EXPLORING THE GINI INDEX OF INEQUALITY WITH DERIVE

I. Introduction:

The Gini index of income or resource inequality is a measure of the degree to which a population shares that resource unequally. It is based on the statistical notion known in the literature as the "mean difference" of a population.[1] The index is scaled to vary from a minimum of zero to a maximum of one, zero representing no inequality and one representing a maximum possible degree of inequality.

In order to begin a derivation of the Gini index, consider the lowest 20% of the population, ranked by per capita income, and ask what portion of the total income is attributible to this 20%? If the corresponding proportion of total income as a percentage is also 20% we will call this fair. If it is less than 20% we will say there is income inequality. It cannot be more than 20%. In general, to measure this, we define a function, $g(\alpha)$, to be the fraction of the total value of a certain resource belonging to the lowest (100α)% of the population as ranked by per-capita ownership of that resource. This curve is defined on the interval [0,1] and is referred to as the Lorenz curve of the resource distribution. Here I'll always convert the resource into money, usually dollars. Then the Gini index of inequality is a measure of the difference between $g(\alpha)$ and the ideal which is assumed to be α, (ie. same percentage of the resource as portion of the population).

As a discrete example, in 1960 the small Norwegian city of Moss exhibited the following data[2] . The last column is an estimate of the function $g(\alpha)$ using the center of each income bracket as represen- tative of that bracket.

Income (in Krona)	# of Males	% of Males (cumulative)	Estimated % of income (cumulative)
0 - 5000	753	11.0	2.0
5000 - 10000	967	25.2	9.6
10000 - 15000	2,347	59.7	40.4
15000 - 20000	1,786	85.8	73.3
20000 - 25000	493	93.1	84.9
25000 - 30000	202	96.0	90.8
30000 - 35000	270	100.0	100.0
	6818		

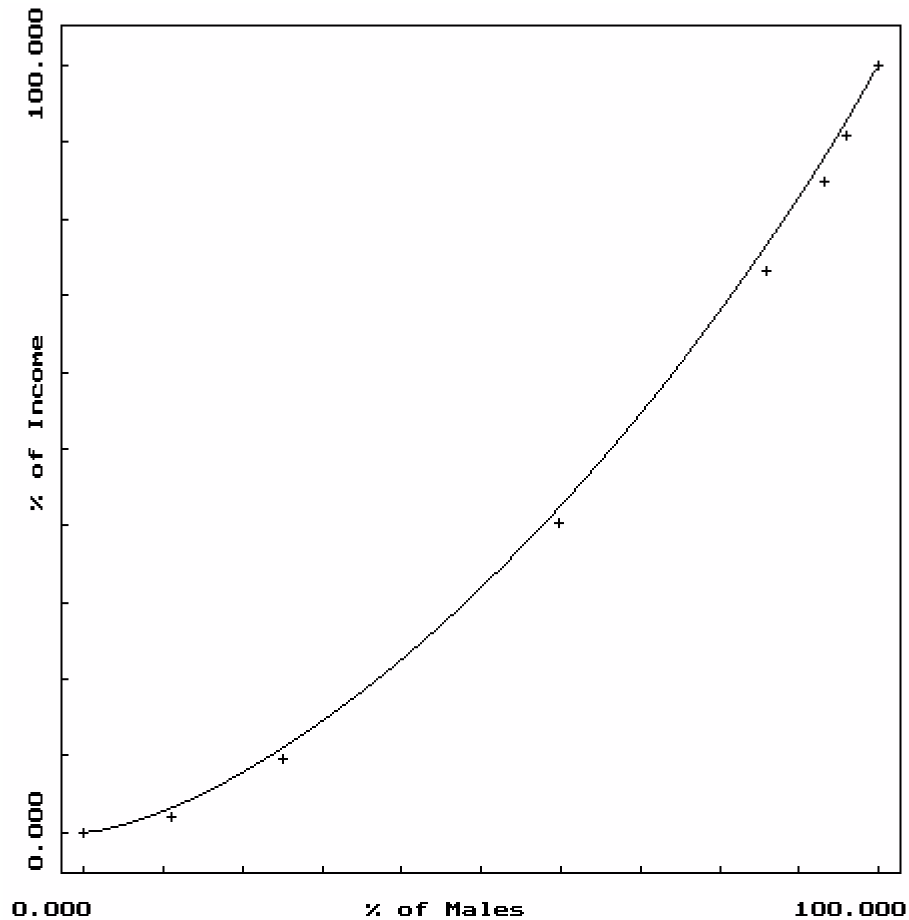
As can be seen from the last two columns the cumulative percentages are different indicating that there is uneveness in the income distribution. In particular, for example, the bottom 25.2% of the men have only 9.6% of the income, while the top 4% of the men have 9.2% of the income.

