

The Distribution of the Ratio of Two Independent Dagum Random Variables

Angiola Pollastri, Giovanni Zambruno

Abstract. In this paper we propose an estimation procedure for the distribution of the ratio between two independent Dagum random variables. Such an issue is of remarkable importance when analyzing the characteristics of ratios of economic variables which can be described by the Dagum model. The distribution and density functions are computed via numerical procedures; a numerical method is also proposed in order to make the distribution computation easier and faster. Finally, some empirical investigations are reported in order to establish the model effectiveness, and an application is presented concerning the estimation of the distribution of the ratio between the expenditures of a 2-member and a 1-member household, based on the Banca d'Italia 2006 survey .

Keywords: Dagum distribution, ratio of independent Dagum r.v.'s, distribution of the ratio of expenditures of households differing in size, estimation of distribution functions.

1 Introduction

In this paper we analyze some distributional characteristics of the ratio of two independent Dagum r.v.'s with three parameters.

The model proposed by Dagum fulfils many properties considered relevant for an income distribution model: model specifications exploit the economic framework, the convergence to the Pareto law and the economic significance of the parameters. In the present paper, the choice of the Dagum model is also supported by the fact that it provides a good fit to both extreme sides of the observed income distribution in Italy

Angiola Pollastri, Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali, Università di Milano Bicocca, e-mail:
angiola.pollastri@unimib.it

Giovanni Zambruno, Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali, Università di Milano Bicocca, e-mail:
giovanni.zambruno@unimib.it

(Latorre, 1989). The model has been successfully used also to describe the size distribution of business firms (Bisante, Fiori, 2009).

The importance of this study is tied to the distribution of ratios between two economic variables.

Some empirical studies confirm the validity of the model proposed.

An application regarding the estimation of the distribution function of the ratio of r.v.'s describing the expenditures of families with different number of components is also considered.

2 The Theoretical Framework

Let X be a r.v. distributed as Dagum type I (see Dagum (1977, 1990), Dancelli (1986)). Its cumulative distribution function (cdf) is

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

Our purpose is to obtain the distribution function of the r.v.

$$U = \frac{X}{Y}$$

defined as the ratio of two r.v.'s X and Y both following a Dagum distribution with parameters:

$$X \approx D(\beta_1, \delta_1, \lambda_1), Y \approx D(\beta_2, \delta_2, \lambda_2)$$

We will develop our analysis under the assumption that these two r.v.'s are independent. Following Mood, Graybill and Boes [1974, p. 187] the density function of the variable U is

$$f_U(u) = \int_{-\infty}^{+\infty} |y| f_{X,Y}(uy, y) dy$$

which, according to the independence assumption, and accounting for the support of Y , becomes

$$f_U(u) = \int_0^{+\infty} y f_X(uy) f_Y(y) dy$$

Substituting for the expressions of the Dagum densities yields

$$f_U(u) = \int_0^{+\infty} y \left\{ \beta_1 \lambda_1 (uy)^{-\delta_1-1} \left[1 + \lambda_1 (uy)^{-\delta_1} \right]^{-\beta_1-1} \right\} \left\{ \beta_2 \lambda_2 (y)^{-\delta_2-1} \left[1 + \lambda_2 (y)^{-\delta_2} \right]^{-\beta_2-1} \right\} dy$$

and the cumulative probability distribution, after some minor rearrangements, takes the expression

$$\begin{aligned}
F_U(u) &= \Pr[U \leq u] = \\
&= \beta_1 \beta_2 \lambda_1 \lambda_2 \int_0^u t^{-\delta_1-1} \int_0^{+\infty} y^{-\delta_1-\delta_2-1} \left[1 + \lambda_1(ty)^{-\delta_1}\right]^{-\beta_1-1} \left[1 + \lambda_2(y)^{-\delta_2}\right]^{-\beta_2-1} dy dt
\end{aligned}$$

In plain words, the above expression can be read as follows. Apart from some constants left aside, we assign a fixed value t to the r.v. U and sum the probabilities of all values of y and the corresponding $x=ty$ such that their ratio equals t . Then the outer integral provides an estimate of the probability that the ratio takes a value not exceeding the upper bound u .

The expression in the inner integral doesn't match any form reported in the currently available tables (e.g. Gradshteyn – Ryzhik 1994), nor does a change in variable appear to be suitable in order to reduce it to known forms: therefore a numeric quadrature is advisable. As is well-known, the goodness of the approximation depends largely on the more or less appropriate choice of the points where the integrand must be evaluated: on one side, the number of points should be kept reasonably low for computational efficiency, on the other their location should span the whole integration interval, with a higher “crowding” around the most significant values.

