Calculating the Gini Coefficient

Contents

1. Introduction
2. Definite Integrals
3. Our Strategy
4. Household Income
5. Total Income and Cumulative Income
6. Household Income as a Percentage of the Total, and Cumulative Percentage Income
7. The Lorenz Curve
8. Area A: the Area between the Perfect-Equality Line and the Lorenz Curve
9. The Gini Coefficient

Introduction

The Gini Coefficient is a very widely used measure of a terribly important problem, income inequality. But it’s quite difficult to calculate ... unless you know some calculus. To calculate the Coefficient we need to know the area between the two curves in this diagram. If we had a way of calculating the area under the Lorenz curve (the curved line) and if then we could subtract it from the area of the triangle formed by the diagonal and the axes, we would only need to divide the resulting area by the area of the triangle, and then we’d be done.

Definite Integrals

To take an integral is to "undo" taking a derivative. What is also interesting about integrals is that they can be used to measure the area under a curve. This animation might be illustrative

Riemann Integral Animation

If you have a rectangle, you can easily calculate its area. So if you fill the area under a curve with many rectangles, and then sum the areas of those rectangles, you get pretty close to the actual area under the curve. What a Riemann integral does is
to put an infinite number of little rectangles under the curve, getting an exact answer.

"Definite integrals" such as \( \int_{a}^{b} f(x) \, dx \) are used to calculate areas under curves, where \( f(x) \) is the curve, \( a \) is the smallest value of \( x \) we consider and \( b \) is the largest value of \( x \) we consider. The symbol \( \int \) comes from the idea of summation, so we can also use integrals to add a bunch of numbers.

I'm told in that in a FoxTrot cartoon, Jason found the area of a 5x6 rectangle by using a definite integral

\[
\int_{0}^{5} 6 \, dx = 6x \bigg|_{0}^{5} = 6 \times 5 - 6 \times 0 = 30
\]

Our Strategy

The Lorenz curve is a relation between percentage cumulative population and percentage cumulative household income. You'll see that it's very easy to get "cumulative population", but getting cumulative household income requires several steps:

* Find a formula for household income.
* Find a formula to add up a) total income ... and to add up b) income "up to the \( X^{th} \)-richest person".
* Trivially, find "percentage cumulative income" by b) by a).
* Plot the Lorenz curve.

Household Income

First we'll need a relation between households and income. Typically what we have is a table, such as table 6.1. From such a table, we use this formula
\[ G = \left| 1 - \sum_{k=1}^{n} [(X_k - X_{k-1})(Y_k + Y_{k-1})] \right| \]

where \(X_k\) is the cumulative proportion of the population variable, for \(k = 0, \ldots, n\), with \(X_0 = 0\), \(X_n = 1\) and \(Y_k\) cumulative proportion of the income variable, for \(k = 0, \ldots, n\), with \(Y_0 = 0\), \(Y_n = 1\). For example,

<table>
<thead>
<tr>
<th>Cumulative population (k)</th>
<th>Cumulative Percentage of Population</th>
<th>Income</th>
<th>Cumulative Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.25</td>
<td>7.5</td>
<td>0.075</td>
</tr>
<tr>
<td>2</td>
<td>0.50</td>
<td>7.5</td>
<td>0.15</td>
</tr>
<tr>
<td>3</td>
<td>0.75</td>
<td>42.5</td>
<td>0.575</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>42.5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Another possibility is to use econometrics or some computer program to come up with some kind of equation that approximately describes how much income \(m\) does a household \(x\) have.

Suppose we say that \(m\) is income and \(x\) is the position of the household on a list of households, ranked by income (so \(x = 1\) would be the poorest, \(x = 2\) would be the second poorest, etc. If there are 20 households, \(x = 20\) represents the richest household).

One example of such a function is

\[ m = 1.4^x \]

It’d be pretty easy to know what the poorest person’s income is: \(m = 1.4^1 = 1.4\), or the richest person’s income: \(m = 1.4^{20} = 836.68\).

