Inference on Income Distributions

by

Russell Davidson

Department of Economics and CIREQ
McGill University
Montréal, Québec, Canada
H3A 2T7

GREQAM
Centre de la Vieille Charité
2 Rue de la Charité
13236 Marseille cedex 02, France

russell.davidson@mcgill.ca

Abstract

This paper attempts to provide a synthetic view of varied techniques available for performing inference on income distributions. Two main approaches can be distinguished: one in which the object of interest is some index of income inequality or poverty, the other based on notions of stochastic dominance. From the statistical point of view, many techniques are common to both approaches, although of course some are specific to one of them. I assume throughout that inference about population quantities is to be based on a sample or samples, and, formally, all randomness is due to that of the sampling process. Inference can be either asymptotic or bootstrap-based. In principle, the bootstrap is an ideal tool, since in this paper I ignore issues of complex sampling schemes, and suppose that observations are IID. However both bootstrap inference, and, to a considerably greater extent, asymptotic inference can fall foul of difficulties associated with the heavy right-hand tails observed with many income distributions. I mention some recent attempts to circumvent these difficulties.

Keywords: Income distribution, delta method, asymptotic inference, bootstrap, influence function, empirical process.

JEL codes: C12, C13, C81, D31, I32

This research was supported by the Canada Research Chair program (Chair in Economics, McGill University) and by grants from the Social Sciences and Humanities Research Council of Canada, and the Fonds Québécois de Recherche sur la Société et la Culture. The first version of this paper was prepared for the Innis Lecture given at the 44th Annual Conference of the Canadian Economics Association, held in May 2010 at Quebec City. I thank Jean-Yves Duclos for his valuable comments on the first version.

May 2010

1. Introduction

In this paper, I survey many of the statistical techniques used in order to perform inference on income-distribution data. By interspersing the review with a few new results, I hope to show to what extent the problems that arise still provide an active research agenda. Questions relating to income inequality and poverty have been of interest to economists for a long time. One of the earliest analyses of inequality dates back to Gini (1912), in which Gini proposes the celebrated index that bears his name. More recently, Atkinson (1970) is a pioneering paper on the measurement of inequality, and Sen (1976) extends ideas to the measurement of poverty.

Whereas Gini's focus was statistical, the approach of Atkinson, Sen, and their many followers is instead axiomatic and normative. They attempt to characterise indices of inequality and poverty, including Gini's of course, by their properties with respect to various axioms thought to be desirable on ethical grounds. Canadians have contributed importantly to this literature: here I may cite Blackorby and Donaldson (1978) and (1980), papers that have had a profound impact on our subsequent thinking about the ethics of income distribution.

With the notable exception of early work by Gastwirth and his co-authors, in particular Gastwirth (1974), little attention was paid to statistical issues, and in particular statistical inference, in the analysis of income distribution until a paper by Beach and Davidson (1983), in which the asymptotic distribution of estimated Lorenz curve ordinates was worked out. After that, a very abundant literature followed. The references I give here constitute at best a representative sample: Bishop, Chakraborti, and Thistle (1989). Chow (1989), Cowell (1989), McFadden (1989), Preston (1995), Anderson (1996), Dardanoni and Forcina (1999), Davidson and Duclos (2000). These papers trace something of the history of statistical research on income distribution.

There are two rather different strands to this literature. One is concerned with inference on particular indices of inequality and poverty; the other with inference on stochastic dominance. The techniques required are considerably different for the two strands. In both, though, the usual approach is to assume that the object of inference is some aspect of a population distribution or distributions, and that inference is to be performed on the basis of a random sample drawn from the population(s). The sole source of randomness is that of the sampling process. A few recent papers deal with the intricacies of complex sampling designs – a notable example is Bhattacharya (2007) – but the vast majority of papers assume that the sample is drawn completely at random from the population. Issues relating to measurement error seem not so far to have attracted attention, despite their obvious importance in practice.

A statistical tool that is showing its worth in income distribution studies is the bootstrap. I will have a good deal to say about that in this paper, for it has become abundantly clear that bootstrap-based inference is capable of yielding much more reliable results than conventional asymptotic inference. The latter is by no means rendered useless by the bootstrap, since the bootstrap works best with statistics of which the asymptotic variance is known; see, among many other sources, Horowitz (1997).

In the next section, I look at three approaches that are commonly used in the study of the asymptotic distributions of inequality and poverty indices, and give a few examples of how they work, and the tradeoffs involved in selecting a suitable method for any given problem. In Section 3, I develop further one of the three approaches, namely that based on empirical process theory. I illustrate its virtues and its disadvantages relative to other methods. Section 4 discusses the role of the bootstrap in inference, and points out that our best efforts often fail when the distributions we work with have heavy right-hand tails. In Section 5, I discuss some of the literature on inference on stochastic dominance, mentioning in particular the relative virtues of testing a null hypothesis of dominance or non-dominance. Some very recent work is presented in Section 6, where I look at measures of the divergence between two distributions. An important application of this is to measure the goodness of fit of a sample to a specified distribution or parametric family of distributions. The empirical-process approach turns out to be crucial here, and is used to study the asymptotic distributions of various goodness-of-fit measures. A few concluding remarks are found in Section 7.

2. Inequality and Poverty Indices

Inference on the Gini index and related indices, such as the S-Gini and E-Gini indices, has become a rather popular topic in the literature of the last five years. See, among others, Barrett and Pendakur (1995), Zitikis and Gastwirth (2002), Deltas (2003), Xu (2007), Barrett and Donald (2009), and Davidson (2009a). As far as I can tell, no two of these papers adopt the same technical approach to find an expression for the asymptotic variance of the sample indices, and to obtain computable estimates of it. By sample indices, I mean plugin estimators of the population quantities, in which the population cumulative distribution function (CDF) is systematically replaced by the sample empirical distribution function (EDF).

The approaches used falls roughly into three categories: those based on the delta method, on influence functions, or on empirical process theory. My preferred approach until now has been based on the delta method, which seems intuitive and tractable. However, it is clear that different methods are best adapted to different problems, as I will illustrate. In this section, I extend the approach used in my 2009 paper (reference 2009a) for the usual Gini index to the case of the family of S-Gini indices. I hope thereby to illustrate the virtues of the method for the many indices that share the same basic properties with the Gini index.

Throughout, I denote by F the population CDF , and by $\hat{F}$ the EDF, defined for an IID sample of size n as

$$\hat{F}(x) = n^{-1} \sum_{i=1}^{n} \mathrm{I}(y_i \leq x), \quad x \geq 0, \tag{1}$$

where the y_i, $i = 1, \ldots, n$ are the observations of the sample, and I is the indicator function. The value of $\hat{F}(x)$ is just the proportion of incomes in the sample that are less than or equal to x. Note that according to this definition, $\hat{F}$ is **cadlag** (continue à droite avec limite à gauche), which, although arbitrary, is the usual convention. The argument x is restricted to non-negative values. Negative incomes do occur, especially in times of financial crisis, and at any time for some self-employed people. But their presence is excluded by the very definitions of some indices, and it may render the interpretation of others counter-intuitive, including those of the extended Gini family. For instance, a Gini index greater than 1 can arise with negative incomes.

There are several ways in which the Gini index can be expressed as a functional of the distribution F. The one I use here (slightly different from that used in my 2009a paper) is

$$G \equiv 1 - \frac{2}{\mu} \int_0^\infty y\big(1 - F(y)\big)\, \mathrm{d}F(y). \tag{2}$$

where μ is mean income. This is a **relative** Gini index; the **absolute** index is this times μ. The relative index is thus scale invariant, and measures pure inequality, while the absolute index takes account of the overall income level. The S-Gini indices are parametrised by a positive parameter α, and reduce to the usual Gini index when $\alpha = 2$. The relative index is defined as

$$R_\alpha = 1 - \frac{\alpha}{\mu} \int_0^\infty y\big(1 - F(y)\big)^{\alpha-1} \mathrm{d}F(y),$$

the absolute index once again being this times μ. The greater the value of α, the more weight is given to the bottom of the distribution. The sample indices have exactly the same expression, with F replaced by $\hat{F}$ and μ by $\hat{\mu}$.

Consider the integral

$$H_\alpha \equiv \alpha \int_0^\infty y\big(1 - F(y)\big)^{\alpha-1} \mathrm{d}F(y)$$

and its plugin estimate

$$\begin{aligned} \hat{H}_\alpha &= \alpha \int_0^\infty y\big(1 - \hat{F}(y)\big)^{\alpha-1} \mathrm{d}\hat{F}(y) \\ &= -\int_0^\infty y \, \mathrm{d}\big[\big(1 - \hat{F}(y)\big)^\alpha\big]. \end{aligned}$$

The integrator function $\big(1 - \hat{F}(y)\big)^\alpha$ is a step function, with discontinuities at the observed incomes y_i, $i = 1, \ldots, n$, and so the integral in the second line above becomes a sum of contributions from each of these discontinuities. Let the **order statistics** of the sample be denoted by $y_{(i)}$, $i = 1, \ldots, n$, with $0 \le y_{(1)} < y_{(2)} < \ldots < y_{(n)}$. (For ease I ignore the possibility of ties.) The value of $\hat{F}$ at $y_{(i)}$ is just i/n. Thus the discontinuity of the integrator at $y_{(i)}$ is

$$\Big(1 - \frac{i}{n}\Big)^\alpha - \Big(1 - \frac{i-1}{n}\Big)^\alpha = n^{-\alpha}\big((n-i)^\alpha - (n-i+1)^\alpha\big). \tag{3}$$

Consequently,

$$\hat{H}_\alpha = n^{-\alpha} \sum_{i=1}^{n} y_{(i)} \big((n-i+1)^\alpha - (n-i)^\alpha\big). \tag{4}$$

This shows that $\hat{H}_\alpha$ is an **L-statistic**, that is, a linear combination of the order statistics.

