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.

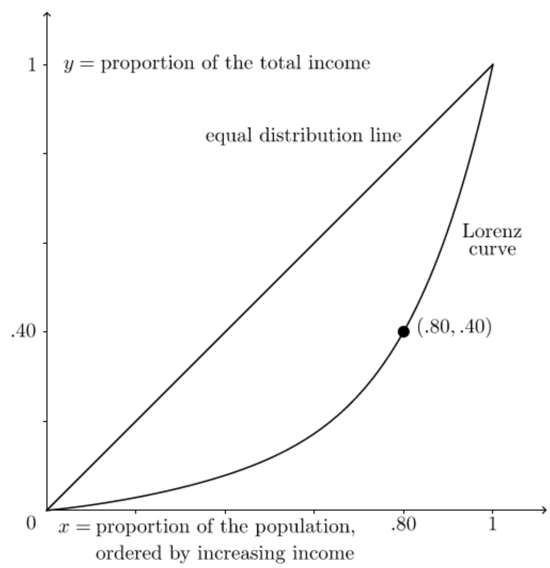


Figure 1

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

> The point $(0, 0)$ is on the Lorenz Curve, because the lowest 0% of earners is 0 people, whose total income is hence 0% of the population's total. The point $(1, 1)$ is on the curve, because 100% of the earners (that is, proportion 1 of them) earn 100% of the income.

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

If the Lorenz Curve were to coincide with the line $I(x) = x$, this would mean that for any proportion x of earners, the lowest x make x fraction of the total income of the population. In particular, if the population is n, then the lowest earner, who represents $\frac{1}{n}$ of the population, makes $\frac{1}{n}$ of the income. Since they are the lowest earner, everyone must make at least that amount. Since the sum of their proportional incomes must be 1, and in particular cannot exceed it, everyone has to make exactly $\frac{1}{n}$ of the income (that is, the income is *equally distributed*).

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 lowest 25% of earners cannot make 75% of the income, since in that case, they would not be the *lowest* 25% of earners. More precisely, and more generally, suppose (x, y) lies on the Lorenz Curve, so the lowest x proportion of earners makes proportion y of the income. Consider then the remaining $1-x$ of the population, who make the remaining proportion $1-y$ of the income. Since they consist of earners who make at least as much those in the lower proportion, the $1-y$ they earn must be at least $(1-x)\frac{y}{x}$, as this is the amount they would make if they earned the same on average as the bottom x, who on average make $\frac{y}{x}$ proportion of the income. Thus

$$1 - y \geq (1 - x)\frac{y}{x},$$

and hence

$$(1 - y)x \geq (1 - x)y,$$

that is,

$$x - yx \geq y - xy,$$

which simplifies to $y \leq x$. Since (x, y) was an arbitrary point on the Lorenz Curve, this shows that that curve must lie on or below the equal distribution line $I(x) = x$.

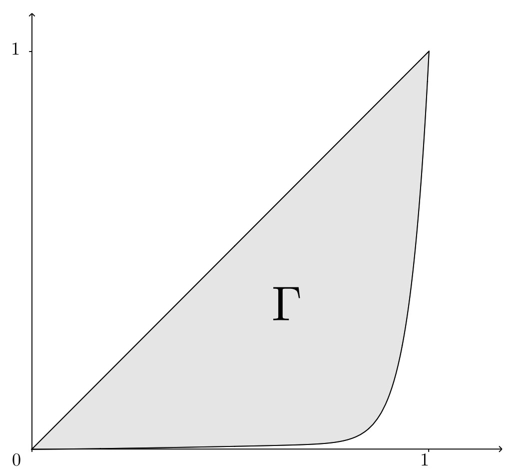
The Lorenz Curve describes the distribution of income among the members of the community, but how can you compare, for instance, the income distribution of the United States to that of Mexico? In 1912, the Italian statistician Corrado Gini proposed a way to describe the distribution of income by a single 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 $\frac{1}{2}$.

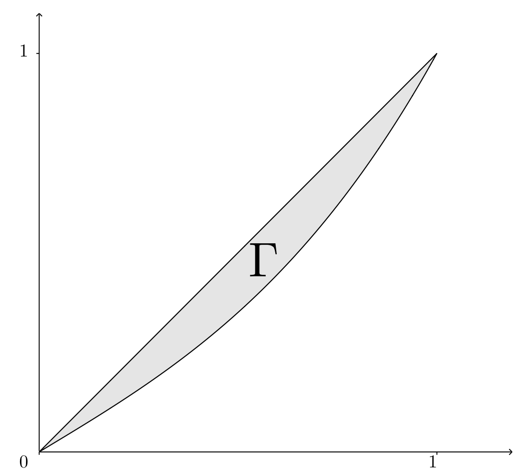
Exercise 4. What distribution of income would make the area of Γ equal to $1/2$?

