A vector valued Bivariate Gini index for truncated distributions

E. I. Abdul-Sathar¹, R. P. Suresh¹, K. R. Muraleedharan Nair²

¹ Indian Institute of Management Kozhikode, Kunnamangalam P.O., Calicut - 673 571, India; (e-mail: rps@iimk.ac.in)
² Department of Statistics, Cochin University of Science and Technology, Cochin - 682 022, India

Received: April 1, 2004; revised version: September 15, 2005

Abstract Gini index is widely used in the study of inequality of income distribution. In the present paper we give a definition of the Gini index in the Bivariate set-up and look into the problem of characterizing probability distributions based on some relationship between this index and various other commonly used measures. We also generalized the Gini index to a situation where several attributes of the population are considered.

Key words Disparity measurement, Gini index, Vector valued failure rate, Truncated distributions, Characterization of probability distributions

1 Introduction

To measure the inequality of an income distribution the Gini mean difference and its scale invariant version, the Gini index, are most widely used. The Gini index is closely connected to the Lorenz curve. It amounts to twice the area between the Lorenz curve and the diagonal of the unit square. In other words, it equals the area between the Lorenz curve and its dual. Many authors have stressed the need for including more than one attribute in the analysis of economic inequality. The works of Atkinson and Bourguignon (1982,1989), Koim (1977), Maasoumi (1986), Maasoumi and Nickelsburg (1988), Mosler (1994), Rietveld (1990) and Slottje (1987) proceed in this direction. Taguchi (1972a) defined the concentration surface of a two-dimensional random vector and in (1972b) he extended the notions of concentration surface to complete surface (which he called the Lorenz manifold). Also in 1981 he extended the concepts of mean difference, mean deviation and Gini index to two-dimensional random variables.
Arnold (1987) proposed a much more simple definition to the Lorenz curve for the Bivariate random variable. Another work by Koshevoy and Mosler (1996, 1997) gave two extensions of the univariate Gini index. The first definition is based on the expected distance between two-independent vectors from the same distribution and the second definition is based on the volume of the lift zonoid formed from more than two independent vectors. Ord, Patil and Taillie (1983) proposed a measure of truncated income inequality in the univariate case having the property of truncation invariance and is given below.

**Definition 1** Let $X \geq 0$ be a non-negative random variable with distribution function $F(x)$ and survival function $\overline{F}(x)$. The truncated form of the Gini index is defined by

$$G(t) = 2 \int_t^\infty F_t(x) dF_{1,t}(x) - 1$$

where $F_t(x)$ is the distribution function of $X_1(t) = X | X > t$, defined by $F_t(x) = \int_t^x \frac{f(y)}{F(t)} dy$ and $F_{1,t}(x)$ is the first moment distribution given by

$$F_{1,t}(x) = \frac{\int_t^x \frac{yf(y)}{F(t)} dy}{\int_t^\infty \frac{yf(y)}{F(t)} dy}$$

On the same lines, in the present paper, we propose a new measure of income inequality to Bivariate probability distributions and study some characterization results. In Section 2, we define the vector valued Gini index and compute the same for some well-known distributions. In Section 3, we derive certain relationships between the vector valued Gini index and other analogous measures. We also define a measure of income Gap ratio for this random variable. In Section 4, we give an illustration of the vector valued Gini index using a data on U.S. income and wealth. In Section 5, we look into the problem of extending the concept to higher dimensions when several attributes of the population are under consideration.

**2 Vector valued Gini index**

Consider the study of income inequality using multivariate distribution, when we observe more than one attributes of the population. To start with, we restrict our attention to Bivariate distributions and in the end we extend this to multivariate distributions.

Most of the economic measures are subjected to threshold values and the low values are either missing or unreliable. For example, much of the data on income come from income tax returns and most countries have a threshold