Clearly, $R_\alpha = 1 - H_\alpha/\mu$, and so we define $\hat{R}_\alpha = 1 - \hat{H}_\alpha/\hat{\mu}$, where $\hat{\mu}$ is the sample mean, that is, the expectation of the empirical distribution. We then have

$$\hat{R}_\alpha - R_\alpha = \frac{H_\alpha}{\mu} - \frac{\hat{H}_\alpha}{\hat{\mu}} = \frac{1}{\mu\hat{\mu}}\big(H_\alpha(\hat{\mu}-\mu) - \mu(\hat{H}_\alpha - H_\alpha)\big).$$

If $\hat{R}$ is to be consistent, then, as $n \to \infty$, the above expression must tend to zero. The asymptotic variance is by definition, the variance of the limiting distribution of the sequence with typical element $n^{1/2}(\hat{R}_\alpha - R_\alpha)$. This sequence has the same limiting distribution as that with element

$$\mu^{-2}\big(H_\alpha n^{1/2}(\hat{\mu}-\mu) - \mu n^{1/2}(\hat{H}_\alpha - H_\alpha)\big),$$

since, with very mild regularity, $n^{1/2}(\hat{\mu}-\mu)$ and $n^{1/2}(\hat{H}_\alpha - H_\alpha)$ are both $O_p(1)$ as $n \to \infty$.

Now for the delta method. Trivially, we see that

$$\hat{\mu} - \mu = n^{-1}\sum_{i=1}^{n}(y_i - \mu).$$

For the other small quantity, we find that

$$\begin{aligned}\alpha^{-1}(\hat{H}_\alpha - H_\alpha) &= \int_0^\infty y\big(1-F(y)\big)^{\alpha-1} \mathrm{d}(\hat{F}-F)(y) \\ &+ \int_0^\infty y\big[\big(1-\hat{F}(y)\big)^{\alpha-1} - \big(1-F(y)\big)^{\alpha-1}\big]\, \mathrm{d}F(y) \\ &+ \int_0^\infty y\big[\big(1-\hat{F}(y)\big)^{\alpha-1} - \big(1-F(y)\big)^{\alpha-1}\big]\, \mathrm{d}(\hat{F}-F)(y).\end{aligned}$$

The first two terms on the r.h.s are $O_p(n^{-1/2})$, while the last is $O_p(n^{-1})$, and so is ignored for the purposes of the asymptotic distribution. The first term is

$$n^{-1}\sum_{i=1}^{n}\big[y_i\big(1-F(y_i)\big)^{\alpha-1} - H_\alpha/\alpha\big],$$

while the second can be seen by a little algebra to be

$$n^{-1}(\alpha-1)\sum_{i=1}^{n}\big[m(y_i) - H_\alpha/\alpha\big],$$

where we have defined

$$m(y) = \int_0^y x\bigl(1 - F(x)\bigr)^{\alpha-2} \mathrm{d}F(x). \tag{5}$$

It can be seen that $\mathrm{E}\bigl(m(Y)\bigr) = H_\alpha/\alpha$, where Y denotes the random variable of which the observations are realisations.

We end up with the result that, to leading order asymptotically,

$$n^{1/2}(\hat{R}_\alpha - R_\alpha) = n^{-1/2}\mu^{-2}\sum_{i=1}^{n}\Bigl[H_\alpha(y_i - \mu) \\ - \mu\alpha\bigl\{(\alpha-1)m(y_i) + y_i\bigl(1 - F(y_i)\bigr)^{\alpha-1} - H_\alpha\bigr\}\Bigr]$$

This can be rewritten as follows so as to see that the expectation of the limiting distribution is zero:

$$n^{1/2}(\hat{R}_\alpha - R_\alpha) = n^{-1/2}\mu^{-2}\sum_{i=1}^{n}\Bigl[H_\alpha(y_i - \mu) \\ - \mu\alpha\Bigl\{(\alpha-1)\bigl(m(y_i) - H_\alpha/\alpha\bigr) + \Bigl(y_i\bigl(1 - F(y_i)\bigr)^{\alpha-1} - H_\alpha/\alpha\Bigr)\Bigr\}\Bigr]. \tag{6}$$

The summands in the above expression are IID, from which the asymptotic normality of $\hat{R}_\alpha$ follows immediately. Replacing y_i in the expression of the summand by Y yields the **influence function** of the index R_α. What we have seen here is that the delta method provides a convenient way to compute the influence function in a form that is useful for inference, as we will now see. Make the following definition of the random variable Z:

$$Z = H_\alpha Y - \mu\alpha\bigl\{(\alpha-1)m(Y) + Y\bigl(1 - F(Y)\bigr)^{\alpha-1}\bigr\}.$$

The summands in (6) are clearly IID realisations of Z minus its expectation. Thus the asymptotic variance of $\hat{R}_\alpha$ is μ^{-4} times the variance of Z.

The summands in (6) are not directly observable, because the functions m and F are unknown. However, it is easy enough to estimate them. In particular, we can use $\hat{F}$ to estimate F. For m, the easiest approach is to estimate $m(y_{(i)})$ for each order statistic $y_{(i)}$. From (5),

$$m(y) = \mathrm{E}\Bigl(Y\bigl(1 - F(Y)\bigr)^{\alpha-2}\mathrm{I}(Y \le y)\Bigr).$$

The estimator $\hat{F}$ that is most convenient here is not (1), for which $\hat{F}(y_{(i)}) = i/n$, but rather

$$\hat{F}(y_{(i)}) = (2i-1)/(2n),$$

which "splits the difference" between a right- and left-continuous EDF. With this, we can make the definition

$$\hat{m}(y_{(i)}) = n^{-1} \sum_{j=1}^{i} y_{(j)} \Big(\frac{2(n-j)+1}{2n}\Big)^{\alpha-2}. \tag{7}$$

This allows us to compute estimates of the realisations of Z, as follows:

$$\hat{Z}_i = \hat{H}_\alpha y_{(i)} - \hat{\mu}\alpha\Big\{(\alpha-1)\hat{m}(y_{(i)}) + y_{(i)}\Big(\frac{2(n-i)+1}{2n}\Big)^{\alpha-1}\Big\}. \tag{8}$$

Although the estimate $\hat{Z}_i$ depends on the i^{th} order statistic, and the order statistics are by no means mutually independent, each $\hat{Z}_i$ is an estimate of the corresponding Z_i, and the Z_i *are* IID. Thus, if $\bar{Z}$ is the mean of the $\hat{Z}_i$, the estimate of the variance of $\hat{R}_\alpha$ is

$$n^{-1}\mu^{-4}\sum_{i=1}^{n}(\hat{Z}_i - \bar{Z})^2. \tag{9}$$

This estimate is easy to compute on the basis of the sorted sample:

- Compute $\hat{\mu}$ as the sample mean;
- Compute $\hat{H}_\alpha$ by the formula (4);
- For each $i = 1, \ldots, n$, compute $\hat{m}(y_{(i)})$ using (7);
- With the results thus obtained, compute the $\hat{Z}_i$ using (8);
- Finally, compute the estimated variance of $\hat{R}_\alpha$ from (9).

3. The Empirical Process Approach

We have seen that the approaches based on the delta method and influence functions are closely related. The other popular approach is based on empirical process theory. Many examples of the use of empirical process theory in econometrics can be found in the Handbook article of Andrews (1994). It usually gives results in a different form from those of the other approaches. In addition, it is often a good deal harder to implement estimates of the expressions found for asymptotic variances. This can be illustrated poignantly by the case of the sample mean.