This equation generates the numbers for the "highly unequal" distribution that I showed in class. (In class, I scaled this function so that the income of household 1 was equal to that of the example in the book, so I multiplied the above equation by 0.8/1.4).

**Total Income and Cumulative Income**

To calculate a person’s income as a percentage of total income, we need to know total income. To do this we simply sum all of 20 person’s incomes.

But the Gini coefficient is the area under a curve, and a curve is continuous. If we only plot 20 points, we won’t get a very good answer.
We need to have a continuous line ... but we have it! That's \( m = 1.4^x \). We just need to assume that "people are continuous", or, more likely, that instead of 20 people we’re talking about 20,000,000 people. In that case we can use an integral.

We want an operation that will sum all the values of \( m \) over a continuous series of households. This integral

\[
\int_0^{20} m \, dx
\]

sums \( m \) over 20 (continuous) people. \( dx \) indicates that \( x \) is what changes (household 1, household 2, etc.). \( 0 \) is the lowest bound of \( x \) and it gets excluded in the calculation. \( 20 \) is the highest possible value of \( x \), and it gets included.

Because \( m = 1.4^x \),

\[
\int_0^{20} m \, dx = \int_0^{20} 1.4^x \, dx.
\]
Using my computer, I calculate

\[ \text{Total Income} = \int_{0}^{20} 1.4^x \, dx = 2483.7. \]

If \( m = 1.4^x \) is the income of person \( x \), what is the \textit{cumulative} income of person \( x \)? That is, if \( x = 10 \), we want to sum the incomes of persons 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10. In general we want to add the incomes of person 1, person 2, ..., person \( x \). That’s very easy: just calculate this integral:

\[ \text{Cumulative income} = \int_{0}^{x} m \, dx = \int_{0}^{x} 1.4^x \, dx \]

(the 0 is excluded). So, concretely

- Cumulative income of person 1 = \( \int_{0}^{1} 1.4^x \, dx = 1.1888 \)
- Cumulative income of person 2 = \( \int_{0}^{2} 1.4^x \, dx = 2.8531 \)
- Cumulative income of person 5 = \( \int_{0}^{5} 1.4^x \, dx = 13.012 \)
- Cumulative income of person 10 = \( \int_{0}^{10} 1.4^x \, dx = 82.995 \)
- Cumulative income of person 15 = \( \int_{0}^{15} 1.4^x \, dx = 459.38 \)
- Cumulative income of person 18 = \( \int_{0}^{18} 1.4^x \, dx = 1265.7 \)
- Cumulative income of person 19 = \( \int_{0}^{19} 1.4^x \, dx = 1773.2 \)
- Cumulative income of person 20 = \( \int_{0}^{20} 1.4^x \, dx = 2483.7 \)

Household Income as a Percentage of the Total, and

Cumulative Percentage Income

What about percentages? Well, person’s \( x \) income, as a percentage of total income, is just

\[ \text{Percentage of total} = \frac{1.4^x}{\left( \int_{0}^{20} 1.4^x \, dx \right)}. \]

Cumulative percentage income is easily defined as

\[ \text{Cumulative percentage income} = \int_{0}^{x} \frac{1.4^x}{\left( \int_{0}^{20} 1.4^x \, dx \right)} \, dx \]
Person 1’s income, as a percentage of total $= \int_{0}^{1} \frac{1.4^x}{\int_{0}^{20} 1.4^x \, dx} \, dx \quad 0.047\%$

Person 5’s income, as a percentage of total $= \int_{0}^{5} \frac{1.4^x}{\int_{0}^{20} 1.4^x \, dx} \, dx \quad 0.52\%$

Person 10’s income, as a percentage of total $= \int_{0}^{10} \frac{1.4^x}{\int_{0}^{20} 1.4^x \, dx} \, dx \quad 3.34\%$

Person 19’s income, as a percentage of total $= \int_{0}^{19} \frac{1.4^x}{\int_{0}^{20} 1.4^x \, dx} \, dx \quad 71.39\%$

Person 20’s income, as a percentage of total $= \int_{0}^{20} \frac{1.4^x}{\int_{0}^{20} 1.4^x \, dx} \, dx \quad 100.0\%$

This simply sums Percentage of total from 0 (excluded) to $x$.

