THE GAMMA-DAGUM DISTRIBUTION: DEFINITION, PROPERTIES AND APPLICATION

Jailson Araujo Rodrigues
Instituto Federal da Bahia - Campus de Feira de Santana
jailsondearaujo@yahoo.com.br

Ana Paula Coelho Madeira Silva
Universidade Federal de São João Del Rei - Campus de Sete Lagoas
apcmadeira@hotmail.com

ABSTRACT

In this work, a new distribution called the gamma-Dagum is introduced. Some of the main properties of this distribution are derived, including, kth moment, mean, variance, skewness and kurtosis. The estimation of parameters using the methods of moments and maximum likelihood is also discussed. The flexibility of this distribution is illustrated in an application to a real data set.

RESUMO

Neste trabalho, é introduzida uma nova distribuição denominada gama-Dagum. Algumas das principais propriedades dessa distribuição são deduzidas, incluindo o momento de ordem k, média, variância, coeficiente de assimetria e o coeficiente de curtose. A estimação dos parâmetros utilizando o método da máxima verossimilhança e o método dos momentos também é discutida. A flexibilidade dessa distribuição é ilustrada em uma aplicação a um conjunto de dados reais.

Palavras-chave: Critério de informação de Akaike, distribuição Dagum, distribuição gama, momentos.

1 INTRODUCTION

In recent years, generalized distributions have been widely studied in statistics as they possess flexibility in applications. This is justified because the traditional distributions often do not provide good fit in relation to the real data set studied. For example, Akinsete [1] proposed the beta-Pareto distribution and discussed various properties. Alzaatreh [2] introduced and studied the gamma-Pareto distribution. Alzaatreh [3] proposed the Weibull-Pareto distribution. Condino [4] studied the beta-Dagum distribution. Famoye [5] presented the beta-Weibull distribution. Paranaíba [7] defined a five-parameter beta Burr XII distribution and discussed various properties. Pascoa [8] introduced the Kumaraswamy generalized gamma distribution. Pinho [9] studied the gamma-exponentiated Weibull distribution. Rodrigues [12] proposed a five-parameter distribution, the so-called Pareto confluente hypergeometric distribution and Silva [13] studied the beta modified Weibull distribution.

In this work, is presented a new distribution called the gamma-Dagum distribution. Some of the main properties of this distribution are derived. We discuss the estimation of the model parameters by methods of moments and maximum likelihood. Moreover, the proposed distribution is applied to fit a real data set.

The papper is organized as follows. In Section 2 is defined the gamma-Dagum distribution and some special sub-models are discussed. The kth moment, mean, variance, skewness and kurtosis are derived in Section 3. The estimation of parameters using the methods of moments and maximum likelihood is discussed in Section 4. Finally, in Section 5 an application on a real data set is reported.

The calculations of this note involve the gamma function defined by

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} \exp(-t) dt \tag{1}$$

the incomplete gamma function defined by

$$\gamma(\alpha, x) = \int_0^x t^{\alpha-1} \exp(-t) dt \tag{2}$$

and the beta function defined by

$$B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt \tag{3}$$

2 THE MODEL

Let $F(x)$ be the cumulative distribution function (CDF) of any random variable X. The CDF of the gamma-G family of distributions defined by Ristic [11] is given by

$$G(x) = 1 - \frac{1}{\Gamma(\alpha)} \int_0^{-\log F(x)} t^{\alpha-1} \exp^{-t} dt \tag{4}$$

for $x > 0$ and $\alpha > 0$. The probability density function (PDF) associated with (4) is

$$g(x) = \frac{1}{\Gamma(\alpha)} [-\log F(x)]^{\alpha-1} f(x) \tag{5}$$

If $\alpha = n \in \mathbb{N}$, Ristic [11] states that (5) is the nth lower record value of sequence of i.i.d variable from a population with density $f(x)$.

