Soc Choice Welfare (2002) 19: 789–794

Social Choice and Welfare

A characterization of income distributions in terms of generalized Gini coefficients

Christian Kleiber[1], **Samuel Kotz**[2]

[1] Fachbereich Statistik, Universität Dortmund, D-44221 Dortmund, Germany
(e-mail: Kleiber@statistik.uni-dortmund.de)
[2] Department of Engineering Management and Systems Engineering,
George Washington University, Washington, DC, USA (e-mail: kotz@seas.gwu.edu)

Received: 8 March 2001/Accepted: 31 July 2001

Abstract. Most commonly used parametric models for the size distribution of incomes possess only a few finite moments, and hence cannot be characterized by the sequence of their moments. However, all income distributions with a finite mean can be characterized by the sequence of first moments of the order statistics. This is an attractive feature since the generalized Gini coefficients of Kakwani (1980), Donaldson and Weymark (1980, 1983) and Yitzhaki (1983) are simple functions of expectations of sample minima. We present results which streamline these characterizations motivated by Aaberge (2000).

1 Introduction

Parametric models for the size distribution of personal incomes have been of interest to economists and applied statisticians for more than a hundred years. A basic feature of empirical income distributions is that they are well approximated by models with polynomially decreasing tails, i.e., $1 - F(x) \sim x^{-\alpha}$, for some $\alpha > 0$. This is known as the weak Pareto law (Mandelbrot 1960) and implies, inter alia, that models obeying it, such as the Pareto or the Singh-Maddala (1976) distributions, cannot be characterized in terms of their moments, since only a few of the moments exist. It is well known that another popular income distribution, the lognormal, is not characterized by the sequence of its moments, although all the moments are finite (see, e.g., Heyde 1963, and Leipnik 1991).

The work of C. Kleiber was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 475. The authors would like to thank two anonymous referees and associate editor John A. Weymark for valuable comments and suggestions.

Recently Aaberge (2000), in an important communication, observed that any income distribution with a finite mean is characterized up to a scale by the moments of its 'Lorenz curve distribution'. These moments turn out to be affine functions of the generalized Gini coefficients of Kakwani (1980), Donaldson and Weymark (1980, 1983) and Yitzhaki (1983). The present paper shows that income distributions can be characterized directly in terms of these generalized Gini coefficients. This is connected with the moment problem of order statistics.

2 The moment problem of order statistics

Let $X_1, \dots, X_n$ be a sample of size n from a distribution with the c.d.f. F and define the order statistics $X_{k:n}$ in the ascending order by

$$X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}.$$

The cumulative distribution function (c.d.f.) $F_{k:n}$ of $X_{k:n}$ can be written as

$$F_{k:n}(x) = \sum_{j=k}^{n} \binom{n}{j} \{F(x)\}^j \{1 - F(x)\}^{n-j}$$

and its first moment is

$$E(X_{k:n}) = k \binom{n}{k} \int_0^1 F^{-1}(p) p^{k-1} (1-p)^{n-k} \, dp, \tag{1}$$

where $F^{-1}(p) := \sup\{x \,|\, F(x) \leq p\}$ is the quantile function of F (see, e.g., David 1981).

The moment problem of order statistics inquires to what extent the c.d.f. F is uniquely determined by (a subset of) the first moments of all of its order statistics

$$\{E(X_{k:n}) \,|\, k = 1, 2, \dots, n; n = 1, 2, 3, \dots\}. \tag{2}$$

It follows from (1) that $E|X_{k:n}| \leq c \cdot E|X|$, for some $c > 0$; thus a finite mean of the parent distribution assures the existence of the first moment of any order statistic. This implies that characterizations in terms of the moments of order statistics are of interest for heavy-tailed distributions of the Pareto type, for which only a few moments exist and, consequently, no characterization in terms of (ordinary) moments is feasible. Many parametric models for the size distribution of income are of this type, e.g. the Pareto or Singh-Maddala (1976) distributions.

In view of the well-known recurrence relation (see, e.g., David 1981, p. 46)

$$(n-k)E(X_{k:n}) + kE(X_{k+1:n}) = nE(X_{k:n-1})$$

it is not necessary to know the whole array (2), one merely requires to have one moment for each sample size, e.g. the sequence of expectations of minima $E(X_{1:n})$ will do. The basic characterization result is thus as follows:

Lemma 1. *Let* $E|X| < \infty$. *For* $n = 1, 2, 3, \ldots$, *let* $k(n)$ *be an integer with* $1 \leq k(n) \leq n$. *Then,* F *is uniquely determined by the sequence* $\{E(X_{k(n):n}) \,|\, n = 1, 2, 3, \ldots\}$.

The most natural choices for $k(n)$ are either 1 or n. Many extensions and refinements of this basic result are available in the literature, see e.g., Galambos and Kotz (1978), Huang (1989) or Kamps (1998) for further details. Huang (1989) provides a detailed survey of the results up to the late eighties. Lemma 1 serves as the basis for Theorem 2 below; some further characterizations are briefly discussed in Sect. 4.