Γ having area equal to $1/2$ corresponds to a Lorenz Curve with $L(x) = 0$ for all $0 \leq x < 1$, and $L(1) = 1$. One might hope to achieve this by concentrating all of the income on a single individual; in fact, in a finite population, the Lorenz Curve is only defined at finitely many values of x, and the area of Γ depends on how we connect those points on the curve. This is addressed in more detail later in the worksheet, where the convention is established that line segments will connect consecutive points on the Lorenz Curve. Under this convention, it is in fact impossible to have Γ with area equal to $1/2$ in a finite population. However, it is true that if all income is concentrated on a single individual, then the area of Γ will approach $1/2$ as the size of the population tends to infinity.

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

The first figure below shows a Lorenz Curve with large income inequality, and the second depicts a Lorenz Curve with 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.

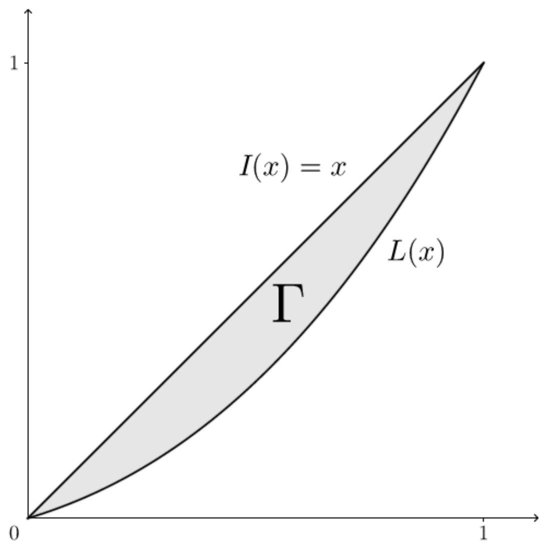


Figure 2

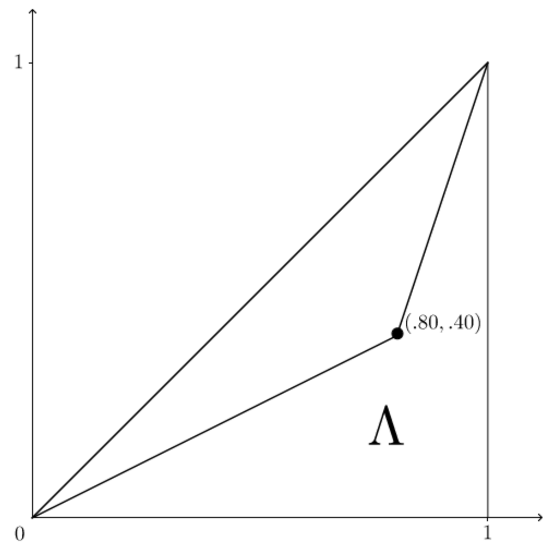


Figure 3

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.

We can break Λ up into two triangles and a rectangle, in which case we find its area to be

$$\text{Area}(\Lambda) = \frac{1}{2}(.80)(.40) + (1 - .80)(.40) + \frac{1}{2}(1 - .80)(1 - .40) = .3,$$

and hence the corresponding Gini Index is

$$G = 1 - 2\text{Area}(\Lambda) = 1 - 2(.3) = .4.$$

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.

If we again break Λ into two triangles and a rectangle, we find its area to be

$$\text{Area}(\Lambda) = \frac{1}{2}ab + (1-a)b + \frac{1}{2}(1-a)(1-b) = \frac{1}{2}ab + b - ab + \frac{1}{2}(1-a-b+ab) = \frac{1}{2} - \frac{1}{2}a + \frac{1}{2}b$$

