

# Expansão da Função de Densidade de Probabilidade da Distribuição Marshall-Olkin Exponencial Generalizada

Geovano Francisco Felipe<sup>1</sup>, Saullo Wesley Soares da Silva<sup>2</sup>,  
André Pereira do Nascimento<sup>3</sup>, Kleber N. N. O. Barros<sup>4</sup>

Departamento de Estatística, Universidade Estadual da Paraíba-UEPB, Campus I, Campina Grande-PB. E-mails: geovano\_felipe@hotmail.com<sup>1</sup>, saulowesley@gmail.com<sup>2</sup>, andre.pn82@yahoo.com<sup>3</sup>, kleberbarros@cct.uepb.edu.br<sup>4</sup> Telefone: (83) 3315 3342

## RESUMO

Os dados estatísticos se apresentam nas mais diversas formas, desde contagens, dados contínuos ou de tempos de vida ou falhas, que são modelados, em geral, por distribuições de um ou dois parâmetros. Com o advento da computação de alto desempenho e de ferramentas de computação analítica muitos modelos novos vem sendo desenvolvidos. Neste contexto, distribuições mais flexíveis com a incorporação de um ou dois parâmetros vem sendo cada vez mais estudadas, denominando-a de família de distribuição Marshal-Olkin generalizada. No presente trabalho, explora-se a densidade da distribuição de Marshal-Olkin exponencial generalizada (MOEG), uma distribuição triparamétrica competindo com a distribuição beta exponencial e distribuição gama generalizada e de Weibull generalizada. Integrais de funções desta densidade podem demandar alto custo computacional. Para contornar este problema, a partir de séries provenientes do teorema binomial, expande-se a densidade, para obter resultados computacionalmente tratáveis para funções, tais como esperança, variância, curtose e moda da nova distribuição, como funções da densidade exponencial, que são facilmente obtidas. Os resultados mostram que bastam poucas somas, do ponto de vista computacional, para se obter boas aproximações usando expansões.

**Palavras-chave:** Distribuições Generalizadas, Distribuição Marshal-Olkin exponencial generalizada, Expansões Binomiais.

# INTRODUÇÃO

A estatística paramétrica é o principal e mais promissor ramo da inferência estatística. Nos últimos anos a quantidade de novos modelos paramétricos vem aumentando com o maior acesso à ferramentas computacionais. Distribuições como normal, exponencial, gama, beta, de Laplace, normal-inversa, de Rayleigh, de Weibull, Gumbel, Weibull estendida, Fréchet, entre outras vem sendo generalizadas pela adição de parâmetros que possibilitam uma maior flexibilidade.

Gupta e Kundu (1999) propuseram a distribuição exponencial generalizada (EG) como alternativas às distribuições gama e de Weibull. O princípio de tais distribuições é simples, dada uma distribuição com função de distribuição contínua  $G(y)$ , sua generalização ou exponencialização é obtida por

$$F(y; \alpha) = G(y)^\alpha, \quad y \in \mathbb{R}, \alpha > 0. \quad (1)$$

Quando  $G(y) = 1 - e^{-\lambda y}$  e  $G(y) = 1 - e^{-(\lambda y)^\gamma}$ , tem-se a distribuição exponencial generalizada e Weibull generalizada, respectivamente.

As distribuições da família beta foram obtidas por Nadarajah e Kotz (Nadarajah & Kotz, 2006) com a inserção da função de distribuição acumulada da distribuição de exponencial na integral da distribuição beta, formando-se a distribuição beta exponencial (BE). Assim, seja uma distribuição com função de distribuição contínua  $G(y)$ , a distribuição beta-G será

$$F(y; a, b) = \frac{1}{B(a, b)} \int_0^{G(y)} w^{a-1} (1-w)^{b-1} dw,$$

com  $y \in \mathbb{R}$ ,  $a > 0$  e  $b > 0$ , em que

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

Uma alternativa (Jones, 2008; Cordeiro & Castro, 2010) é utilizar a distribuição de Kumaraswamy, definindo-se as distribuições K-G por

$$F(y; a, b) = 1 - (1 - G(y)^a)^b, \quad y \in \mathbb{R}, a > 0, b > 0,$$

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

primitiva  $G$ .

