A LOOK AT EQUITY

Carole Siegel, Morris Meisner, Eugene Laska,

Research Foundation for Mental Hygiene
Rockland State Hospital

I. In the Nichomachean Ethics, Aristotle considered the question of what is a fair or equitable apportionment of a resource. He considered the distribution of a resource to be equitable if the apportionment to each individual is in direct proportion to his worth to society. Many would take issue with this definition, and indeed, philosophical and ideological debate about an acceptable definition of equity continues. The issues are not simple. In a recent article in Commentary [8], the readership was challenged to define a "fair" distribution of income. Income is not the only resource for which an equitable distribution is of concern. For example, some recent work [2] has discussed equity of political representation (one man-one vote) and school integration.

Equity questions may be posed in different contexts. For example, "Is the resource distribution of a given population equitable?" Another context may pose the question "Is the resource equitably distributed among the subgroups of a population?" The authors' interests in this problem arises from concern for equity in the delivery of mental health services to population subgroups [12].

Statisticians as a rule do not attempt to make a value judgment as to what is equitable but have worked on methodology - defining curves and indices and giving their properties - to describe the dispersion of a resource over the members of a population. The principal indices which have been developed for this purpose have been based on Lorenz curves. Section II of this paper collects and mathematically organizes properties of Lorenz curves and briefly surveys some associated measures of equity. In Section III a new theorem relating the behavior of Lorenz curves to monotone hazard rates is given. Specifically, it is shown that the Lorenz curves corresponding to distribution functions admitting increasing (decreasing) hazard functions lie above (below) the Lorenz curve of the exponential distribution. Section IV introduces an equity function and an equity index which may be used to study the relative behavior of the distribution of a resource over a subpopulation with respect to the total population.

II. Lorenz curves introduced in 1905 [9] have traditionally been used by economists to describe the equity of distribution of income across a population. The curve at a fixed point u measures the percentage of total income accounted for by the uth percentile of the population ordered according to increasing income.

Let Y denote the nonnegative random variable representing the resource, G its absolutely continuous distribution function and g its density function. For those random variables Y having a finite mean E[Y], the Lorenz curve, $\ell_Y(u)$, is defined by the equation

$$\ell_Y(u) = \frac{\int_0^{G^{-1}(u)} y g(y) dy}{E[Y]} \qquad (1)$$

In this setting $\ell(u)$ may roughly be thought of as the proportion of E[Y] accounted for by the uth percentile of the distribution of Y.

By differentiating (1), we obtain an alternate definition of $\ell(u)$, [4].

$$\ell_Y(u) = \int_0^u \frac{G^{-1}(x)}{E} dx. \qquad (2)$$

We observe that a distribution function gives rise to a Lorenz curve, and conversely since

$$E \cdot \ell'(u) = G^{-1}(u),$$

G may be recaptured from knowledge of $\ell(u)$ and E. However more than one distribution function gives rise to the same Lorenz curve. In fact, it is easily seen that for $a > 0$,

$$\ell_{aY}(u) = \ell_Y(u).$$

These definitions have the disadvantage that the so called curve of equal distribution, $\ell(u)=u$, cannot be obtained. However, one can produce Lorenz curves arbitrarily close to the curve of equal distribution. This follows from the observation that for $a > 0$,

$$\ell_{Y+a}(u) = \frac{E[Y]\ell_Y(u) + au}{E[Y] + a} . \qquad (3)$$

Letting $a \to \infty$ the Lorenz curve approaches u.

The following properties of $\ell(u)$ are immediately obvious.

1. $\ell(0) = 0$, $\ell(1) = 1$.
2. $\ell(u)$ is monotone nondecreasing. ($\ell'>0$)
3. $\ell(u) \leq u$.
4. $\ell(u)$ is convex.

