THE GINI INDEX AND OTHER MEASURES OF INEQUALITY

by

Richard F. Green

TR 2002-1

Department of Mathematics and Statistics
140 Solon Campus Center
University of Minnesota Duluth
Duluth, Minnesota 55812

February, 2002

THE GINI INDEX AND OTHER MEASURES OF INEQUALITY

Richard F. Green

Department of Mathematics and Statistics

University of Minnesota-Duluth

Duluth, MN 55812

Abstract

This paper mentions several scale-invariant measures of inequality and discusses one, the Gini index, in some detail. The Gini index, which is sometimes referred to as the "index of mean difference," is the most commonly used measure of income inequality. In this paper the Gini index is defined in terms of mean difference and in terms of the Lorenz curve, which is often used to illustrate income distributions. The Gini index is calculated for a number of particular (theoretical) income distributions. An appendix describes how to find the Lorenz curve and the Gini index given particular income distributions and also how to find income distributions that produce particular Lorenz curves. The calculations in this paper are my own, but I assume that they have been done before by other people. Perhaps eventually I will find who else has done the calculations that I have done here and where their results have been published. However, doing calculations on one's own is a useful exercise. I have prepared this technical report in order to have the nicer results of my calculations and the references to a few papers at hand when I lose my notes and forget what I have done.

INTRODUCTION

Inequality in general, and income inequality in particular, can be measured in a number of ways. Statisticians usually measure the variability of a population by the variance, which is the average squared deviation from the mean, or by the standard deviation, which is the square root of the variance. The standard deviation has the useful property that if all the population values are stretched out, perhaps by multiplying all the values by the same thing (by two, say), then the standard deviation is multiplied by the same thing (by two, say).

If all incomes were multiplied by a constant (perhaps by changing from dollars to yen) the standard deviation would change. But a sensible measure of inequality would not change with such changes in scale. What is wanted is a measure that is invariant to changes in scale. All the measures that I will describe are invariant under scale changes. Allison (1978) discusses a number of inequality measures. He lists five invariant measures of inequality: (1) the coefficient of variation, (2) the relative mean deviation, (3) Gini's coefficient of mean difference, (4) Theil's information measure, and (5) the variance of the logarithms. Allison added another desired property, the *principle of transfers*, the idea suggested by Dalton (1920) that an inequality measure should increase if income is transferred from a poorer person to a richer person. Both the relative mean deviation and the variance of logarithms violate the principle of transfers and Allison concentrated his attention of the coefficient of variation, Gini's index, and Theil's measure.

The coefficient of variation, which is the standard deviation divided by the mean,

$$V = \sigma/\mu,$$

is perhaps the measure of inequality best known to statisticians. The Gini index is the average absolute difference between observations [Kendall and Stuart (1969), p. 46 call this the "coefficient of mean difference"], standardized by dividing by twice the mean:

$$G = (1/n^2)\, \Sigma\Sigma\, |x_i - x_j| / (2\mu). \quad (1)$$

The Gini index [Kendall and Stuart, p. 47, call this "Gini's coefficient of concentration"] is the measure most widely used by economists and it appeals to me. Theil's measure,

$$T = (1/n)\, \Sigma\, (x_i/\mu) \log (x_i/\mu),$$

resembles the Shannon-Wiener measure of information. Theil's T has the important advantage that it can be partitioned into parts, for example, overall inequality is the sum of inequality within countries and inequality between countries. However, I think that the interpretation of information in the Theil measure is obscure at best. (Perhaps it is a measure of information obtained by knowing into whose income each unit of money flows.)

Allison (1978) preferred the coefficient of variation and the Theil measure to the Gini index because they showed a nicer sensitivity to transfers. If a certain amount of income is moved from a poorer person to a person a certain amount richer, the change in the coefficient of variation is insensitive to the level from which the income is moved. The Theil measure is more sensitive to changes from smaller values than from larger ones, while the Gini index is more sensitive to changes from the part of the distribution where the density is highest (around the middle for bell-shaped distributions).

THE LORENZ CURVE AND THE GINI INDEX

Allison (1978) gives the definition of the Gini index [named for the Italian statistician and polymath, Carlo Gini (1912)] as the ratio of the mean difference divided by twice the mean (Eqn 1), and he mentions an equivalent expression due to Dasgupta *et al.* (1973) which he says is easier to calculate:

$$G = (2/\mu n^2)\ \Sigma\ ix_i - (n + 1)/n, \qquad (2)$$

where x_i is the i<u>th</u> smallest observation.