and hence the corresponding Gini Index is

$$G = 1 - 2\text{Area}(\Lambda) = 1 - (1-a+b) = a - b.$$

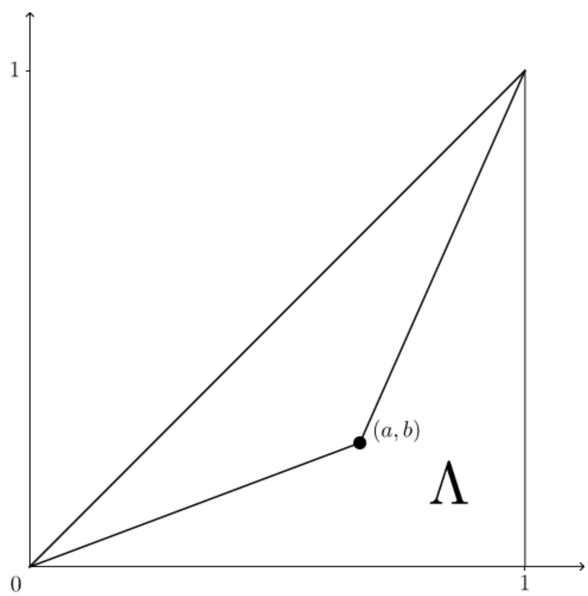


Figure 4

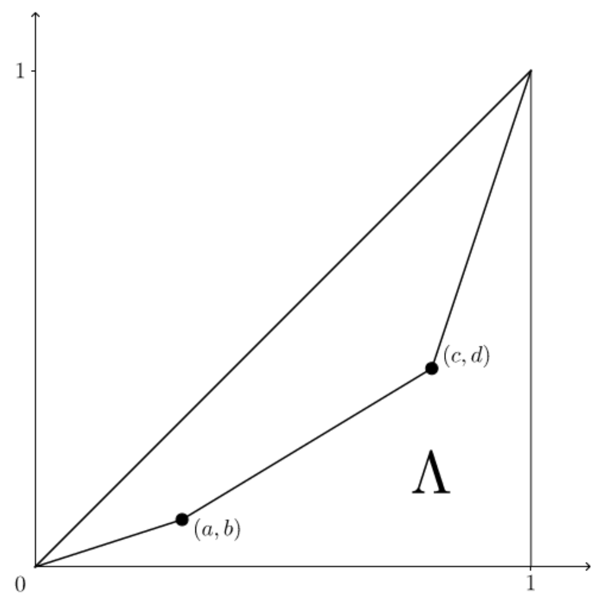


Figure 5

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.

For simplicity, this time, let's split Λ into one triangle and two trapezoids by dropping vertical lines down from the points (a, b) and (c, d). Since the area of a trapezoid with height h and bases b_1 and b_2 is given by $\frac{1}{2}h(b_1 + b_2)$, we find the area of Λ to be

$$\begin{aligned}\text{Area}(\Lambda) &= \frac{1}{2}ab + \frac{1}{2}(c-a)(b+d) + \frac{1}{2}(1-c)(d+1) \\ &= \frac{1}{2}[ab + (cb + cd - ab - ad) + (d + 1 - cd - c)] \\ &= \frac{1}{2}[cb - ad + d + 1 - c],\end{aligned}$$

and hence the corresponding Gini Index is

$$G = 1 - 2\text{Area}(\Lambda) = 1 - (cb - ad + d + 1 - c) = c - d + ad - bc.$$

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.

So in this case,

$$(a, b) = (.80, .40), \quad \text{and} \quad (c, d) = (1 - .01, 1 - .20) = (.99, .80),$$

so that by the previous exercise, the corresponding Gini Index is

$$G = c - d + ad - bc = .99 - .80 + (.80)(.80) - (.40)(.99) = .434.$$

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?

(a) If the Lorenz Curve is a straight line between (a, b) and (c, d), this means that the income earned by the individuals represented by in the $c - a$ proportion of earners is equally distributed amongst them. This is justified rigorously below at the end of the solution to part (b).

(b) Since the slope of the line segment connecting (a, b) and (c, d) is given by $m = \frac{d-b}{c-a}$, that line segment is given by

$$\ell(x) = b + (x - a)m = b + (x - a)\frac{d - b}{c - a} \quad \text{for} \quad x \in [a, c].$$

So supposing that (x, y) lies on the Lorenz Curve with $a \leq x \leq b$, our task is to show that $y \leq \ell(x)$.