Conversely any function ℓ satisfying these properties will be thought of as a Lorenz curve. Properties (1) and (2) imply that ℓ itself is a distribution function whose support is the unit interval. The kth moment associated with the distribution function ℓ is given by

$$\frac{E[YG^k(Y)]}{E[Y]} .$$

An interesting observation is that if these moments are known for all k along with E[Y] then $\ell(u)$, and hence G(x), can be recaptured.

It is known that if G is lognormal then $\ell(u)$ is symmetric about the line $\ell = 1-u$. Kendall [7] and more recently Al-Atraqchi [1] have shown that the converse is false and have derived necessary conditions on G for the symmetry conditions to hold.

Various measures of equity of distribution have been derived on the basis of the Lorenz curve. One such measure is the fair share coefficient which searches for the total proportion of the population whose values of the resource are less than the mean value, i.e., G(E). In terms of the Lorenz curve finding the fair share coefficient is equivalent to finding the value of u at which $\ell'(u) = 1$.

A more commonly used measure, the Gini coefficient, γ, [5], is the normalized (to have maximum value of one) area between $\ell(u)$ and the curve of equal distribution,

$$\gamma = 2 \int_0^1 (u-\ell(u))du.$$

For example in a recent Fortune article [10], it is noted that the Gini coefficient for the distribution of income in the U. S. has fallen over the past 35 years from .44 to .35. It is well known that γ is related to the measure of concentration also introduced by Gini, $\Delta=E[|Y_1-Y_2|]$ where Y_1 and Y_2 are independent and identically distributed. The relationship is given by the equation

$$\gamma = \frac{\Delta}{2E} \tag{4}$$

Another measure of inequity is given by the maximum vertical distance between the curve of equal distribution and $\ell(u)$, [11]. It is easily seen that this quantity occurs at u = G(E), the fair share coefficient. This distance is referred to as the Schutz coefficient and is equal to $G(E) - \ell(G(E))$.

III. The exponential distribution gives rise to the Lorenz curve

$$\ell(u) = u + (1-u)\log(1-u).$$

γ, in general, ranges between 0 and 1, and for the exponential $\gamma = \frac{1}{2}$. In some sense this makes the exponential a dividing distribution in terms of γ. In this section we show that it is a dividing distribution for the class of distribution functions admitting monotone hazard rates. The hazard rate for a distribution G is defined as

$$h(t) = \frac{-d}{dt} \log (1-G(t)).$$

Theorem:[1] Let Y have an increasing (decreasing) hazard. Then

$$\ell_Y(u) \underset{(\le)}{\ge} u + (1-u)\log(1-u). \tag{5}$$

Proof:

The Lorenz curve for a random variable X having the exponential distribution may be rewritten in terms of the distribution function G of the random variable Y as

$$\ell_X(u) = -\int_0^{G^{-1}(u)} g(t)\log(1-G(t))dt.$$

To prove the theorem, it is equivalent to show that

$$\ell_Y(u) = \int_0^{G^{-1}(u)} \frac{tg(t)}{E} dt \ge$$

$$-\int_0^{G^{-1}(u)} g(t)\log(1-G(t))dt = \ell_X(u). \tag{6}$$

According to a result from reliability theory [3], the tail of a distribution having a monotone hazard rate crosses the tail of an exponential with the same mean only once. For increasing (decreasing) hazard we have

$$1 - G(t) \underset{(\le)}{\ge} e^{-t/E}, \ t \le t^*$$
$$1 - G(t) \underset{(\ge)}{\le} e^{-t/E}, \ t \ge t^* \tag{7}$$

where t^* is the point at which they cross. Thus for any u such that $G^{-1}(u) \le t^*$, the inequality (6) is true because it is pointwise true. Consider now a u for which $G^{-1}(u) > t^*$. The inequality is no longer pointwise true. However the integral inequality holds. The proof hinges on the fact that the integrands cross once at t^* and $\ell_X(1) = \ell_Y(1) = 1$. Each side of (6) can be rewritten as a sum of integrals from 0 to t^* and t^* to $G^{-1}(u)$. The integrals from 0 to t^* may be written as 1 minus the integrals from t^* to ∞. Thus we have to show

