12 Lorenz line and Gini index

In 1905, in the paper *Methods of measuring the concentration of wealth*, Max Otto Lorenz, an American economist, introduced a polygonal line which starts at the point (0,0) and goes through points

$$L_k := \left(\frac{k}{N}, \frac{\sum_{j=1}^{k} z_j}{\sum_{j=1}^{N} z_j} \right)$$

determined by an ordeence (i.e., a non-decreaslingly ordered sample)

$$z = (\, z_1, z_2, \ldots, z_N \,).$$

Any of these points is called a **Lorenz point**, and the line is referred to as a **Lorenz line** (defined on equidistant set, i.e., with abscissas $h_j = j/N$, $j = 0...N$). It is easy to see that this line is contained in the square <0, 1>×<0, 1> marked in the Cartesian coordinate system *Oxy*. More precisely, it lies in the right lower half of this square.

As the title of his paper says, Lorenz used this curve to express numerically how the wealth is distributed among people. An extreme case holds true if there is no differentiation, every member has the same wealth what any other does ($z_j = z_k$ for all values j and k). Then there is the ideal equality, all ordinates are equal to abscissas, and, in consequence, the Lorenz line is the diagonal of the square (this is sometimes referred to as a **line of perfect equality**). Otherwise it does not hold; for instance, the point (0.8, 0.2) says that eighty percent of employees receives a fifth of the total salaries paid (and, at the same time, 20% of best paid employees takes 80% of wages).

Example. Let's work again with the enterprise *We20*, where the payroll is

$$z = (\,2.0, 2.0, 2.1, 2.2, 2.2, 2.9, 2.9, 2.9, 2.9, 3.1, 3.3, 3.3, 3.3, 3.5, 3.5, 3.8, 4.3, 6.4, 7.0, 10.0).$$

It produces Lorenz points:

$$L_1 = \left(\frac{1}{20}, \frac{2.0}{73.6} \right) = (0.05, 0.027),$$

$$L_2 = \left(\frac{2}{20}, \frac{2.0+2.0}{73.6} \right) = (0.10, 0.054),$$

$$L_3 = \left(\frac{3}{20}, \frac{2.0+2.0+2.1}{73.6} \right) = (0.15, 0.082),$$

$$L_3 = \left(\frac{4}{20}, \frac{2.0+2.0+2.1+2.2}{73.6} \right) = (0.20, 0.113),$$

$\ldots, L_{20} = (1, 1).$

The complete sequence of Lorenz points is

(0, 0), (0.05, 0.027), (0.10, 0.054), (0.15, 0.082), (0.20, 0.113), (0.25, 0.143), (0.30, 0.179), (0.35, 0.218), (0.40, 0.248), (0.45, 0.297), (0.50, 0.339), (0.55, 0.384), (0.60, 0.474), (0.65, 0.479), (0.70, 0.521), (0.75, 0.569), (0.80, 0.621), (0.85, 0.681), (0.90, 0.769), (0.95, 0.864), (1, 1).

Marking them and connecting successively we get the Lorenz line presented in the figure below.

