

The Beta-Dagum Distribution

Francesca Condino and Filippo Domma

Abstract In this paper, we propose a new distribution called Beta-Dagum distribution which includes some important distributions as special case, such as Dagum, Beta-Burr III and Beta-Fisk. We show that the probability density function can be expressed as a mixture of densities of Dagum random variables. This is important to obtain some mathematical properties of the new distribution. The usefulness of new distribution is illustrated through real data set of Bank of Italy on income and consumption.

Key words: Statistical Modeling, Beta-Generated Distribution, Maximum Likelihood Estimation

1 Introduction

The Dagum (1977, 1980) distribution has been used in studies of income and wage distribution as well as wealth distribution. In this context its features have been extensively analyzed by many authors, for excellent survey on the genesis and on empirical applications see Kleiber and Kotz (2003) and Kleiber (2008). The distribution function (*df*) and probability density function (*pdf*) of Dagum random variable (*rv*), respectively, are

$$G(x; \boldsymbol{\theta}) = \left(1 + \lambda x^{-\delta}\right)^{-\beta} \quad (1)$$

and

Francesca Condino
Institute of Neurological Science, CNR, Cosenza, Italy, e-mail: f.condino@isn.cnr.it

Filippo Domma
Department of Economics and Statistics, University of Calabria, Italy e-mail: f.domma@unical.it

$$g(x; \boldsymbol{\theta}) = \beta \lambda \delta x^{-\delta-1} \left(1 + \lambda x^{-\delta}\right)^{-\beta-1} \quad (2)$$

with $x > 0$, $\boldsymbol{\theta} = (\beta, \lambda, \delta)$, and $\beta > 0$, $\lambda > 0$ and $\delta > 0$. The parameter λ is a scale parameter, while β and δ are shape parameters. Throughout this paper, the Dagum distribution with parameters β , λ and δ will be denoted by $Da(\beta, \lambda, \delta)$. The Dagum distribution has positive asymmetry, it is unimodal for $\beta\delta > 1$ and zero-modal for $\beta\delta \leq 1$. Moreover, the q -th quantile of the Dagum distribution is

$$x(q) = \lambda^{\frac{1}{\delta}} (q^{-\frac{1}{\beta}} - 1)^{-\frac{1}{\delta}} \quad (3)$$

and the r -th moment is given by:

$$E(X^r; \beta, \lambda, \delta) = \beta \lambda^{\frac{r}{\delta}} B\left(\beta + \frac{r}{\delta}, 1 - \frac{r}{\delta}\right) \quad (4)$$

for $\delta > r$, where $B(.,.)$ is the mathematical function Beta.

An important characteristic of a Dagum distribution is that, according to the values of parameters, its hazard rate can be monotonically decreasing, upside-down bathtub and, finally, bathtub and then upside-down bathtub (Domma, 2002). This behavior has led several authors to study the model in different fields. In fact, recently, the Dagum distribution has been studied from a reliability point of view and used to analyse survival data (see Domma et al. (2009 a,b)). In the proposed model in this paper, it is possible to verify that the hazard rate it is more flexible than that of Dagum distribution.

In order to have greater flexibility in modeling observed data, Eugene et al. (2002) defined a class of beta-generated distribution as follows

$$F_G(x) = \frac{1}{B(a, b)} \int_0^{G(x)} z^{a-1} (1-z)^{b-1} dz \quad (5)$$

where $G(x)$ is the *df* of a rv X , with *pdf* $g(x)$. The probability density function corresponding to (5) is

$$f_G(x) = \frac{1}{B(a, b)} [G(x)]^{a-1} [1 - G(x)]^{b-1} g(x). \quad (6)$$

This class of generalized distributions has been receiving considerable attention over the last years. In particular, taking $G(x)$ to be the *df* of normal distribution, Eugene et al. (2002) defined and studied the Beta-Normal distribution, highlighting great flexibility in modeling not only symmetric heavy-tailed distribution, but also skewed and bimodal distribution, see also Razzaghi (2008). Akinsete et al. (2008) introduced the Beta-Pareto distribution, Barreto-Souza et al. (2009) proposed and studied the properties of the Beta-Generalized Exponential distribution and the Beta-Generalized Half-Normal distribution has been introduced by Pescim et al. (2010).

In this paper, we introduce the five parameters Beta-Dagum distribution using the class of distribution defined by Eugene et al. (2002). The new distribution can

be express as mixture of Dagum *rv*'s and shows greater flexibility of the Dagum distribution. Moreover, the proposed distribution gives a better fit to real data set regarding income and consumption than the Dagum distribution.

