\alpha}{G(x)^\alpha + \beta[1 - G(x)^\gamma]^\delta} \right\}^{1+\theta}}{1 - \left\{ \frac{G(x)^\alpha}{G(x)^\alpha + \beta[1 - G(x)^\gamma]^\delta} \right\}^\theta} \quad (7)$$

Assim como no caso da densidade, utilizando a distribuição de base (2) na equação (7) acima, é possível se produzir gráficos para a função de risco da distribuição *MOGEW*. A Figura 2 expressa diversos formatos da função de risco dessa distribuição. Observamos comportamentos constantes, crescentes e decrescentes, como também os formatos de *U* (curva preta tracejada) e de *U* invertido (curvas cheias verde e vermelha), bastante requeridos em Análise de Sobrevida [1]. Também é possível se notar os dois comportamentos na mesma função (isto é, risco decrescente-crescente-decrescente: curva verde tracejada). Isso, outra uma vez, mostra a flexibilidade do modelo.

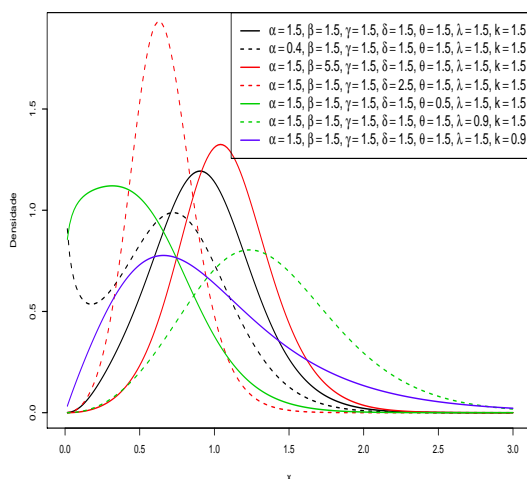


Figura 1: Densidade da distribuição proposta para alguns valores dos parâmetros.

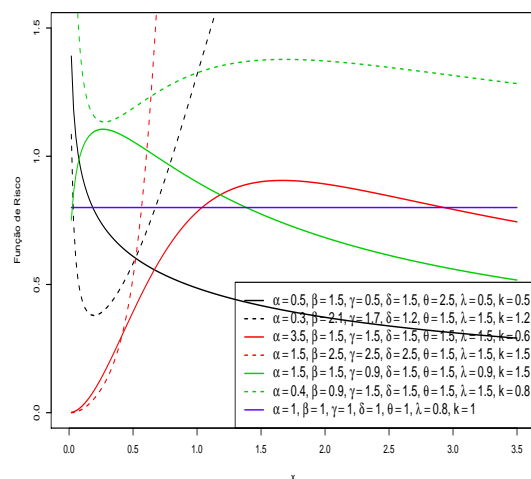


Figura 2: Função de risco da distribuição proposta para alguns valores dos parâmetros.

Seja $X \sim MOGE(\alpha, \beta, \gamma, \delta, \theta)$, a função de distribuição (4) pode ser escrita como uma combinação linear da distribuição de base $G(x)$. Para $\beta > \frac{G^\alpha(x)}{[1-G^\gamma(x)]^\delta}$ temos que

$$F(x) = \sum_{j,k=0}^{\infty} \nu_{j,k} G^{\alpha(\theta+j)+\gamma k}(x) \tag{8}$$

em que $\nu_{j,k} = \nu_{j,k}(\beta, \delta, \theta) = \frac{\Gamma(\theta+j)\Gamma(\delta(\theta+j)+k)(-1)^j}{\Gamma(\theta)\Gamma(\delta(\theta+j))j!k!\beta^{\theta+j}}$.

Para $\beta < \frac{G^\alpha(x)}{[1-G^\gamma(x)]^\delta}$ temos que

$$F(x) = \sum_{j,k=0}^{\infty} \omega_{j,k} G^{\gamma k-\alpha j}(x) \tag{9}$$

com $\omega_{j,k} = \omega_{j,k}(\beta, \delta, \theta) = \binom{\delta j}{k} (-1)^{j+k} \frac{\Gamma(\theta+j)\beta^j}{\Gamma(\theta)j!}$.

Derivando a função (8) em relação a x , temos

$$f(x) = \frac{dF(x)}{dx} = \sum_{j,k=0}^{\infty} \nu_{j,k} g(x) G^{\alpha(\theta+j)+\gamma k-1}(x) \tag{10}$$

em que $\nu_{j,k} = \nu_{j,k}(\beta, \delta, \theta) = \frac{\Gamma(\theta+j)\Gamma(\delta(\theta+j)+k)(-1)^j[\alpha(\theta+j)+\gamma k]}{\Gamma(\theta)\Gamma(\delta(\theta+j))j!k!\beta^{\theta+j}}$.