Since (a, b) and (x, y) lie on the Lorenz Curve, the lowest a proportion of earners makes proportion b of the income, and the lowest x proportion of earners makes proportion y of the income. Hence the earners represented by the proportion $x - a$ make $y - b$ proportion of the income. The next $c - x$ proportion of earners, consisting of earners who make at least as much those in the lower proportion, must in total earn at least $(c - x)\frac{y-b}{x-a}$ (this is the amount they would make if they earned the same on average as the $x - a$ proportion directly below them in income, who on average make $\frac{y-b}{x-a}$ proportion of the income). Since the sum of the income earned by the $x - a$ proportion and the $c - x$ proportions must be $d - b$, we have

$$d - b \geq y - b + (c - x)\frac{y - b}{x - a},$$

and hence

$$(d - b)(x - a) \geq (y - b)(x - a) + (c - x)(y - b) = (y - b)(x - a + c - x) = (y - b)(c - a),$$

so that

$$y - b \leq (x - a)\frac{d - b}{c - a},$$

and hence

$$y \leq b + (x - a)\frac{d - b}{c - a} = \ell(x),$$

as desired.

Returning to part (a), note that equality occurs in the above computation precisely when the $c - x$ proportion of earners make exactly $(c - x)\frac{y-b}{x-a}$ of the income, that is, when they make on average the same amount as the $x - a$ proportion of earners. Under the assumption of part (a), this occurs for every x in the interval $[a, c]$. In particular, taking $x - a$ to include only the lowest earner of those represented in $(c - a)$, we see that all earners in that interval have the same income as the lowest amongst them, implying that all earners in that interval in fact earn the same income.

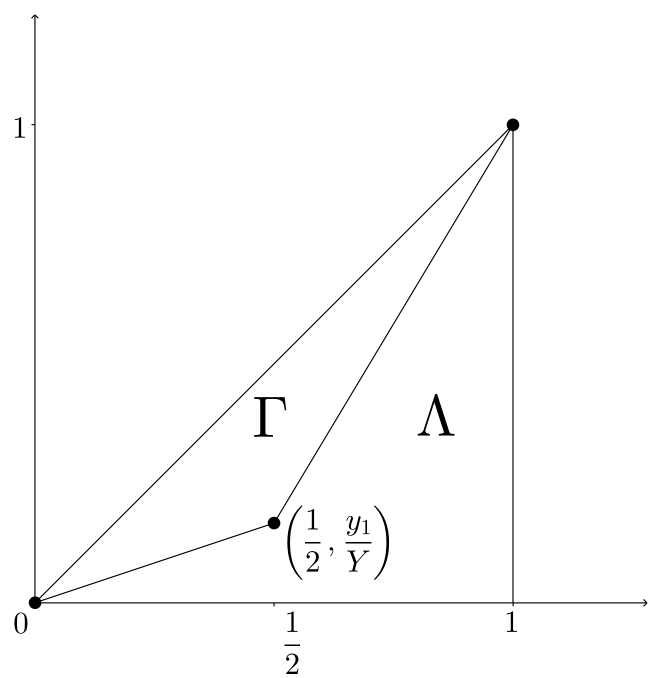
(c) The one-point and two-point estimates are always less than or equal to the actual Gini Index. This is because they both approximate the Lorenz Curve by line segments, and as shown in part (b), the actual Lorenz Curve must lie on or below these segments.

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.

Since there are only two earners in this population, there are points on the Lorenz Curve for the lowest 0%, 50%, and 100% of earners, that is, proportions $0, \frac{1}{2}$, and 1. The proportional income of the lower earner is y_1/Y, so the Lorenz Curve looks as follows:



Since this Lorenz Curve in fact agrees with the 1-point estimate with $(a, b) = (\frac{1}{2}, \frac{y_1}{Y})$, we can use our formula from Exercise 7 to find that

$$G = a - b = \frac{1}{2} - \frac{y_1}{Y}.$$

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.