3 A characterization of income distributions

An income distribution has the property that its c.d.f. F is supported on the positive halfline, i.e. $\operatorname{supp}(F) \subseteq [0, \infty)$. Let $L(p) := \frac{1}{E(X)} \int_0^p F^{-1}(t)\, dt$, $p \in [0, 1]$, denote the Lorenz curve of F and define $\mathscr{L} = \{F \,|\, 0 < \int x\, dF(x) < \infty\}$, the set of all (income) distributions admitting a Lorenz curve. The most well-known measure of income inequality is the venerated Gini coefficient. One of its definitions is given as twice the area between the Lorenz curve and the "equality line":

$$G = 2 \int_0^1 (p - L(p))\, dp = 1 - 2 \int_0^1 L(p)\, dp.$$

The Gini index is indeed a relative measure of income inequality since it depends only on income shares. A sizable number of alternative representations of the Gini coefficient are available in the literature. For our purposes the expression

$$G = 1 - \frac{E(X_{1:2})}{E(X)} = 1 - \frac{\int_0^\infty (1 - F(x))^2\, dx}{E(X)} \tag{3}$$

is the most appropriate. This formula is presumably due to Arnold and Laguna (1977), at least in the non-Italian literature. It has been independently rediscovered in the economics literature by Dorfman (1979).

Kakwani (1980) proposed a one-parameter family of generalized Gini indices by introducing different weighting functions for the area under the Lorenz curve,

$$G_n = 1 - n(n-1) \int_0^1 L(p)(1 - p)^{n-2}\, dp.$$

The traditional Gini coefficient is obtained for $n = 2$. Donaldson and Weymark (1980, 1983) have arrived at the same family from different considerations. These authors also defined a family of 'equally-distributed-equivalent-income functions' of the form

$$\Xi_n = -\int_0^\infty x\, d\{(1 - F(x))^n\},$$

which may be rewritten as $\Xi_n = \int_0^\infty (1 - F(x))^n\, dx$, and an 'absolute' Gini index of the form $G_n^A = E(X) - \Xi_n$. Muliere and Scarsini (1989) observed that Ξ_n, which they also call an absolute Gini index, equals $E(X_{1:n})$ and that

$$G_n = 1 - \frac{\Xi_n}{E(X)} = 1 - \frac{E(X_{1:n})}{E(X)}. \tag{4}$$

Equation (4) is a straightforward generalization of (3).

The following characterization is an immediate consequence of Lemma 1 via the moment problem of order statistics:

Theorem 2. *Any $F \in \mathscr{L}$ is characterized by its sequence of absolute Gini indices, $\{\Xi_n\}$.*

It is more common to measure inequality by relative indices, for which the following variant of Theorem 2 is valid:

Corollary 3. *Any $F \in \mathscr{L}$ is characterized (up to a scale) by its sequence of relative Gini indices, $\{G_n\}$.*

Aaberge (2000) recently obtained a similar result using a different approach. Considering the Lorenz curve as a distribution function, he observes that this 'Lorenz curve distribution' has bounded support and is therefore characterized by the sequence of its moments. These moments, as well as Aaberge's new 'Lorenz measures' of inequality, turn out to be affine functions of the generalized Gini coefficients presented above. In particular, the first Lorenz curve moment $\int_0^1 p\, dL(p)$ equals $G/2 + 1/2$, where G is the "traditional" Gini coefficient. Consequently our Corollary 3 also follows from Aaberge's result. In addition, Aaberge has a representation of the Lorenz measures of inequality in terms of expectations of sample maxima (his Eq. (2.6)), and therefore a characterization of income distributions in terms of expectations of sample maxima. Thus, his results can also be considered as a consequence of Lemma 1.

We shall clarify the results presented above by means of a number of examples of specific distributions:

Examples:

(a) For the uniform distribution on $[0, 1]$, $X_{k:n} \sim \text{Beta}(k, n-k+1)$, so that $E(X_{1:n}) = 1/(n+1) = \Xi_n$, and this sequence characterizes the $U[0,1]$ distribution. Similarly, any uniform distribution on $[0, b]$ is characterized up to a scale by the sequence $\{G_n \mid G_n = 1 - 2/(n+1)\}$.

(b) For exponential F, with scale parameter equal to unity, $E(X_{1:n}) = 1/n$, so that $\{G_n \mid G_n = 1 - 1/n\}$ characterizes the family of exponential distributions.

(c) For the Pareto distribution, with $1 - F(x) = (x/\sigma)^{-\alpha}$, where $x \geq \sigma > 0$ and $\alpha > 0$, one has $E(X_{1:n}) = \sigma \cdot \frac{\alpha n}{\alpha n - 1}$, and therefore $G_n = (n-1)/(\alpha n - 1)$, a sequence that characterizes Pareto distributions up to a scale.

