Statistics & Probability Letters 4 (1986) 47–49
North-Holland

January 1986

LORENZ ORDERING OF MEANS AND MEDIANS

Barry C. ARNOLD
University of California, Riverside, CA, USA

Jose A. VILLASEÑOR
Colegio de Postgraduados, Chapingo, Mexico

Received February 1985
Revised April 1985

Abstract: Increasing sample size decreases the inequality of the sample mean as measured by the Lorenz order. A similar result occurs in the case of sample medians from symmetric distributions. In certain cases the sample median can be expected to be Lorenz ordered with respect to the sample mean. The problem of determining sufficient conditions for such Lorenz ordering remains open.

Keywords: Lorenz order, density crossings, U-statistics, means, medians.

1. Introduction

Let $X_1, X_2, \ldots$ be i.i.d. random variables with common distribution function F. Denote the corresponding order statistics by $X_{i:n}$, $i = 1, 2, \ldots, n$, $n = 1, 2, \ldots$ and the corresponding means by $\bar{X}_n = (\sum_{i=1}^n X_i)/n$, $n = 1, 2, \ldots$. In many situations it is common knowledge that the sample means are more stable, in the sense of having smaller variance, than the corresponding sample medians. In addition, variability typically decreases as sample size increases. An alternative scale free variability ordering is that provided by the Lorenz order. In the present note we present some preliminary results relating to the Lorenz ordering on sample medians and means. In order to have a well defined Lorenz curve we restrict attention to non-negative random variables with positive finite expectations.

2. Definitions

Let $\mathscr{L}$ be the class of all non-negative random variables with positive finite expectations. For a random variable X in $\mathscr{L}$ with distribution function F_X we define its inverse distribution function F_X^{-1} by

$$F_X^{-1}(y) = \sup\{x\colon F_X(x) \leqslant y\}. \tag{1}$$

The Lorenz curve associated with X is then defined by

$$L_X(u) = \left[\int_0^u F_X^{-1}(y)\,\mathrm{d}y\right] \Big/ \left[\int_0^1 F_X^{-1}(y)\,\mathrm{d}y\right], \quad u \in [0, 1]. \tag{2}$$

The Lorenz partial order, $\leqslant_L$, on $\mathscr{L}$ is defined by

$$X \leqslant_L Y \iff L_X(u) \geqslant L_Y(u), \quad \forall u \in [0, 1].$$

If $X \leqslant_L Y$, then X exhibits less inequality than Y in the Lorenz sense.

There are several sufficient conditions for Lorenz ordering scattered in the literature. A brief list, together with examples, may be found in Arnold and Villaseñor (1984). The following conditions will be used in the present note.

Theorem 1 (Hardy, Littlewood and Polya (1929)). *Suppose that X, $Y \in \mathscr{L}$ and $E(X) = E(Y)$. A necessary and sufficient condition that $X \leqslant_L Y$ is that $E[h(X)] \leqslant E[h(Y)]$ for every convex function h for which the expectations exist.*

Note. This result is included in Theorem 10 of Hardy, Littlewood and Polya (1929, p. 152), provided we write $E[h(X)]$ as $\int_0^1 h(F_X^{-1}(u))\,du$ and $E[h(Y)]$ as $\int_0^1 h(F_Y^{-1}(u))\,du$.

Theorem 2 (Strassen (1965)). *Suppose X, $Y \in \mathscr{L}$ and $E(X) = E(Y)$. A necessary and sufficient condition that $E[h(X)] \leqslant E[h(Y)]$ for every convex h for which the expectations exist, is that there exist random variables Y' and Z' defined on some probability space such that $Y \overset{d}{=} Y'$ and $X \overset{d}{=} E(Y' \mid Z')$. (Here $\overset{d}{=}$ denotes 'has the same distribution as'.)*

Note. The correspondence with Strassen's more general theorem hinges on the fact that $(E(Y' \mid Z'), Y')$ is an (albeit short) martingale sequence.