The paper is organized as follows. In *Section 2*, we define the Beta-Dagum distribution and we highlight some special cases. The moments, the hazard rate and the mean residual life of Beta-Dagum distribution are derived in *Section 3*. The point inequality measures most used in the literature related to the study of income distribution for the proposed model are described in *Section 4*. The maximum likelihood estimation are discussed in *Section 5* and an application on a real data set is reported in *Section 6*.

2 The Beta-Dagum Distribution

Replacing (1) in (5) we obtain a new distribution, called Beta-Dagum, with distribution function given by

$$F_G(x; \xi) = \frac{1}{B(a, b)} \int_0^{G(x; \theta)} z^{a-1} (1-z)^{b-1} dz \quad (7)$$

and probability density function

$$f_G(x; \xi) = \frac{1}{B(a, b)} [G(x; \theta)]^{a-1} [1 - G(x; \theta)]^{b-1} g(x; \theta). \quad (8)$$

Throughout this paper, the Beta-Dagum distribution with parameters $\xi = (\beta, \lambda, \delta, a, b)$ will be denoted by $BDa(\beta, \lambda, \delta, a, b)$.

Observed that, for $b > 0$ real non-integer and $|z| < 1$,

$$(1-z)^{b-1} = \sum_{j=0}^{\infty} \frac{(-1)^j \Gamma(b)}{\Gamma(b-j) j!} z^j,$$

we can express (7) as a mixture of Dagum distributions as follows

$$F_G(x; \xi) = \sum_{j=0}^{\infty} \frac{\Gamma(a+b)(-1)^j}{\Gamma(a)\Gamma(b-j)j!(a+j)} G(x; \theta)^{a+j} = \sum_{j=0}^{\infty} p_j G(x; \beta(a+j), \lambda, \delta) \quad (9)$$

given that $G(x; \theta)^{a+j} = (1 + \lambda x^{-\delta})^{-\beta(a+j)} = G(x; \beta(a+j), \lambda, \delta)$, where $p_j = \frac{\Gamma(a+b)(-1)^j}{\Gamma(a)\Gamma(b-j)j!(a+j)}$. The probability density function corresponding to equation (9) is

$$f_G(x; \xi) = \sum_{j=0}^{\infty} p_j g(x; \beta(a+j), \lambda, \delta). \quad (10)$$

If b is an integer then (7) and (8) become

$$F_G(x; \boldsymbol{\xi}) = \sum_{j=0}^{b-1} p_j G(x; \beta(a+j), \lambda, \delta) \quad , \quad f_G(x; \boldsymbol{\xi}) = \sum_{j=0}^{b-1} p_j g(x; \beta(a+j), \lambda, \delta).$$

Following Zografos and Balakrishnan (2009), we can simulate observations from a $B\text{Da}(\beta, \lambda, \delta, a, b)$ using the relationship: $x(B) = \lambda^{\frac{1}{\delta}} (B^{-\frac{1}{\beta}} - 1)^{-\frac{1}{\delta}}$, where B is a Beta random variable with parameters $a > 0$ and $b > 0$. By (9), we have the follows special cases: if $a = b = 1$ then X is a $\text{Da}(\beta, \lambda, \delta)$; if $a > 0$ and $b = 1$ then X is a $\text{Da}(a\beta, \lambda, \delta)$; if $a = 1$ and $b > 1$ then X is a $B\text{Da}(1, b, \beta, \lambda, \delta)$; if $\lambda = 1$ then X is a Beta-Burr III distribution and if $\beta = 1$ then X is a Beta-Fisk distribution.

3 Properties of Beta-Dagum rv

We can to obtain the r th moment of a Beta-Dagum random variable using the well-known properties of mixture distribution. In fact, keeping in mind the r th moment of a Dagum random variable, (4), it is simple to verify that

$$E[X^r; \boldsymbol{\xi}] = \sum_{j=0}^{\infty} p_j E[X^r; \beta(a+j), \lambda, \delta] = \sum_{j=0}^{\infty} p_j \beta(a+j) \lambda^{\frac{r}{\delta}} B\left(\beta(a+j) + \frac{r}{\delta}, 1 - \frac{r}{\delta}\right). \quad (11)$$

It is simple to verify that the r th incomplete moment for a Beta-Dagum distribution is given by

$$\begin{aligned} E_{X \leq x}[X^r; \boldsymbol{\xi}] &= \int_0^x t^r f(t) dt = \sum_{j=0}^{\infty} p_j \int_0^x t^r g(t; \beta(a+j), \lambda, \delta) dt \\ &= \sum_{j=0}^{\infty} p_j \beta(a+j) \lambda^{\frac{r}{\delta}} B\left(y^*; \beta(a+j) + \frac{r}{\delta}, 1 - \frac{r}{\delta}\right) \end{aligned}$$

where $y^* = (1 + \lambda x^{-\delta})^{-1} < 1$ and $B(z; q_1, q_2) = \int_0^z w^{q_1-1} (1-w)^{q_2-1} dw$. This result will be used in the coming study of the mean residual life and of the income inequality measures.

