The Gini Index

A Measure of Social Inequality

Adapted from notes by Robert Brown,
with contributions by Dillon Zhi, Matthew Gherman, and Adam Lott.

In 1905, the economist Max Lorenz introduced an income inequality curve, coined the Lorenz Curve. Let the x values between 0 and 1 correspond to the proportion of the population of a given country. On the y-axis, Lorenz placed the proportion of the total income of the population that was received by the bottom x-proportion of the population. In 2006, the highest-earning 20 percent of the American population earned about 60 percent of all income so the bottom 80 percent, represented by x=0.80, received 40 percent. Thus the point $(0.8, 0.4)$ appears on the Lorenz Curve. Figure 1 shows a typical Lorenz Curve. We also include the line segment connecting $(0, 0)$ and $(1, 1)$. We will call this line the equal distribution line.

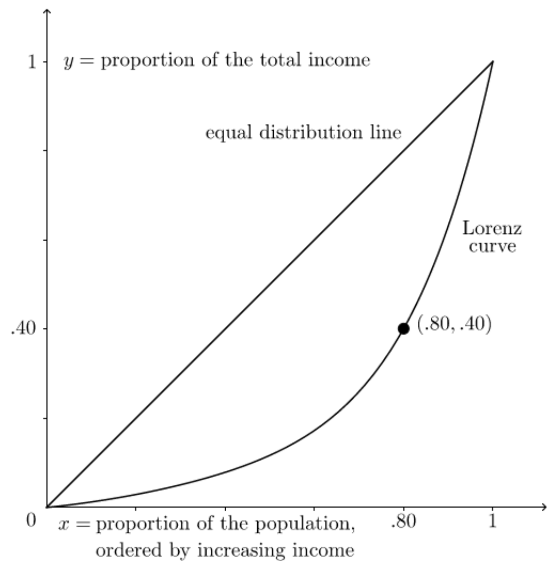

Figure 1

Exercise 1. Why do the points $(0, 0)$ and $(1, 1)$ lie on the Lorenz Curve?

Exercise 2. Explain why the line segment from $(0, 0)$ to $(1, 1)$ should be called the "equal distribution line".

Exercise 3. Why does the Lorenz Curve always lie on or below the equal distribution line? For example, why can't $(.25, .75)$ lie on the Lorenz Curve?

The Lorenz Curve describes the distribution of income among the members of the community, but how can you compare, for instance, the income distribution of the United States to that of Mexico? In 1912, the Italian statistician Corrado Gini proposed a way to describe the distribution of income by a single number. In Figure 2 the 45-degree line is labeled as (part of) the graph of the identity function $I(x) = x$ and the Lorenz Curve is the graph of some function we'll call $L(x)$. The region of the plane between these two curves is labeled by Γ, the Greek capital letter "G", because it was the region of interest to Gini. Since the region beneath the 45-degree line is a triangle with height and base equal to one, and therefore its area is $\frac{1}{2}$, we can see that the area of Γ is no greater than $\frac{1}{2}$.

Exercise 4. What distribution of income would make the area of Γ equal to 1/2?

Exercise 5. Draw an example Lorenz Curve where there is large income inequality. Draw an example where there is little income inequality.

In order to present the area of Γ as a proportion of the possible area, that is $\frac{1}{2}$, Gini divided the area by $\frac{1}{2}$, which multiplies it by 2, so it is on a scale that runs between 0 and 1. The result came to be called the "Gini Index" so, formally,

$$\text{Gini Index} = 2 \cdot \text{area}(\Gamma).$$

The region below the Lorenz Curve, which we have labeled with Λ, the Greek "L", that is, what is "left over" after taking away the Gini region Γ. We can calculate the Gini Index

$$G = 2 \cdot \text{area}(\Gamma)$$

if we know the area of Λ because

$$\text{area}(\Gamma) + \text{area}(\Lambda) = \frac{1}{2}$$

and therefore

$$G = 1 - 2 \cdot \text{area}(\Lambda).$$

We can make a rough estimate of the Gini Index even from a single observation. Suppose it is estimated that, in some country, the top 20 percent of income earners receive 60 percent of the total income for that country. Thus the other 80 percent of the population shares the remaining 40 percent of the income and we know that the point (.80, .40) lies on the Lorenz Curve. Since (0, 0) and (1, 1) also lie on that curve, we'll connect these three points by line segments, as in Figure 3.