The CDF and PDF of Dagum random variable, respectively, are

$$F(x) = \left(1 + \lambda x^{-\delta}\right)^{-\beta} \tag{6}$$

and

$$f(x) = \beta\lambda\delta x^{-\delta-1} \left(1 + \lambda x^{-\delta}\right)^{-\beta-1} \tag{7}$$

where $x > 0$, $\beta > 0$, $\lambda > 0$ and $\delta > 0$. The parameter λ is a scale parameter, while β and δ are shape parameters. Replacing (6) in (4), we obtain a new distribution, called gamma-Dagum (GD), with CDF, for interger $\alpha > 0$, given by

$$G(x) = 1 - \frac{1}{\Gamma(\alpha)} \int_0^{\beta \log\left(1+\lambda x^{-\delta}\right)} t^{\alpha-1} \exp^{-t} dt \tag{8}$$

or equivalently

$$G(x) = 1 - \frac{1}{\Gamma(\alpha)} \gamma\left(\alpha, \beta \log\left(1 + \lambda x^{-\delta}\right)\right) \tag{9}$$

The corresponding PDF of the GD distribution is

$$g(x) = \frac{\lambda\delta\beta^\alpha}{\Gamma(\alpha)} x^{-\delta-1} \left(1 + \lambda x^{-\delta}\right)^{-\beta-1} \left[\log\left(1 + \lambda x^{-\delta}\right)\right]^{\alpha-1} \tag{10}$$

The failure rate function, also known as the hazard rate function, of the GD distribution (10) is defined as

$$h(x) = \frac{\lambda\delta\beta^{\alpha}x^{-\delta-1}\left(1+\lambda x^{-\delta}\right)^{-\beta-1}\left[\log\left(1+\lambda x^{-\delta}\right)\right]^{\alpha-1}}{\displaystyle\int_0^{\beta\log\left(1+\lambda x^{-\delta}\right)} t^{\alpha-1}\exp^{-t}\,dt} \quad (11)$$

Observing the expansion of the exponential function,

$$\exp^{-z} = \sum_{j=0}^{\infty}\frac{(-1)^j z^j}{j!} \quad (12)$$

We can express the CDF (8) as follows

$$G(x) = 1 - \frac{1}{\Gamma(\alpha)}\sum_{j=0}^{\infty}\frac{(-1)^j\beta^{\alpha+j}}{(\alpha+j)j!}\left[\log\left(1+\lambda x^{-\delta}\right)\right]^{\alpha+j} \quad (13)$$

Otherwise, the PDF (10) is given by

$$g(x) = \frac{\lambda\delta x^{-\delta-1}}{\left(1+\lambda x^{-\delta}\right)\Gamma(\alpha)}\sum_{j=0}^{\infty}\frac{(-1)^j\beta^{\alpha+j}}{(\alpha+j)j!}\left[\log\left(1+\lambda x^{-\delta}\right)\right]^{\alpha+j-1} \quad (14)$$

The GD distribution represents a generalization of some distributions. If $\lambda = 1$ the model reduces to the gamma-Burr III and the gamma-Fisk distribution when $\beta = 1$.

Some of the possible shapes of the GD density (10) and failure rate function (11) are illustrated in Figure 1.