**The Lorenz Curve**

We can use this result to write a general equation of *cumulative percentage income* as a function of $x$, the rank of the household. For the same reason, $x$ is also a measure of "cumulative population". To get this equation, instead of putting a specific number at the top of the distribution, we just ask "what is the cumulative percentage income of household $x$?" This is the Lorenz curve. The Lorenz curve is a relation between percentage cumulative population and percentage cumulative household income.

\[
\text{Lorenz curve} = \int_{0}^{x} \frac{1.4^x}{\left(\int_{0}^{20} 1.4^x \, dx\right)} \, dx = \frac{1.1966 \times 10^{-3}}{5.0^{1.0x}} 7.0^x - 1.1966 \times 10^{-3}
\]

Evaluating this integral (using a computer, but it’s not hard by hand) we get the equation of the Lorenz curve, if income is $m = 1.4^x$. Oh my goodness! Now, lets try it.

\[
\int_{0}^{20} 1.4^x \left(\int_{0}^{20} 1.4^x \, dx\right) \, dx = 1.0
\]

\[
\int_{0}^{19} 1.4^x \left(\int_{0}^{20} 1.4^x \, dx\right) \, dx = 0.71394
\]

\[
\int_{0}^{10} 1.4^x \left(\int_{0}^{20} 1.4^x \, dx\right) \, dx = 3.3416 \times 10^{-2}
\]

Alternatively, using the equation for the Lorenz Curve we found,
\[
\left[ \frac{1.1966 \times 10^{-3}}{5.0^{1.0x}} \right] 7.0x - 1.1966 \times 10^{-3} \bigg|_{x=20} = 0.99998 \\
\left[ \frac{1.1966 \times 10^{-3}}{5.0^{1.0x}} \right] 7.0x - 1.1966 \times 10^{-3} \bigg|_{x=19} = 0.71393 \\
\left[ \frac{1.1966 \times 10^{-3}}{5.0^{1.0x}} \right] 7.0x - 1.1966 \times 10^{-3} \bigg|_{x=10} = 3.3416 \times 10^{-2}
\]

which is the same except for rounding error.

Let's graph the Lorenz curve:

In the horizontal axis: \(x/20\) In the vertical axis: \(\frac{1.1966 \times 10^{-3}}{5.0^{1.0x}} 7.0x - 1.1966 \times 10^{-3}\)

\(x/20\) is the "cumulative percentage of population". Why? Take the richest person, \(x = 20\). How many people are poorer than him? 19. Adding himself, the cumulative population is 20. What about the guy right in the middle, \(x = 10\)? Below him there are 9 people. Counting him, the cumulative population is 10.

And now for the "cumulative percentage of population". If there are 20 people in total, and the "cumulative population" the richest guy is \(x = 20\), then the cumulative percentage of people up to the richest guy is 20/20, or 100%. The cumulative percentage of population up to \(x = 10\) is 10/20, or \(x/20 = 50\%\). So the "cumulative percentage of population" is found by calculating \(x/20\).

Up to this point, what do we have? We have obtained
* A formula that relates household rank with income
* A formula that tells us the household's cumulative percentage income (as an integral)
* A formula that gives us the Lorenz curve.