$$-\int_{t^*}^{\infty} \frac{tg(t)}{E} dt + \int_{t^*}^{G^{-1}(u)} \frac{tg(t)}{E} dt \ge$$
$$\int_{t^*}^{\infty} g(t)\log(1-G(T))dt - \int_{t^*}^{G^{-1}(u)} g(t)\log(1-G(t))dt$$

or equivalently that

$$-\int_{G^{-1}(u)}^{\infty} \frac{tg(t)}{E} \ge \int_{G^{-1}(u)}^{\infty} g(t)\log(1-G(t))dt.$$

In this form, the inequality is pointwise true. Hence the lemma is true for all u.

The intuitive interpretation of this result is based upon inequality (7) which indicates that for increasing hazard there is less mass in the tails of the distribution function than in the tails of the exponential. Hence there is a smaller proportion of the population accounting for high values of the resource leading to a more equitable (higher) $\ell(u)$. Similar intuition holds for the decreasing hazard case.

Corollary: Let Y_i, $i=1,2$ be independent and identically distributed with monotone hazard rate. Then

$$E[|Y_1 - Y_2|] \underset{(\leq)}{\geq} E[Y_1]$$

if the hazard is increasing (decreasing).

Proof: The proof follows from the theorem and equation (4).

If the Lorenz curve of a distribution lies above (below) that of the exponential, we cannot, however, conclude that the distribution has an increasing (decreasing) hazard rate. That is, the converse of the theorem is not true. We can easily see this from (3) since if we choose Y with an arbitrary hazard function, 'a' can be chosen so that the Lorenz curve of Y+a lies above that of the exponential. One can also construct examples for which the random variable's Lorenz curve lies below that of the exponential, while its hazard function is not decreasing (e.g., $\ell(u) = u^{2k}$, k large).

IV. If one wishes to compare distinct populations with respect to equitable apportionment of a variable such as income, computing separate Lorenz curves (and their associated Gini coefficients) for each population is a method which is reasonable. However, in the particular application of concern to the authors this approach did not seem appropriate. The application involved comparing the apportionment of a resource over the total population with its apportionment over substratas of the population. A comparison of the Lorenz curve of the total with the Lorenz curve of the substrata does not indicate whether the substrata receives its fair share with respect to the total. For example, suppose the Lorenz curve of a substrata was identical to the Lorenz curve for the total population. This indicates that the apportionment of the total resource among the entire population is identical to the apportionment of the resource restricted to the substrata. However, this gives no immediate insight into the relative position of the substrata among the entire population. Hence to describe equity of a substrata, A, with respect to the total population, an equity function $B_A(u)$ is introduced whose value at a fixed point u represents the proportion of the substrata, A, whose values are less than or equal to the uth percentile for the total population.

More formally, we consider the underlying population Ω to be decomposable into two measurable disjoint, exhaustive subsets A and $\overline{A}$. Let $Y(\omega)$, with distribution function G, be the variable of interest. Introduce the indicator function of A i.e.,

$$I(\omega) = 1, \ \omega\epsilon A$$

$$I(\omega) = 0, \ \omega\epsilon \overline{A}$$

for any $\omega\epsilon\Omega$. Let

$$W(y) = P(Y \leq y | I=1).$$

Thus, W represents the distribution function of Y restricted to members of the population in A. The equity function for the substrata A described above is formally written as

$$B_A(u) = P(Y \leq G^{-1}(u) | I=1) = W(G^{-1}(u)), \ 0 \leq u \leq 1. \quad (8)$$

(Note that since B is a distribution function it is monotonically increasing and lies between zero and one.)