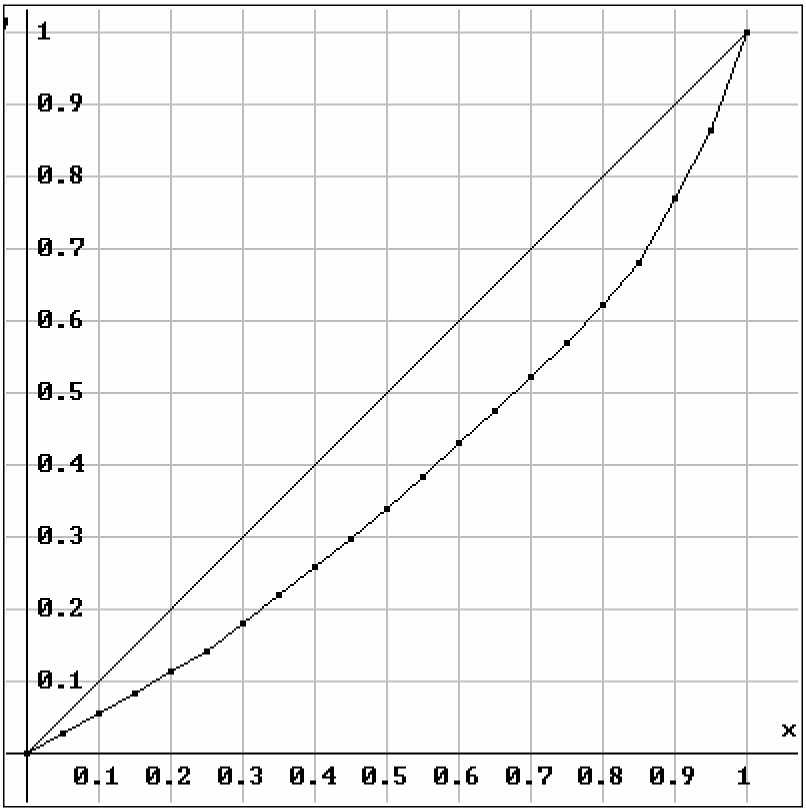


Fig. . The line of perfect equality and the Lorenz line for the payroll in *We20*. It reads that, for example,

25% of lowest paid workers get 14.3% of wages,

and 25% of best paid employees takes 100–62.1=37.9% of the wage fund.

Thus the inequality between these two groups is 37.9/14.3 = 2.65.

Notice that the extreme disproportion is greater (by the factor 1.12, or by 12%):

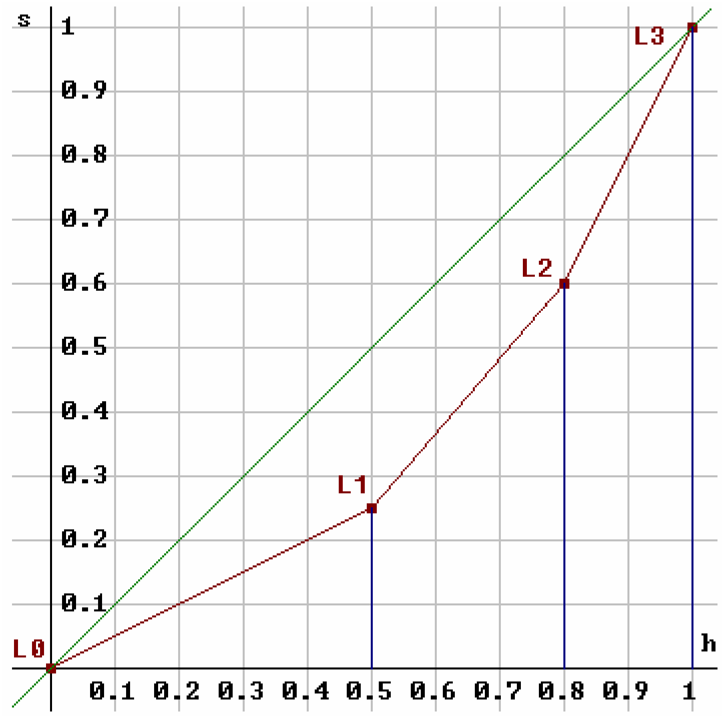
five lowest paid workers gain together 10.5 thousand zlotys,

five best remunerated employees are granted with 31.7 thousand zlotys,

and it results with the disproportion 31.7/10.5 = 2.96 = 1.12·2.65.

DRAFT VERSION

The Lorenz line represents graphically the distribution of the total wealth among the sample at hand (in our example: the wealth is the total month salary going to employees in the firm *We20*). It sits in the square $<0, 1>^2$, and the diagonal of this square is – as we already noticed – a graphical representation of the perfect equality. The area of the region bounded by these two lines, namely by the Lorenz line and the diagonal, is commonly accepted measure of the unevenness/irregularity a good/wealth is distributed.



Lorenz points and Lorenz line determined by data considered in the example. There are also traced the diagonal and the verticals separating trapeziums determined by Lorenz points

Example. Let's work, for example (see figure …), with the Lorenz points

$$L_0 := (0, 0), L_1 := (0.5, 0.25), L_2 := (0.8, 0.6), L_3 := (1, 1).$$

They describe/report the following distribution of the wealth: the poorest half of the society owns 25% of the total wealth, 40% of the total wealth is owned by the richest 20% of the society.