The graphs of these cumulative percentages appear below with a fitted curve:

---

[1] M.G. Kendall & A Stuart, The Advanced Theory of Statistics, Vol.1,2nd Ed, Hafner Publishing Co., NY, 1963.

[2] Lee Soltow, Toward Income Equality in Norway, University of Wisconsin Press, 1965, p.10.



It might prove to be interesting to investigate what would happen with a more uniform income distribution. If each income category had the same number of individuals for example, then the cumulative distribution of individuals would be linear from 0 to 100% , and the cumulative distribution of income would correspond as follows for the same seven categories:

Income (in Krona)	# of Males	% of Males (cumulative)	Estimated % of Income (cumulative)
0 - 5000	974	14.2	2.0
5000 - 10000	974	28.6	8.2
10000 - 15000	974	42.9	18.4
15000 - 20000	974	57.1	32.7
20000 - 25000	974	71.4	51.0
25000 - 30000	974	85.7	73.5
30000 - 35000	974	100.0	100.0
	6818		

So although the income distribution is "uniform", the two cumulative distributions are not the same. In fact the actual data for Moss shows less disparity in the two cumulative measures than the above. So a “uniform” income distribution is definitely not fair as reflected in this measure. Indeed, eventhough the higher income categories have the same proportion of individuals as do lower income categories they obviously have a far larger share of the whole resource. Thus the measure we are discussing also reflects the dispersion of the resource distribution as well. In the case of a distribution with a small variance we will find that the Gini measure is in fact close to zero.

II. **Definitions and Theorems**:

Formally we begin a population income density function, f(x), defined for $x \geq 0$, and F(x), the corresponding cumulative distribution function where x is per capita income in dollars. Recall that $F'(x) = f(x)$. Many calculus textbooks introduce these ideas as applications of differentiation and integration. They certainly appear again in the calculus based statistics course.

Then for each number α in the interval [0,1] the Lorenz function, $g(\alpha)$, is defined as that fraction of the total income which is attributible to the poorest $(100\alpha)\%$ of the given population ranked according to per capita income, x.

Note that $x \cdot (f(x) \cdot \Delta x)$ approximates a value proportional to the sum total of income attributible to individuals with incomes between x and $x+\Delta x$, and only proportional because we don’t have the total number of individuals in the population. Thus $\int_0^t x \cdot f(x)dx$ is proportional to the sum total of income attributible to those with incomes less than or equal to t. Hence the fraction of total income attributible to those with incomes less than or equal to t is:

$$\frac{1}{\mu}\int_0^t x \cdot f(x) \cdot dx$$, where μ is, by definition, the mean income.

Now if $(100\alpha)\%$ is the percentage of the population with income less than or equal to t, ie. the poorest $(100\alpha)\%$ of the population, then $\alpha = F(t)$ and we can assert finally that $g(\alpha)$, the fraction of total income attributible to the poorest $(100\alpha)\%$ of the population, is:

Theorem: $$g(\alpha) = \frac{1}{\mu}\int_0^t x \cdot f(x)dx,$$ where $\alpha = F(t)$ and μ is mean income.

One can show the following inequality for all income distributions.

Theorem: $g(\alpha) \leq \alpha$.

Proof: Let $h(t) = \mu\int_0^t f(x)dx - \int_0^t x \cdot f(x)dx$. If $h(t) \geq 0$ for all t, then the theorem is proved. First note that $h(0) = 0$ and $h(t) \to 0$ as $t \to \infty$. Next we have that $h'(t) = \mu f(t) - t f(t) = (\mu - t)f(t)$. Hence $h'(t) \geq 0$ for $t \leq \mu$ and $h'(t) \leq 0$ for $t \geq \mu$. Since $f(t) \geq 0$ for all t, it thus follows that h(t) is non-decreasing on $[0,\mu]$, while non-increasing on $[\mu,\infty)$. Clearly then $h(t) \geq 0$ on $[0,\mu]$. If h(t) were negative at some point of $[\mu,\infty)$, then we could not have $h(t) \to 0$ as $t \to \infty$.

The formal definition of the Gini index of inequality is:

Definition (Gini): $2\int_0^1 [\alpha - g(\alpha)]d\alpha$

Clearly this is always non-negative. In the case with $g(\alpha)=0$, which would be the worst case of income inequality, the index is 1, while in the case with $g(\alpha)=\alpha$, the index is 0 and thus corresponds to no inequality or perfect equality. The scale factor of 2 simply insures that the index will range between 0 and 1, rather than 0 and ½.

III. Exercise:

Here is an interesting exercise in which Derive is very useful:

Use Derive to determine the Gini index of inequality for a population income distribution with density function given as:

$$f(x) = x \cdot e^{-x/b} / b^2 \quad \text{for } x \geq 0 \text{ and } b>0.$$

This is an example of the ordinary Gamma distribution, with mean 2b and variance $2b^2$. The gamma distribution is often used to simulate an income distribution. An interesting fact here is that the Gini index of inequality does not depend on b!!