If $B_A(u) \geq u$ for all u then $W(\cdot) \geq G(\cdot)$. This implies that the proportion of the substrata whose resource values are less than $y = G^{-1}(u)$ is larger than the corresponding proportion in the total population. Analogous conclusions follow for the case $B_A(u) \leq u$. Thus there is a certain inequity with respect to the resource in the substrata. If $B_A(u) = u$, for all u the resource values among the substrata are distributed exactly as over the total population. We can in fact analytically show the following lemma.

Lemma: $B_A(u) = u$ for all u if and only if Y and I are independent.

Proof: Using (8), $B(u) = u$ implies $W = G$. Hence

$$P(Y \leq y \mid I = 1) = P(Y \leq y) ;$$

i.e., Y and I are independent. The converse is also immediate. The case $B_A(u) \equiv u$ may be thought of as "Equity in Distribution."

If we let K(y) be the distribution function of Y restricted to the substrata $\overline{A}$ and $\alpha = P(I=1)$ then

$$G(y) = \alpha W(y) + (1-\alpha) K(y). \quad (9)$$

The relationship

$$u = \alpha B_A(u) + (1-\alpha) B_{\overline{A}}(u)$$

immediately follows from equation (9). Hence if $B_A(u) > u$ then $B_{\overline{A}}(u) \leq u$.

At the beginning of this section, an argument was given for the inappropriateness of direct examination of the Lorenz curve of the substrata as a measure of equity of the substrata with respect to the total population. However, knowledge of the curve $B_A(u)$ provides no more information than does knowledge of the Lorenz curves of the total population and the substrata's along with the respective means since this enables one to generate both G and W and hence $B_A(u) = W(G^{-1}(u)) = W(E\ell'(u))$. However since these

Lorenz curves do not indicate the relative amount of the total resource available to the substrata, A, it is advisable to plot in any application the Lorenz curve of the total population and $B_A(u)$.

In an actual application one may observe N independent observations of Y of which n are observations from the substrata A. The distribution function of Y will typically be unknown. Thus one must estimate $B_A(u)$. For points u=i/N, B_A may be estimated as the proportion of the n observations of the substrata whose values are below that of the ith smallest of the total. Since $\hat{B}_A(u)$ is a step function its values for other u are given by the equation

$$\hat{B}_A(u) = \hat{B}_A(i/N) \quad i/N \leq u < (i+1)/N.$$

Using the lemma one may test for "Equity in Distribution" by testing for independence of the random variables Y and I. To test $B_A(u) \equiv u$ one may use standard two sample tests since the hypothesis is equivalent to $W \equiv G$ which is equivalent to $W \equiv K$. Hence if the distribution functions over A and $\bar{A}$ are not found to differ then one cannot reject "Equity in Distribution".

In the case where B(u) crosses u, one may test for equity at a given point u_o. This is done by using Fisher's exact probability test on the 2 x 2 contingency table whose rows are the proportions of the population whose resource values are less than and more than $G^{-1}(u_o)$ and whose columns refer to the substratas A and $\bar{A}$. Thus the four entries are $B_A(u_o)$, $B_{\bar{A}}(u_o)$, $1-B_A(u_o)$, $1-B_{\bar{A}}(u_o)$.

In this context the estimator of the quantity

$$J_A(u_o) = \frac{W(G^{-1}(u_o))}{1-W(G^{-1}(u_o))} \Bigg/ \frac{K(G^{-1}(u_o))}{1-K(G^{-1}(u_o))},$$

has been introduced as a measure of association and large sample distribution theory is well known. The assumption of equity at the point u_o corresponds to the rows and columns being independent which corresponds alternatively to the hypothesis that $J_A(u_o) = 1$. The result $J_A(u_o)>1$ corresponds to $B_A(u_o) > u_o$, i.e., inequity towards the substrata A. Hence J_A may be viewed as an index of equity of the substrata at the point u_o. If u_o is chosen as the fair share coefficient of the total population, G(E), then J has a simple interpretation. The numerator is the relative odds of being in the under fair share group of the total population, conditional on being a member of A. The denominator is defined analogously and J is the ratio of the relative odds.