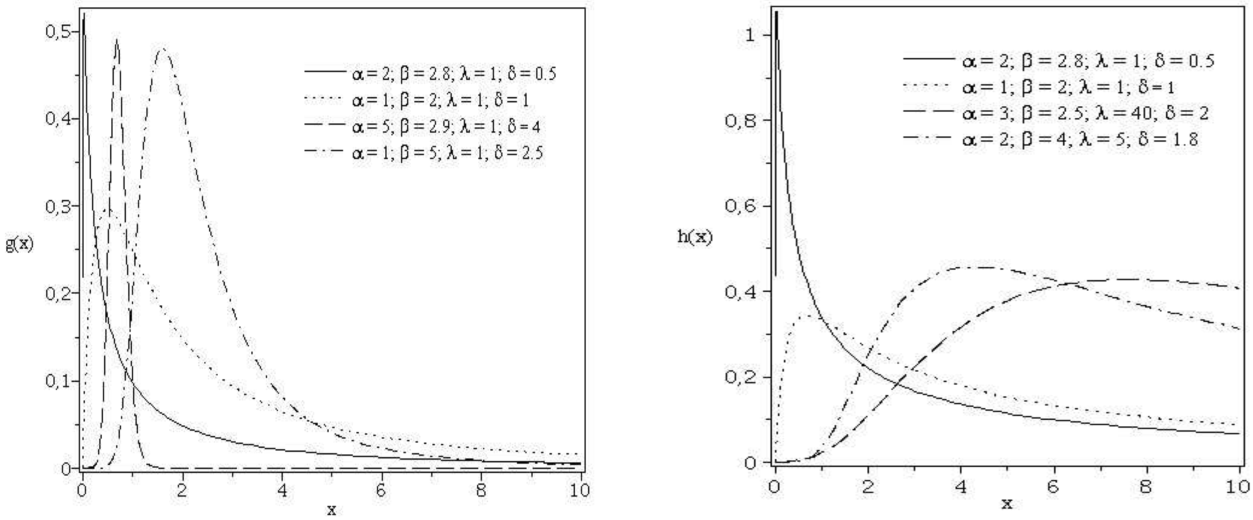

FIGURA 1: Plots of the GD density (10) and hazard rate function (11) for selected parameter values.

3 MOMENTS

Many of the interesting characteristics and features of a distribution can be studied through its moments (e.g. tendency, dispersion, skewness and kurtosis). Therefore, it is customary to derive the moments when a new distribution is proposed.

Lema 3.1: *(Equation (2.6.10.48), [10]). For* $Re\ \alpha > 0$, $Re\ (\rho-\alpha) > 0$ *and* $|arg\ (cx+d)| < \pi$,

$$\int_0^{\infty} x^{\alpha-1}(cx+d)^{-\rho}\log^n(cx+d)\,dx = (-1)^n\left(\frac{d}{c}\right)^{\alpha}\frac{\partial^n}{\partial\rho^n}\left[d^{-\rho}B(\alpha,\rho-\alpha)\right].$$

Teorema 3.1: *If X has the GD PDF (10), then its kth moment can be written as*

$$E\left(X^k\right)=\frac{(-1)^{\alpha-1}\beta^{\alpha}\lambda^{\frac{k}{\delta}}}{\Gamma(\alpha)}\frac{\partial^{\alpha-1}}{\partial(\beta+1)^{\alpha-1}}B\left(1-\frac{k}{\delta},\beta+\frac{k}{\delta}\right) \tag{15}$$

for $\delta > k$.

Proof:

The kth moment about origin of the GD distribution has been obtained as

$$E\left(X^k\right)=\frac{\lambda\delta\beta^{\alpha}}{\Gamma(\alpha)}\int_0^{\infty}x^{k-\delta-1}\left(1+\lambda x^{-\delta}\right)^{-\beta-1}\left[\log\left(1+\lambda x^{-\delta}\right)\right]^{\alpha-1}dx \tag{16}$$

substituting $t=x^{-\delta}$, the integral (16) can be rewritten as

$$E\left(X^k\right)=\frac{\lambda\beta^{\alpha}}{\Gamma(\alpha)}\int_0^{\infty}t^{-\frac{k}{\delta}}\left(1+\lambda t\right)^{-\beta-1}\left[\log\left(1+\lambda t\right)\right]^{\alpha-1}dx \tag{17}$$

If $\delta > k$, the direct application of Lemma 3.1 shows that (17) can be rewritten as

$$E\left(X^k\right)=\frac{(-1)^{\alpha-1}\beta^{\alpha}\lambda^{\frac{k}{\delta}}}{\Gamma(\alpha)}\frac{\partial^{\alpha-1}}{\partial(\beta+1)^{\alpha-1}}B\left(1-\frac{k}{\delta},\beta+\frac{k}{\delta}\right) \tag{18}$$