The Lorenz line form the triangle (which can be seen as a degenerated trapezium) and two trapezia. All these figures stay on the horizontal axis, their vertices are L_0 and L_1, L_1 and L_2, and L_2 and L_3, resp. Their areas [1] are

[1] the area of a trapezium is $(a+b)/2 \cdot h$, where a and b are length of its bases, and h is the length of its height, in oor considerations the bases are perpendicular to the horizontal axis, and the height is measured along the horizontal axis)

$$\frac{0+0.25}{2}\cdot 0.5 = 0.0625\,,\ \frac{0.25+0.6}{2}\cdot 0.3 = 0.1275 \text{ and } \frac{0.6+1}{2}\cdot 0.2 = 0.16\,,$$

resp., so the region under the Lorenz line is of area

$$0.0625 + 0.1275 + 0.16 = 0.35.$$

In consequence, the Lorenz line and the diagonal bound the region of area

$$0.5–0.35 = 0.15,$$

and it is the third tenths (0.15/0.5 = 0.3), or 30%, of the area of the lower right triangle (i.e., the triangle which stays on the horizontal axis and is bounded form above by the diagonal). Accordingly to the definition we give below, 0.3 is the Gini coefficient determined by Lorenz points at hand.

End of Example.

As it was said, the quantity, 0.3, we calculated in the above example is called the Gini coefficient. In general case, i.e, when there are given two ordeences,

$$h = (h_j)_{j=1..N} \text{ and } z = (z_j)\, j_{=1..N},$$

and $0 = h_0 < h_1 < h_2 < \ldots < h_{N-1} < h_N = 1,$
$0 = s_0 < s_1 < s_2 < \ldots < s_{N-1} < s_N = 1,\ N \geq 2,$

the value $r_j = zj$

zle zle zle zle DRAFT VERSION

----------------------------//////////////-

//////?????? áßáß© ß

sum $s_j := z_1 + z_2 + \ldots + z_j$ is called a j-th cumulative sum generated by z,

the point $L_j := (h_j, s_j)$ is referred to as a j-th Lorenz point

a **Gini index**, or a **Gini coefficient**, determined by the pair (h, z) is defined as the doubled area of the region bounded by the corresponding Lorenz line and the diagonal. It can be denoted by Gini(L), Gini(h, s), where $L = (\ L_j\ )\, j$=0..N,

when there are given Lorenz points

$$L_j = (\ h_j,\ s_j\ ),\ j = 0..N,$$

with $L_0 = (0, 0)$ and $L_N = (1, 1)$,

where $s_j = z_1 + z_2 + \ldots + z_j$ for j =1..N

$L_0 = (0, 0)$, $L_1 = (h_1, s_1)$, $L_2 = (h_2, s_2)$, …, $L_{N-1} = (h_{N-1}, s_{N-1})$, $L_N = (h_N, s_N) = (1, 1)$,

a **Gini index**, or a **Gini coefficient**, determined by them is defined as the doubled area of the region bounded by the corresponding Lorenz line and the diagonal. It can be denoted by Gini(L), Gini(h, s), where $L = (L_j)\, j{=}0..N$,
Above there were considered the Lorenz line and Gini index for the ordeence

$$z = (z_1, z_2, ..., z_N)$$

corresponding to the equidistant case, i.e., when $h_j = j/N$.
This is generalized to the case where h_j's are not equidistantly distributed, and we start to treat it now. So we have two ordeences,

$$h = (h_j)_{j=1..N} \text{ and } s = (s_j)\, j_{=1..N},$$

and $0 = h_0 < h_1 < h_2 < \ldots < h_{N-1} < h_N = 1$,
$0 = s_0 < s_1 < s_2 < \ldots < s_{N-1} < s_N = 1$,
with $N \geq 2$.
For any $j \in \{1, 2, \ldots, N-1\}$,

- h_j divides the whole population (taken to be 1, i.e., 100%) in two parts, they are referred to as a **poorer part** and a **richer part**, every member of a poorer part is poorer than any member of a richer part,
- s_j says that the j-th poorer part has s_j-th part of the total wealth the whole population owns,
- $L_j := (h_j, s_j)$ is, as above, referred to as a j-th **Lorenz point**.