Por outra linha, Marshall e Olkin (1997) propõem o seguinte método de adição de parâmetros para uma dada distribuição  $G$ ,

$$F(y; \beta) = \frac{G(y)}{G(y) + \beta(1 - G(y))}, \quad y \in \mathbb{R}, \beta > 0. \quad (2)$$

Aqui, introduz-se um segundo parâmetro à família Marshall-Olkin, ficando com a forma mais geral

$$F(y; \alpha, \beta) = \frac{G(y)^\alpha}{G(y)^\alpha + \beta(1 - G(y)^\alpha)}, \quad y \in \mathbb{R}, \alpha > 0, \beta > 0. \quad (3)$$

cuja derivada em relação a  $y$ , após alguma manipulação algébrica, é

$$f(y; \alpha, \beta) = \frac{\alpha\beta G(y)^{\alpha-1} g(y)}{[1 - \beta G(y)^\alpha]^2}, \quad y \in \mathbb{R}, \alpha > 0, \beta > 0. \quad (4)$$

sendo  $\bar{u} = 1 - u$ . Naturalmente, esta nova classe engloba a classe de distribuições de Marshall-Olkin (2) quando  $\alpha = 1$  e a família de distribuições generalizadas (1) quando  $\beta = 1$ , o que é bastante conveniente a título de comparação das propriedades. Se  $G(y) = 1 - e^{-\lambda y}$ , definimos a distribuição Marshall-Olkin exponencial generalizada (MOEG), como uma distribuição triparamétrica competindo com a distribuição beta exponencial e distribuição Weibull generalizada.

## RESULTADOS E DISCUSSÃO

Seja a Equação (3) da família de distribuições de Marshall-Olkin generalizada, obtida pela adição de um expoente  $\alpha > 0$  em cada termo  $G(y)$  na Equação (2). Escrevendo-se  $G(y) = 1 - \exp(-\lambda y)$ , com  $y > 0, \lambda > 0$ , obtém-se a seguinte função

$$F(y; \alpha, \beta, \lambda) = \frac{(1 - e^{-\lambda y})^\alpha}{(1 - e^{-\lambda y})^\alpha + \beta [1 - (1 - e^{-\lambda y})^\alpha]}, \quad (5)$$

que é a função de distribuição para a distribuição de probabilidade Marshall-Olkin exponencial generalizada. Derivando a Equação (5) em relação a  $y$ , obtém-se a respectiva função densidade de probabilidade

$$f(y; \alpha, \beta, \lambda) = \frac{\alpha\beta\lambda(1 - e^{-\lambda y})^{\alpha-1}e^{-\lambda y}}{\{1 - \bar{\beta}(1 - e^{-\alpha\lambda y})\}^2}. \quad (6)$$

Avaliar a integrais de funções da densidade (6) pode demandar alto custo computacional em máquinas de configuração razoável. Assim, manipulações algébricas podem facilitar os cálculos, ao se encontrar expressões de funções conhecidas. Para  $a > 0$  real e  $z \in (0, 1)$ , podemos utilizar a representação em séries

$$(1 - z)^{a-1} = \sum_{i=0}^{\infty} (-1)^i \binom{a-1}{i} z^i \quad (7)$$

e para  $|z| < 1$ , temos

$$(1 - z)^{-a} = \sum_{j=0}^{\infty} \frac{\Gamma(a+j)}{\Gamma(a)j!} z^j \quad (8)$$

em que  $\Gamma(u) = \int_0^{\infty} x^{u-1}e^{-x}dx$ , com  $u > 0$  é a função gama. Aplicando-se as funções (7) e (8) na função (6), após alguns passos, para  $\alpha > 1$ , chega-se a expressão:

$$\begin{aligned} f(y; \alpha, \beta, \lambda) &= \alpha\beta\lambda \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^j \sum_{s=0}^{\alpha k} \bar{\beta}(j+1)(-1)^{i+k+s} \binom{\alpha-1}{i} \binom{j}{k} \binom{\alpha k}{s} e^{-\lambda(i+s+1)y} \\ &= \sum_{i,j,k,s} \frac{\alpha\beta\bar{\beta}}{i+s+1} (j+1)(-1)^{i+k+s} \binom{\alpha-1}{i} \binom{j}{k} \binom{\alpha k}{s} \cdot \lambda(i+s+1)e^{-\lambda(i+s+1)y} \\ &= \sum_{i,j,k,s} \nu \cdot f_{\lambda(i+s+1)}(y) \end{aligned} \quad (9)$$

onde foi utilizada a seguinte notação

$$\sum_{i,j,k,s} = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^j \sum_{s=0}^{\alpha k}$$