Suppose that the random variables U_i, $i = 1, \ldots, n$, are IID with common distribution the uniform distribution U(0,1). The corresponding order statistics are denoted by $U_{(i)}$, and their joint distribution is well known. Under the assumption that F is an absolutely continuous distribution, the order statistics $Y_{(i)}$ of a set of IID variables with common distribution F have the same distribution as the set of variables $F^{-1}(U_{(i)})$, or, to state this result in a form that is usually more useful, the random variables $F(Y_{(i)})$ follow the same joint distribution as the uniform order statistics $U_{(i)}$. The key result concerning the $U_{(i)}$ is that their joint distribution tends to a Brownian bridge as $n \to \infty$. Specifically, the following stochastic process, indexed by the continuous variable $t \in [0,1]$,

$$n^{1/2}\Big(U_{(\lceil nt\rceil)} - \frac{\lceil nt\rceil}{n+1}\Big)$$

tends in distribution (converges weakly to) the Brownian bridge process $B(t)$. Here, the notation $\lceil x\rceil$ means the smallest integer not smaller than x. It corresponds to "rounding up" x to the next integer. The Brownian bridge is a mean-zero Gaussian process with covariance function

$$\operatorname{cov}\big(B(s), B(t)\big) = s(1-t), \quad s \le t. \tag{10}$$

The variables $F(Y_{(i)})$ have the same distribution as the $U_{(i)}$, and so the process $n^{1/2}\big(F(Y_{(\lceil nt\rceil)}) - \lceil nt\rceil/(n+1)\big)$ also converges weakly to a Brownian bridge. Let $\lceil nt\rceil = i$. Then, by Taylor expansion,

$$F(Y_{(\lceil nt\rceil)}) - F\Big(F^{-1}\Big(\frac{\lceil nt\rceil}{n+1}\Big)\Big) = f\Big(F^{-1}\Big(\frac{\lceil nt\rceil}{n+1}\Big)\Big)\Big(Y_{(\lceil nt\rceil)} - F^{-1}\Big(\frac{\lceil nt\rceil}{n+1}\Big)\Big) + o_p(n^{-1/2}),$$

where $f = F'$ is the density corresponding to F. Thus

$$n^{1/2}\Big(Y_{(\lceil nt\rceil)} - F^{-1}\Big(\frac{\lceil nt\rceil}{n+1}\Big)\Big) \rightsquigarrow \frac{B(t)}{f\big(F^{-1}(\lceil nt\rceil/(n+1)\big)}, \tag{11}$$

where the symbol $\rightsquigarrow$ represents the weak convergence of a stochastic process to a limiting process.

The sample mean $\hat\mu$ is $n^{-1}\sum_{i=1}^n y_{(i)}$ – no harm is done by summing over the order statistics. As a random variable, rather than as a realisation, $\hat\mu$ can thus be written as $n^{-1}\sum_{i=1}^n Y_{(i)}$, so that

$$n^{1/2}(\hat\mu - \mu) = n^{-1/2}\sum_{i=1}^n\Big[F^{-1}\Big(\frac{i}{n+1}\Big) + n^{-1/2}\frac{B(t)}{f\big(F^{-1}(i/(n+1))\big)}\Big] - n^{1/2}\mu$$

But

$$\mu = \int_{-\infty}^{\infty} y\,\mathrm{d}F(y) = \int_0^1 F^{-1}(t)\,\mathrm{d}t,$$

and this last integral is the limit of the Riemann sum

$$n^{-1}\sum_{i=1}^n F^{-1}\Big(\frac{i}{n+1}\Big),$$

which lets us see that $n^{1/2}(\hat\mu - \mu)$ converges to the same limit as

$$n^{-1}\sum_{i=1}^n \frac{B(t)}{f\big(F^{-1}(i/(n+1))\big)}.$$

This expression can also be interpreted as a Riemann sum, which converges to

$$\int_0^1 \frac{B(t)\,\mathrm{d}t}{f\bigl(F^{-1}(t)\bigr)}.$$

This random variable, which has the same distribution as the limiting distribution of $n^{1/2}(\hat\mu - \mu)$, has expectation zero, and, from (10), variance

$$2\int_0^1\int_0^t \frac{s(1-t)}{f\bigl(F^{-1}(t)\bigr)f\bigl(F^{-1}(s)\bigr)}\,\mathrm{d}s\,\mathrm{d}t.$$

By changing variables by the relations $s = F(y)$ and $t = F(x)$, this double integral becomes

$$2\int_{-\infty}^{\infty}\frac{1-F(x)}{f(x)}\int_{-\infty}^{x}\frac{F(y)}{f(y)}\,\mathrm{d}F(y)\,\mathrm{d}F(x) = 2\int_{-\infty}^{\infty}\bigl(1-F(x)\bigr)\int_{-\infty}^{x}F(y)\,\mathrm{d}y\,\mathrm{d}x. \quad (12)$$

A messy calculation, which interested readers can find in the Appendix, shows that this last expression is actually the variance of the distribution F.

The approach has given the right answer! But it can hardly be accused of taking the shortest road to it. Further, suppose I wish to estimate the rightmost expression in (12) by replacing F by $\hat F$. The resulting double integral can be evaluated, of course, but it is far from obvious that its value is the sample variance, although the proof in the Appendix shows that it must be. This example highlights the pitfalls of the empirical process approach, although it does not display another important property of the approach, namely its generality. Later, we will see another circumstance in which it is invaluable.

I now give a more substantive illustration of the empirical process approach, no more complicated than the one above for the mean, and one that allows us to make contact with the literature. The object of interest is the absolute Gini index, that is, the relative Gini (2) times the mean μ. The plugin estimator is

$$\hat\mu\hat G = \hat\mu - 2\int_0^{\infty} y\bigl(1-\hat F(y)\bigr)\,\mathrm{d}\hat F(y) = \hat\mu + \int_0^{\infty} y\,\mathrm{d}\bigl[\bigl(1-\hat F(y)\bigr)^2\bigr].$$

From the result (3) with $\alpha = 2$, we obtain a suitable expression for the discontinuity of the integrator function in the last expression above, and we find that

$$\hat\mu\hat G = n^{-1}\sum_{i=1}^{n} y_{(i)}\Bigl(\frac{2i-1}{n} - 1\Bigr).$$

Using (11), we can approximate this by

$$n^{-1}\sum_{i=1}^{n}\Bigl(\frac{2i-1}{n} - 1\Bigr)\Bigl[F^{-1}\Bigl(\frac{i}{n+1}\Bigr) + n^{-1/2}\frac{B\bigl(i/(n+1)\bigr)}{f\bigl(F^{-1}\bigl(i/(n+1)\bigr)\bigr)}\Bigr],$$

which, interpreted as a Riemann sum, is an approximation to the integral

$$\int_0^1 (2t-1)\Big[F^{-1}(t) + n^{-1/2}\frac{B(t)}{f\big(F^{-1}(t)\big)}\Big]\,\mathrm{d}t.$$

It is easy to check that

$$\mu G = \mu - 2\int_0^\infty y\big(1-F(y)\big)\,\mathrm{d}F(y) = \int_0^1 (2t-1)F^{-1}(t)\,\mathrm{d}t,$$

where the last step follows on changing the integration variable by $t = F(y)$. Thus

$$n^{1/2}(\hat\mu\hat G - \mu G) \rightsquigarrow \int_0^1 \frac{(2t-1)B(t)}{f\big(F^{-1}(t)\big)}\,\mathrm{d}t.$$

The variance of this integral is the asymptotic variance of $\hat\mu\hat G$. From (10), we see that the variance is

$$2\int_0^1 \frac{(2t-1)(1-t)}{f\big(F^{-1}(t)\big)}\int_0^t \frac{s(2s-1)}{f\big(F^{-1}(s)\big)}\,\mathrm{d}s\,\mathrm{d}t.$$

The same change of variables as used to give (12) converts this expression to

$$2\int_0^1 \big(2F(x)-1\big)\big(1-F(x)\big)\int_0^x F(y)\big(2F(y)-1\big)\,\mathrm{d}y\,\mathrm{d}x.$$

This expression has already appeared in the literature; see Zitikis and Gastwirth (2002), theorem 2, where the expression is generalised to teh case of the S-Gini index. Implementing it by replacing F by $\hat F$ is possible, but not as simple as the procedure of the previous section.

4. The Bootstrap

All the approaches we have considered so far lead to more or less tractable expressions for the asymptotic variances of a wide variety of inequality and poverty indices. Thus they make it possible to perform asymptotic inference. This fact suggests that the well-known deficiencies of asymptotic inference in finite samples can be alleviated by use of the bootstrap. Suppose we are interested in an index I with sample counterpart $\hat I$. Bootstrap inference can be based on the asymptotically pivotal function $\hat\tau \equiv (\hat I - I)/\big(\widehat{\mathrm{Var}}(\hat I)\big)^{1/2}$, and such inference should benefit from the asymptotic refinements that come with the use of asymptotic pivots; see Beran (1988) and Horowitz (1997). A **pivot** is a statistic the distribution of which does not depend on unknown parameters; if this is true only of the asymptotic distribution, we have an asymptotic pivot. Since $\hat\tau$ is asymptotically standard normal, it is an asymptotic pivot.

It may be useful to be a bit more specific about two bootstrap procedures, the first the construction of a confidence interval, the second a hypothesis test. Since the sample is assumed to be IID, the most obvious non-parametric bootstrap DGP to use is a straightforward resampling DGP. Thus the CDF of the bootstrap "population" is the EDF, for which the true value of the index is $\hat{I}$. For the j^{th} resample, $j = 1, \ldots, B$, we compute the index I_j^* and its estimated variance $\text{Var}^*(I_j^*)$, and form the approximately pivotal quantity $\tau_j^* \equiv (I_j^* - \hat{I})/\big(\text{Var}^*(I_j^*)\big)^{1/2}$. The empirical distribution of the B bootstrap realisations τ_j^* is our estimate of the distribution of $(\hat{I} - I)/\big(\widehat{\text{Var}}(\hat{I})\big)^{1/2}$.