In particular,

$$E\left(X\right)=\frac{(-1)^{\alpha-1}\beta^{\alpha}\lambda^{\frac{1}{\delta}}}{\Gamma(\alpha)}\frac{\partial^{\alpha-1}}{\partial(\beta+1)^{\alpha-1}}B\left(1-\frac{1}{\delta},\beta+\frac{1}{\delta}\right) \tag{19}$$

$$E\left(X^2\right)=\frac{(-1)^{\alpha-1}\beta^{\alpha}\lambda^{\frac{2}{\delta}}}{\Gamma(\alpha)}\frac{\partial^{\alpha-1}}{\partial(\beta+1)^{\alpha-1}}B\left(1-\frac{2}{\delta},\beta+\frac{2}{\delta}\right) \tag{20}$$

$$E\left(X^3\right)=\frac{(-1)^{\alpha-1}\beta^{\alpha}\lambda^{\frac{3}{\delta}}}{\Gamma(\alpha)}\frac{\partial^{\alpha-1}}{\partial(\beta+1)^{\alpha-1}}B\left(1-\frac{3}{\delta},\beta+\frac{3}{\delta}\right) \tag{21}$$

$$E\left(X^4\right)=\frac{(-1)^{\alpha-1}\beta^{\alpha}\lambda^{\frac{4}{\delta}}}{\Gamma(\alpha)}\frac{\partial^{\alpha-1}}{\partial(\beta+1)^{\alpha-1}}B\left(1-\frac{4}{\delta},\beta+\frac{4}{\delta}\right) \tag{22}$$

The variance, skewness, and kurtosis measures can now be calculated using the relations

$$Var\left(X\right)=E\left(X^2\right)-E^2\left(X\right) \tag{23}$$

$$Skewness\left(X\right)=\frac{E\left(X^3\right)-3E\left(X\right)E^2\left(X\right)+2E^3\left(X\right)}{Var^{3/2}\left(X\right)} \tag{24}$$

$$Kurtosis\left(X\right)=\frac{E\left(X^4\right)-4E\left(X\right)E^3\left(X\right)+6E\left(X^2\right)E^2\left(X\right)-3E^4\left(X\right)}{Var^2\left(X\right)} \tag{25}$$

4 ESTIMATION

In this section, we consider estimation of the four parameters by method of moments and the maximum likelihood of the gamma-Dagum distribution. Let $x_1,\ldots,x_n$ be a random sample of size n from the GD distribution given by (10). Under the method of moments, equating $E\left(X^j\right)$ with the corresponding sample moment,

$$M_j=\frac{1}{n}\sum_{i=1}^{n}x_i^j,\quad j=1,\ldots,4 \tag{26}$$

respectively, one obtains the system of equations

$$\frac{(-1)^{\alpha-1}\beta^{\alpha}\lambda^{\frac{j}{\delta}}}{\Gamma(\alpha)}\frac{\partial^{\alpha-1}}{\partial(\beta+1)^{\alpha-1}}B\left(1-\frac{j}{\delta},\beta+\frac{j}{\delta}\right)=M_j,\quad j=1,\ldots,4 \tag{27}$$

which can be solved simultaneously to give estimates for α, β, λ and δ.

The log-likelihood for a random sample $x_1,\ldots,x_n$ from the GD distribution given by (10) is:

$$\log L(\alpha,\beta,\lambda,\delta)=n\log\lambda+n\log\delta+n\alpha\log\beta+(\alpha-1)\sum_{i=1}^{n}\log\left(\log\left(1+\lambda x_i^{-\delta}\right)\right)-$$
$$-(\delta+1)\sum_{i=1}^{n}\log x_i-n\log\Gamma(\alpha)-(\beta+1)\sum_{i=1}^{n}\log\left(1+\lambda x_i^{-\delta}\right) \tag{28}$$

The derivatives of this log-likelihood with respect to α, β, λ and δ are:

$$\frac{\partial\log L}{\partial\alpha}=n\log\beta+\sum_{i=1}^{n}\log\left(\log\left(1+\lambda x_i^{-\delta}\right)\right)-\psi(\alpha) \tag{29}$$

$$\frac{\partial\log L}{\partial\beta}=\frac{n\alpha}{\beta}-\beta\sum_{i=1}^{n}\log\left(1+\lambda x_i^{-\delta}\right) \tag{30}$$

$$\frac{\partial\log L}{\partial\lambda}=\frac{n}{\lambda}-\sum_{i=1}^{n}\frac{\lambda\delta(\alpha-1)x_i^{-\delta-1}}{\left(1+\lambda x_i^{-\delta}\right)\log\left(1+\lambda x_i^{-\delta}\right)}-(\beta+1)\sum_{i=1}^{n}\frac{x_i^{-\delta}}{1+\lambda x_i^{-\delta}} \tag{31}$$