As above, there are also taken into account two more points (namely points corresponding to $j = 0$ and $j = N$):

$L_0 := (h_0, s_0) = (0, 0)$,
$L_N := (h_N, s_N) = (1, 1)$.

This completes N+1 Lorenz points $L_0, L_1, L_2, \ldots, L_N$. Their sequence, $(L_j)_{j=1..N}$, describes the distribution of the wealth. As above, the polygonal line starting at $L_0 = (0, 0)$, finishing at $L_N = (1, 1)$ and joining successive Lorenz points is called a **Lorenz line**.

The diagonal of the square $<0, 1>^2$ is the graphical
Now we can calculate the area of the surface bounded by the Lorenz line and the diagonal of the square. sum

Lorenz points and Lorenz line are also defined for an arbitrary set; then they are determined by a frequence (h, s), where $h = (h_j)_{j=0..n+1}$ and $s = (s_j)_{j=0..n+1}$ are increasingly ordered sequences of numbers from <0, 1> with $h_0 = s_0 = 0$ and $h_{n+1} = s_{n+1} = 1$. Now $L_j = (h_j, s_j)$.

A particular case is when $n = 1$, $h_1 = 0.8$, $s_j = 0.2$, is related to a **80/20 rule**, aka a **law of the vital few**, a **Pareto principle**. In 1890's Italian civil engineer, economist and sociologist Vilfredo Pareto noticed that circa 80% of homes are owned by 20% of families.
A Pareto principle is an example of a **rule of thumb**, this is a simple rule, which accurate and reliable enough in the most of situations. For example, in the monetary market there is a **rule of 70**, as a mental math shortcut saying that the number of years required to double the asset is $70/R$ provided that the annual percentage rate $r = R\,\% = R/100$, is 1–5% and the interest is compounded. For example, if $r = 2\%$ (so $R = 2$), then the time, $T = T(r)$, at which the asset is doubled is circa 70/2 = 35 years.
The exact formula reads $(1+r)^T = 2$, so $T = \ln 2/\ln(1+r)$. Since $\ln(1+r) \approx r$, so $T \approx \ln 2/r \approx 0.6931/r$ and, in consequence $T \approx 70/R$.
When r is 5–10%, the better approximation is provided by $72/R$; this is referred to as a **rule of 72**, and was first presented, as in Luca Paccioli's treatise *Summa de arithmetica* (1494). The first Taylor polynomial around 0.08 is
$\ln 0.08 + (r–0.08)/1.008 \approx 0.9259r+0.00289$,
so $\frac{\ln 2}{\ln(1+r)} \approx \frac{0.6931}{0.9258r + 0.00289} \approx \frac{0.72}{r}$.

In 1912, in the paper *Variabilità e mutabilità*, Conrado Gini introduced the quantity which is now called a **Gini coeffcient**, **Gini index, Gini ratio**. This is also called a **concentration index**, an **inequality coefficient**, both these names directly refer to what the Gini index does: it measures measures the concentration (of wealth), the inequality (in the distribution of wealth). Given a ordeence

$$z = (z_1, z_2, ..., z_N),$$

Gini index equals [2]

$$\text{Gini}(z) := \frac{2}{N^2\,\bar{z}} \sum_{k=1}^{N} \left(k - \frac{N+1}{2} \right) \cdot z_k \;=\; \frac{2}{N^2\,\bar{z}} \sum_{k=1}^{N} k \cdot z_k \;-\frac{N+1}{N}$$

$$= \frac{1}{N} \left\{ 2 \cdot \frac{\sum_{k=1}^{N} k \cdot z_k}{\sum_{k=1}^{N} z_k} + 1 \right\} - 1 \;=\; \frac{1}{N} \left\{ 2 \cdot \frac{\mu_1}{\mu_0} + 1 \right\} - 1 \;=\; \frac{1}{N} \cdot \left\{ 2 \frac{\mu_1}{\mu_0} + N - 1 \right\}$$

where, as usual, $\bar{z} = \mu_0$ denotes the mean of the ordeence z,

μ_1 is the first raw moment of z, $\mu_1 = \frac{1}{N} \sum_{k=1}^{N} k \cdot z_k$.