A confidence interval with nominal coverage $1-\alpha$ is the set of values I_0 for which a test of the hypothesis that $I = I_0$ is not rejected at nominal level α. A two-tailed equal-tailed test based on the statistic $\hat{\tau}$ and the bootstrap estimate of its distribution rejects if $\hat{\tau}$ lies outside the interval between the $\alpha/2$ and $1 - \alpha/2$ quantiles of the bootstrap distribution. These quantiles are estimated by the order statistics $\lfloor \alpha(B+1)/2 \rfloor$ and $\lfloor (1-\alpha/2)(B+1) \rfloor$ of the τ_j^*. (For α an integer percentage, we want to choose B so that $(B+1)/200$ is an integer. The smallest suitable value is thus $B = 199$.) Denote these quantiles by $q^*_{\alpha/2}$ and $q^*_{1-\alpha/2}$. Then the confidence interval is the set

$$\big\{ I_0 \,|\; q^*_{\alpha/2} \le (\hat{I} - I_0)/\big(\widehat{\text{Var}}(\hat{I})\big)^{1/2} \le q^*_{1-\alpha/2} \big\}.$$

If we let $\hat{\sigma}_I$ denote the square root of $\widehat{\text{Var}}(\hat{I})$, we can write this set as the interval

$$\big[\hat{I} - \hat{\sigma}_I q^*_{1-\alpha/2}, \hat{I} - \hat{\sigma}_I q^*_{\alpha/2}\big].$$

This is the standard percentile-t bootstrap confidence interval; see for instance Hall (1988).

Next, consider a test of the hypothesis that $I = I_0$. This can clearly be implemented by computing the statistic $\hat{\tau} \equiv (\hat{I} - I_0)/\big(\widehat{\text{Var}}(\hat{I})\big)^{1/2}$ and using the same bootstrap DGP as for the confidence interval in order to obtain B bootstrap statistics τ_j^*, $j = 1, \ldots, B$. A bootstrap P value for a two-tailed test is then given by the formula

$$P^* = \frac{1}{B} \sum_{j=1}^{B} \mathrm{I}\big(|\tau_j^*| > \hat{\tau}\big),$$

which is just the proportion of bootstrap statistics that are more extreme than $\hat{\tau}$; see, among many other possible references, Davidson and MacKinnon (2006).

But this approach ignores the second Golden Rule of Bootstrapping, as formulated in Davidson (2007), which requires that the bootstrap DGP should be as good as possible an estimate of the true DGP *under the assumption that the null is true*. However, the true value of I for the usual resampling bootstrap DGP is not I_0, but $\hat{I}$. One way to obtain a bootstrap DGP for which the true I is indeed I_0 is to use weighted resampling. In drawing resamples, the observations no longer have equal probabilities of being resampled. Instead, observation i is assigned a weight p_i, which can be found by solving the empirical likelihood maximisation problem

$$\max_{p_i} \sum_{i=1}^{n} \log p_i \quad \text{subject to} \quad \sum_{i=1}^{n} p_i = 1, \quad p_i \ge 0, \text{ and } I(\hat{F}_{\boldsymbol{p}}) = I_0. \tag{13}$$

The notation $I(\hat{F}_{\boldsymbol{p}})$ means the value of the index for the discrete distribution with support at the observations, with probability p_i assigned to observation i. This technique of weighted resampling is explored in Brown and Newey (2002).

Problem (13) sometimes has no solution. This is a pretty sure indicator that any reasonably reliable test would reject the null. Sometimes there is a solution, but with one or more of the probabilities p_i equal to zero. This would correspond to the log of the empirical likelihood function being equal to minus infinity, and, again, it indicates a rejection of the null.

How reliable is inference based on plugin estimators and estimates of their asymptotic variance? Heavy-tailed distributions are notorious for causing problems for both asymptotic and bootstrap inference, and income distributions are notoriously heavy-tailed. I cite here some evidence taken from Davidson (2009a).

In Figure 1 are shown empirical distributions for the statistic τ for the ordinary relative Gini index, with data generated by the Pareto distribution, of which the CDF is $F_{\text{Pareto}}(x) = 1 - x^{-\lambda}$, $x \geq 1$, $\lambda > 1$. The second moment of the distribution is $\lambda/(\lambda - 2)$, provided that $\lambda > 2$, so that, if $\lambda \leq 2$, no reasonable inference about the Gini index is possible. If $\lambda > 1$, the true Gini index is $1/(2\lambda - 1)$. The plots in the figure are for $n = 100$ and $\lambda = 100, 5, 3, 2$. For values of λ greater than about 50, the distribution does not change much, which implies that there is a distortion of the standard error with the heavy tail even if the tail index λ is large.

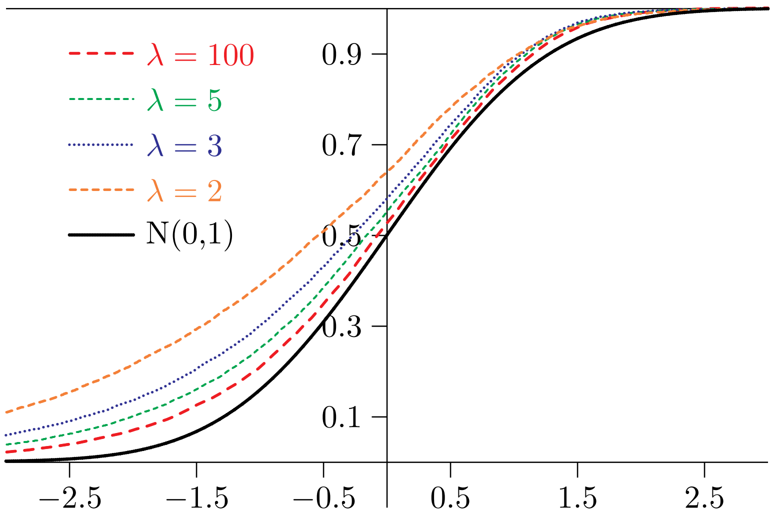

Figure 1. Distribution of τ for the Pareto distribution

Table 1 shows how the bias of τ, its variance, and the greatest absolute deviation of its distribution from standard normal vary with λ.

It is plain that the usual difficulties with heavy-tailed distributions are just as present here as in other circumstances.

The lognormal distribution is not usually considered as heavy-tailed, since it has all its moments. It is nonetheless often used in the modelling of income distributions.

λ	Bias	Variance	Divergence from N(0,1)
100	-0.1940	1.3579	0.0586
20	-0.2170	1.4067	0.0647
10	-0.2503	1.4798	0.0742
5	-0.3362	1.6777	0.0965
4	-0.3910	1.8104	0.1121
3	-0.5046	2.1011	0.1435
2	-0.8477	3.1216	0.2345

Table 1. Summary statistics for Pareto distribution

Since the Gini index is scale invariant, we consider only lognormal variables of the form $e^{\sigma W}$, where W is standard normal. In Figure 2 the distribution of τ is shown for $n = 100$ and $\sigma = 0, 5, 1.0, 1.5$. We can see that, as σ increases, distortion is about as bad as with the genuinely heavy-tailed Pareto distribution. The comparison is perhaps not entirely fair, since, even for the worst case with $\lambda = 2$ for the Pareto distribution, $G = 1/3$. However, for $\sigma = 1$, the index for the lognormal distribution is 0.521, and for $\sigma = 1.5$ there is a great deal of inequality, with $G = 0.711$. For a Pareto distribution and a lognormal one with similar values of G, the distortion is greater with the Pareto.

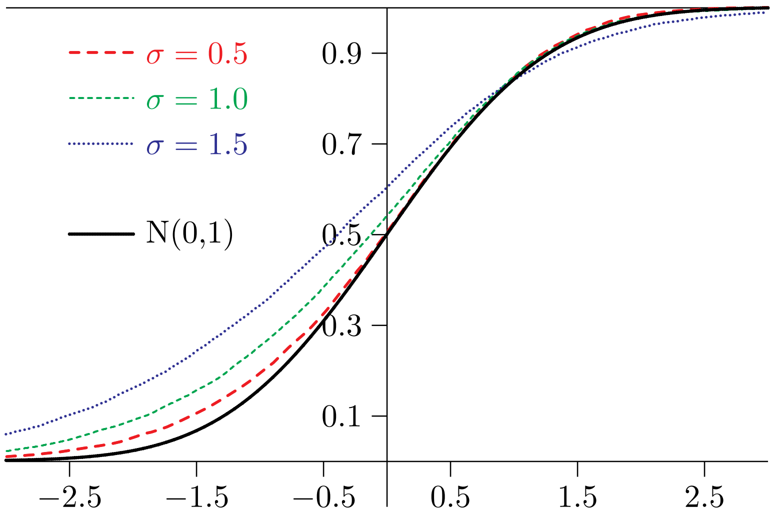

Figure 2. Distribution of τ for the lognormal distribution

What of the bootstrap? In Table 2, coverage rates of percentile-t bootstrap confidence intervals are given for $n = 100$ and for nominal confidence levels from 90% to 99%. The successive rows of the table correspond, first, to the (normally well-behaved) exponential distribution, then to the Pareto distribution for $\lambda = 10, 5, 2$, and finally to the lognormal distribution for $\sigma = 0.5, 1.0, 1.5$. The numbers are based on 10,000 replications with 399 bootstrap repetitions each.

