Physica A 392 (2013) 1398–1408

Contents lists available at SciVerse ScienceDirect

# Physica A

journal homepage: www.elsevier.com/locate/physa

# About the modified Gaussian family of income distributions with applications to individual incomes

José María Sarabia *, Faustino Prieto, Carmen Trueba, Vanesa Jordá

*Department of Economics, University of Cantabria, 39005 Santander, Spain*

ARTICLE INFO

*Article history:*
Received 17 August 2012
Received in revised form 2 November 2012
Available online 23 November 2012

*Keywords:*
Income distribution
First-degree stochastic dominance
Income inequality
Gini index
Individual incomes

ABSTRACT

In a recent paper in this journal [Q. Guo, L. Gao, Distribution of individual incomes in China between 1992 and 2009, Physica A 391 (2012) 5139–5145], a new family of distributions for modeling individual incomes in China was proposed. This family is the so-called Modified Gaussian (MG) distribution, which depends on two parameters. The MG distribution shows a satisfactory fit for the individual income data between 1992 and 2009. However, for the practical use of this model with individual incomes, it is necessary to know its probabilistic and statistical properties, especially the corresponding inequality measures. In this paper, probabilistic functions and inequality measures of the MG distribution are obtained in closed form, including the normalizing constant, probability functions, moments, first-degree stochastic dominance conditions, relationships with other families of distributions and standard tools for inequality measurement (Lorenz and generalized Lorenz curves and Gini, Donaldson–Weymark–Kakwani and Pietra indices). Several methods for parameter estimation are also discussed. In order to illustrate all the previous formulations, we have fitted individual incomes of Spain for three years using the European community household panel survey, concluding a static pattern of inequality, since the Gini index and other inequality measures remain constant over the study period.



## 1. Introduction

The modeling of data from income and wealth distributions is one of the central research topics in econophysics (see Refs. [1–4]).

Since Pareto's (1897) work, the list of probability distribution functions for modeling income and wealth distributions has increased considerably. This list includes classical distributions such as the log-normal, gamma, beta, Singh–Maddala, Mandelbrot, Pareto and generalized versions of each. A comprehensive survey of these distributions can be found in Refs. [5,6]. Other relevant parametric models have also been recently proposed. These new models include the $\kappa$-generalized distribution (see Ref. [7]), the Gompertz–Pareto income distribution (see Refs. [8,9]) and the Pareto Positive Stable distribution (see Ref. [10]). Typically economical systems (but also several physical systems) present power-law tails (see for instance [11]), and many of these families present this kind of tail.

One of the most important advantages of all these parametric models is that the main probabilistic measures (e.g. moments) and inequality tools (e.g. Gini index) are available in closed form. This fact provides a correct description of the parametric family of income and wealth distributions and allows us to compute all these indicators in an exact form (see Refs. [12,13]).

In a recent paper in this journal (see Ref. [14]), a new family of distributions for modeling individual incomes in China was proposed. This family is the so-called modified Gaussian (MG) distribution, which depends on two parameters. The

* Corresponding author. Tel.: +34 942201635; fax: +34 942201603.
*E-mail address:* sarabiaj@unican.es (J.M. Sarabia).

0378-4371/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2012.11.025

MG distribution shows a satisfactory fit for the individual income data between 1992 and 2009. However, for the practical use of this model, it is necessary to know its probabilistic and statistical properties, especially the corresponding inequality measures. In this paper, probabilistic functions and inequality measures of the MG distribution are obtained in a closed form, including the normalizing constant, probability functions, moments and standard tools for inequality measurement. Several methods for parameter estimation are also discussed. In order to illustrate all the previous formulations, we have fitted individual incomes of Spain for three years using the European community household panel survey, concluding a static pattern of inequality, since the Gini index and other inequality measures remain constant over the study period.

The contents of this paper are as follows. In Section 2 we present the probabilistic properties of the MG distribution: the normalizing constant, a simple interpretation in terms of weighted distributions, the cumulative distribution, survival and quantile functions, moments and related quantities, first-degree stochastic dominance conditions and the relationships with other families of distributions (chi-square, stretched exponential and Weibull distributions). The different tools for inequality measurement (Lorenz curve, generalized Lorenz curve, Gini index, Donaldson–Weymark–Kakwani index and Pietra index) are obtained in Section 3. Estimation methods (moments and maximum likelihood) are discussed in Section 4. An empirical application with individual incomes of Spain for three years using the European community household panel survey is included in Section 5. Some conclusions are given in Section 6.

## 2. The modified Gaussian distribution

According to Guo and Gao in Ref. [14], their distribution is composed of two factors. The first factor is the variable factor $(x-\mu)$ if $x \geq \mu$ and the second factor is related to the planned economic system income, which is $\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$. Then, the modified Gaussian distribution (see Ref. [14]) is defined in terms of the probability density function (PDF) by,

$$f(x;\mu,\sigma) = K(x-\mu)\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \quad x \geq \mu \tag{1}$$

and $f(x;\mu,\sigma) = 0$ if $x < \mu$, where $\mu, \sigma > 0$ are parameters and $K$ is the normalizing constant.

### 2.1. The normalizing constant

Since (see the Appendix),

$$\int_{-\infty}^{\infty} f(x;\mu,\sigma)\mathrm{d}x = K\sigma^2,$$

the value of the normalizing constant is $K = \frac{1}{\sigma^2}$.

### 2.2. Interpretation of the MG distribution

The PDF of the MG distribution defined in Eq. (1) can be seen as a weighted distribution of the form

$$f_w(x) = \frac{w(x)f(x)}{E_f[w(X)]},$$

where $f(x)$ is the PDF of the classical Gaussian distribution with mean $\mu$ and standard deviation $\sigma$, and $w(x)$ is the weighted function defined as $w(x) = (x-\mu)$ if $x \geq \mu$ and $w(x) = 0$ if $x < \mu$.

The new PDF $f_w(x)$ is called the weighted version of $X$, and its distribution in relation to that of $X$ is called the weighted distribution with weight function $w$. In our case, because $w(x)$ is linear $f_w(x)$ is called the length-biased or size-biased version of $f$, and the corresponding observational mechanism is called length- or size-biased sampling (see Refs. [15,16]). In the case of income distributions, this mechanism provides different weights to the different incomes.

### 2.3. Cumulative distribution, survival and quantile functions

The cumulative distribution function (CDF) is defined by $F(x) = \Pr(X \leq x)$. Then $F(x) = 0$ is $x < \mu$. If $x \geq \mu$ we have

$$\begin{aligned} F(x) &= \Pr(X \leq x) \\ &= \int_{\mu}^{x}\left(\frac{t-\mu}{\sigma^2}\right)\exp\left\{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right\}\mathrm{d}t \\ &= 1-\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}. \end{aligned} \tag{2}$$