Gini index assumes values in the interval <0, 1>. It is often expressed in percent, then 0% corresponds to an ideal equality, and it becomes greater as the inequality is bigger. It is standard to use it when there are discussed the inequalities in various countries. For example, after data provided by World Bank (24.11.2011), Gini index is equal to

25% in Sweden, Japan and Czech Republic,
29% in Austria and in Germany,
31% in Slovenia, 33% in France, 34% in Poland and Spain,
41% in USA, 42% in Russia,
53% in Lesotho, 60% in Haiti, 74% in Namibia.

Geometrically the Gini index is interpreted as the measure how much the Lorenz line is deviated from the diagonal of the square $<0, 1>^2$.

[2]] we will produce this formula further on, this is a particular (namely the equidistant) case of a general expression for a Gini indem.

An extreme case holds for two-element ordeence (z_1, z_2). It produces three Lorenz points, $L_0 = (0, 0)$, $L_1 = \left(\frac{1}{2}, \frac{z_1}{z_1 + z_2}\right)$, $L_2 = (1, 1)$. For instance, with $z_1 = 2$, $z_2 = 8$ we have the relation of absolute wealth equal to 8:2 = 4, and we have

$$L_0 = (0, 0), L_1 = (0.5, 0.2), L_2 = (1, 1).$$

It says that the poorest half has 20% of wealth, and the richest half owns 80% of the wealth. The area contained in the square $<0, 1>^2$ and laying below the Lorenz line is

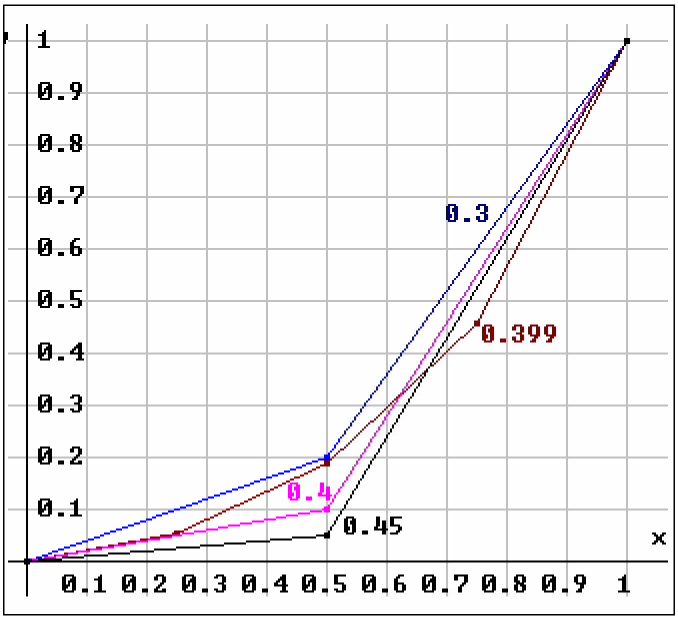
$$\frac{1}{2}\cdot 0.5\cdot 0.2 + 0.5\cdot 0.2 + \frac{1}{2}\cdot 0.5\cdot 0.8 = 0.35,$$

so the area between this line and the line of perfect equality is 0.5–0.35 = 0.15, and Gini index is 2·0.15 = 0.30. Notice that it is more than the real disproportion 0.20/0.80 = 0.25.

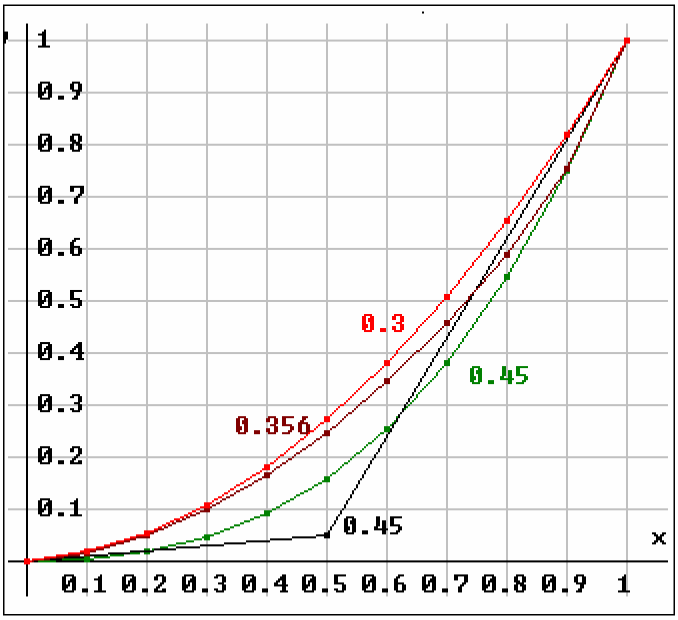
Lorenz points and Lorenz line are also defined for an arbitrary set, i.e., can b