However there is a substantial difference in the two integrals to evaluate. The inner integrand admits an analytical representation: therefore one can efficiently use any of the numeric quadrature subroutines available. In this case we have used the subroutine QDAGI of the IMSL library, which reportedly is designed to accommodate unbounded integration intervals. In contrast, the outer integral's argument is expressed in tabular form, although the user has some choice on the most effective values of t . For this purpose, we observe that values close to either support boundary have an extremely low probability of occurrence (and even more so the upper one), while we expect that a considerable probability is concentrated around the value $v = E(X)/E(Y)$ (of course, this value has the only purpose of providing a very rough approximation for the mean of the ratio, whose value we do not know yet).

Therefore a good choice for the sequence of points

$$\{t_1, t_2, \dots, t_n\}$$

where to evaluate the inner integral would be, for instance, to have them fairly clustered around the value v and more and more scattered as long as we move away from it. A way to produce this result would be, for instance, to consider a function defined on $(0,1)$, continuous and strictly increasing, with vertical asymptotes at both domain boundaries, and one inflexion point: a good example of it is $x \rightarrow g(x) = -\ln^{-1}(x)$. Then take a suitable sequence of equally spaced values $p_i, i=1, \dots, n$ in $(0,1)$ with p_1 as close as possible to 0 and p_n to 1; finally define

$$t_i = g(\kappa p_i)$$

where the scale coefficient κ is such that the inflexion point of g corresponds to the value v .

These values are then fed into each inner integral, whose value is in turn an input for the outer one, which is computed through the standard trapezoidal rule.

An example of this procedure is presented in the next section.

3 Some Applications

The purpose of this section is to illustrate, by way of simple applications based on real data, how the method now proposed performs.

Before starting, we want to ascertain whether our methods works in replicating the true distribution of observations: in other words, in this preliminary step we want to avoid all problems related with parameters estimation.

For this purpose we consider two Dagum distributions with fixed parameters ($\lambda_1 = 0.7, \beta_1 = 1.2, \delta_1 = 4$) and resp. ($\lambda_2 = 0.9, \beta_2 = 1.4, \delta_2 = 6$).

From the (0,1) uniform distribution we draw at random two series, of 800 numbers each: call them p_i and q_j ($i, j = 1, \dots, 800$). Then we compute the numbers x_{p_i}, y_{q_j} interpreted as the p_i and resp. q_j quantiles of the two Dagum, that is

$$x_{p_i} = \lambda_1 (p_i^{-1/\beta_1} - 1)^{-1/\delta_1}, y_{q_j} = \lambda_2 (q_j^{-1/\beta_2} - 1)^{-1/\delta_2}$$

Next, we form all possible ratios that can be obtained by taking any x_{p_i} as the numerator and any y_{q_j} as the denominator. In this way we obtain a series of $800 \times 800 = 640000$ observations from which we can form a frequency distribution. This is plotted in Fig. 1 together with the distribution obtained by running the procedure presented earlier, with the same set of parameters. The close matching is quite apparent.

Then we have computed the two Cumulative Distribution Function reported in Fig.2. Also in this case the two distributions are very similar.

The Data Set

A very important source of microdata about expenditure, income and wealth in Italy is provided by the Banca d'Italia survey which consists of a series of interviews. The sampling unit is the household and the survey population is the whole set of households dwelling in Italy. In the 2006 survey, the sample size was 7768 households. Each family is randomly drawn from a two stage sample. In the questionnaire there is also an item regarding the average monthly expenditure on all kinds of consumption. Our goal is to analyze such data.

Figure 1: Empirical and computed density functions of ratios drawn from Dagum with parameters $(\lambda_1 = 0.7, \beta_1 = 1.2, \delta_1 = 4)$ and $(\lambda_2 = 0.9, \beta_2 = 1.4, \delta_2 = 6)$