$$\frac{\partial\log L}{\partial\delta}=\frac{n}{\delta}-\sum_{i=1}^{n}\frac{\lambda(\alpha-1)x_i^{-\delta}\log x_i}{\left(1+\lambda x_i^{-\delta}\right)\log\left(1+\lambda x_i^{-\delta}\right)}+\sum_{i=1}^{n}\frac{\lambda(\beta+1)x_i^{-\delta}\log x_i}{1+\lambda x_i^{-\delta}} \tag{32}$$

where $\psi(x)=d\log\Gamma(x)/dx$ is the digamma function. Setting these expressions to zero and solving them simultaneously yields the maximum-likelihood estimates of the four parameters.

5 APPLICATION

In this section, we use a real data set to compare the fits of the GD distribution (10) with three generalized distributions: beta-Dagum (BD), beta-Pareto (BP) and Pareto confluente hypergeometric (PCH) distributions. The data set given in Table 1 represents the failure times of the air conditioning system of an airplane reported in Linhart [6].

TABELA 1: Failure times of the air conditioning system of an airplane.

1	3	5	7	11	11	11	12	14	14
14	16	16	20	21	23	42	47	52	62
71	71	87	90	95	120	120	225	246	261

The maximum likelihood estimates (MLE) and the Akaike Information Criterion(AIC) values for the fitted distributions are reported in Table 2. The results show that the gamma-Dagum distribution provides a significantly better fit than the other three models.

Figure 2 display the probability plots and supports the results shown in Table 2. The fitted distributions superimposed to the histogram of the data in Figure 3 reinforce the results found for the gamma-Dagum.

TABELA 2: The MLE and AIC of the models based on data set.

Distribution	MLE	AIC
BD	$\widehat{a} = 2.97,\ \widehat{b} = 17.91,\ \widehat{\beta} = 2.20,\ \widehat{\delta} = 0.33,\ \widehat{\lambda} = 4.68$	312.93
BP	$\widehat{a} = 5.02,\ \widehat{b} = 2.90,\ \widehat{k} = 0.35,\ \widehat{s} = 0.55$	323.96
GD	$\widehat{\alpha} = 12.01,\ \widehat{\beta} = 12.49,\ \widehat{\delta} = 0.35,\ \widehat{\lambda} = 4.94$	311.16
PCH	$\widehat{a} = 5.16,\ \widehat{b} = 2.50,\ \widehat{c} = 0.24,\ \widehat{k} = 0.31,\ \widehat{s} = 0.55$	322.30

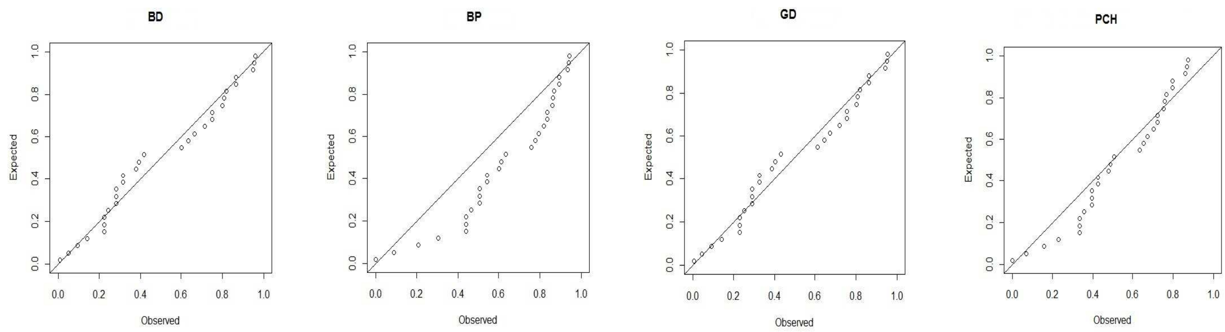

FIGURA 2: Probability plots from the fitted BD, BP, GD and PCH for the failure times of the air conditioning system data.