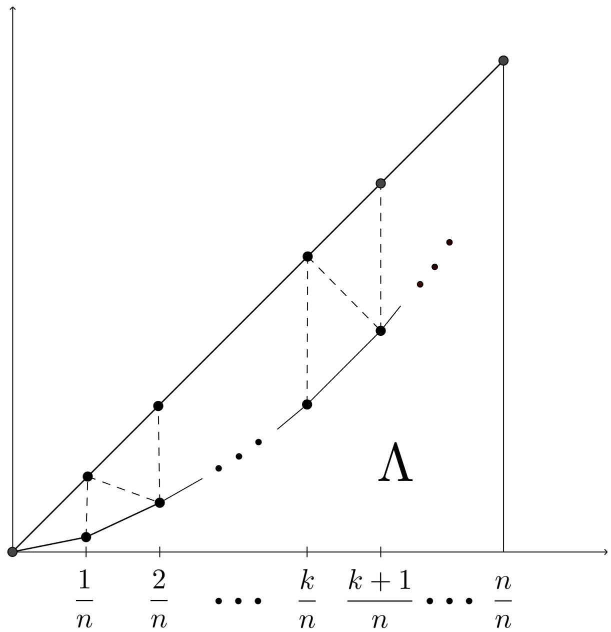


Figure 6

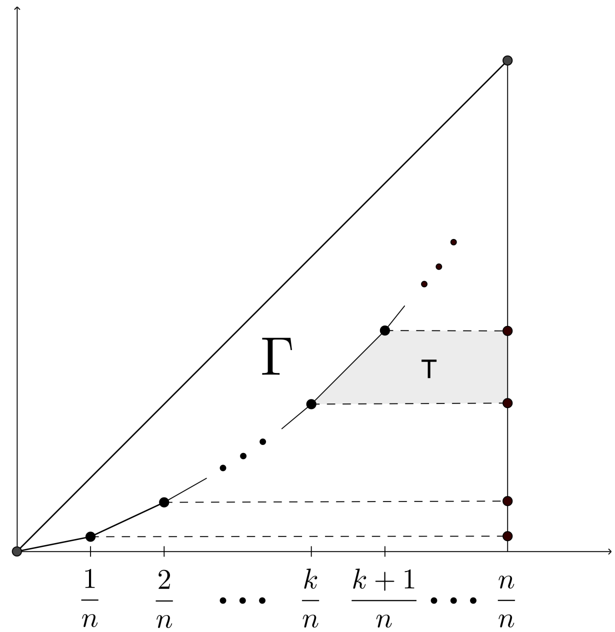
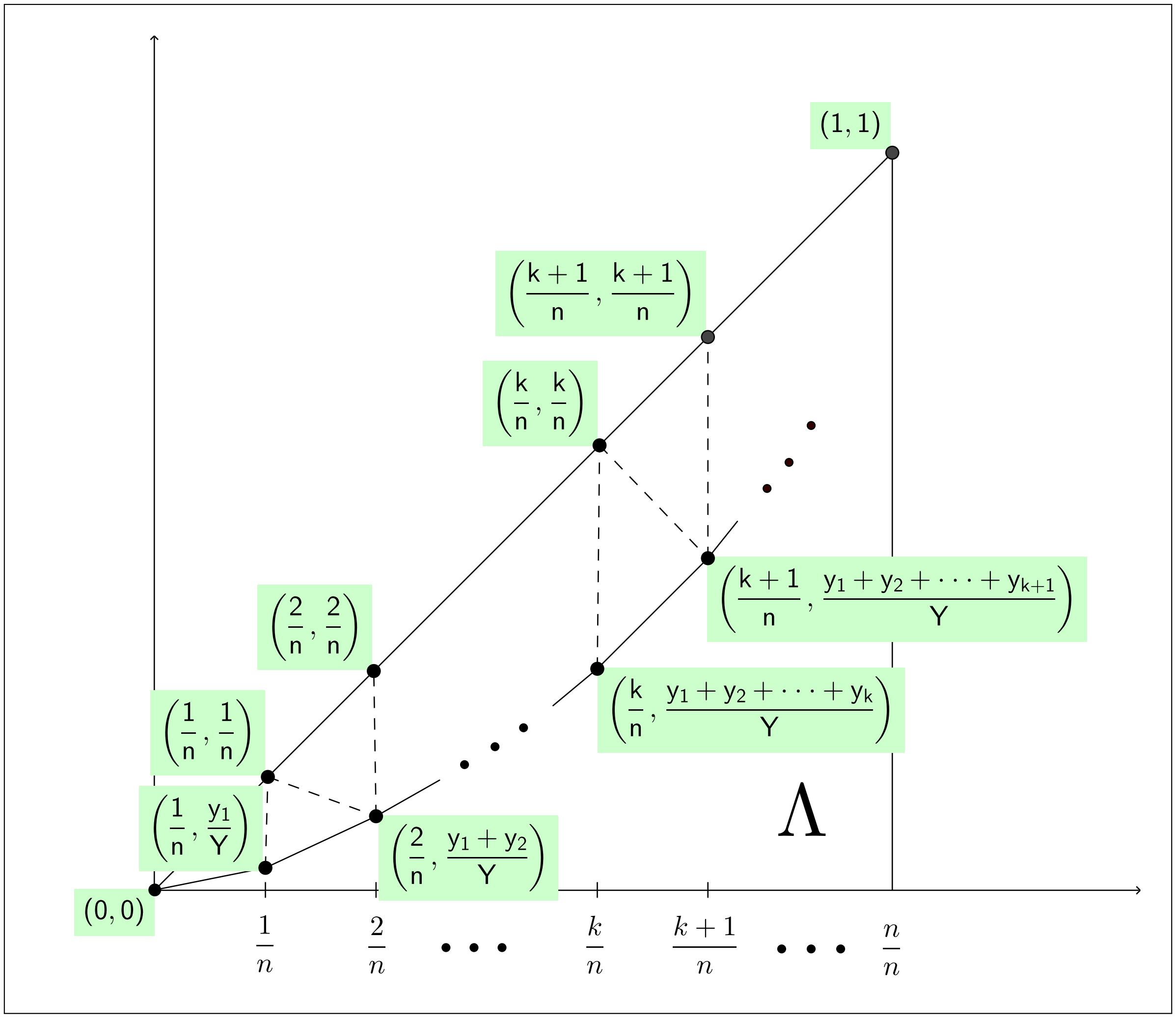


