Entropy maximization under the constraints on the generalized Gini index and its application in modeling income distributions

A. Khosravi Tanak, G.R. Mohtashami Borzadaran *, J. Ahmadi

Department of Statistics, Ferdowsi University of Mashhad, P.O. Box 1159 Mashhad 91775, Iran

HIGHLIGHTS

- The maximum entropy distribution with a given mean and generalized Gini index is found.
- The obtained and GB2 family distributions are fitted to the US family income data.
- Some of the obtained distributions provided better fits to the data.

ARTICLE INFO

ABSTRACT

In economics and social sciences, the inequality measures such as Gini index, Pietra index etc., are commonly used to measure the statistical dispersion. There is a generalization of Gini index which includes it as special case. In this paper, we use principle of maximum entropy to approximate the model of income distribution with a given mean and generalized Gini index. Many distributions have been used as descriptive models for the distribution of income. The most widely known of these models are the generalized beta of second kind and its subclass distributions. The obtained maximum entropy distributions are fitted to the US family total money income in 2009, 2011 and 2013 and their relative performances with respect to generalized beta of second kind family are compared.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

In economics and social sciences, approximating income distribution with regard to income inequality in society is of interest. Inequality can be defined as the dispersion of the distribution of income or some other welfare indicator. There are various ways to measure inequality. The Lorenz curve developed by Lorenz [1] is perhaps the most fundamental tool used to measure income inequality. Graphically, the Lorenz curve gives the proportion of total societal income accruing to the lowest earning proportion of income earners. Let X denote a random variable with cumulative distribution function (cdf) F supported in $(0, \infty)$ and mean $E(X) = \mu$. The Lorenz curve is defined as follows:

$$L_F(u) = \frac{1}{\mu} \int_0^u F^{-1}(x)\mathrm{d}x, \quad 0 \le u \le 1, \tag{1}$$

where $F^{-1}(x) = \inf\{t : F(t) \ge x\}$. If F is income distribution, then $L_F(u)$ denotes the fraction of total income which is in the hands of the uth fraction of population possessing the lowest income. This representation forms the basis of many common

---

* Corresponding author.
    E-mail addresses: khosravi.a_66@yahoo.com (A. Khosravi Tanak), grmohtashami@um.ac.ir (G.R. Mohtashami Borzadaran), ahmadi-j@um.ac.ir (J. Ahmadi).

inequality measures; among them the Gini index which was proposed by Gini [2] is a famous and well-known measure. The Gini index, $G(F)$, is defined as twice the area between the considered Lorenz curve and the line of perfect equality $L_F(u) = u$,

$$G(F) = 2 \int_0^1 (u - L_F(u))du = 1 - 2 \int_0^1 L_F(u)du. \tag{2}$$

The Gini index takes values in the unit interval [0, 1]. A low Gini index indicates more equal income distribution, while a high Gini index indicates more unequal distribution. Various generalizations of the Gini index have already been suggested in the literature. Kakwani [3], Donaldson and Weymark [4,5], and Yitzhaki [6] proposed a family of generalized Gini indices by introducing different weighting functions for the area under the Lorenz curve

$$G_\nu(F) = 1 - \int_0^1 \nu(\nu - 1)(1 - u)^{\nu-2} L_F(u)du, \quad \nu > 1. \tag{3}$$

In the case of $\nu = 2$, we have the Gini index. When ν increases, higher weights are attached to small incomes. Statistical inference based on generalized Gini index has been studied by several researchers. For example, Kleiber and Kotz [7] showed that the income distributions can be characterized by the sequence of the associated generalized Gini indices (when ν takes integer values), provided the mean is finite. The asymptotic properties of this index were developed by Barrett and Pendakur [8] and Zitikis and Gastwirth [9]. Xu [10] obtained confidence intervals for generalized Gini index using the bootstrap method. Pietra [11] proposed another inequality measure that is most useful and appropriate in the case of asymmetric and skewed probability distributions. The Pietra index, $P(F)$, is defined as the maximal vertical deviation between the Lorenz curve and the equality line,

$$P(F) = \max_{0 \leq u \leq 1} \{u - L_F(u)\}. \tag{4}$$

See Ref. [12] and references therein for more details about inequality measures.

Suppose that X is a random variable having a continuous cdf F with probability density function (pdf) f. The basic uncertainty measure for distribution F is defined as

$$H(f) = - \int_{-\infty}^{\infty} f(x) \ln f(x)dx,$$

provided the integral exists. It was originally introduced by Shannon [13]. In the literature, $H(f)$ is often referred to as the entropy of X or Shannon's information about F. We refer the reader to Verdu [14], Cover and Thomas [15] and references therein for more details on the theory, extension and applications of $H(f)$. Approximation of distributions is a fundamental problem in statistical data analysis. The maximum entropy principle proposed by Jaynes [16], gives us a general way of approximating a distribution. According to this principle, the best approach is to ensure that the approximation satisfies any constraints on the unknown distribution that we are aware of, and that subject to those constraints, the distribution should have maximum entropy. The problem of maximizing entropy subject to some constraint such as moments has been studied by many authors. For example, see Refs. [17,18]. Recently, some works have been done in the subject of entropy maximization under the constraints on the inequality measures. Eliazar and Sokolov [19] found the distribution that maximizes entropy subject to a given mean and Gini index. Also, Eliazar and Sokolov [20] obtained the distribution that maximizes entropy subject to a given mean and Pietra index. In this paper, we intend to develop their results in terms of generalized Gini index.

Throughout the paper, we consider non-negative income random variables. The rest of the paper is as follows: Section 2 contains some preliminaries and the basic tools which will be used in the next sections. In Section 3, we review some results on entropy maximization under some moment constraints and inequality measures constraints. Section 4 is devoted to our result in maximization of entropy with a given mean and generalized Gini index. In Section 5, we compare maximum entropy distributions under the constraints on the generalized Gini index with alternative distributions from generalized beta family in fitting to the US family income data in 2009, 2011 and 2013.

2. Basis tools

One of the most well-known integral functionals that has been studied in variational calculus is the Lagrange functional

$$L(y) = \int_a^b G(y(x), y'(x), x)dx, \tag{5}$$

where the given function G is continuous and has continuous first partial derivatives in each of its arguments. The basic variational problem can be stated as follows: "Find the curve $y = y(x)$ for which the functional $L(y)$ has an extremum". There is a general solution which is stated in the following theorem.

Theorem 1 (*Gelfand and Fomin [21, p. 15]*)**.** *Let $L(y)$ be a functional of the form* (5) *defined on the set of functions $y(x)$ which have continuous first derivatives in $[a, b]$ and satisfy the boundary conditions $y(a) = A$, $y(b) = B$. Then, a necessary condition*