(d) The Singh-Maddala (1976) distribution, called the Pareto (IV) distribution by Arnold and Laguna (1977), is perhaps one of the most popular income distribution models, given by the c.d.f. $F(x) = 1 - (1 + (x/\sigma)^{\gamma})^{-\alpha}$, where $x \geq 0$ and all parameters are positive. Arnold and Laguna show that $E(X_{1:n}) = \sigma\Gamma(n\alpha - 1/\gamma)\Gamma(1 + 1/\gamma)/\Gamma(n\alpha)$. Therefore $G_n = 1 - \Gamma(n\alpha - 1/\gamma)\Gamma(\alpha)/(\Gamma(n\alpha)\Gamma(\alpha - 1/\gamma))$ and this sequence characterizes Singh-Maddala distributions up to a scale.

We observe that the distributions described in (a) and (b) possess the ordinary moments of all orders which is not the case for the distributions in (c) and (d). It is worth noting that the lognormal distribution, a distribution known not to be determined by the sequence of its moments (see e.g., Heyde 1963), can be characterized in terms of generalized Gini coefficients. Unfortunately there are no simple expressions for moments of lognormal order statistics (see, e.g., Gupta et al. 1974).

4 Extensions

A natural question which arises in connection with Lemma 1 is as follows: are the expectations of all minima $E(X_{1:n})$ required to determine F, or will certain subsequences be sufficient? In view of (1) this problem reduces to the question of L^1-completeness of the sequence of functions

$$\{(1-p)^n \mid p \in [0,1]; n = n_1, n_2, n_3, \ldots\}. \tag{5}$$

It is well-known from Müntz's (1914) theorem that a necessary and sufficient condition for L^1-completeness, and therefore determinacy of F, is:

$$\sum_{j=1}^{\infty} \frac{1}{n_j} = \infty,$$

see e.g., Huang (1989). This leads to

Theorem 4. (a) *Any $F \in \mathscr{L}$ is determined by any subsequence of absolute Gini indices, $\{\Xi_n\}$, for which Müntz's condition holds.*
(b) *Any $F \in \mathscr{L}$ is determined up to a scale by any subsequence of relative Gini indices, $\{G_n\}$, for which Müntz's condition holds.*

An example of a sequence of integers satisfying the Müntz condition is the sequence $\{p_n\}$ of all prime numbers.

Evidently it is also possible to obtain characterizations in terms of expectations of moment differences, e.g. differences of successive generalized Gini coefficients, or Donaldson and Weymark's (1980) absolute Gini indices, $\{G_n^A\}$. Details are omitted and an interested reader is referred to Kamps (1998).

References

Aaberge R (2000) Characterizations of Lorenz curves and income distributions. Soc Choice Welfare 17: 639–653
Arnold BC, Laguna L (1977) On generalized Pareto distributions with applications to income data. International Studies in Economics no. 10, Dept. of Economics, Iowa State University, Ames, Iowa
David HA (1981) Order Statistics, 2nd ed. Wiley, New York
Donaldson D, Weymark JA (1980) A single-parameter generalization of the Gini index of inequality. J Econ Theory 22: 67–86
Donaldson D, Weymark JA (1983) Ethically flexible Gini indices for income distributions in the continuum. J Econ Theory 29: 353–358
Dorfman R (1979) A formula for the Gini coefficient. Rev Econ Stat 61: 146–149
Galambos J, Kotz S (1978) Characterizations of Probability Distributions. Springer, Berlin Heidelberg New York
Gupta SS, McDonald GC, Galarneau DI (1974) Moments, product moments and percentage points of the order statistics from the lognormal distribution for samples of size twenty and less. Sankhyā B 36: 230–260
Heyde CC (1963) On a property of the lognormal distribution. J Royal Stat Soc B 25: 392–393
Huang JS (1989) Moment problem of order statistics: A review. Int Stat Rev 57: 59–66
Kakwani N (1980) On a class of poverty measures. Econometrica 48: 437–446
Kamps U (1998) Characterizations of distributions by recurrence relations and identities for moments of order statistics. In: Balakrishnan N, Rao CR (eds) Handbook of Statistics, vol. 16. Elsevier, Amsterdam
Leipnik RB (1991) On lognormal random variables. I: The characteristic function. J Austral Math Soc B 32: 327–347
Mandelbrot B (1960) The Pareto-Lévy law and the distribution of income. Int Econ Rev 1: 79–106
Muliere P, Scarsini M (1989) A note on stochastic dominance and inequality measures. J Econ Theory 49: 314–323
Müntz CH (1914) Über den Approximationssatz von Weierstrass. In: Festschrift für H.A. Schwarz. Springer, Berlin Heidelberg New York, pp 303–312
Singh SK, Maddala GS (1976) A function for the size distribution of incomes. Econometrica 44: 963–970
Yitzhaki S (1983) On an extension of the Gini inequality index. Int Econ Rev 24: 617–628