In order to calculate the hazard rate of the Beta-Dagum distribution, we observe that

$$F_G(x; \boldsymbol{\xi}) = \frac{B(G(x; \boldsymbol{\theta}); a, b)}{B(a, b)} = I_{G(x; \boldsymbol{\theta})}(a, b) \quad (12)$$

is the incomplete Beta function ratio. Now, let $I_{1-G(x; \boldsymbol{\theta})}(a, b) = \int_{1-G(x; \boldsymbol{\theta})}^1 z^{a-1} (1-z)^{b-1} dz$. Then it is immediate to verify that $1 = I_{G(x; \boldsymbol{\theta})}(a, b) + I_{1-G(x; \boldsymbol{\theta})}(a, b)$ and, therefore, we can write the survival function as follows: $S_G(x; \boldsymbol{\xi}) = 1 - F_G(x; \boldsymbol{\xi}) = I_{1-G(x; \boldsymbol{\theta})}(a, b)$. Consequently, the hazard function is given by

$$h_G(x; \boldsymbol{\xi}) = \frac{G(x; \boldsymbol{\theta})^{a-1} [1 - G(x; \boldsymbol{\theta})]^{b-1} g(x; \boldsymbol{\theta})}{B(a, b) I_{1-G(x; \boldsymbol{\theta})}(a, b)}. \quad (13)$$

The mean residual life, denoted by $\mu(x; \xi)$, is

$$\begin{aligned}\mu(x; \xi) &= E[X - x | X \geq x] = \frac{1}{S_G(x; \xi)} \int_x^\infty (u - x) f_G(u; \xi) du \\ &= \frac{E(X; \xi) - E_{X \leq x}(X; \xi)}{S_G(x; \xi)} - x.\end{aligned}$$

In the case of Beta-Dagum distribution, we have:

$$\mu_G(x; \xi) = \frac{\sum_{j=0}^\infty p_j \beta(a+j) \lambda^{\frac{1}{\delta}} C}{S_G(x; \xi)} - x.$$

where $C = [B(\beta(a+j) + \frac{1}{\delta}, 1 - \frac{1}{\delta}) - B(y^*; \beta(a+j) + \frac{1}{\delta}, 1 - \frac{1}{\delta})]$. It is clear that the hazard rate and the mean residual life are complicated function of x . However, from plots not reported in this work for save space purpose, we can state that $h_G(x; \xi)$ is much more flexible than the hazard rate of Dagum distribution.

4 Inequality Measures

In the literature related to study of income and wealth distribution, the most widely used point inequality measures are the Lorenz and the Bonferroni curves. Recently, Zenga (2007) proposed an alternative curve based on the ratios of lower and upper means. In this section, we derive the Lorenz, Bonferroni and Zenga curves for Beta-Dagum distribution.

The Lorenz, Bonferroni and Zenga curves are defined, respectively, by

$$L_F(x) = \frac{\int_0^x t f(t) dt}{E(X)} = \frac{E_{X \leq x}(X)}{E(X)}, \quad B(F(x)) = \frac{\int_0^x t f(t) dt}{F(x)E(X)} = \frac{E_{X \leq x}(X)}{F(x)E(X)} = \frac{L_F(x)}{F(x)}$$

and

$$A(x) = 1 - \frac{\mu^-(x)}{\mu^+(x)}$$

where $\mu^-(x) = \frac{\int_0^x t f(t) dt}{F(x)} = \frac{E_{X \leq x}(X)}{F(x)}$ and $\mu^+(x) = \frac{\int_x^\infty t f(t) dt}{1-F(x)} = \frac{E(X) - E_{X \leq x}(X)}{1-F(x)}$ are the lower and upper means, respectively. In the case of Beta-Dagum distribution, using the results of above paragraphs, we obtain