Se $\beta < \frac{G^\alpha(x)}{[1-G^\gamma(x)]^\delta}$, utilizamos a função (9) e derivamos

$$f(x) = \frac{dF(x)}{dx} = \sum_{j,k=0}^{\infty} w_{j,k} g(x) G^{\gamma k-\alpha j-1}(x) \tag{11}$$

em que $w_{j,k} = w_{j,k}(\beta, \delta, \theta) = \binom{\delta j}{k} (-1)^{j+k} \frac{\Gamma(\theta+j)\beta^j(\gamma k-\alpha j)}{\Gamma(\theta)j!}$.

Seja a densidade na forma expandida (10) para $\beta > \frac{[1-G^\gamma(x)]^\delta}{G^\alpha(x)}$. Então o s -ésimo momento é

$$\begin{aligned} E(X^s) &= \int_{-\infty}^{\infty} x^s f(x) dx \\ &= \sum_{j,k=0}^{\infty} \nu_{j,k} \tau_{s,\alpha(\theta+j)+\gamma k-1}. \end{aligned} \tag{12}$$

em que $\tau_{s,r} = E[X^s F^r(x)] = \int_{-\infty}^{\infty} x^s F^r(x) f(x) dx$. Por outro lado, se $\beta < \frac{G^\alpha(x)}{[1-G^\gamma(x)]^\delta}$ utilizamos a função densidade (11) e encontramos

$$E(X^s) = \sum_{j,k=0}^{\infty} w_{j,k} \tau_{s,\gamma k-\alpha j-1}. \tag{13}$$

As equações (12) e (13) permitem encontrar os momentos centrais, particularmente, média, variância, coeficiente de assimetria e de curtose.

3 Resultados e Discussão

Seja $\xi = (\alpha, \beta, \gamma, \delta, \theta, \lambda, k)'$ os parâmetros da distribuição $MOGEW$. Com o intuito de testar a adequação da distribuição proposta foram simulados 150 valores da distribuição $MOGEW$ com $\alpha = 1, 5$, $\beta = 1, 2$, $\gamma = 1, 5$, $\delta = 1, 0$, $\theta = 0, 3$, $\lambda = 1, 7$ e $k = 2, 5$. Os dados foram subdivididos

em dois conjuntos, um de **treino** ou ajuste com 100 valores e um segundo de **validação** com 50 valores. Foi ajustado, aos dados de treino, a densidade com os parâmetros verdadeiros ξ acima, sua estimativa de máxima verossimilhança (EMV) $\hat{\xi}$ via algoritmo de estimação *BFGS*. Também se repetiu a simulação de máxima verossimilhança utilizando reamostragem bootstrap ξ^* e bootstrap corrigido $\xi^{**} = \xi^* - \text{viés}(\xi^*)$. Na Figura 3 podemos observar o histograma e os respectivos ajustes para os dados de treino. Já na Figura 4 estão dispostos o histograma e os respectivos ajustes para os dados de validação.

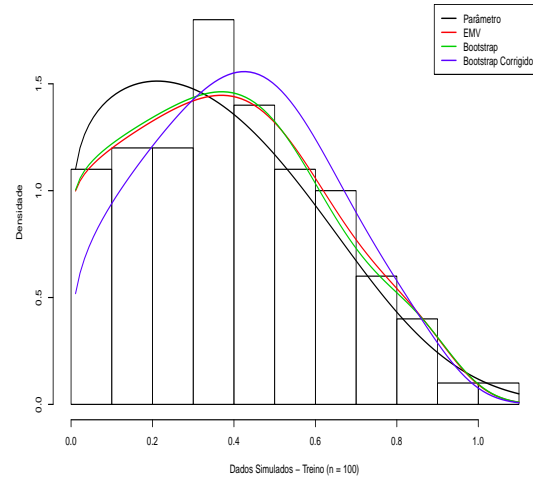
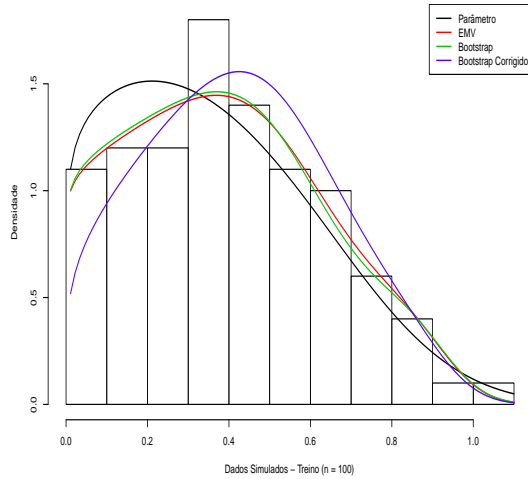


Figura 3: Histograma para dados de treino.

Figura 4: Histograma para dados de validação.