However, the Gini index is usually interpreted in terms of the Lorenz curve [named for the American statistician, Max Otto Lorenz (1905)], in which the proportion of all income amounts, y[x(t)], less than or equal to t is plotted against x(t), the proportion of all incomes that are less than or equal to t. Kendall and Stuart (1969, p. 48) refer to the Lorenz curve as the "curve of concentration."

The Lorenz curve is standardized [y(1) = 1], so the income mean—or the units used to measure income—is not relevant, but, except for scale (for μ), all the information in the distribution of incomes is contained in the Lorenz curve. Simple measures, such as the proportion of total income earned by people below the median, y(.5), can be read from the Lorenz curve. Wilkinson (1996) considered a number of such measures, including y(.7), the proportion of all income earned by the bottom 70% of earners. Wilkinson was interested in the relationship between inequality and health and wanted to know what would be the best value of x if one wanted to use a measure of the form y(x). He was not sure which was best. One thing that might be considered is the average (over x) of all

$y(x)$. This average, $A_2 = \int y(x)\,dx$, is equivalent to G in a sense. That is, $G = 1 - 2A_2$. [See Figure 1.]

Another measure of inequality, called the "Robin Hood index," is the proportion of all income that must be taken from the rich and given to the poor in order to make everyone equal. It turns out that for the purposes of Robin Hood the "rich" are everyone above average and the "poor" are everyone below average. Using the Lorenz curve, the Robin Hood index is the maximum distance that the Lorenz curve is below the 45°-line, that is, the maximum of $x - y(x)$, which will be $x(\mu) - y[x(\mu)]$.

<u>Values of the Gini index, G, for various forms of the income distribution, f(t)</u>

The Gini index is a measure of how much the income distribution differs from perfect equality. Its value varies from 0 to 1, with 0 being perfect equality and 1 being perfect inequality. In principle, the Gini index may be calculated for any income distribution, but there are a few distributions for which the results are easy to obtain and interesting.

a) If all incomes are equal, $G = 0$.
b) If a single individual earns all the money, $G = 1$ [or rather, $G = 1 - 1/n$, where n is the population size.]
c) If incomes are uniformly distributed [if $f(t) = 1$, for $0 < t < 1$, say], $G = 1/3$.
d) If incomes are exponentially distributed [if $f(t) = \exp(-t)$, for $t > 0$], $G = 1/2$.
e) If incomes have a beta distribution with $\alpha = n$ and $\beta = 1$ [that is, if $f(t) = n\,t^{n-1}$, for $0 < t < 1$],

$$G = 1/(2n + 1). \qquad (3)$$

f) If incomes have a beta distribution with $\alpha = 1$ and $\beta = n$ [that is, if $f(t) = n(1-t)^{n-1}$, for $0 < t < 1$],

$$G = n/(2n + 1). \tag{4}$$

Notice that if n = 1, both (3) and (4) yield G = 1/3, which is what we expect, since the beta distribution with $\alpha = 1$ and $\beta = 1$ is the uniform distribution.

g) If incomes have a Pareto distribution [that is, $f(t) = (\alpha - 1)(t + 1)^{-\alpha}$, for $t > 0$],

$$G = (\alpha - 1)/(2\alpha - 3). \tag{5}$$

For G to be meaningful, μ, the mean of the incomes, must exist, and this requires that $\alpha > 2$. Note that for α only slightly larger than 2, G is only slightly less than 1, which means that incomes are concentrated in only a few hands. On the other hand, as $\alpha \to \infty$, $G \uparrow 1/2$. This is understandable, because as the Pareto parameter α increases the Pareto distribution converges to exponential.

Similarly, beta distributions with $\alpha = 1$ and $\beta = n$ converge to exponential as $n \to \infty$, and $G = n/(2n + 1) \downarrow 1/2$. Thus two simple classes of right-skewed distribution functions converge to exponential and their Gini indices converge (from below for beta distributions, from above for Pareto distributions) to 1/2.

h) The family of gamma distributions might be more realistic economically than the other examples that I have used, but I don't have analytic solutions for the Gini index except for the case $\alpha = 1$ (exponential). Approximate numerical solutions are given in Table 1 for $\alpha = 0.5$ to 10.0 in steps of 0.5 and from 11 to 20. Gini values for incomes in the fifty states of the United States ranged from 0.38 to 0.48 in 1990 [Table 1 in Kawachi and Kennedy (1997), reprinted in Kawachi, Kennedy and Wilkinson (1999)], which correspond to gamma distributions with α values between 1 and 2.