e realizadas as seguintes simplificações

$$\nu = \nu_{i,j,k,s}(\alpha, \beta) = \frac{\alpha\beta\bar{\beta}}{i+s+1} (j+1)(-1)^{i+k+s} \binom{\alpha-1}{i} \binom{j}{k} \binom{\alpha k}{s}$$

e

$$f_{\lambda(i+s+1)}(y) = \lambda(i+s+1)e^{-\lambda(i+s+1)y}$$

A expressão (9) é bastante conveniente, uma vez que  $\nu$  envolve apenas coeficientes e  $f_{\lambda(i+s+1)}(y)$  segue uma distribuição exponencial com parâmetro  $\lambda(i+s+1)$ . Assim, qualquer função de interesse para a distribuição MOEG pode ser avaliada pelas propriedades conhecidas da distribuição exponencial. Por exemplo, a esperança e a variância podem ser obtidas por:

$$\begin{aligned} E(y) &= \sum_{i,j,k,s} \nu \int_0^{\infty} y f_{\lambda(i+s+1)}(y) dy \\ &= \sum_{i,j,k,s} \frac{\nu}{\lambda(i+s+1)} \end{aligned} \quad (10)$$

e

$$\begin{aligned} Var(y) &= \sum_{i,j,k,s} \nu \int_0^{\infty} [y - E(y)]^2 f_{\lambda(i+s+1)}(y) dy \\ &= \sum_{i,j,k,s} \frac{\nu}{[\lambda(i+s+1)]^2} \end{aligned} \quad (11)$$

As expressões acima podem ser calculadas limitando os somatórios a uma quantidade finita, que se denominará de agora em diante de *IMAX* (valor limite para os índices  $i, j, k - s$  é função  $\alpha$  e  $k$ ), com um pequeno erro, uma vez que se dispõe de tempo computacional finito.

Outros resultados de interesse, em Estatística, são a assimetria, a curtose e a mediana, cujos resultados para a distribuição MOEG são, respectivamente:

$$Skew(y) = 2, \quad Kurt(y) = 6, \quad MD(y) = \frac{\log 2}{\lambda}$$

Porém, os resultados acima, uma vez que o somatório  $\sum_{i,j,k,s} \nu = 1$ , são exatamente os mesmos obtidos pela distribuição exponencial, isto é, a distribuição MOEG e a distribuição exponencial têm a mesma assimetria, curtose e mediana. Assim, faz-se desnecessária qualquer menção aos mesmos nos resultados deste trabalho.

Para obter resultados numéricos para as expressões (10) e (11) foram utilizados os valores 1, 2 e 3 para  $\alpha$ , para  $\beta$  se utilizou 0,5, 0,9 e 2, enquanto que para  $\lambda$  se utilizou 0,5, 1 e 2. Ainda foram utilizados os valores 10 e 20 como índices máximos (*IMAX*) em  $i, j$  e  $k$ . Na **Tabela 1** aparecem os resultados para esperança e variância para cada  $\alpha, \beta$  e  $\lambda$  acima. Note que não há grandes divergências entre os valores calculados para *IMAX* = 10 ou 20. Para o caso em que  $\alpha = 1$  avaliamos  $i \cdot j \cdot k \cdot s = 10000$  e  $16000$  somas no caso em que o valor o máximo dos índices

é 10 e 20, respectivamente. Isto é, a simples mudança de  $IMAX$  de 10 para 20 resulta num tempo computacional no mínimo 16 vezes superior.

Tabela 1: Estimativas das esperanças e variâncias para alguns valores de  $\alpha$ ,  $\beta$ ,  $\lambda$  e  $IMAX$ .