Figure 7



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

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

Examining the figure above, we see that Γ can be broken into triangles, each with (horizontal) altitude $\frac{1}{n}$. Each of the $n-1$ vertical dotted lines determines the base of two triangles, so there are $2(n-1)$ triangles. From part (a), we see that the kth vertical line has length

$$\frac{k}{n} - \frac{y_1 + y_2 + \cdots + y_k}{Y},$$

so the area of each the two triangles whose base it determines is

$$\frac{1}{2} \cdot \frac{1}{n} \cdot \left[\frac{k}{n} - \frac{y_1 + y_2 + \cdots + y_k}{Y}\right],$$

and the pair together has area

$$\frac{1}{n}\left[\frac{k}{n} - \frac{y_1 + y_2 + \cdots + y_k}{Y}\right].$$

Summing k over those $n-1$ vertical dotted lines, we see that

$$\text{Area}(\Gamma) = \sum_{k=1}^{n-1} \frac{1}{n}\left(\frac{k}{n} - \frac{y_1 + y_2 + \cdots + y_k}{Y}\right) = \frac{1}{n}\sum_{k=1}^{n-1}\left(\frac{k}{n} - \frac{y_1 + y_2 + \cdots + y_k}{Y}\right),$$

so that

$$G = 2\text{Area}(\Gamma) = \frac{2}{n}\sum_{k=1}^{n-1}\left(\frac{k}{n} - \frac{y_1 + y_2 + \cdots + y_k}{Y}\right).$$

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

Using part (a), we see that T has height

$$\frac{y_1 + y_2 + \cdots + y_{k+1}}{Y} - \frac{y_1 + y_2 + \cdots + y_k}{Y} = \frac{y_{k+1}}{Y},$$

and bases $1 - \frac{k}{n}$ and $1 - \frac{k+1}{n}$, so that its area is

$$\frac{1}{2}\frac{y_{k+1}}{Y}\left(1 - \frac{k}{n} + 1 - \frac{k+1}{n}\right) = \frac{y_{k+1}}{2nY}(2n - 2k - 1)$$

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

This is the trapezoid directly below T, so we can find its area by replacing k by $k-1$ in our formula from part (c), in which case we have its area to be

$$\frac{y_k}{2nY}(2n - 2(k-1) - 1) = \frac{y_k}{2nY}(2n - 2k + 1)$$

(e) Use part (d) to show that

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

Using part (d), and summing over the n trapezoids, we find that

$$\text{Area}(\Lambda) = \sum_{k=1}^{n}\frac{y_k}{2nY}(2n-2k+1) = \frac{1}{2nY}\sum_{k=1}^{n}(2n-2k+1)y_k,$$

and hence

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

(f) Prove that

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

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

Combining part (e) with $1 = \frac{1}{Y}\sum_{k=1}^{n} y_k$, we have

$$\begin{aligned} G &= 1 - \frac{1}{nY}\sum_{k=1}^{n}(2n-2k+1)y_k \\ &= \frac{1}{Y}\sum_{k=1}^{n} y_k - \frac{1}{nY}\sum_{k=1}^{n}(2n-2k+1)y_k \\ &= \frac{1}{nY}\sum_{k=1}^{n} y_k[n-(2n-2k+1)] \\ &= \frac{1}{nY}\sum_{k=1}^{n}(2k-n-1)y_k. \end{aligned}$$

(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).