For $z = (10, 25, 50, 100)$ Lorenz points and Gini index are (0, 0), (0.25, 0.054), (0.5, 0.189), (0.75, 0.459), (1, 1) and 0.399, resp.

Both Lorenz lines, and two more ones, are drawn in figure below.



Lorenz lines for two ordeences (1, 19), (1, 9), (1, 4) and (1, 2.5, 5, 10) and their Gini inexes 0.45, 0.4, 0.3 and 0.399, resp.



Lorenz lines for ordeences (1, 19), (1, 3, 6, 10, 15, 21, 28, 36, 45, 55) and (1, 2, 3, 4, 5, 6, 7, 8, 10, 15), (1, 2, 3, 4, 5, 6, 7, 8, 9, 10) and their Gini indexes 0.45, 0.45, 0.356 and 0.3, resp.

Above there were considered the Lorenz line and Gini index for the ordeence

$$z = (z_1, z_2, ..., z_N)$$

corresponding to the equidistant case, i.e., when $h_j = j/N$.

This is generalized to the case where h_j's are not equidistantly distributed, and we start to treat it now. So we have two ordeences,

$$h = (h_j)_{j=1..N} \text{ and } s = (s_j)\, j_{=1..N},$$

and $0 = h_0 < h_1 < h_2 < \ldots < h_{N-1} < h_N = 1$,

$0 = s_0 < s_1 < s_2 < \ldots < s_{N-1} < s_N = 1$,

with $N \geq 2$.

For any $j \in \{1, 2, \ldots, N-1\}$,

h_j divides the whole population (taken to be 1, i.e., 100%) in two parts, they are referred to as a **poorer part** and a **richer part**, every member of a poorer part is poorer than any member of a richer part,

s_j says that the j-th poorer part has s_j-th part of the total wealth the whole population owns,

$L_j := (h_j, s_j)$ is, as above, referred to as a j-th **Lorenz point**.

As above, there are also taken into account two more points (namely points corresponding to $j = 0$ and $j = N$):

$L_0 := (h_0, s_0) = (0, 0)$,

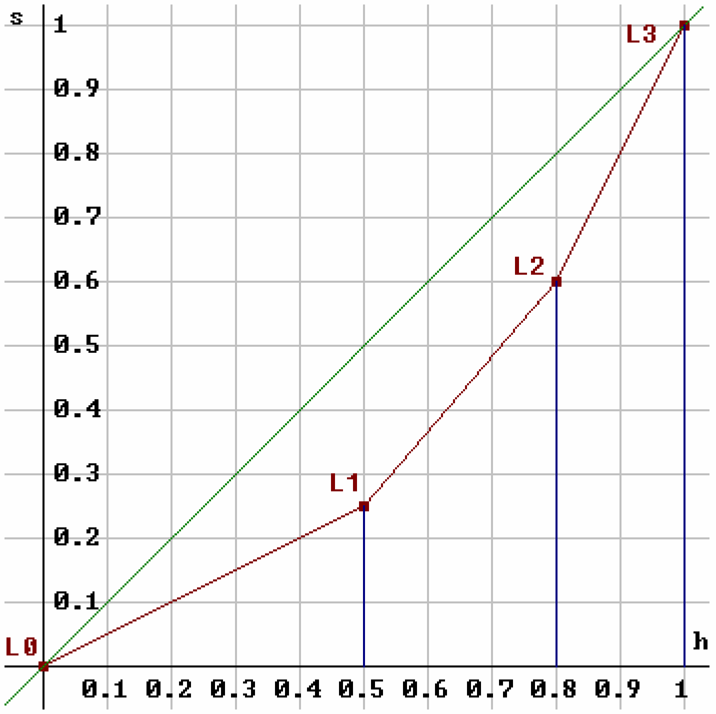
$L_N := (h_N, s_N) = (1, 1)$.