			$IMAX = 10$		$IMAX = 20$	
$\alpha$	$\beta$	$\lambda$	$E(y)$	$Var(y)$	$E(y)$	$Var(y)$
1,0	0,5	0,5	1,38622	2,32895	1,38629	2,32896
		1,0	0,693109	0,582238	0,693147	0,582241
		2,0	0,346555	0,145559	0,346574	0,14556
	0,9	0,5	1,89649	3,69424	1,89649	3,69424
		1	0,948245	0,92356	0,948245	0,92356
		2,0	0,474122	0,23089	0,474122	0,23089
	2,0	0,5	2,94618	6,60981	2,86556	6,58838
		1,0	1,47309	1,65245	1,43278	1,64709
		2,0	0,736544	0,413113	0,71639	0,411773
2,0	0,5	0,5	2,26366	4,3909	2,26394	4,39101
		1,0	1,13183	1,09773	1,13197	1,09775
		2,0	0,565916	0,274431	0,565986	0,274438
	0,9	0,5	2,87875	6,53607	2,87875	6,53607
		1,0	1,43937	1,63402	1,43937	1,63402
		2,0	0,719687	0,408504	0,719687	0,408504
	2,0	0,5	4,49398	11,0127	4,31015	10,8808
		1,0	2,24699	2,75317	2,15504	2,7202
		2,0	1,12349	0,688292	1,07752	0,680049
3,0	0,5	0,5	2,88002	6,19493	2,88052	6,19522
		1,0	1,44001	1,54873	1,44026	1,5488
		2,0	0,720004	0,387183	0,720131	0,387201
	0,9	0,5	3,53829	8,87497	3,53829	8,87497
		1,0	1,76915	2,21874	1,76915	2,21874
		2,0	0,884574	0,554685	0,884574	0,554685
	2,0	0,5	5,62286	14,6131	151368	19,8942
		1,0	2,81143	3,65327	2,69077	3,58518
		2,0	1,40572	0,913318	1,34539	0,896295

Pode-se observar, na **Tabela 2**, os tempos de execução, para  $IMAX = 10$  e 20, bem como o verdadeiro valor ( $REAL$ ) das esperanças e variâncias, fixando  $\alpha = 1$ ,  $\beta = 0,5$  e fazendo  $\lambda = 0,5$  e 2. Observe que os valores não diferem entre si, como antes, e aproximam-se bastante valor verdadeiro. No entanto, o tempo de execução para aumenta drasticamente, quando se aumenta o índice. Pode-se notar que o tempo de para se calcular as integrais, sem utilizar expansões ( $REAL$ ) muito maior do que quando se utiliza expansões para  $IMAX = 20$ , chegando a ser aproximadamente vinte vezes maior para  $\lambda = 2$  para a esperança, e quatro vezes maior para a variância. Pode-se observar, ainda, que o viés da esperança é bem menor que o viés da

variância, possivelmente por envolver a esperança que já tem algum viés e por envolver um termo quadrático.

Tabela 2: Estimativas e valores reais das esperanças e variâncias com seus respectivos tempos de execução.

Método	$\lambda$	$E(y)$	Tempo (seg)	$Var(y)$	Tempo (seg)
<i>IMAX</i> = 10	0,5	1,38622	0,093	2,32895	0,11
	2	0,346555	0,109	0,145559	0,109
<i>IMAX</i> = 20	0,5	1,38629	1,185	2,32896	1,248
	2	0,346574	1,155	0,14556	1,217
REAL	0,5	1,38629	12,87	2,73611	4,009
	2	0,346574	24,866	0,171007	4,025

Do exposto conclui-se que com um pequeno viés, é possível estimar a esperança da densidade (6), bastando-se utilizar *IMAX* = 10.

## CONCLUSÕES

Neste trabalho, investigou-se a esperança, variância, assimetria, curtose e moda da densidade da distribuição tripamétrica de Marshal-Olkin exponencial generalizada (MOEG). Utilizando expansões, observou-se que a assimetria, curtose e moda desta distribuição, coincidem com as correspondentes funções para a distribuição exponencial, não sendo necessária avaliação das correspondentes estimativa. Porém, para a média e variância, as expansões dependem dos índices, de forma que para valores maiores dos mesmos, pode-se demandar um tempo computacional bastante maior com um ganho pequeno. Os resultados mostram que basta uma soma de poucas parcelas (da ordem de dezenas de milhares), do ponto de vista computacional, com tempo da ordem de décimos de segundos, para se obter boas aproximações para a esperança e a variância usando expansões.

## REFERÊNCIAS

CORDEIRO, G. M., CASTRO, M. *A new family of generalized distributions*. Journal of Statistical Computation and Simulation, 2010. 2009.

GUPTA, R. D., KUNDU, D. *Generalized Exponential Distributions*. Australian and New Zealand Journal of Statistics, Oxford, 1999.

JONES, M. C. *Kumaraswamy's distribution: A beta-type distribution with some tractability advantages*. Statistical Methodology 6, 70 – 81, 2008.

MARSHALL, A.W., OLKIN, I. *A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families*. Biometrika, 1997.

NADARAJAH, S., KOTZ, S. *The Beta Exponential Distribution*. Reliability Engineering and System Safety, 2006.