A useful formula

There is a useful formula for finding the Gini index if a constant amount is added to all incomes. If one knows the value of the Gini index, G_{old}, where the mean income is μ, and if amount m more is added to each income, the Gini index for the new income distribution is given by,

$$G_{new} = [\mu/(\mu + m)]\, G_{old}. \qquad (6)$$

I used this formula to simplify the calculation of equation (5). For a simple illustration of equation (6), find the Gini index for incomes uniformly distributed between 1 and 5 [say that incomes are uniformly distributed with the highest income five times the lowest]. Without loss of generality, I could assume that incomes are uniformly distributed between 1/4 and 5/4, which could be obtained by adding a constant amount m = 1/4 to every income where we began with a uniform distribution the interval, (0,1), for which G = 1/3. The mean of a uniform distribution on (0,1) is 1/2. Therefore, the Gini index for a uniform distribution on (1,5) is given by

$$G_{new} = [.5/(.5 + .25)]\,(1/3) = (2/3)\,(1/3) = 2/9.$$

In fact, a Gini index of 2/9 would be quite low even for wages [see Galbraith (1998), Figure 15.1], which are less variable than incomes.

The Gini index when the largest income is *r* times the smallest

More generally, if the highest income is r times the lowest, and incomes are uniformly distributed between the lowest and highest, then

$$G = [(r - 1)/(r + 1)]\,(1/3). \tag{7}$$

The highest possible value of the Gini index when the highest income is r times the lowest would occur when incomes are concentrated at the highest and lowest values, with frequencies in the ratio, $r^{1/2} : 1$, for example, income distribution: $f(1) = r^{1/2}/(r^{1/2} + 1)$, $f(r) = 1/(r^{1/2} + 1)$. For this distribution, the Gini index is,

$$G = (r - r^{1/2})^2/[r(r - 1)]. \tag{8}$$

Notice that for $r = 4$, $G = 1/3$ and for $r = 9$, $G = 1/2$.

An idea that is suggested by equation (6) is that one might lower the Gini index (inequality) by adding a fixed amount to everyone's income. The question is, how much do we have to add to everyone's income in order to reduce the Gini index to some desired value? This question is easy to answer using equation (6).

Finding the Gini index for real income distributions

While Allison (1978) mentioned that the Gini index was difficult to calculate, this is not a serious problem in this time of great computing power and cheap computer memories. Obtaining income data is several orders of magnitude more expensive than computing indices of inequality once the data are available. Data are not always presented in ways that make it possible to calculate indices accurately, but the simple rough summary of income data given in the Statistical Abstract of the United States [U.

S. Census Bureau (2000)] permits an easy, quite accurate calculation of the Gini index. Table No. 745 (p. 471) provides the following data for family incomes for 1998: bottom 5^{th}, 4.2% of all incomes; second 5^{th}, 9.9%; third 5^{th}, 15.7%; fourth 5^{th}, 23.0%, highest 5^{th}, 47.3%, and top 5%, 20.7%. (In order to make the five 5ths add up to 100% I changed the highest 5^{th} to 47.2%.) Using these data, one knows that the Lorenz curve has the following values: y(.2) = .042, y(.4) = .141, y(.6) = .298, y(.8) = .528, and y(.95) = .793. These values are the basis of the Lorenz curve illustrated in Figure 2.

A first approximation to the Gini index can be obtained by subtracting the y-values from their arguments for x = .2, .4, .6 and .8, dividing by five (the distance between the x-values) and adding. One has (1/5)[(.2 - .042) + (.4 - .141) + (.6 - .298) + (.8 - .528)] = .1982, the area between the broken line representing the Lorenz curve and the diagonal [y = x] line. This area is doubled to get the Gini index. In fact, the Lorenz curve falls on or (usually) below the given points, and .1982 is a lower estimate of the desired area. A slight improvement can be obtained by taking into account the y(.95) value, which is .089 = .207 – (1 - .528)/4 below the broken line. The width of the triangle (whose height is .089) is 1/5, so its area is .0089, which may be added to 0.1982, yielding .2071. This is doubled (divided by 1/2) to get the Gini index, G = .4142. This value is a slight underestimate of the true Gini index for U. S. family incomes. The Gini index is quite high for the United States compared to other modern, industrialized countries. The United States has more income inequality than other industrialized countries, and as a consequence—it has been argued (Wilkinson 1996)—we are less healthy than citizens of many countries with which we like to compare ourselves.