This completes N+1 Lorenz points $L_0, L_1, L_2, \ldots, L_N$. Their sequence, $(L_j)_{j=1..N}$, describes the distribution of the wealth. As above, the polygonal line starting at $L_0 = (0, 0)$, finishing at $L_N = (1, 1)$ and joining successive Lorenz points is called a **Lorenz line**.

For example (see figure …), the points

$$L_0 := (0, 0),\ L_1 := (0.5, 0.25),\ L_2 := (0.8, 0.6),\ L_3 := (1, 1)$$

describe/report the following distribution of the wealth: the poorest half of the society owns 25% of the total wealth, 40% of the total wealth is owned by the richest 20% of the society.



Lorenz points and Lorenz line determined by data considered in the example. There are also traced the diagonal and the verticals separating trapeziums determined by Lorenz points

Obviously, the Lorenz line lays in the right lower half of the square <0, 1>×<0, 1> and is increasing. Let A denote the region the Lorenz line bounds from above in this square. This region is the sum of N trapezia $A_1, A_2, \ldots, A_N$ (with the first trapezium, A_1, being a triangle), so the area $|A|$ of the region A is

$$\begin{aligned}|A| &= |A_1| + |A_2| + \ldots + |A_N| \\ &= \frac{1}{2}h_1 s_1 + \frac{(s_1+s_2)\cdot(h_2-h_1)}{2} + \ldots + \frac{(s_{N-1}+s_N)\cdot(h_N-h_{N-1})}{2} \\ &= \frac{1}{2}\sum_{k=1}^{N}(s_{k-1}+s_k)\cdot(h_k-h_{k-1}).\end{aligned}$$

Let B stands for the region bounded by the Lorenz line and the diagonal. Its area is

$$|B| = 0.5 - |A| = \frac{1}{2} \cdot \left(1 - \sum_{k=1}^{N} (h_k - h_{k-1}) \cdot (s_{k-1} + s_k) \right).$$

Obviously, $0 < |B| < 0.5$. Multiplying it by 2 we get a quantity assuming values between 0 and 1 (or, equivalently, between 0 and 100%). This is called a **Gini coefficient**, or **Gini index** (determined by sequences h and s, or by Lorenz points $L_1, L_2, \ldots, L_{N-1}$),

$$\mathrm{Gini}(h, s) := 1 - \sum_{k=1}^{N} (h_k - h_{k-1}) \cdot (s_{k-1} + s_k) .$$

With above data, $h = (0, 0.5, 0.8, 1)$ and $s = (0, 0.25, 0.6, 1)$, it is

$$\mathrm{Gini}(h, s) = 1 - \{ (0.5-0)\cdot(0+0.25) + (0.8-0.5)\cdot(0.25+0.6) + (1-0.8)\cdot(0.6+1) \}$$
$$= 1 - 0.4125 = 0.5875 = 58.75\%.$$

It is said that the inequality in the wealth distribution is 58.75%, or (in practice, where a high accuracy is not needed, or even can not be achieved) that the inequality in the wealth distribution is 59%.

DRAFT VERSION

When $s_k = h_k$ for each k, then the Lorenz line coincides with the diagonal of the square $<0, 1> \times <0, 1>$, so $|A| = 0.5$ and $\mathrm{Gini}(h, s) = 0$. This is also obtained from the formula defining Gini index, namely there is

$$\mathrm{Gini}(h, h) = 1 - \sum_{k=1}^{N} (h_k - h_{k-1}) \cdot (h_{k-1} + h_k) = 1 - \sum_{k=1}^{N} (h_k{}^2 - h_{k-1}{}^2)$$

$$= 1 - \left(\sum_{k=1}^{N} h_k{}^2 - \sum_{k=1}^{N} h_{k-1}{}^2 \right) = 1 - (\, h_N^2 - \, h_0^2) = 1 - (\, 1 - 0 \,) = 0.$$

The case $s = h$ is referred to as a **perfect equality**, and the respective Lorenz line – a line of perfect equality.