Visualmente temos que as estimativas por máxima verossimilhança e via bootstrap parecem ser mais condizentes com o formato dos histogramas. Uma medida de ajuste indicada nesse caso é a log-verossimilhança definida por

$$\ell(\xi) = \sum_{j=1}^n \log f(\xi; x_j).$$

O melhor modelo deve ser aquele que produz a maior log-verossimilhança. A medida de AIC, difundida da literatura, produziria a mesma discriminação, uma vez que o número de parâmetros a ser testado aqui é o mesmo (estamos testando métodos de ajustes na mesma distribuição) e portanto será desnecessária.

Na Tabela 3 encontramos os valores obtidos dos ajustes bem como as log-verossimilhanças produzidas para os dados de treino e de validação. Observamos que tanto no conjunto de treino, quanto no de validação as log-verossimilhanças do ajuste de máxima verossimilhança via bootstrap são maiores. Concluímos que as estimativas ajuste via bootstrap são mais consistentes para este conjunto de dados.

Tabela 1: Estimativas dos parâmetros e logverossimilhança para os dados de treino e de validação

| | α | β | γ | δ | θ | λ | k | ℓ - treino | ℓ - valid. |
|-----------------|----------|---------|----------|----------|----------|-----------|-------|-----------------|-----------------|
| Parâmetros | 1,500 | 1,200 | 1,500 | 1,000 | 0,300 | 1,700 | 2,500 | 7,295 | 2,147 |
| EMV | 0,687 | 2,820 | 0,085 | 0,473 | 0,285 | 1,496 | 5,333 | 9,085 | 2,981 |
| Bootstrap | 0,820 | 2,832 | 0,098 | 0,416 | 0,244 | 1,537 | 5,301 | 9,091 | 2,988 |
| Boot. Corrigido | 0,639 | 3,496 | 0,066 | 0,600 | 0,365 | 1,462 | 5,170 | 7,550 | 2,745 |

Do exposto observamos que a distribuição de Marshall-Olkin Generalizada Exponenciada de Weibull é uma distribuição flexível e capaz de modelar dados de Análise de Sobrevivência. Em trabalhos suplementares já se tem resultados promissores em que a distribuição proposta tem

ajuste superior às concorrentes com as bases de dados e medidas utilizadas nos trabalhos já publicados. Também já foram obtidas diversas quantidades não expostas aqui, tais como função geradora de momentos, função quantílica (a qual se utilizou para gerar os dados deste trabalho), mediana, entre outras.

4 Conclusão

Neste trabalho foi introduzida a distribuição de Marshall-Olkin Generalizada Exponenciada de Weibull através de uma generalização da família de Marshall-Olkin. Foram obtidas as densidades e funções de risco da mesma, além de se mostrar os gráficos destas funções para diversos valores dos parâmetros. Ainda foram obtidas expansões da função de sobrevivência e da densidade de probabilidade e momentos de ordem s . Por último, foi simulado um conjunto de dados de treino e outro de validação, aos quais ajustou-se diversas estimações de densidades, observando-se que o melhor ajuste foi produzido pela estimativa de máxima verossimilhança combinada com o método de reamostragem bootstrap pelo medida da log-verossimilhança. Assim, a distribuição *MOGEW* é uma distribuição flexível e capaz de modelar dados de Análise de Sobrevivência.

Referências

- [1] E. A. Colosimo; S. R. Giolo, *Análise de Sobrevivência Aplicada*. [S.l.]: Associação Brasileira de Estatística, 2006.
- [2] G. M. Cordeiro; M. Castro, A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, v. 81, n. 7, p. 883 - 898, 2011.
- [3] G. M. Cordeiro; A. B. Simas; B. D. Stosic, Closed form expressions for moments of the beta weibull distribution. *Annals of the Brazilian Academy of Sciences*, v. 83, p. 357 - 373, 2011.
- [4] M. Ghitany; F. Al-Awadhi; L. Alkhalfan, Marshall-olkin extended lomax distribution and its application to censored data. *Communications in Statistics - Theory and Methods*, Taylor & Francis, v. 36, n. 10, p. 1855 - 1866, 2007.
- [5] R. D. Gupta; D. Kundu, Generalized exponential distributions. *Australian and New Zealand Journal of Statistics*, 1999.
- [6] G. James; D. Witten; T. Hastie; R. Tibshirani, "An Introduction to Statistical Learning - with applications in R", Springer, [S.l.], 2013.
- [7] A. W. Marshall; I. Olkin, A new method for adding a parameter to a family of distributions with application to the exponential and weibull families. *Biometrika*, v. 84, n. 3, p. 641 - 652, 1997.
- [8] G. S. Mudholkar; D. K. Srivastava; M. Freimer, The exponentiated weibull family. *Technometrics*, v. 37, p. 436 - 445, 1995.
- [9] S. Nadarajah; S. Kotz, The beta exponential distribution. *Reliability Engineering and System Safety*, 2006.
- [10] T. Zhang; M. Xie, Failure data analysis with extended weibull distribution. *Communications in Statistics Simulation and Computation*, Taylor & Francis, v. 36, n. 3, p. 579 - 592, 2007.