REFERENCES

Allison, P. D. 1978. Measures of inequality. *American Sociological Review* 43:865-880.

Dalton, H. 1920. The measurement of the inequality in incomes. *Economic Journal* 30:348-361.

Dasgupta, P., Sen, A. and Starrett, D. 1973. Notes on the measurement of inequality. *Journal of Economic Theory* 6:180-187.

Galbraith, J. K. 1998. *Created Unequal.* University of Chicago Press, Chicago.

Gini, C. 1912. Variabilità e mutabilità, contributo allo studio della distribuzioni e relazioni statistiche. *Studi Economico Giuridici dell' Univ. di Cagliari* 3:1-158

Kawachi, I., and Kennedy, B. P. 1997. The relationship of income inequality to mortality—Does the choice of indicator matter? *Social Science & Medicine* 45:1121-1127.

Kawachi, I., Kennedy, B. P., and Wilkinson, R. G. 1999. *The Population and Health Reader, Volume I: Income Inequality and Health.* The New Press, New York.

Kendall, M. G. and Stuart, A. 1969. *The Advanced Theory of Statistic. Volume 1: Distribution Theory. 3rd Edition.* Hafner, New York.

Lorenz, M. O. 1905. Methods of measuring the concentration of wealth. *Publications of the American Statistical Association* 9:209-219.

U. S. Census Bureau. 2000. *Statistical Abstract of the United States: 2000 (120th edition).* Washington, D. C.

Wilkinson, R. G. 1996. *Unhealthy Societies: The Afflictions of Inequality.* Routledge, London.

APPENDIX—ANALYTICAL CALCULATIONS

Finding the Lorenz curve and the Gini index from income distributions and *vice versa*

A. Finding the Lorenz curve given an income distribution.

Given an explicit expression, f(t), for the distribution of incomes, t, we can find the Lorenz curve, y(x), by getting parametric equations for x and y in terms of t. X(t) is the cumulative distribution of incomes: $x(t) = F(t) = P(T < t) = \int f(u)du$, and y(t) is the incomplete first moment: $y(t) = \Phi(t) = (1/\mu)\int uf(u)du$. [The integration required to find x(t) and y(t) is from 0 to t.]

Example 1. Incomes are uniformly distributed, for example, $f(t) = 1$ for $0 < t < 1$. Then, $x(t) = t$ and $y(t) = t^2$, which implies that $y(x) = x^2$. The Gini index may be calculated from y(x) by integrating from $x = 0$ to $x = 1$ to obtain a lower area, A_2, which yields: $A_2 = 1/3$. The upper area, $A_1 = 1/2 - A_2$, and the Gini index is given by $G = 1 - 2A_2 = 1 - 2/3 = 1/3$. This is the example illustrated in Figure 1.

Example 2. Incomes are exponentially distributed, for example, $f(t) = e^{-t}$ for $t > 0$. Then, $x(t) = 1 - e^{-t}$ and $y(t) = 1 - e^{-t} - t\,e^{-t}$. The Lorenz curve, $y(x) = x + (1 - x)\ln(1 - x)$, and the Gini index is $G = 1/2$. The Lorenz curve for exponential income distribution is plotted in Figure 3 (heavy line), where it is compared with $Y = X^3$ (dotted line), which yields the same Gini value.

B. Finding the income distribution from the Lorenz curve

Note: The income distribution is not *uniquely* determined by the Lorenz curve, because changes in scale of income leave the Lorenz curve unchanged, but the Lorenz curve does determine income distribution uniquely up to a scale parameter. Without loss of generality we can set the mean of the income distribution, μ, equal to one. This is equivalent to saying that incomes are measured in terms of the average, for example, this income is half the mean (t = 1/2), that income is three times the mean (t = 3). [The mean must exist for us to use the Lorenz curve or the Gini index, because the mean is required for standardization.]