Theorem 3 (Shaked (1980)). *Suppose the X and Y are absolutely continuous members of $\mathscr{L}$ with $E(X) = E(Y)$ and densities f_X and f_Y. A sufficient conditions for $X \leqslant_L Y$ is that $f_X(x) - f_Y(x)$ changes sign twice on $(0, \infty)$ and the sequence of signs for $f_X - f_Y$ is $-, +$.*

Note. Theorem 3 can be argued as follows. The given sign change sequence for $f_X - f_Y$ implies that $F_X - F_Y$ changes sign once with sign sequence $-, +$. This implies $F_X^{-1} - F_Y^{-1}$ changes sign once with sign sequence $+, -$ from which Lorenz ordering is readily deduced.

Observe that the Lorenz ordering is invariant under scale changes, i.e. $X \leqslant_L Y \Leftrightarrow cX \leqslant_L dY$ for any $c, d > 0$. Theorems 1, 2 and 3 apply to random variables with equal means. In practice we frequently apply them to $E(Y)X$ and $E(X)Y$ or $X/E(X)$ and $Y/E(Y))$ in order to have the equal means condition automatically satisfied.

3. Sample means

Our first observation is that it is indeed true that sample means are Lorenz ordered.

Theorem 4. *For any n, $\overline{X}_n \leqslant_L \overline{X}_{n-1}$.*

Proof. By exchangeability $E(X_i \mid X_1 + \cdots + X_n) = (X_1 + \cdots + X_n)/n = \overline{X}_n$ for every i and consequently $E(X_1 + \cdots + X_{n-1} \mid X_1 + \cdots + X_n) = (n-1)\overline{X}_n$. Thus $E(\overline{X}_{n-1} \mid \overline{X}_n) + \overline{X}_n$ and by Theorems 1 and 2, since $E(\overline{X}_{n-1}) = E(\overline{X}_n)$, we have $\overline{X}_n \leqslant_L \overline{X}_{n-1}$.

Note. The argument does not require that the X_i's be i.i.d., merely that they be exchangeable. The argument extends readily to cover the case of any U-statistic. Thus if $g(x_1, x_2, \ldots, x_m)$ is a function of m variables such that $E(g(X_1, \ldots, X_m))$ exists and is positive then, if we define the U-statistic

$$U_n = \frac{(n-m)!}{n!} \sum g(X_{i_1}, X_{i_2}, \ldots, X_{i_m})$$

where the summation is over all permutations of the n subscripts of the X_i's taken m at a time, we find as above

$U_n \leqslant_L U_{n-1}$.

4. Sample medians

In Arnold and Villaseñor (1984) it was shown that sample medians are not necessarily Lorenz ordered. In order to avoid potential ambiguities of definition we restrict attention to odd sample sizes. A sufficient condition is symmetry. We have from Arnold and Villaseñor (1984)

Theorem 5. *If the common density of the X_i's is symmetric on the interval $[0, c]$ then $X_{n+2:2n+3} \leqslant_L X_{n+1:2n+1}$, for any n.*

Proof. By symmetry we have $E(X_{n+2:2n+3}) = E(X_{n+1:2n+1})$. The ratio of the density of $X_{n+2:2n+3}$ and $X_{n+1:2n+1}$ is $F(x)(1 - F(x))(4n+6)/(n+1)$. This is clearly greater than 1 for intermediate values of x and the result follows from Theorem 3.