Finally we note that several more global measures of inequity suggest themselves. We propose two which correspond somewhat to the definitions given in Section II.

The first measure (analogous to the Gini coefficient) proposed is given by the equation

$$\left[\sigma = \log_2 \frac{(1-\int^{+}(B_A(u)-u)^{+}\, du)}{(1-\int^{-}(B_A(u)-u)\, du)} \right]$$

where $\int^{+}(\int^{-})$ is integrated over those values of u such that $B_A(u) \underset{(<)}{>} u$. The range of σ is between -1 and +1, with negative values indicating inequity towards the substrata A. The value σ = 0 does not imply equity in distribution although it does in some sense imply that inequities toward A are compensated by inequities toward $\bar{A}$, albeit at different values in the range of the resource values.

Analogous to the Schutz coefficient, we can define the maximum vertical distance between the forty five degree line and $B_A(u)$. This maximum may occur at several points u. In a given application where N independent observations of the resource are made, quantity $D_N = \sup_u |\hat{B}_A(u)-u|$ where $\hat{B}_A(u)$ is the step function whose estimates at the points u = i/N are described in Section III. The asymptotic distribution of D_N (which is the Kolmogov-Smirnov statistic) is well known. Since the equation $B_A(u) = u$ is equivalent to the hypothesis G = W, this statistic may also be used as a test of the two sample problem. Hajek [6] has commented on the relationship between rank order tests and the Kolmogorov-Smirnov test. Further for any given confidence level α, the quantity d(α) obtained from the asymptotic distribution of D_N such that

$$\lim_{N\to\infty} P\{D_N \geq d(\alpha)\} = \alpha$$

may be used to form an α level confidence band for $B_A(u)$, namely

$$(\hat{B}_A(u) - d(\alpha),\ \hat{B}_A(u) + d(\alpha)) .$$

1. A referee has pointed out that this result is a corollary of theorem 7 of The Estimation of the Lorenz Curve and Gini Index by J.Gastwirth appearing in The Review of Economics and Statistics, 2/72.

References

1. Al-Atraqchi, M.A. "The Symmetry of the Lorenz Curve Generated by the Log Normal Distribution Revisited", The New York Statistician, Vol. 24, No. 1, 1972, pp. 4-5.

2. Alker, Hayward, R., Jr. "Measuring Socio-Political Inequity", Tanur et al, Editors, "Statistics" A Guide to the Unknown", Holden-Day, 1972, pp. 343-351.

3. Barlow, Richard E. and Proschan, Frank "Mathematical Theory of Reliability", John Wiley and Sons, 1965, pp. 28-29.

4. Gastwirth, Joseph, L. "A General Definition of the Lorenz Curve", Econometrika, Vol. 39, No. 6, pp. 1037-1039.

5. Gini, Corrado "Measurement of Inequality of Incomes", Econ. Journal, March, 1921, pp. 124-126.

6. Hajek, J. and Sydak, Z. "Theory of Rank Tests", Academic Press, 1967, p. 90.

7. Kendall, M. G. "Discussion on paper by Hart and Prais", Journal of Royal Statistical Society, Series A, Vol. 119, Part 2, pp. 184-185.

8. Kristol, Irving, "About Equality", Commentary, Vol. 54, No. 5, November, 1972, pp. 41-47.

9. Lorenz, M. C. "Methods of Measuring the Concentration of Wealth", Publication of A.S.A., Vol. 9, 1905, pp. 209-219.

10. Rose, Sanford "The Truth about Income Inequality in the U.S.", Fortune, December, 1972.

11. Schutz, Robert "On the Measurement of Income Inequality", Am. Econ. Review, Vol. 41, May 1951, pp. 107-122.

12. The National Institute of Mental Health, Contract No. HSM-42-72-212.