Assume that the Lorenz curve is differentiable. [This need not be the case, but we are interested in simple cases in which calculations are easy.]

The derivative of the Lorenz curve, dy/dx = t, because the amount added to the mean for a small change in the cumulative distribution is just the density function of income. But, if y(x) has a nice form, we can equate dy/dx = y'(x) with t. Then, with luck, we can solve for x in terms of t: x(t), which is F(t). The desired function is f(t), the density function of income, which is the derivative of F(t).

Example 1. The Lorenz curve has the particular form, $y(x) = x^2$. We already know the answer (up to a scale constant) from Example 1 above, but here we are using the method describe above.

We have dy/dx = 2x = t, which implies that x(t) = F(t) = t/2 and

$$f(t) = dx(t)/dt = 1/2, \text{ for } 0 < t < 2, \tag{9}$$

since x takes values between 0 and 1. Notice that income distribution f(t) is uniform, but here it is uniform on (0,2) instead of (0,1) since I am requiring that the mean equals 1, which it does for U(0,2), but not for U(0,1).

Example 2. The Lorenz curve has a general form, $y(x) = x^n$, where n is a number greater than one. [We want y(0) = 0 and y(1) = 1.]

We have $dy/dx = n\, x^{n-1} = t$, which implies that $x(t) = t^{1/(n-1)}/n^{1/(n-1)}$, and

$$f(t) = dx/dt = [1/(n-1)]\, t^{(2-n)/(n-1)}/n^{1/(n-1)}, \text{ for } 0 < t < n, \tag{10}$$

since x takes values between 0 and 1. Notice that income, f(t) = dx/dt, has the form of a beta distribution, with $\alpha = 1/(n-1)$ and $\beta = 1$. Notice that this distribution is cut off sharply on the right (at t = n in our case).

TABLE 1. Gini indices for Gamma distributed incomes for various values of alpha.

α	Gini
0.5	.632
1.0	.500
1.5	.424
2.0	.375
2.5	.339
3.0	.312
3.5	.291
4.0	.273
4.5	.258
5.0	.246
5.5	.235
6.0	.225
6.5	.217
7.0	.209
7.5	.202
8.0	.196
8.5	.191
9.0	.185
9.5	.180
10.0	.176
11.0	.168
12.0	.161
13.0	.155
14.0	.149
15.0	.144
16.0	.140
17.0	.136
18.0	.132
19.0	.128
20.0	.125

Note: These values are rounded to three decimal places. The only Gini value that I have calculated analytically is $G = 0.5$ for $\alpha = 1$, for which I obtained the value $G = 0.49987$ with steps of .0002 when I integrated numerically from 0 to 50. I am not sure that the three places that I give are correct for $\alpha = 0.5$ and for some of the larger α values, but I think that my values are pretty close to the correct ones. I would guess that the Gini index for a gamma distribution with $\alpha = 2$ is exactly $0.375 = 3/8$, but I can't see any values other than that for $\alpha = 1$ which are as nice.