5. Medians and means

For a fixed sample size we may ask about the relative variability (as measured by the Lorenz order) of the sample mean and the sample median. In the case of an exponential parent distribution we have

$$\overline{X}_{2n+1} \leqslant_L X_{n+1:2n+1}, \quad n = 1, 2, \ldots. \tag{5.1}$$

To verify this result we may, without loss of generality, assume $E(\overline{X}_{2n+1}) = 1$. Then

$$\begin{aligned} E(X_{n+1:2n+1}) &= \frac{1}{n+1} + \frac{1}{n+2} \\ &\quad + \cdots + \frac{1}{2n+1} \\ &= \alpha_n < 1. \end{aligned}$$

The ratio of the densities of $\overline{X}_{2n+1}$ and $\alpha_n^{-1} \times X_{n+1:2n+1}$ is proportional to

$$\left[\frac{e^{\gamma x} - e^{\delta x}}{x^2}\right]^{-n} \tag{5.2}$$

where

$$\gamma = \left(2 + \frac{1}{n}\right) - \left(1 + \frac{1}{n}\right)\alpha_n \quad \text{and}$$

$$\delta = \left(2 + \frac{1}{n}\right) - \left(2 + \frac{1}{n}\right)\alpha_n.$$

Since $\gamma > \delta > 0$ it is readily verified that (5.2) is large for intermediate values of x and (5.1) follows for Theorem 3.

By direct computation of the corresponding Lorenz curve one may verify that (5.1) is true for a Bernoulli parent distribution (i.e. $P(X_i = a) = p = 1 - (P(X_i = b))$. A third example in which (5.1) is conjectured to be true, is the case of a uniform parent distribution. Here a density crossing argument verifies the result for $n = 1, 2, 3$. Thanks to the symmetry of the uniform distribution we have a little simplification in the argument since $E(X_{n+1:2n+1}) = E(\overline{X}_{2n+1})$. However verification of (5.1) for $n \geqslant 4$, assuming a uniform parent distribution, appears to be difficult. The problem is not with the median, which has a simple density (proportional to $[x(1-x)]^n$), but with the mean which, as n increases, has an increasingly more complex density. For example, when $n = 2$ we already have the complicated form

$$\begin{aligned} &f_{\overline{X}_5}(x) \\ &= \tfrac{5}{24}x^4, && 0 < x < \tfrac{1}{5}, \\ &= \tfrac{5}{24}\left[x^4 - 5\left(x - \tfrac{1}{5}\right)^3\right], && \tfrac{1}{5} < x < \tfrac{2}{5}, \\ &= \tfrac{5}{24}\left[x^4 - 5\left(x - \tfrac{1}{5}\right)^3 + 10\left(x - \tfrac{2}{5}\right)^2\right], && \tfrac{2}{5} < x \leqslant \tfrac{1}{2}, \\ &= f_{\overline{X}_5}(1 - x), && \tfrac{1}{2} < x < 1. \end{aligned}$$

This crosses the density $f_{X_{3:5}}(x) = 30[x(1-x)]^2$, $0 < x < 1$ once in in the interval $(\frac{1}{5}, \frac{2}{5})$ and again in the interval $(\frac{3}{5}, \frac{4}{5})$.

If we consider parent distributions of the form $P(X = 0) = P(X = 2) = p$, $P(X = 1) = 1 - 2p$ then by direct computation we can show that (5.1) holds when $p = \frac{3}{7}$ but that $X_{2:3}$ and $\overline{X}_3$ are not Lorenz comparable when $p = \frac{1}{7}$. As of this moment we have not encountered a parent distribution for which the inequality (5.1) is reversed.

The problem of determining sufficient conditions on the parent distribution to ensure that (5.1) holds remains open.

Acknowledgement

We are indebted to C.A. Robertson for helpful discussions and suggestions. Constructive criticism was also provided by an anonymous referee.

References

Arnold, B.C. and J.A. Villaseñor (1984), Lorenz ordering of order statistics, Technical Report #131, Statistics Department, University of California, Riverside.

Hardy, G.H., J.E. Littlewood and G. Polya (1929), Some simple inequalities satisfied by convex functions, *Messenger of Mathematics* **58**, 148–152.

Shaked, M. (1980), On mixtures from exponential families, *Journal of the Royal Statistical Society* **B42**, 192–198.

Strassen, V. (1965), The existence of probability measures with given marginals, *Annals of Mathematical Statistics* **36**, 423–439.