Level	90%	92%	95%	97%	99%
Exponential	0.889	0.912	0.943	0.965	0.989
$\lambda = 10$	0.890	0.910	0.942	0.964	0.984
$\lambda = 5$	0.880	0.905	0.937	0.957	0.982
$\lambda = 2$	0.831	0.855	0.891	0.918	0.954
$\sigma = 0.5$	0.895	0.918	0.949	0.969	0.989
$\sigma = 1.0$	0.876	0.898	0.932	0.956	0.981
$\sigma = 1.5$	0.829	0.851	0.888	0.914	0.951

Table 2. Coverage of percentile-t confidence intervals

Apart from the expected serious distortions when $\lambda = 2$, and when $\sigma = 1.5$, the coverage rate of these confidence intervals is remarkably close to nominal. Given the large distortions reported in Table 1 for small λ or large σ, it seems that, unless the tails are very heavy indeed, or the Gini index itself large, the bootstrap can yield acceptably reliable inference in circumstances in which the asymptotic distribution does not.

5. Stochastic Dominance

Indices are very specific things. Any particular index can be thought of as a welfare index, or alternatively may rank populations in the same way as some social welfare function. This index or welfare function may or may not have desirable ethical properties, however. People have tried to get around this by looking at entire income distributions, or distributions restricted to the poor, rather than focusing on one index or even any finite number of indices. The best-known example of this is when one chooses to study a Lorenz curve rather than just a Gini index.

If we consider two distributions, A and B, say, then we can compute any of the Gini family of indices for both, and determine which is greater. For many purposes, this is enough for an investigator to declare than there is more inequality in the distribution with the greater Gini. But other indices might well lead to the opposite conclusion. If, however, we see that one distribution **Lorenz-dominates** the other, by which we mean that the Lorenz curve of the dominating distribution is everywhere (weakly) above that of the dominated one, then this fact implies that the ranking of the two distributions is unanimous over a wide class of indices.

The notion of stochastic dominance embodies these ideas. We have numerous theorems telling us what class of indices is such that its members unanimously rank two distributions in the same way, under the condition that one stochastically dominates the other at some given order. For instance, generalised Lorenz dominance, which differs from conventional Lorenz dominance in the same way that an absolute Gini index differs from the corresponding relative index, is equivalent to second-order stochastic dominance, as was shown in Foster and Shorrocks (1988).

The issues that have to be faced when performing statistical inference on stochastic dominance are quite different from those that arise with a single index. What has to be estimated from a given sample is an entire function, in a way perfectly analogous to the estimation of a CDF by an EDF. First-order stochastic dominance of a distribution A by another distribution B is the requirement that

$$F_A(y) \geq F_B(y) \quad \text{for all} \quad y \geq 0. \tag{14}$$

The theoretical literature distinguishes weak from strong dominance, but in the statistical context the distinction is meaningless, and so I write all inequalities as weak.

Sample dominance is defined analogously to (14): it requires that $\hat{F}_A(y) \geq \hat{F}_B(y)$ for all $y \geq 0$. What inferences can be drawn about population dominance or non-dominance from the corresponding sample properties? A first point is that one can never reject a hypothesis of dominance in the population if there is dominance in the sample, and, conversely, sample non-dominance can never lead to rejection of population non-dominance. To reject dominance in the population, therefore, the sample non-dominance must be statistically **significant**, and analogously for a rejection of non-dominance.

In order to clarify these issues, it may be useful to consider a very simple case with two distributions A and B with the same support, consisting of three points, $y_1 < y_2 < y_3$. Since $F_A(y_3) = F_B(y_3) = 1$, inference on stochastic dominance can be based on just two quantities, $\hat{d}_i \equiv \hat{F}_A(y_i) - \hat{F}_B(y_i)$, for $i = 1, 2$. Distribution B dominates distribution A if, in the population, $d_i \geq 0$.

Figure 3 shows a two-dimensional plot of $\hat{d}_1$ and $\hat{d}_2$. The first quadrant corresponds to dominance of A by B in the sample. In order to reject a hypothesis of dominance, therefore, the observed $\hat{d}_1$ and $\hat{d}_2$ must lie significantly far away from the first quadrant, for example, in the area marked as "B does not dominate A" separated from the first quadrant by an L-shaped band. For a rejection of non-dominance, on the other hand, the observed sample point must lie "far enough" inside the first quadrant that it is significantly removed from the area of non-dominance, as in the area marked "B dominates A". The zone between the rejection regions for the two possible null hypotheses of dominance and non-dominance corresponds to situations in which neither hypothesis can be rejected. We see that this happens when one of the $\hat{d}_i$ is close to zero and the other is positive. Note also from the figure that inferring dominance by rejecting the hypothesis of non-dominance is more demanding than failing to reject the hypothesis of dominance, since, for dominance, both statistics must have the same sign and be statistically significant.

Tests for dominance and non-dominance can thus be seen as complementary. Positing a null of dominance cannot be used to infer dominance; it can however serve to infer

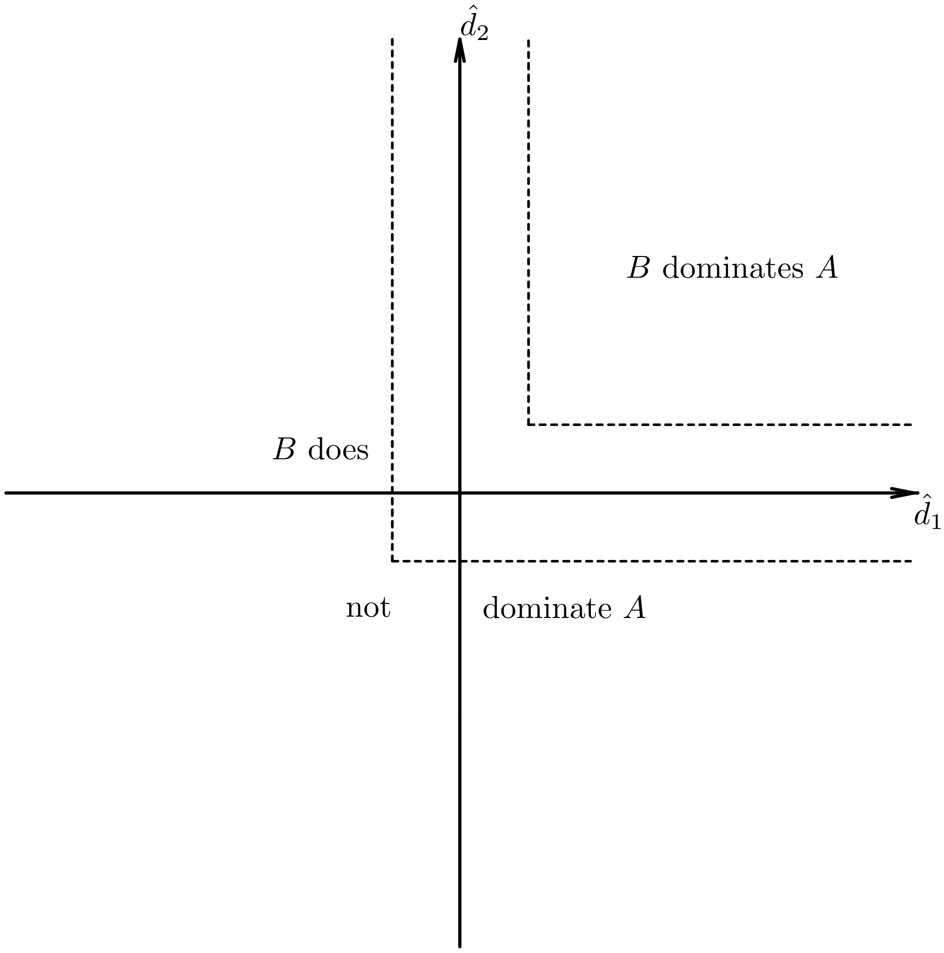

Figure 3: Tests of dominance and non-dominance

non-dominance. Positing a null of non-dominance cannot serve to infer non-dominance; it can however lead to inferring dominance.

In Kaur, Prakasa Rao, and Singh (1994), a test is proposed based on the minimum of the t statistic for the hypothesis that $F_A(z) - F_B(z) = 0$, computed for each value of z in some closed interval contained in the interior of the union of the supports of A and B. The minimum value is used as the test statistic for the null of non-dominance against the alternative of dominance. The test can be interpreted as an **intersection-union test**. It is shown that the probability of rejection of the null when it is true is asymptotically bounded by the nominal level of a test based on the standard normal distribution.