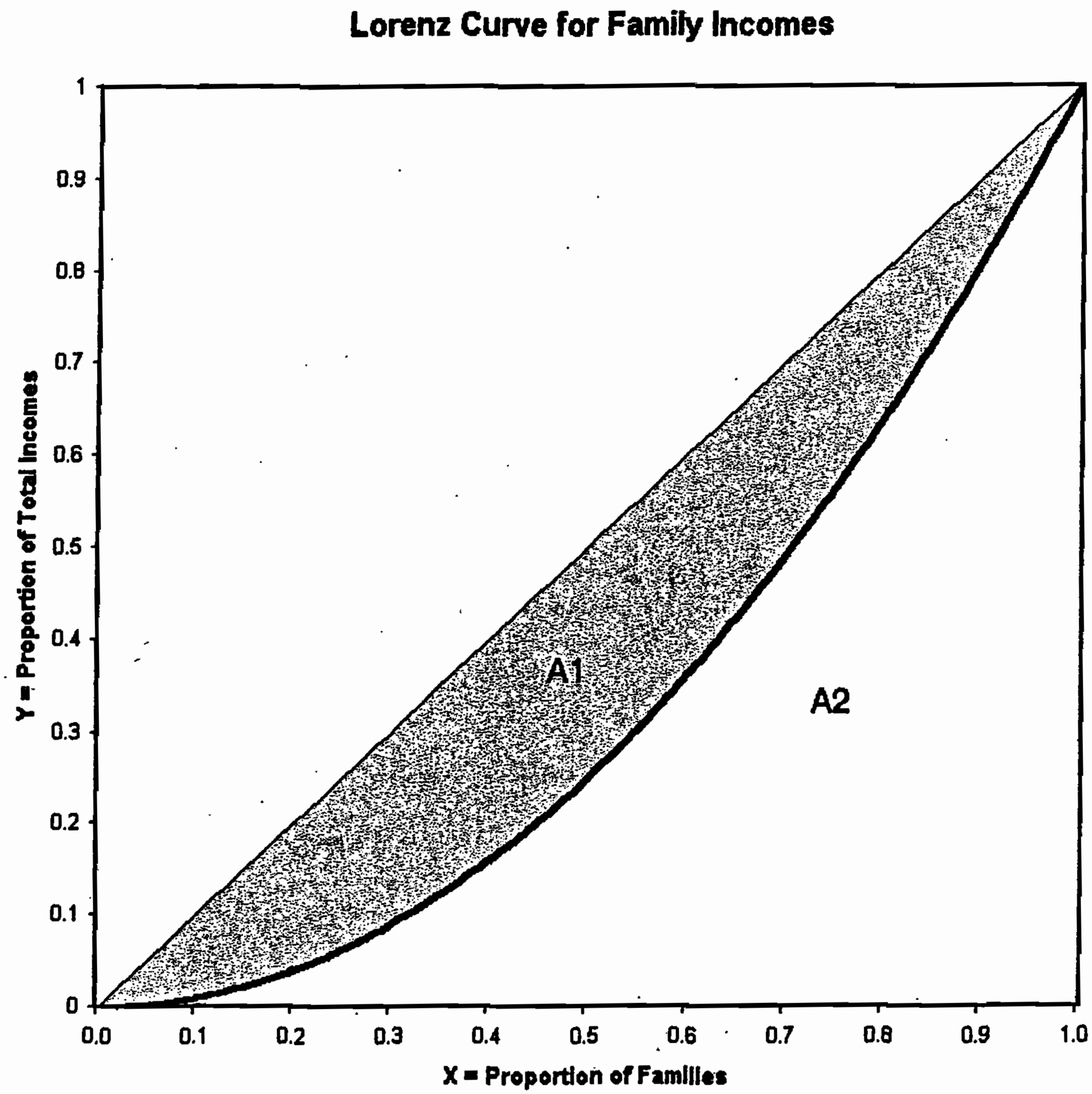

Figure 1. The Lorenz curve and the Gini index

The heavy line is the Lorenz curve, here corresponding to an income distribution that is uniform on the interval (0,1). The Gini index is the shaded area A1 divided by the sum of the areas A1 and A2, or A1/(1/2) = 2 A1. It is sometimes easiest to calculate 1 – 2 A2.