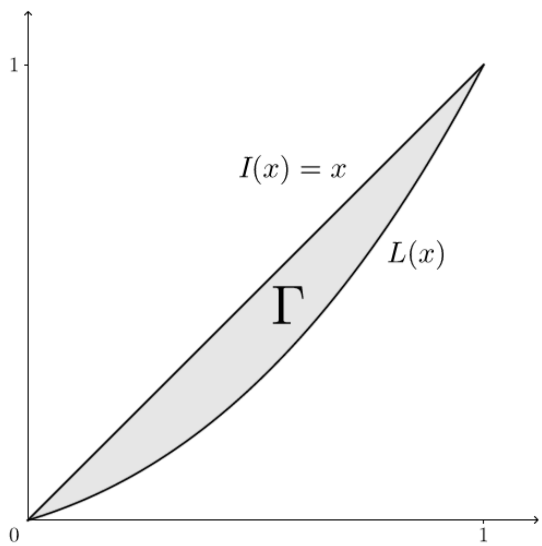

Figure 2

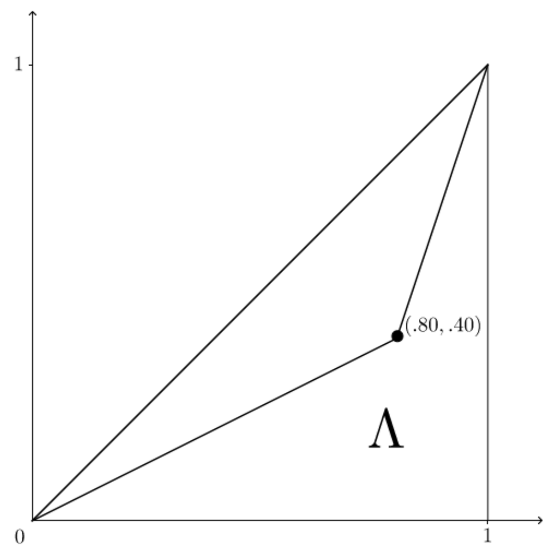

Figure 3

Exercise 6. Given the Lorenz Curve in Figure 3 consisting of two line segments, calculate the area of Λ and use that to compute the Gini Index.

In general, if we know that the proportion a of the lowest earners receives a proportion b of the total income, that means that the point (a, b) lies on the Lorenz Curve and we can approximate that curve by line segments as in Figure 4.

Exercise 7. Calculate the area of Λ in Figure 4 and then show that $G = a - b$ gives a general formula for the one-point estimate of the Gini Index.

We can get a more accurate estimate of the Gini Index if we know two points on the Lorenz Curve. In Figure 5 we connected the four points of the Lorenz Curve by line segments assuming we know points (a, b) and (c, d).

Exercise 8. Calculate the area of Λ in Figure 5 and then show that $G = c - d + ad - bc$ gives general formula for the two-point estimate of the Gini Index.