As discussed in the hint, it is enough to prove the case where $y_1, y_2, \ldots, y_k$ are in fact in ascending order. Given any index k with $1 \leq k \leq n$, we know that in the double sum $\sum_{i,j=1}^{n} |y_i - y_j|$, the term y_k can appear as either $|y_k - y_j|$ or $|y_i - y_k|$. Note that for $1 \leq j < k$ we have $|y_k - y_j| = y_k - y_j$ (since then $y_k - y_j \geq 0$), and for $k < j \leq n$ we have $|y_k - y_j| = y_j - y_k$ (since then $y_k - y_j \leq 0$). Similarly $|y_i - y_k| \leq y_k - y_i$ for $1 \leq i < k$ and $|y_i - y_k| = y_i - y_k$ for $k < i \leq n$.

Hence in the double sum, y_k appears $2(k-1)$ times (when $1 \leq j < k$ or $1 \leq i < k$, as above) whereas $-y_k$ appears $2(n-k)$ times (when $k < j \leq n$ or $k < i \leq n$, as above). So if we expand $\sum_{j=1}^{n} |y_k - y_j|$, the coefficient on y_k will be $2(k-1) - 2(n-k) = 2(2k - n - 1)$. Therefore, doing this for each k, we find that

$$\frac{1}{2nY} \sum_{i,j=1}^{n} |y_i - y_j| = \frac{1}{2nY} \sum_{k=1}^{n} 2(2k - n - 1)y_k = \frac{1}{nY} \sum_{k=1}^{n} (2k - n - 1)y_k,$$

which we know is equal to the Gini Index by the result of 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.

We will use the result of part (g), so we begin by noting that $n = 4$ and $Y = 20000 + 40000 + 60000 + 80000 = 200000 = 2 \cdot 10^5$. Noting that for each pairwise difference with $i \neq j$ appears twice in that formula (once as $|y_i - y_j|$ and once as $|y_j - y_i|$, we have

$$\begin{aligned} G = \frac{1}{2 \cdot 4 \cdot (2 \cdot 10^5)} \big[& 2 \cdot |20000 - 40000| + 2 \cdot |20000 - 60000| + 2 \cdot |20000 - 80000| + 2 \cdot |40000 - 60000| \\ & + 2 \cdot |40000 - 80000| + 2 \cdot |60000 - 80000| \big] \\ & = \frac{1}{4}. \end{aligned}$$

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.

In this case, the Gini Index will decrease. To see this, we will again use the result of Exercise 12(g), which gives that $G = \frac{1}{2nY}\sum_{i,j=1}^{n}|y_i - y_j|$ in the notations there. Note that the quantity $\sum_{i,j=1}^{n}|y_i - y_j|$ does not change if we add B to every income, since

$$\sum_{i,j=1}^{n} |(y_i + B) - (y_j + B)| = \sum_{i,j=1}^{n} |y_i - y_j|.$$

On the other hand, $\frac{1}{2nY}$ decreases, since the total income Y becomes $Y + nB > Y$. Thus the Gini Index $G = \frac{1}{2nY}\sum_{i,j=1}^{n}|y_i - y_j|$ itself decreases, if $B > 0$ is added to every income.

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?

We will again use the result of Exercise 12(g), so we begin by noting that now we have a population size of $\tilde{n} = 2$ households. The household of Alice and Bob has income $\tilde{y}_1 = 20,000 + 40,000 = 60,000$, and the household of Carl and Danielle has income $\tilde{y}_2 = 60,000 + 80,000 = 140,000$. Thus $\tilde{Y} = 60,000 + 140,000 = 200,000 = 2 \cdot 10^5$, and

$$G = \frac{1}{2 \cdot 2 \cdot (2 \cdot 10^5)}\left[2 \cdot |60,000 - 140,000|\right] = \frac{1}{4 \cdot 10^5}\left[80,000\right] = \frac{8}{4 \cdot 10} = \frac{1}{5}.$$

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.

For part (a), consider a population with $n = 2$ earners with incomes $y_1 = 1$ and $y_2 = 2$, so that $Y = 1 + 2 = 3$. Then by Exercise 12(g), the Gini Index is $G = \frac{1}{2 \cdot 2 \cdot Y}\left[2|y_1 - y_2|\right] = \frac{1}{2 \cdot 3} \cdot 1 = \frac{1}{6}$.

But if the incomes now increase to $\tilde{y}_1 = 2$ and $\tilde{y}_2 = 10$, so that $\tilde{Y} = 12$, then by Exercise 12(g), the Gini Index $G = \frac{1}{2 \cdot 2 \cdot Y}\left[2|y_1 - y_2|\right] = \frac{1}{2 \cdot 12} \cdot 8 = \frac{1}{3}$ has increased.