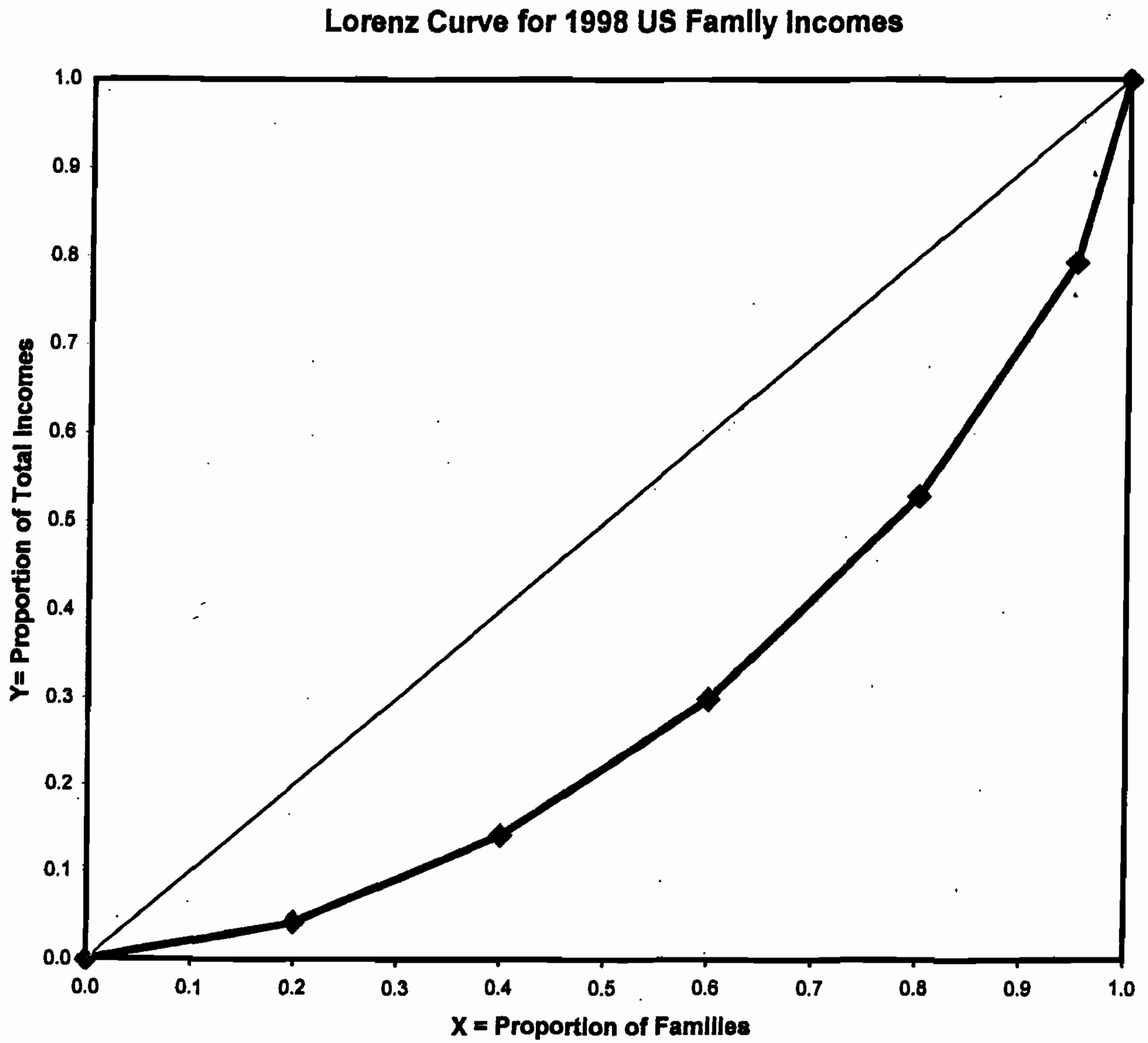

Figure 2. Lorenz curve for 1998 US Family Incomes

The data were summarized by the proportion of all income earned by those in the five quintiles, plus the top 5%. These bits of information are shown in the points marked with diamonds.