Now we want to know the Area under the Lorenz Curve:
Area under Lorenz Curve = \[ \int_{0}^{20} \left( \int_{0}^{x} \frac{1.4^x}{\left( \int_{0}^{20} 1.4^x \, dx \right)} \, dx \right) \, dx \]

= \[ \int_{0}^{20} \left( \frac{1.1966 \times 10^{-3}}{5.0^{1.0x}} \cdot 7.0^x - 1.1966 \times 10^{-3} \right) \, dx \]

= 2.948.

What if the distribution were a little less unequal. Suppose that it was described by the equation

\[ m = x^2 \]

which is drawn as a red line in this graph.

Then the Lorenz Curve would be

\[ \int_{0}^{x} \frac{x^2}{\left( \int_{0}^{20} x^2 \, dx \right)} \, dx = \frac{1}{8000} x^3 \]

and the area under the Lorenz Curve would be

\[ \int_{0}^{20} \frac{1}{8000} x^3 \, dx \]

= 5

**Area A: the Area between the Perfect-Equality Line and the Lorenz Curve**

How can we use this information to calculate area A? By using integrals to calculate an area. We want to calculate the area between the perfect-equality line and the Lorenz curve. First we have to come up with an equation for the perfect-equality line.
This is $x/20$. This is obvious. If everyone had the same income $m$, the income of the $x$th person would simply be $m$. The cumulative income of $1, 2, \ldots x$ would be $xm$. So the cumulative income for the 3rd person would be $3m$; for the 20th person, $20m$. Total income would be $20m$, the level of income that everyone would have times the number of people.

Each person’s income as a percentage of total income would be $m/20m = 1/20$. And since cumulative income is $xm$, the cumulative percentage is $xm/20m = x/20$.

A slightly more snobbish way of finding the same result is to use an integral to add up people’s full-equality income $m$

$$\int_0^{20} mdx = 20m,$$

and then using this result to calculate the cumulative percentage

$$\int_0^x m\left(\int_0^{20} mdx\right)dx = \frac{1}{20} x.$$

Perfect equality line = $\frac{1}{20} x$.

which is also the Lorenz curve of a perfectly equal economy.

**Area A**

So now we want to calculate the area between the perfect-equality line and the Lorenz curve.
An integral gives us the area under a curve. So the easiest way is to calculate the area under the full equality line,

\[
\text{Area under perfect-equality line} = \int_0^{20} \frac{1}{20}x \, dx = 10,
\]

and then the area under the Lorenz curve

\[
\text{Area under Lorenz Curve} = \int_0^{20} \left( \frac{1.1966 \times 10^{-3} \times 7.0^x - 1.1966 \times 10^{-3}}{5.0^{1.0x}} \right) \, dx = 2.948.
\]

We are interested in the area between the two curves, which is the difference between the two:

\[
\text{Area under perfect-equality line} - \text{Area under Lorenz Curve} = \text{Area A}
\]

\[
\int_0^{20} \left( \frac{1}{20}x \right) \, dx - \int_0^{20} \left( \frac{1.1966 \times 10^{-3} \times 7.0^x - 1.1966 \times 10^{-3}}{5.0^{1.0x}} \right) \, dx = 7.0520
\]

This is the value of area A.

**The Gini Coefficient**

What is the Gini coefficient? It’s the ratio of area A to the ratio of the area under the perfect-equality line.
Gini Coefficient = \frac{\text{Area under perfect-equality line}}{\text{Area under perfect equality line}} - \frac{\text{Area under Lorenz Curve}}{\text{Area under perfect equality line}}

Gini Coefficient = \int_0^{20} \left( \frac{1}{20} x \right) dx - \int_0^{20} \left( \frac{1.1966 \times 10^{-3}}{5.010} \right) dx

Gini Coefficient = 0.70520

For the less-unequal distribution I mentioned above, the Gini coefficient is

Gini Coefficient = \frac{\int_0^{20} \left( \frac{1}{20} x \right) dx - \int_0^{20} \left( \frac{1}{x} \right) dx}{\int_0^{20} \frac{1}{20} x dx}

Gini Coefficient = 0.5