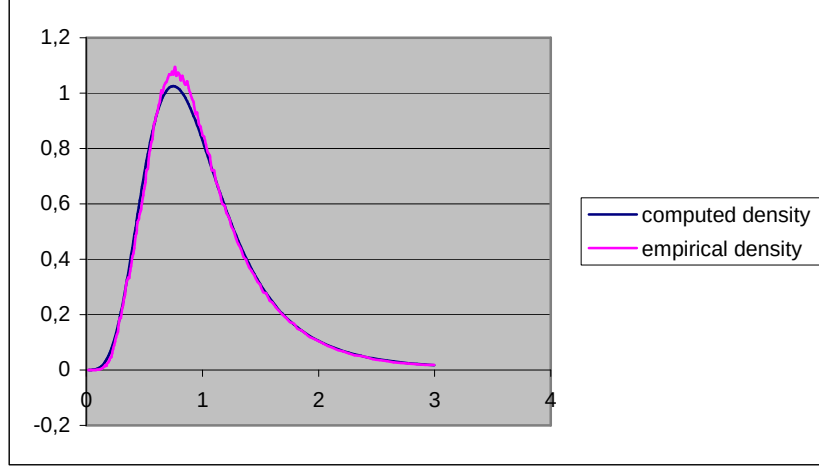
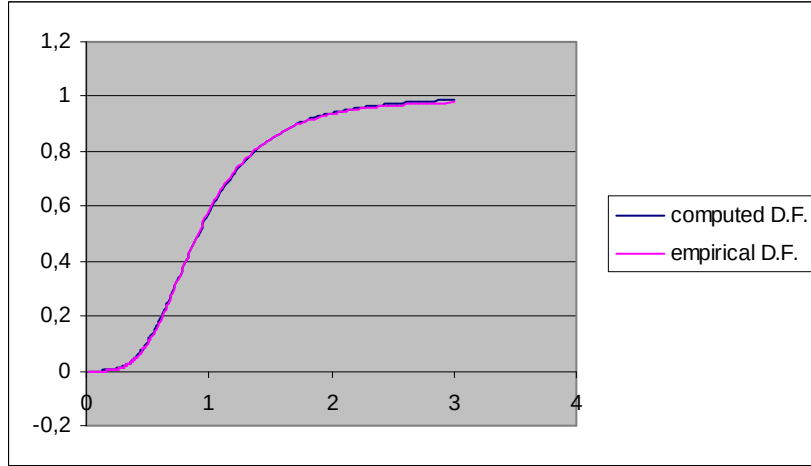


Figure 2: Empirical and computed c.d.f. of ratios drawn from Dagum with parameters $(\lambda_1 = 0.7, \beta_1 = 1.2, \delta_1 = 4)$ and $(\lambda_2 = 0.9, \beta_2 = 1.4, \delta_2 = 6)$



Indeed, for each household size we have collected data on income, whence we can fit a Dagum distribution by estimating the relevant parameters.

Let the r. v. X_r ($r=1, 2, \dots$) describe the income of a household with r members. Assume that each X_r is distributed as a Dagum. Therefore we can estimate the parameters λ_r , β_r , δ_r .

Now consider the r. v. X_r / X_1 as the ratio of two Dagum r.v.'s. It is possible to find the numerical distribution function of this r. v. which represents the result of the experiment of selecting at random a household of one member and a household of r members. The r. v. X_r is independent from the r. v. X_1 . Subsequently we have generated the distribution of all possible ratios of the expenditures of the households with r components and those with one component only.

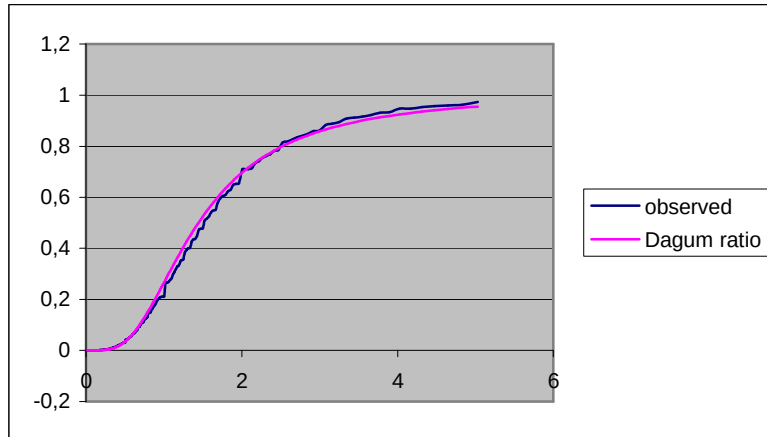
In the present study we estimate the Distribution Function of the ratios of the expenditures of all the households with two components divided by every expenditure of the household with one component. Relative to 2006 the sample size of the households with 2 members is $n_2=2366$ and the sample size for the household of a single component is $n_1=1327$.

In order to obtain the c.d.f. in question, we estimated the parameters of the Dagum distribution for the expenditures of the households with one component and that of the households with two components. We obtained the estimates of the sets of parameters of the variables X_1 and X_2 using the minimum Chi-Square method (see, e.g., Kendall et al., 1973). Through numerical computations, we get the following: $(\hat{\lambda}_N = 0.9, \hat{\beta}_N = 1.24, \hat{\delta}_N = 3.98)$ are the minimum Chi-Square estimates of the parameters of the r.v. X_2 and $(\hat{\lambda}_D = 0.93, \hat{\beta}_D = 0.4, \hat{\delta}_D = 5.59)$ the estimates of the parameters of the r.v. X_1 .