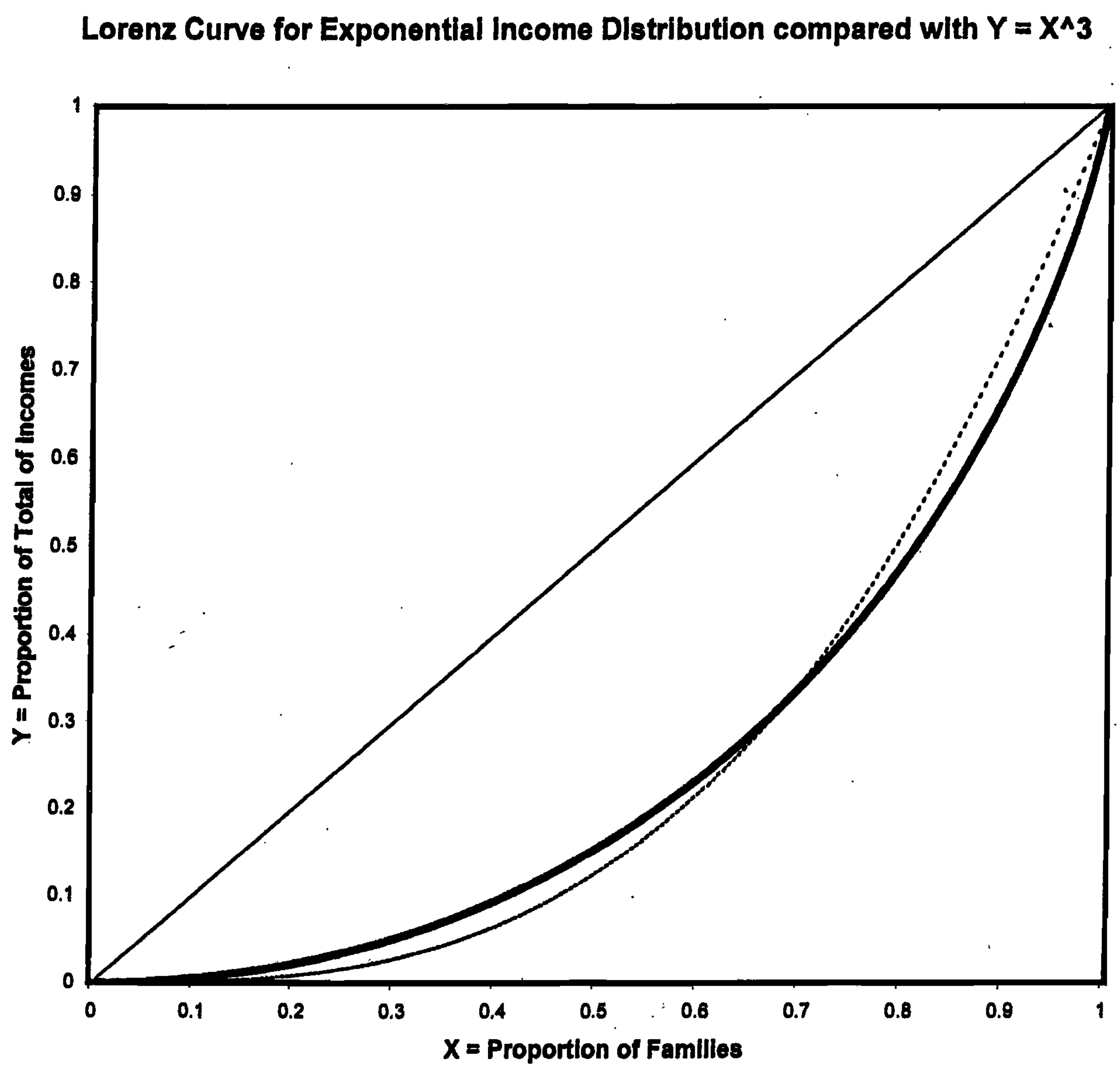

Figure 3. Lorenz curve for Exponential Income Distribution

The bold line is the Lorenz curve for an exponential income distribution. This is compared with the dotted line (a Lorenz curve: $Y = X^3$) which yields the same Gini index, $G = 0.5$. The Lorenz curve, $Y = X^3$, would arise if incomes had a beta distribution with $\alpha = 1/2$, and $\beta = 1$.

THE GINI INDEX WITH TWO INCOMES IN PROPORTION 1:R, FOR R > 1.

An intuitively useful way to look at the Gini index is to compare an actual value (e. g. Gini = 0.435 for U. S. in 2001) is to find the value of r such that if everyone has income 1 or r, with r > 1, the Gini index will equal the actual value. Of course, for any r, if the proportion of families (say) with one income is large enough, the Gini index will be close to 0, which is not interesting. Instead, I use the proportions that maximize the Gini index for a particular r. The Gini index will be maximized when the two incomes, 1 and r, are in proportions $r^{1/2}$:1. The maximized value of the index for a particular income ratio 1:r is given by (8), which is equivalent to:

$$G(r) = (r^{1/2} - 1)/(r^{1/2} + 1) = (r^{1/2} - 1)^2/(r - 1)$$

Outline of proof: Assume that 1 family has income 1 and x (need not be an integer) families have income r. After a bit of algebra, the Gini index is seen to be: $1/(1 + x) - 1/(1 + rx)$, which is maximized for $x = 1/r^{1/2}$.

TABLE 2. Values of the Gini index for various income ratios.

Ratio, r	Gini index, G(r)
2	0.171573
3	0.267949
4	0.333333
5	0.381966
6	0.420204
7	0.451416
8	0.477592
9	0.500000
10	0.519494
16	0.600000
25	0.666667
36	0.714284
49	0.750000
64	0.777778
81	0.800000
100	0.818182