For part (b), we can simply do the reverse of part (a); that is, if we consider a population with $n = 2$ earners with incomes $y_1 = 2$ and $y_2 = 10$, the Gini Index will be $G = \frac{1}{3}$ as shown above, but when the incomes decrease to $\tilde{y}_1 = 1$ and $\tilde{y}_2 = 2$, the Gini Index also decreases to $G = \frac{1}{6}$.

Exercise 16.

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

In fact, we can give a good example using Lorenz Curves which agree with the 1 point approximation of Exercise 7; that is, which consist of a line segment connecting $(0,0)$ to a point (a,b), together with a line segment connecting (a,b) to $(1,1)$. Such a curve is realizable as an actual Lorenz Curve whenever a is a rational number with $0 < a < 1$ and $0 \leq b \leq a$. Exercise 7 showed that the Gini Index in such a case is given by $a - b$.

So if we take $a_0 = \frac{1}{4}$ and $b_0 = 0$, we get the same Gini Index ($G = \frac{1}{4}$) as if we take $a_1 = \frac{99}{100}$ and $b_1 = \frac{74}{100}$. But the former case corresponds to a population in which the bottom quarter of earners make absolutely nothing with income distributed evenly amongst everyone else, whereas the latter case corresponds to an equal distribution of nearly three quarters of the income amongst the lowest 99% of earners, with the top 1% making the rest. Naturally, these are qualitatively very different situations.

(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.

For our first distribution, $L(x)^2$ consists of a segment from $(0,0)$ to $(a_0, b_0^2) = (\frac{1}{4}, 0)$ (it remains a line since $0^2 = 0$) together with a parabolic arc connecting $(a_0, b_0^2) = (\frac{1}{4}, 0)$ to $(1,1)$. Using the given formula, the area under the latter is $\frac{(1-1/4)(1)^2}{3} = \frac{1}{4}$, which is the area of Λ, so that

$$G_2 = 1 - 2\text{Area}(\Lambda) = 1 - 2 \cdot \frac{1}{4} = \frac{1}{2}.$$

In the second distribution, we will have a parabolic arc connecting $(0,0)$ to $(a_1, b_1^2) = (\frac{99}{100}, (\frac{74}{100})^2)$ together with a parabolic arc from there to $(1,1)$. The area under the former is $\frac{a_1 b_1^2}{3}$. The area under the latter can be divided into a rectangular portion of area $(1 - a_1)b_1^2$ together with portion to which the given formula applies with a result of $\frac{(1-a_1)(1-b_1)^2}{3}$. So all together, we find that

$$\begin{aligned}\text{Area}(\Lambda) &= \frac{a_1 b_1^2}{3} + (1 - a_1)b_1^2 + \frac{(1-a_1)(1-b_1)^2}{3} \\ &= \frac{1}{3}\left(\frac{99}{100}\right)\left(\frac{74}{100}\right)^2 + \left(\frac{1}{100}\right)\left(\frac{74}{100}\right)^2 + \frac{1}{3}\left(\frac{1}{100}\right)\left(\frac{26}{100}\right)^2 = 139807/750000 \approx .186\end{aligned}$$

giving a Second Gini Coefficient of

$$G_2 = 1 - 2\text{Area}(\Lambda) = 235193/375000 \approx .627.$$

So even though these distributions have the same Gini Index G, G_2 is able to distinguish them!

(**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.

This problem is meant to give the student a bit of work, and we will not give a complete solution here, but we will get the reader started:

It is indeed possible for two different income distributions to have the same G and the same G_2 simultaneously. Intuitively, this is because the two pieces of information are not enough to fully recover the information contained in the Lorenz Curve whenever the population has size $n \geq 4$ (in fact, for $n = 3$, they are enough). However, it is not possible to give two distinct Lorenz Curves that agree with 1-point approximations and which have the same G and G_2. Intuitively, this is because such a Lorenz Curve is specified by only two parameters (namely, the choice of a and b).

It is possible, however, to find such an example using Lorenz Curves that mimic the 2-point approximation (although the population will have to be greater than $n = 3$, since in that case a and c are predetermined as $a = \frac{1}{3}$ and $c = \frac{2}{3}$). In order to do so, using the given parabolic area formula, the reader can derive a formula for G_2 in this setting. A system of two equations ($G = G'$ and $G_2 = G_2'$) is then determined, and one needs only choose distinct parameters (a, b, c, d) and (a', b', c', d') which solve them while obeying the other constraints of the situation (e.g., that $b \leq a$).