We have estimated the c.d.f. of the variable X_2 / X_1 with the method exposed before. Then, taking all the possible values of the observed expenditures of the household with two members indicated by x_{2i} ($i=1, \dots, 2366$) and all the possible values of the observed expenditures of the household with a single component indicated by x_{1j} ($j=1, \dots, 1327$), we have built the empirical c.d.f. of the ratios x_{2i} / x_{1j} .

In Fig.3 the empirical and theoretical c.d.f. are drawn, they overlap quite closely.

Figure 3: c.d.f. of X_2 / X_1 as the ratio of two Dagum r.v.'s and as estimated from the observed ratios



It is also possible to evaluate the deciles. They are reported in Table 1. The deciles describe many characteristics of the ratio. For instance we can establish that the

estimate of the median of the ratio of expenditures of a household with two component and a household with only one component is 1.44257.

Table 1: Deciles of the ratio of the expenditures of households of two and one components

1	Deciles
1	0.68799
2	0.88385
3	1.05890
4	1.22824
5	1.44257
6	1.68913
7	2.01885
8	2.52895
9	3.55029

4 Conclusions

The present study is a first proposal for the estimation of the distribution of the ratio of two independent r.v.'s having Dagum distribution. The main purpose is to study the distribution of the ratio of two economic variables, often used in economic indexes.

In carrying on this analysis we took some benefit from an improvement of the numerical integration method, consisting in a rule to build the integration grid specifically for this particular situation. This offered the possibility to estimate the percentiles quite easily, and also to compare the estimated distribution with the one computed directly by drawing a sample from actual data.

This technique may prove useful in making inference on a number of economic and financial indexes, whenever they are defined as ratios of two independent random variables distributed according to Dagum.

References

1. BANCA D'ITALIA (2008) I bilanci delle famiglie italiane nell'anno 2006, *Supplementi al Bollettino Statistico*, XVII, Centro Stampa Banca d'Italia, Roma.
2. Bisante E., Fiori A.M. (2009) Firm size distribution e modello di Dagum: un'indagine empirica sull'industria meccanica italiana, *Working paper n. 181*, Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali, Università di Milano Bicocca, Milano.

3. Burr I. W. (1942) Cumulative Frequency Functions, *Annals of Mathematical Statistics*, Vol. 13, pp. 215-232.
4. Dagum C. (1977) A new model for personal income distribution: specification and estimation, *Economie Appliquée*, 30, pp. 413- 437.
5. Dagum C. (1990) Generation and Properties of Income Distribution Functions, *Studies in Contemporary Economics. Income and Wealth Distribution, Inequality and Poverty*, C. Dagum, M. Zenga (Eds), Springer-Verlag, Berlin.
6. Dancelli L.(1986) Tendenza alla massima ed alla minima concentrazione nel modello di distribuzione del reddito personale di Dagum, *Scritti in onore di Francesco Brambilla*, Vol. I, Edizioni di «Bocconi Comunicazione», Milano.
7. Gradshteyn I.S., Ryzhik I.M. (1994) *Table of Integrals, Series, and Products*, Academic Press, Boston.
8. Kendall M.G., Stuart A.(1973) *The Advanced Theory of Statistics*, C. Griffin &Co., London.
9. Kot S.M. (2002a) The Estimation of the Social Welfare Functions, Inequality Aversion, and Equivalence Scale. *International Workshop 'Income Distribution and Welfare'*, May 30th – June 1st, Università Bocconi, Milano, Italy.
10. Kot S.M. (2002b) On the estimation and calibration of the social welfare function, *Quality of Life Research*, W. Ostasiewicz (Ed.), Chapter 4, pp. 61-71, Yang's Scientific Press, Tucson (USA).
11. Latorre G. (1988) Proprietà Campionarie del Modello di Dagum per la distribuzione dei redditi, *Statistica*, XLVIII, n. 1-2, pp. 15-27.
12. Latorre G. (1989) Asymptotic Distributions of Indices of Concentration: Empirical Verification and Application, in: *Studies in Contemporary Economics. Income and Wealth Distribution, Inequality and Poverty*, C. Dagum, M. Zenga (Eds), Springer-Verlag, Berlin.
13. Mood A. M., Graybill F. G., Boes D. C. (1974) *Introduction to the theory of Statistics*, Wiley, New York.
14. Pollastri A. (2003) Scale di equivalenza tramite l'impiego della distribuzione di Dagum, *Working paper n. 62*, Dipartimento di Metodi Quantitativi per le Scienze Economiche e Aziendali, Università di Milano-Bicocca.
15. Pollastri, A. (2007) Estimation of equivalence scales in Italy based on income distribution. *Statistica & Applicazioni*, V(2), 131-140.