$$L_{FG}(x; \xi) = \frac{\sum_{j=0}^\infty p_j \beta(a+j) \lambda^{\frac{1}{\delta}} B(y^*; \beta(a+j) + \frac{1}{\delta}, 1 - \frac{1}{\delta})}{\sum_{j=0}^\infty p_j \beta(a+j) \lambda^{\frac{1}{\delta}} B(\beta(a+j) + \frac{1}{\delta}, 1 - \frac{1}{\delta})},$$

$$B(F_G(x; \boldsymbol{\xi})) = \frac{\sum_{j=0}^{\infty} p_j \beta(a+j) \lambda^{\frac{1}{\delta}} B(y^*; \beta(a+j) + \frac{1}{\delta}, 1 - \frac{1}{\delta})}{F_G(x; \boldsymbol{\xi}) \sum_{j=0}^{\infty} p_j \beta(a+j) \lambda^{\frac{1}{\delta}} B(\beta(a+j) + \frac{1}{\delta}, 1 - \frac{1}{\delta})}$$

and

$$A(x; \boldsymbol{\xi}) = 1 - \frac{S_G(x; \boldsymbol{\xi}) \sum_{j=0}^{\infty} p_j \beta(a+j) \lambda^{\frac{1}{\delta}} B(y^*; \beta(a+j) + \frac{1}{\delta}, 1 - \frac{1}{\delta})}{F_G(x; \boldsymbol{\xi}) \sum_{j=0}^{\infty} p_j \beta(a+j) \lambda^{\frac{1}{\delta}} C},$$

where $y^* = (1 + \lambda x^{-\delta})^{-1} < 1$.

5 Inference

In order to estimate the parameters $\boldsymbol{\xi} = (\beta, \lambda, \delta, a, b)$ of the Beta-Dagum distribution, we use the method of maximum likelihood. To such aim, we suppose that $\mathbf{x} = (x_1, \dots, x_n)$ is a random sample from $f_G(x; \boldsymbol{\xi})$. Then the log-likelihood function is:

$$\begin{aligned} \ell(\boldsymbol{\xi}; \mathbf{x}) = & -n \ln B(a, b) + n \ln(\beta \lambda \delta) - (\delta + 1) \sum_{i=1}^n \ln(x_i) + \\ & + (a + \frac{1}{\beta}) \sum_{i=1}^n \ln G(x_i; \boldsymbol{\theta}) + (b - 1) \sum_{i=1}^n \ln [1 - G(x_i; \boldsymbol{\theta})]. \end{aligned}$$

Differentiating $\ell(\boldsymbol{\xi}; \mathbf{x})$ with respect to β , λ , δ , a and b , respectively, and setting the results equal to zero, we obtain the system of MLE equations which does not admit any explicit solution, therefore the ML estimates $\hat{\boldsymbol{\xi}}_n = (\hat{\beta}_n, \hat{\lambda}_n, \hat{\delta}_n, \hat{a}_n, \hat{b}_n)$ can be obtained only by means of numerical procedures. Under the usual regularity conditions, the well-known asymptotic properties of the maximum likelihood method ensure that $\sqrt{n} (\hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}})$, where $\boldsymbol{\Sigma}_{\boldsymbol{\xi}} = [\mathbf{I}(\boldsymbol{\xi})]^{-1}$ is the asymptotic variance-covariance matrix and $\mathbf{I}(\boldsymbol{\xi})$ is the Fisher Information Matrix, whose elements are determined but not reported for brevity.

6 Application

In this section, we fit $BDa(\beta, \lambda, \delta, a, b)$ model to a real data set. The data, from Survey of Households Income and Wealth of Bank of Italy (2006), regard income and consumption of 7768 households. Furthermore, we compare the obtained results with the analogous obtained for $Da(\beta, \lambda, \delta)$ distribution. In Table 1 MLEs of the parameters, the maximized log-likelihood and the Akaike Information Criterion (AIC) for BDa and Da models are reported. The plot of the estimated densities for

Table 1 MLEs of *B**D**a* and *D**a* model parameters

	Consumption*		Income**	
	<i>B</i> <i>D</i> <i>a</i>	<i>D</i> <i>a</i>	<i>B</i> <i>D</i> <i>a</i>	<i>D</i> <i>a</i>
$\hat{\beta}$	1.805	1.145	4.746	0.783
$\hat{\lambda}$	4.370	8.499	15.970	27.686
$\hat{\delta}$	2.358	3.264	2.619	3.045
$\hat{a}$	1.005		0.201	
$\hat{b}$	1.660		1.310	
$l(\hat{\xi}, x)$	-11538.9	-11546.3	-14931.6	-14938.2
AIC	23087.75	23098.66	29873.12	29882.5

* LRT test: $-2\ln\Lambda(x) = 14.910$; (p -value = 5.79×10^{-4})

** LRT test: $-2\ln\Lambda(x) = 13.381$; (p -value = 1.243×10^{-3})

*B**D**a* and *D**a* distributions given in Figure 1 shows that the *B**D**a* distribution gives a better fit than the *D**a* one. The LR statistics confirm the superiority of the *B**D**a* model both for income and consumption data.