When the null is dominance, a technique originally designed by Richmond (1982) can be used; see also Beach and Richmond (1985). It provides simultaneous confidence intervals for a set of variables asymptotically distributed as multivariate normal with known or consistently estimated asymptotic covariance matrix. It was extended by Bishop, Formby, and Thistle (1992), who suggested a **union-intersection test** of the hypothesis that one set of Lorenz curve decile ordinates dominates another. For a test of stochastic dominance, one can use the t statistics for the hypotheses that the individual differences d_j, $j = 1, \ldots, m$, for a set of m points are zero. The null hypothesis, which implies that they are all non-negative, is rejected against the alternative if any of the t statistics is significant with the wrong sign (that is, in the direction of dominance of B by A), where significance is determined asymptotically by the critical values of the Studentised Maximum Modulus (SMM) distribution with m and an infinite number of degrees of freedom. If the distributions being compared are continuous, empirical-process methods can be used to find the asymptotic distribution of the most extreme t statistic, or of the maximum difference between the two CDFs; see among others Linton, Maasoumi, and Whang (2005). Usually, though, simulation is needed to obtain asymptotic critical values.

Both the intersection-union and union-intersection tests rely on conventional t statistics for their implementation. Asymptotic inference makes use of their asymptotic distributions, which may be poor approximations in finite samples, and so one is led to consider bootstrap tests of dominance or non-dominance. Most, but by no means all, empirical studies of stochastic dominance adopt a **distribution-free** approach, with no parametric specification of the distributions studied. This fact implies that one wants a non-parametric bootstrap DGP. The Golden Rules of bootstrapping assert that, for reliable inference, the bootstrap DGP must satisfy the null, and that it should be as good an estimate as possible of the true DGP, assuming that the latter itself satisfies the null. It can be somewhat of a challenge to satisfy these rules in tests of stochastic dominance.

Suppose first that the null is dominance. If there is dominance in the sample, we cannot reject the null, and we are finished. If not, then we cannot use an ordinary resampling DGP for the bootstrap, since it does not satisfy the null. A similar situation arises with a test of non-dominance. Once again, a viable solution is to solve an empirical-likelihood problem in order to obtain unequal probabilities that put the weighted resampling DGP back into the null. But the nulls of either dominance or non-dominance are huge. General considerations show that, in order to maintain asymptotic control of Type I error, the bootstrap DGP should lie on the *frontier* of the set of DGPs that represents the null. A frontier DGP is one where there is dominance of one function by the other except at one point, at which the two functions touch. Which one of the frontier DGPs can then be determined by maximising the empirical likelihood. This problem is extensively explored in an as yet unpublished paper, available as a working paper: Davidson and Duclos (2006), and in Davidson (2009b).

Duclos and I have strongly advocated using a null of non-dominance, since, when it is rejected, we can draw the strong conclusion of dominance. But here again the possibility of heavy tails limits our ability to carry out tests of unrestricted non-dominance, on account of the phenomenon of micronumerosity, a term invented by the late Arthur Goldberger; see Kiefer (1989). In an extended right-hand tail, observations are few and far between, and so it is most unlikely that any statistically significant effect can be found there. But an intersection-union test looks at the *least extreme* of the set of statistics on which it is based, and so can never reject if any of them is insignificant. The way out of this impasse, if one wishes to have the possibility of inferring some sort of dominance, is to limit attention to some interval in the middle of the distributions. Restricted dominance or non-dominance is defined only on this interval. Of course, if a hypothesis of restricted non-dominance is rejected, all we can infer is dominance restricted to the chosen interval. Another approach is to *estimate* the interval over which dominance can be inferred.

There is an important relation between stochastic dominance and measures of poverty. For a full discussion of the statistical issues involved, see Davidson and Duclos (2000), and the copious bibliography in that paper.

6. Measures of Divergence

A question still more general than that of stochastic dominance is whether two distributions differ significantly in any way. In addition to questions relating to differences in distributions across time and space, it is often useful to compare an empirical distribution with a parametric family of distributions, as a sort of goodness-of-fit test. More than a quarter of a century ago, Frank Cowell proposed a measure of the divergence between two distributions on the basis of an axiomatic approach; see Cowell (1980a) and Cowell (1980b). It can be written as

$$J_\alpha(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{n\alpha(\alpha-1)} \sum_{i=1}^{n} \left(\left[\frac{x_{(i)}}{\hat{\mu}_1} \right]^{\alpha} \left[\frac{y_{(i)}}{\hat{\mu}_2} \right]^{1-\alpha} - 1 \right), \tag{15}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are vectors the elements of which are observations from the two populations to be compared, the $x_{(i)}$ and $y_{(i)}$ are the order statistics, $\hat{\mu}_1$ and $\hat{\mu}_2$ are the means of the elements of $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively, and the parameter α takes any real value. Note that $J_\alpha(\boldsymbol{x}, \boldsymbol{y}) \geq 0$ for arbitrary $\boldsymbol{x}$ and $\boldsymbol{y}$.

In this section, I limit attention to a comparison of a sample, the observations of which are the elements of $\boldsymbol{x}$, and a single specified distribution F. The "order statistics" $y_{(i)}$ are the non-random quantities $F^{-1}\big(i/(n+1)\big)$, and $\hat{\mu}_2$ is not the expectation of distribution F, but rather the mean of the $y_{(i)}$ so defined. The question now is: what is the distribution of $J_\alpha(\boldsymbol{x}, \boldsymbol{y})$ when $\boldsymbol{x}$ is indeed an IID sample from F?

The delta method gives no immediate answer to the question, because it turns out that J_α is quadratic in $\hat{F} - F$, and so the first derivatives used by the conventional form of the delta method vanish. A second-order expansion can be worked out, and it shows that the asymptotic null distribution of J_α, if it exists, depends on F. This means that J_α is not an asymptotically pivotal quantity, and so is ill-adapted to bootstrapping. Further, it is not at all clear how to deduce the asymptotic distribution on the basis of the second-order expansion.

Re-enter the empirical process approach, which does give us something useful. It is not surprising that what it gives us is not at all the asymptotically normal distribution that we found with the Gini indices, but rather something that requires simulation in order to be evaluated. It also makes clear that the very existence of the asymptotic distribution of J_α is not guaranteed. In fact, even if F is the well-behaved exponential distribution, the distribution of J_α diverges as $n \to \infty$. With heavy-tailed distributions, things are still worse. Only when a distribution has compact support is one sure of the existence of the asymptotic distribution.

Let $x_{(i)} = y_{(i)} + z_i$, where, under the null, $z_i = O_p(n^{-1/2})$. Specifically,

$$n^{1/2} z_i \rightsquigarrow \frac{B\big(i/(n+1)\big)}{f\big(F^{-1}(1/(n+1))\big)}, \tag{16}$$

where $f = F'$ is the density. Now

$$\sum_{i=1}^n x_{(i)}^\alpha y_{(i)}^{1-\alpha} = \sum_{i=1}^n y_{(i)}(1 + z_i/y_{(i)})^\alpha.$$

Let $\mu = n^{-1}\sum y_{(i)}$ – this corresponds to μ_2 in (15). Then

$$\sum_{i=1}^n (x_{(i)}^\alpha y_{(i)}^{1-\alpha} - \hat\mu^\alpha \mu^{1-\alpha}) = \sum_{i=1}^n y_{(i)}(1 + z_i/y_{(i)})^\alpha - n\mu\Big(1 + (n\mu)^{-1}\sum_{i=1}^n z_i\Big)^\alpha.$$

By the binomial theorem,

$$n^{-1}\sum_{i=1}^n y_{(i)}(1 + z_i/y_{(i)})^\alpha = \mu + \alpha\sum_{n=1}^n z_i + n^{-1}\frac{\alpha(\alpha-1)}{2}\sum_{i=1}^n \frac{z_i^2}{y_{(i)}} + O_p(n^{-3/2}),$$

and

$$\Big(1 + \frac{1}{n\mu}\sum_i z_i\Big)^\alpha = 1 + \frac{\alpha}{n\mu}\sum_i z_i + \frac{\alpha(\alpha-1)}{2n^2\mu^2}\Big(\sum_i z_i\Big)^2 + O_p(n^{-3/2}).$$

Then, since $\hat\mu^\alpha \mu^{1-\alpha} = \mu + O_p(n^{-1/2})$, we see that

$$nJ_\alpha = \frac{1}{2\mu}\Big(\sum_i (z_i^2/y_{(i)}) - \frac{1}{n\mu}\Big(\sum_i z_i\Big)^2\Big) + o_p(1). \tag{17}$$

It appears that it is nJ_α that is $O_p(1)$ as $n \to \infty$, not J_α itself. This is due to the vanishing of all first-order terms in the expansions that lead to (17).