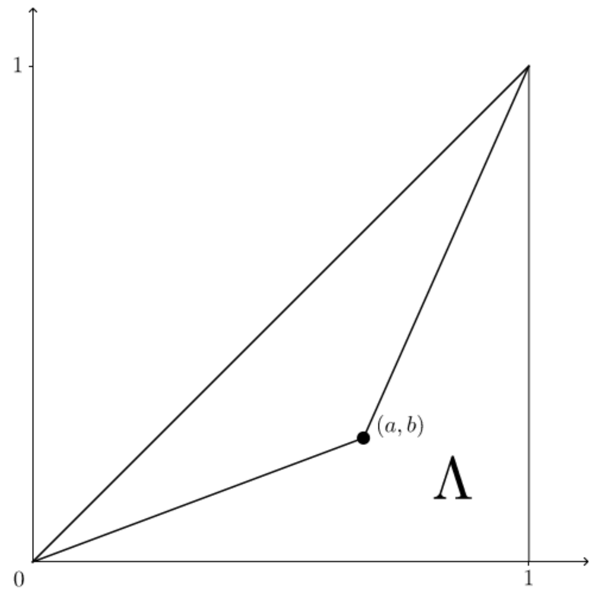

Figure 4

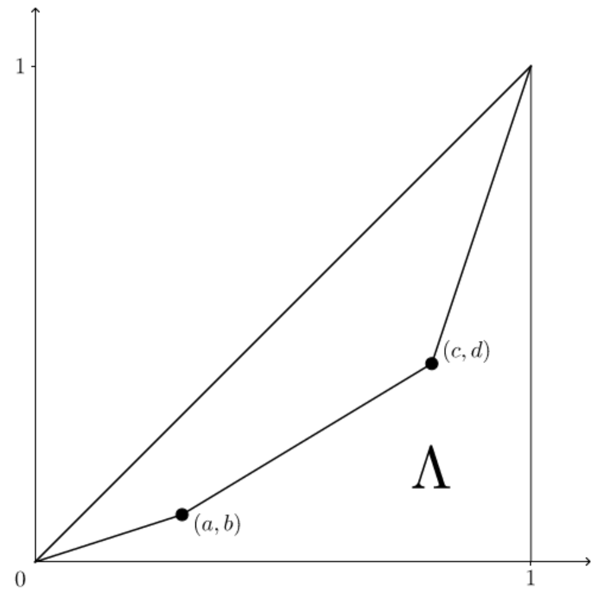

Figure 5

Exercise 9. Consider a population where the bottom 80% of earners make 40% of the total income, while the top 1% earn 20% of the total income. Sketch the two-point approximation to the Lorenz Curve and use the previous exercise to estimate the Gini Index.

Exercise 10.

(a) What can we say about incomes if there is a section of the Lorenz Curve which is *actually* a straight line, say between some points (a, b) and (c, d)?

(b) Show that if (a, b) and (c, d) are points on a Lorenz Curve, then for any $a \leq x \leq c$, the point $(x, L(x))$ cannot lie above the line connecting (a, b) and (c, d). This shows that the Lorenz Curve is a *convex function.*

(c) Is the one-point estimate larger or smaller than the actual Gini Index? Why? What about the two point estimate?

In Exercises 7 and 8 above, we approximated the Lorenz Curve by line segments, but how much information does the *actual* Lorenz Curve $L(x)$ carry? In fact, from our definition above, $L(x)$ doesn't make sense for every value of x between 0 and 1; for example, in a population of only 2 people, what does it mean to consider the total income earned by the bottom 25% of earners in the population? Without considering fractional people, this Lorenz curve only has three points on it!

In order to alleviate this problem, we connect consecutive points on the Lorenz Curve by straight line segments. In fact, in a very large population, any reasonable way of connecting consecutive points will give approximately the same Lorenz Curve and Gini Index. But with our chosen convention, we can now derive some formulas for the exact Gini Index in a population of n people.

Exercise 11. Consider a country with $n = 2$ people in it, and suppose their incomes are y_1 and y_2, with $y_1 \leq y_2$. Let $Y = y_1 + y_2$. Draw the Lorenz Curve for this (very small!) country, and calculate its Gini Index in terms of y_1, y_2, and Y.

Exercise 12. Consider a country with n people. Let the incomes of all the people in the country be $y_1, \ldots, y_n$, sorted in ascending order, and let $Y = y_1 + \ldots + y_n$. In this exercise you will derive useful formulas for the Gini Index in this setting.

(a) Find the coordinates of each of the points plotted on the line $I(x) = x$ and the Lorenz Curve $L(x)$ pictured in Figure 6. Since there are n individuals, those points represent the entire Lorenz Curve, and we've connected them by line segments in keeping with the convention chosen above.