References

1. Akinsete, A., Famoye F. and Lee, C.: The beta-Pareto distribution. *Statistics*, **42:6**, pp. 547–563 (2008).
2. Barreto-Souza, W., Santos, A.H.S. and Cordeiro, G.M.: The beta generalized exponential distribution. *Journal of Statistical Computation and Simulation*, **80:2**, pp.159–172 (2009).
3. Dagum, C.: A new model of personal income distribution: specification and estimation. *Economie Appliquée*, **XXX**, pp. 413–437 (1977).
4. Dagum, C.: The generation and distribution of income, the Lorenz curve and the Gini ratio. *Economie Appliquée*, **XXXIII**, pp. 327–367 (1980).
5. Domma, F.: L'andamento della Hazard function nel modello di Dagum a tre parametri, *Quaderni di Statistica*, **4**, pp. 103–114 (2002).
6. Domma F., G. Latorre and M. Zenga: Reliability studies of Dagum distribution. *Submitted*(2009, a).
7. Domma, F., Giordano S. and Zenga M.: Maximum Likelihood Estimation in Dagum Distribution from Censored Samples. *Submitted*(2009, b).
8. Eugene N., Lee C., Famoye F.: Beta-Normal distribution and its application. *Communication in Statistics - Theory and Methods*, **4**, pp. 497–512, (2002).
9. Kleiber, C.: A guide to the Dagum distribution. In: Duangkamon, C. (ed.) *Modeling Income Distributions and Lorenz Curves Series: Economic Studies in Inequality, Social Exclusion and Well-Being*, **5**, Springer, New York (2008).
10. Kleiber, C. and Kotz, S.: *Statistical Size Distribution in Economics and Actuarial Sciences*. John Wiley & Sons, Inc (2003).
11. Pescim R.R., Demétrio C.G.B., Cordeiro G.M., Ortega E.M.M., Urbano M.R.: The beta generalized half-normal distribution. *Computational Statistics and Data Analysis*, **54**, pp. 945–957 (2010).

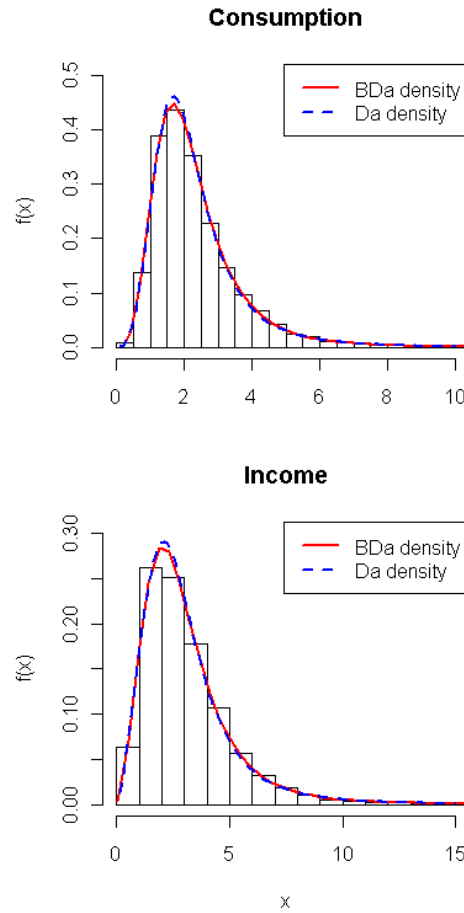


Fig. 1 Empirical and fitted distributions for Consumption and Income data

12. Razzaghi M.: Beta-normal distribution in dose-response modeling and risk assessment for quantitative responses. *Environ. Ecol. Stat.*, **16**, pp. 25–36 (2008).
13. Zagrafos K. and Balakrishnan N.: On families of beta- and generalized gamma-generated distributions and associate inference. *Statistical Methodology*, **6**, 344–362 (2009).
14. Zenga M.: Inequality curve and inequality index based on the ratios between lower and upper arithmetic means. *Statistica & Applicazioni*, **5**, 3–27 (2007).