Expression (17) could be considered as the result of an unconventional, second-order, application of the delta method. But now we need the result (11) from the empirical-process approach in order to study the asymptotic distribution of (17). From (16), we see that

$$\sum_i z_i^2/y_{(i)} \rightsquigarrow \int_0^1 \frac{B^2(t)\,\mathrm{d}t}{F^{-1}(t) f^2\big(F^{-1}(t)\big)}$$

and

$$\frac{1}{n\mu}\Bigl(\sum_i z_i\Bigr)^2 \rightsquigarrow \frac{1}{\mu}\Bigl(\int_0^1 \frac{B(t)\,\mathrm{d}t}{f\bigl(F^{-1}(t)\bigr)}\Bigr)^2,$$

so that

$$nJ_\alpha \rightsquigarrow \frac{1}{2\mu}\left[\int_0^1 \frac{B^2(t)\,\mathrm{d}t}{F^{-1}(t)f^2\bigl(F^{-1}(t)\bigr)} - \frac{1}{\mu}\Bigl(\int_0^1 \frac{B(t)\,\mathrm{d}t}{f\bigl(F^{-1}(t)\bigr)}\Bigr)^2\right]. \tag{18}$$

It is clear that this asymptotic expression depends on F. Interestingly, it does *not* depend on α. However, it is not clear that the integrals are convergent. Indeed, for the exponential distribution, for which $F(y) = 1 - \mathrm{e}^{-y}$, $y \geq 0$, it can be seen that the expectation of (18), which does not in fact exist, would involve a divergent integral. On the other hand, the uniform distribution U(0,1), with $F(y) = y$, $0 \leq y \leq 1$, which has compact support, leads to a well-defined limiting variable for nJ_α.

When heavy tails are a problem, a recommendation that is sometimes made is to base all inference not on moments, but on quantiles. That turns out to solve the problem we have just encountered. If the x_i are drawings from the distribution F, then the quantities $F(x_i)$ follow the U(0,1) distribution, and the $F(x_{(i)})$ have the joint distribution of the order statistics of an IID sample from U(0,1). We can modify the definitions so far made as follows. Instead of the $x_{(i)}$, we use the quantities $F(x_{(i)})$. The sample mean $\hat{\mu}$ becomes the mean of the $F(x_{(i)})$. The "order statistic" $y_{(i)}$, previously defined as $F^{-1}\bigl(i/(n+1)\bigr)$ now becomes simply $t_i \equiv i/(n+1)$, since the CDF of U(0,1) is the identity function on $[0,1]$. The density f is uniformly equal to 1 on $[0,1]$, and F^{-1} is, like F, the identity function. Finally, μ, the mean of the $y_{(i)}$, is now

$$\mu = n^{-1}\sum_{i=1}^n t_i = \frac{1}{n(n+1)}\sum_{i=1}^n i = \frac{1}{(n(n+1)}\frac{n(n+1)}{2} = \frac{1}{2}.$$

With these modifications, what we can call the quantile version of the statistic becomes

$$j_\alpha = \frac{1}{\alpha(\alpha-1)}\sum_{i=1}^n \left(\left[\frac{F(x_{(i)}}{\hat{\mu}}\right]^\alpha (2t_i)^{1-\alpha} - 1\right), \tag{19}$$

where $\hat{\mu} = n^{-1}\sum_i F(x_{(i)})$. Its distribution converges to that of (18) with our modifications, which is

$$\int_0^1 \frac{B^2(t)\,\mathrm{d}t}{t} - 2\Bigl(\int_0^1 B(t)\,\mathrm{d}t\Bigr)^2. \tag{20}$$

Although the first integral above looks as though it might diverge at 0, it does not, since $\mathrm{E}B^2(t) = t(1-t)$. Thus

$$\mathrm{E}\int_0^1 \frac{B^2(t)\,\mathrm{d}t}{t} = \int_0^1 (1-t)\,\mathrm{d}t = \frac{1}{2}.$$

The integral $\int_0^1 B(t)\,\mathrm{d}t$ is normally distributed with expectation 0 and variance

$$2\int_0^1 (1-t)\int_0^t s\,\mathrm{d}s\,\mathrm{d}t = \frac{1}{12}.$$

Thus the expectation of the distribution of (20) is $1/2 - 2/12 = 1/3$. The quantile version of the statistic is now asymptotically pivotal, with an asymptotic distribution that can be evaluated by simulation using (20). In fact, more is true. Since under the null the quantities $F(x_{(i)})$ are IID from U(0,1), $\hat{\mu}$ being their mean, the distribution of (19) depends only on α, although that of (20) does not. Thus, for given α, j_α is actually an exact pivot in finite samples. The finite-sample distribution can therefore be estimated with arbitrary accuracy by simulation, and so exact finite-sample inference is possible. Simulation-based testing using a pivotal statistic is in fact much older than bootstrapping. Such tests are called **Monte Carlo tests**, and were introduced back in the 1950s; see Dwass (1957), and also Dufour and Khalaf (2001) for a more recent discussion. They are implemented in exactly the same way as a parametric bootstrap; see Davidson and MacKinnon (2006).

This fact is less useful than one might hope, because in practice, one usually wants a measure of divergence from a parametric family of distributions rather than from a single distribution. The appropriate modification of the test is to estimate the parameters of the family, preferably but not necessarily by maximum likelihood, and to use the parametric estimate in place of the single distribution F. The result may be asymptotically pivotal, but is not pivotal in finite samples. Now the existence of the asymptotic distribution justifies the use of the bootstrap for testing. Since the null is constituted by a parametric family of distributions, a parametric bootstrap is appropriate – a valid resampling bootstrap would be hard to design. The procedure is as follows for a parametric family $F(\cdot;\boldsymbol{\theta})$.

- Use the sample $\boldsymbol{x}$ to obtain an estimate $\hat{\boldsymbol{\theta}}$ of the parameter vector.
- Sort the elements of $\boldsymbol{x}$ and transform them to obtain the order statistics $F(x_{(i)};\hat{\boldsymbol{\theta}})$.
- Compute the statistic j_α using the $F(x_{(i)};\hat{\boldsymbol{\theta}})$ as the $F(x_{(i)})$.

For each of B bootstrap repetitions:

- Generate a bootstrap sample $\boldsymbol{x}^*$ of size n of IID drawings from $F(\cdot;\hat{\boldsymbol{\theta}})$.
- Obtain the estimated parameters $\boldsymbol{\theta}^*$ from the sample $\boldsymbol{x}^*$.
- Construct the set $F(x^*_{(i)};\boldsymbol{\theta}^*)$ of order statistics just as with the original data.
- Compute the bootstrap statistic j^*_α as with the original data.

Then:

- The bootstrap P value is the proportion of the j^*_α greater than j_α.

It makes sense, of course, to use a one-sided test, since j_α is non-negative and rejects only to the right.

The behaviour of this quantile version of the divergence measure when it is bootstrapped is ongoing research joint with Sanghamitra Bandyopadhyay, Frank Cowell, and Emmanuel Flachaire.

7. Concluding Remarks

The empirical study of income distribution calls for the use of a wide panoply of statistical techniques. In this paper, I have reviewed some of these, stressing the roles of three main approaches to the development of asymptotic results, namely the delta method, the use of influence functions, and the empirical-process approach. Regarding the implementation of inferential techniques, I emphasise the role of the bootstrap in obtaining reliable results, and discuss the details of a couple of bootstrap procedures. It is necessary in any empirical investigation, though, to be aware that the presence of heavy right-hand tails may pose serious problems for any statistical inference. I propose one quantile-based method that can to some extent alleviate this problem.

It is my hope that this paper illustrates convincingly the potential of present-day techniques for inference on income distributions, and to point out the numerous as yet unsolved research problems that await us.

Appendix

We show here that the expression on the right-hand side of (12) is the variance of the distribution F. If X is a random variable that follows this distribution, then the variance is $\mathrm{E}X^2 - (\mathrm{E}X)^2$, which can be written explicitly as

$$\int_0^\infty x^2 \,\mathrm{d}F(x) - \Big(\int_0^\infty x \,\mathrm{d}F(x)\Big)^2 \tag{21}$$

I consider only a distribution defined on the positive real line. For a distribution on the whole real line, the following proof would require separate consideration of the positive and negative parts of the distribution. The treatment of the negative part is quite similar to that of the positive part, and so is omitted.