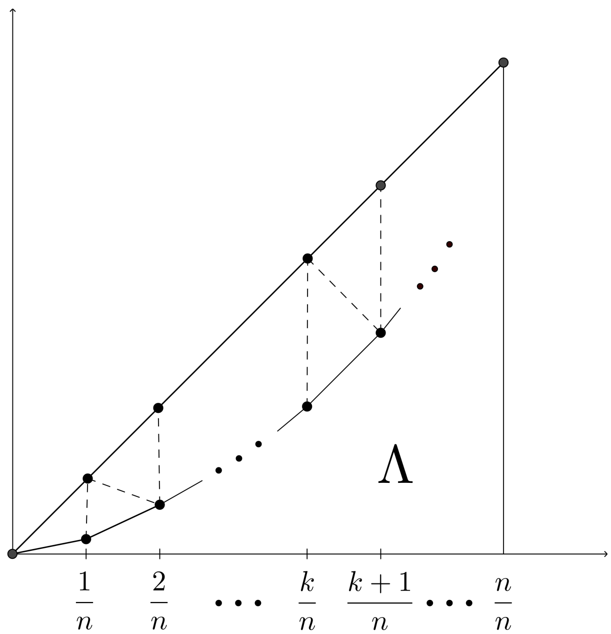

Figure 6

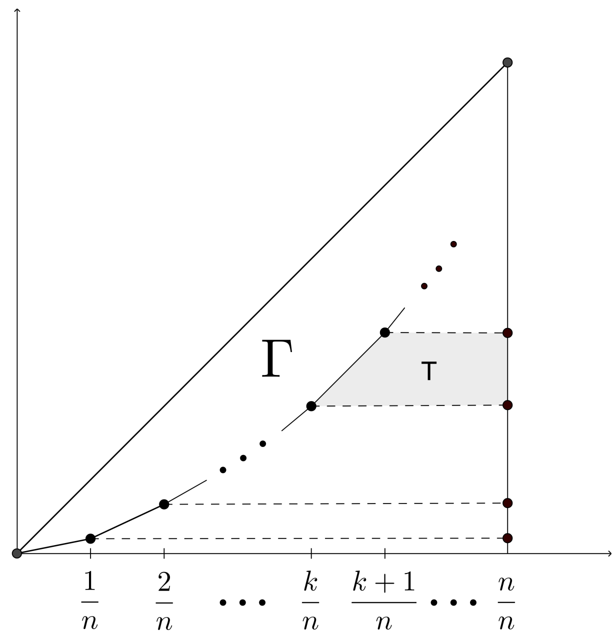

Figure 7

(b) Using part (a), show that the Gini Index is given by

$$G \;=\; \frac{2}{n}\sum_{k=1}^{n-1}\left(\frac{k}{n} - \frac{y_1 + \ldots + y_k}{Y}\right).$$

(c) Find the area of the trapezoid T in Figure 7.

(d) If we partition Λ into trapezoids as in Figure 7, what is the area of the kth trapezoid from the bottom?

(e) Use part (d) to show that

$$G \;=\; 1 - \frac{1}{nY}\sum_{k=1}^{n}(2n - 2k + 1)y_k.$$

(f) Prove that

$$G \;=\; \frac{1}{nY}\sum_{k=1}^{n}(2k - n - 1)y_k.$$

Hint: Combine part (e) with the equality $1 = \frac{1}{Y}\sum_{k=1}^{n} y_k$.

(g) Prove that

$$G = \frac{1}{2nY} \sum_{i,j=1}^{n} |y_i - y_j| \tag{1}$$

even if $y_1, \ldots, y_n$ are not sorted in ascending order.

Hint: Since the right hand side of equation (1) doesn't change if we reorder the y_k, to show that (1) holds in general, it is enough to first prove it under the extra assumption that $y_1, \ldots, y_n$ are indeed sorted in ascending order. Next note that for each i and j, we either have $|y_i - y_j| = y_i - y_j$ or $|y_i - y_j| = y_j - y_i$. Now for each k, count how many times y_k and $-y_k$ appear in the sum in (1), and use part (f).

(h) The city of Villestown has four people: Alice, Bob, Carl, and Danielle. Their annual incomes are \$20,000, \$40,000, \$60,000, and \$80,000 respectively. Find the Gini Index for Villestown.

Exercise 13. Suppose that every person in a population gets a holiday bonus of size $B > 0$ added to their income. Does the Gini Index increase, decrease, or stay the same? Use one of the formulas above to justify your answer.

Limitations of the Gini Index

Exercise 14. In Villestown (see Exercise 12(h)), Alice and Bob form a household and Carl and Danielle form a household. Considering the *household* income distribution rather than the *individual* income distribution, what is the Gini Index?

Exercise 15.

(a) Give an example to show that even if everyone gets richer, the Gini Index may increase.

(b) Give an example to show that even if everyone gets poorer, the Gini Index may decrease.

Exercise 16.

(a) Give an example to show that two income distributions which are qualitatively very different can have the same Gini Index G.

(b) As an ad-hoc definition, let's define the "Second Gini Coefficient" G_2 to be twice the area bounded by the line $I(x) = x$ and function $L(x)^2$. That is, we replace the Lorenz Curve by its square in our computation. Find G_2 for the two income distributions you gave in Part (a).

Hint: If you square the vertical coordinates of a line segment from $(0, 0)$ to a point (a, b) with $a, b > 0$, you end up with a parabolic arc passing from $(0, 0)$ to (a, b^2). If you know integral calculus, show that the area between this parabolic arc and the horizontal axis is given by $\frac{ab^2}{3}$. If you aren't familiar with integral calculus, you may use this formula without justification.

(c) Is it possible for two different income distributions to have the same G and the same G_2 simultaneously? Make sure to establish your answer rigorously.