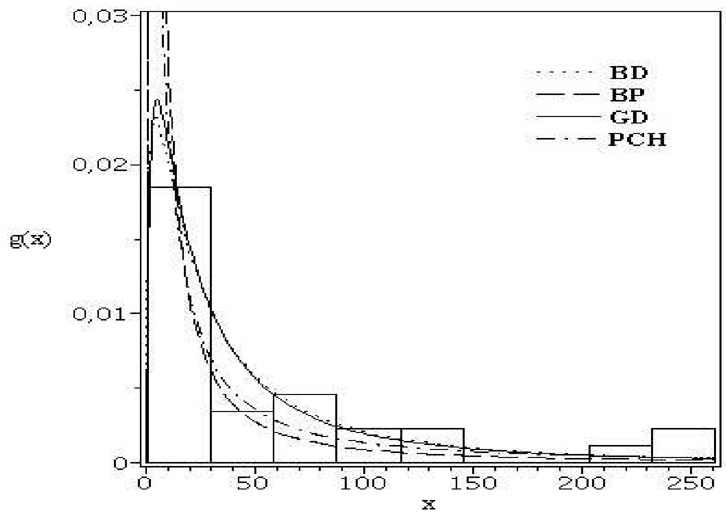

FIGURA 3: Estimated densities of the models from the fitted BD, BP, GD and PCH models for the failure times of the air conditioning system data.

6 CONCLUSION

This article defined the gamma-Dagum distribution. Several properties of the new distribution were investigated, including moments, failure rate function. The estimation of parameters by the method of moments and the maximum likelihood have been discussed.

An application of the gamma-Dagum distribution to real data show that the new distribution can be used quite effectively to provide better fits than the beta-Dagum, beta-Pareto and Pareto confluente hypergeometric distributions.

REFERÊNCIAS

[1] A. Akinsete, F. Famoye e C. Lee: *The beta-Pareto distribution*. Statistics, 42(6):547–563, 2008.

[2] A. Alzaatreh, F. Famoye e C. Lee: *Gamma-Pareto distribution and its applications*. J. Mod. Appl. Stat. Meth., 11(1):78–94, 2012.

[3] A. Alzaatreh, F. Famoye e C. Lee: *Weibull-Pareto distribution and its applications*. Commun. Stat. - Theory and Methods, 42(9):1673–1691, 2013.

[4] F. Condino, F. Domma e C. Lee: *The beta-Dagum distribution*. Commun. Stat. - Theory and Methods, 42(22):4070–4090, 2013.

[5] C. Lee, F. Famoye e O. Olumolade: *Beta-Weibull distribution*. J. Mod. Appl. Stat. Meth., 6(1):173–186, 2007.

[6] H. Linhart e W. Zucchini: *Model Selection*. Wiley, 1986.

[7] P. Paranaíba, E. Ortega, G. Cordeiro e R. Pescim: *The beta Burr XII distribution with application to lifetime data*. Commun. Stat. - Theory and Methods, 55(2):118–1136, 2011.

[8] M. de Pascoa, E. Ortega e G. Cordeiro: *The Kumaraswamy generalized gamma distribution with application in survival analysis*. Commun. Stat. - Theory and Methods, 8(5):411–433, 2011.

[9] L. Pinho, G. Cordeiro e J. Nobre: *The gamma-exponentiated Wiebull distribution*. J. Stat. Theory and Applications, 11(4):379–395, 2012.

[10] A. Prudnikov, Y. Brychkov e O. Marichev: *Integrals and series, vol. 1*. Gordon and Breach, 1998.

[11] M. Ristic e N. Balakrishnan: *The gamma-exponentiated exponential distribution*. J. Stat. Comp. Simulation, 82(8):1191–1206, 2012.

[12] J. Rodrigues, J. S. Filho e L. Chaves: *Pareto confluente hypergeometric distribution*. R. Eletrônica Matemática e Estatística em Foco, 1(1):24–30, 2013.

[13] G. Silva, E. Ortega e G. Cordeiro: *The beta modified Weibull distribution*. Math. and Stat., 16(3):409–430, 2010.