The second term in (21) is

$$\int_0^\infty x \,\mathrm{d}F(x) \int_0^\infty y \,\mathrm{d}F(y) = 2\int_0^\infty x \int_0^x y \,\mathrm{d}F(y)\,\mathrm{d}F(x),$$

by the symmetry of x and y. Thus the variance is

$$\int_0^\infty \Big(x - 2\int_0^x y \,\mathrm{d}F(y)\Big) x \,\mathrm{d}F(x). \tag{22}$$

Now

$$\int_0^\infty x^2 \,\mathrm{d}F(x) = -\int_0^\infty x^2 \,\mathrm{d}\big(1 - F(x)\big) = 2\int_0^\infty \big(1 - F(x)\big) x \,\mathrm{d}x, \tag{23}$$

where the second equality results from an integration by parts. Another integration by parts shows that

$$\int_0^x y\,\mathrm{d}F(y) = xF(x) - \int_0^x F(y)\,\mathrm{d}y. \tag{24}$$

Putting (23) and (24) into (22) gives for the variance 2 times

$$\int_0^\infty x\big(1-F(x)\big)\,\mathrm{d}x - \int_0^\infty x\Big(xF(x) - \int_0^x F(y)\,\mathrm{d}y\Big)\,\mathrm{d}F(x). \tag{25}$$

Consider the remaining double integral in the above expression. It is

$$\begin{aligned}\int_0^\infty x\int_0^x F(y)\,\mathrm{d}y\,\mathrm{d}F(x) &= -\int_0^\infty x\int_0^x F(y)\,\mathrm{d}y\,\mathrm{d}\big(1-F(x)\big)\\ &= \int_0^\infty \big(1-F(x)\big)\,\mathrm{d}\Big[x\int_0^x F(y)\,\mathrm{d}y\Big] \quad \text{(integration by parts)}\\ &= \int_0^\infty \big(1-F(x)\big)\int_0^x F(y)\,\mathrm{d}y\,\mathrm{d}x + \int_0^\infty \big(1-F(x)\big)xF(x)\,\mathrm{d}x. \end{aligned} \tag{26}$$

The first term in the last line above times 2 is exactly expression (12).

Adding the first term in (25) to the second term in (26) gives

$$\int_0^\infty x\big(1-F(x)\big)\big(1+F(x)\big)\,\mathrm{d}x = \int_0^\infty x\big(1-F^2(x)\big)\,\mathrm{d}x. \tag{27}$$

In (25), we still need to take account of the term

$$-\int_0^\infty x^2F(x)\,\mathrm{d}F(x) = \tfrac{1}{2}\int_0^\infty x^2\,\mathrm{d}\big(1-F^2(x)\big) = -\int_0^\infty \big(1-F^2(x)\big)x\,\mathrm{d}x,$$

with another integration by parts. But this cancels with (27), and so the variance is indeed just

$$2\int_0^\infty \big(1-F(x)\big)\int_0^x F(y)\,\mathrm{d}y\,\mathrm{d}x,$$

as we wished to show.

References

Anderson, G. (1996). “Nonparametric Tests of Stochastic Dominance In Income Distributions”, *Econometrica*, **64**, 1183-1193.

Andrews, D. W. K. (1994). *Handbook of Econometrics*, Vol. 4, eds R. F. Engle and D. L. McFadden, Chapter 37, Elsevier.

Atkinson, A. B. (1970). “On the Measurement of Inequality”, *Journal of Economic Theory*, **2**, 244–263.

Barrett, G. F. and S. G. Donald (2009). “Statistical Inference with Generalized Gini Indices of Inequality, Poverty, and Welfare”, *Journal of Business and Economic Statistics*, **27**, 1–17.

Barrett, G. F. and K. Pendakur (1995). “The Asymptotic Distribution of the Generalized Gini Indices of Inequality”, *Canadian Journal of Economics*, **28**, 1042–1055.

Beach, C. M. and R. Davidson (1983). “Distribution-Free Statistical Inference with Lorenz Curves and Income Shares”, *Review of Economic Studies*, **50**, 723–735.

Beach, C. M. and J. Richmond (1985). “Joint Confidence Intervals for Income Shares and Lorenz Curves”, *International Economic Review*, **26**, 439–450.

Beran, R., 1988. “Prepivoting test statistics: A bootstrap view of asymptotic refinements”, *Journal of the American Statistical Association*, Vol. 83, pp. 687–697.

Bhattacharya, D. (2007). “Inference on inequality from household survey data”, *Journal of Econometrics*, Vol. 137, pp. 674–707.

Bishop, J. A., S. Chakraborti and P. D. Thistle (1989). “Asymptotically Distribution–Free Statistical Inference for Generalized Lorenz Curves”, *The Review of Economics and Statistics*, **71**, 725–727.

Bishop, J. A, J. P. Formby, and P. Thistle (1992). “Convergence of the South and Non-South Income Distributions, 1969–1979”, *American Economic Review*, **82**, 262–272.

Blackorby, C. and D. Donaldson (1978). “Measures of Relative Equality and Their Meaning in Terms of Social Welfare”, *Journal of Economic Theory*, **18**, 59–80.

Blackorby, C. and D. Donaldson (1980). “Ethical Indices for the Measurement of Poverty”, *Econometrica*, **48**, 1053–1062.

Brown, B. W. and W. Newey (2002). “Generalized method of moments, efficient bootstrapping, and improved inference”, *Journal of Business and Economic Statistics*, 20, 507–517.

Chow, K. V. (1989). *Statistical Inference for Stochastic Dominance: a Distribution Free Approach*, Ph.D. thesis, University of Alabama.

Cowell, F. A.. (1980a). “Generalized Entropy and the Measurement of Distributional Change”, *European Economic Review*, **13**, 147–159.

Cowell, F. A. (1980b). “On the Structure of Additive Inequality Measures”, *Review of Economic Studies*, **47**, 521–531.

Cowell, F. A. (1989). "Sampling Variance and Decomposable Inequality Measures", *Journal of Econometrics*, **42**, 27–41.

Dardanoni, V. and A. Forcina (1999). "Inference for Lorenz Curve Orderings", *Econometrics Journal*, **2**, 48–74.

Davidson, R. (2007). "Bootstrapping Econometric Models", *Quantile*, **3**, 13–36 (in Russian). English version available at
`http://russell-davidson.arts.mcgill.ca/articles/quantile.pdf`

Davidson, R. (2009a). "Reliable Inference for the Gini Index", *Journal of Econometrics*, **150**, 30–40.

Davidson, R. (2009b). "Testing for Restricted Stochastic Dominance: Some Further Results", *Review of Economic Analysis*, **1**, 34–59.

Davidson, R. and J.-Y. Duclos (2000). "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality", *Econometrica*, **68**, 1435-1464.

Davidson, R. and and J.-Y. Duclos (2006). "Testing for Restricted Stochastic Dominance", IZA discussion paper #2047.

Davidson, R. and J. G. MacKinnon (2006). "Bootstrap Methods in Econometrics", Chapter 23 of *Palgrave Handbook of Econometrics*, Volume 1, *Econometric Theory*, eds T. C. Mills and K. Patterson, Palgrave-Macmillan, London.

Deltas, G. (2003). "The small-sample bias of the Gini coefficient: results and implications for empirical research", *The Review of Economics and Statistics*, Vol. 85, 226–234.

Dufour, J.-M., and L. Khalaf (2001). "Monte Carlo test methods in econometrics", Ch. 23 in *A Companion to Econometric Theory*, ed. B. Baltagi, Oxford, Blackwell Publishers, pp. 494–519.

Dwass, M. (1957). "Modified randomization tests for nonparametric hypotheses", *Annals of Mathematical Statistics*, 28, 181–187.

Foster, J.E. and A.F. Shorrocks (1988a). "Poverty Orderings", *Econometrica*, **56**, 173–177.

Gastwirth, J.L. (1974). "Large Sample Theory of Some Measures of Economic Inequality", *Econometrica*, **42**, 191–196.

Gini, C. (1912). "Variabilità e mutabilità", reprinted in *Memorie di metodologica statistica*, eds. E. Pizetti and T. Salvemini. Rome: Libreria Eredi Virgilio Veschi (1955).

Hall, P. (1988). "Theoretical Comparison of Bootstrap Confidence Intervals", *Annals of Statistics*, **16**, 927–953.

Horowitz, J. L. (1997). "Bootstrap Methods in Econometrics: Theory and Numerical Performance", in David M. Kreps and Kenneth F. Wallis, eds., *Advances in Economics and Econometrics: Theory and Applications*, Vol. 3, pp. 188–222, Cambridge, Cambridge University Press.

Kaur, A., B. L. Prakasa Rao, and H. Singh (1994). "Testing for Second-Order Stochastic Dominance of Two Distributions", *Econometric Theory*, **10**, 849–866.

Kiefer, N. M, (1989). "The ET Interview: Arthur S. Goldberger", *Econometric Theory*, **5**, 133–160.

Linton, O., E. Maasoumi, and Y-J Whang (2005). "Consistent Testing for Stochastic Dominance under General Sampling Schemes", *Review of Economic Studies*, **72**, 735–765.

McFadden, D. (1989). "Testing for Stochastic Dominance" in *Studies in the Economics of Uncertainty*, eds. T. B. Fomby and T.K. Seo, Springer-Verlag, New York.

Preston, I. (1995). "Sampling Distributions of Relative Poverty Statistics", *Applied Statistics*, **44**, 91–99.

Richmond, J. (1982). "A General Method for Constructing Simultaneous Confidence Intervals", *Journal of the Americal Statistical Association*, **77**, 455–460.

Sen A. (1976). "Poverty: an ordinal approach to measurement", *Econometrica*, **44**, 219–231.

Xu, K. (2007). "U-Statistics and their asymptotic results for some inequality and poverty measures", *Econometric Reviews*, Vol. 26, 567–577.

Zitikis, R. and J. L. Gastwirth (2002). "The Asymptotic Distribution of the S-Gini Index", *Australian & New Zealand Journal of Statistics*, **44**, 439–446.