For the equidistant case, $h_k = \dfrac{k}{N}$, there is $h_k - h_{k-1} = \dfrac{1}{N}$ and the Gini index [3] is

$$\mathrm{Gini}(z) = \mathrm{Gini}(h, s) = 1 - \frac{1}{N} \sum_{k=1}^{N} (s_{k-1} + s_k),$$

and it gives the formula we presented above. To see it, notice first that cumulative shares $s_0 = 0, s_1, s_2, \ldots, s_{N-1}, s_N = 0$ determine univocally values $f_j = s_j - s_{j-1}$ and $z_j = S{\cdot}f_j$, where

$$S := s_N = \sum_{j=1}^{k} z_j$$

is the total wealth. Thus the k-th cumulative share is

$$s_k = \frac{1}{S} \sum_{j=1}^{k} z_j \,.$$

Using this expression we have

$$\frac{S}{2} \cdot \sum_{k=1}^{N} (s_{k-1} + s_k) = \frac{S}{2} \cdot (\, s_0 + 2s_1 + 2s_2 + 2s_3 + \ldots + 2s_{N-1} + s_N \,)$$

$$= S{\cdot}s_1 + S{\cdot}s_2 + S{\cdot}s_3 + \ldots + S{\cdot}s_{N-1} + S{\cdot}s_N - \frac{S}{2} \cdot s_N$$

$$= z_1 + (\, z_1 + z_2 \,) + (\, z_1 + z_2 + z_3 \,) + \ldots + (\, z_1 + z_2 + z_3 + \ldots + z_{N-1} \,)$$

$$+ (z_1 + z_2 + z_3 + \ldots + z_{N-1} + z_N \,) - \frac{1}{2} \cdot (\, z_1 + z_2 + z_3 + \ldots + z_{N-1} + z_N)$$

$$= \sum_{k=1}^{N} (N - k + 1) \cdot z_k - \frac{1}{2} \sum_{k=1}^{N} z_k = (N - 1 - \frac{1}{2}) \sum_{k=1}^{N} z_k - \sum_{k=1}^{N} k \cdot z_k$$

---

[3] notice that in the equidistant case we hale $\mathrm{Gini}(z)$, and in arbitrarily distributed hj's we denote $\mathrm{Gini}(h, s)$, where s_k stands for an appropriate cumulative share determined/described by the data $(h, s) = ((h_j, z_j))_{j=1..N}$

$$= (N+\frac{1}{2})\cdot S - \sum_{k=1}^{N} k\cdot z_k \;=\; S\cdot\left\{N+\frac{1}{2}-\frac{1}{N\bar{z}}\sum_{k=1}^{N} k\cdot z_k\right\}$$

and, in consequence,

$$\text{Gini}(h,\, s) = 1-\frac{1}{N}\cdot\frac{2}{S}\cdot\left\{\frac{S}{2}\sum_{k=1}^{N}(s_{k-1}+s_k)\right\}$$

$$= 1-\frac{2}{NS}\cdot S\cdot\left\{N+\frac{1}{2}-\frac{1}{N\bar{z}}\sum_{k=1}^{N} k\cdot z_k\right\} = 1-\frac{2}{N}\cdot\left\{N+\frac{1}{2}-\frac{1}{N\bar{z}}\sum_{k=1}^{N} k\cdot z_k\right\}$$

$$= 1-2-\frac{1}{N}+\cdot\frac{2}{N\bar{z}}\sum_{k=1}^{N} k\cdot z_k \;=\; \frac{2}{N\bar{z}}\sum_{k=1}^{N} k\cdot z_k - \frac{N+1}{N}$$

Since $s_j = \dfrac{z_j}{S}$, the formula for Gini(h, s) can be transformed as follows

$$\text{Gini}(h,\, s) = 1-\frac{1}{N}\sum_{k=1}^{N}\left(\frac{\sum_{j=1}^{k-1} z_j}{S}+\frac{\sum_{j=1}^{k} z_j}{S}\right) = 1-\frac{1}{N\cdot S}\sum_{k=1}^{N}(2z_1+2z_2+\ldots+2z_{k-1}+z_k)$$

$$= 1-\frac{1}{N\cdot S}\sum_{k=1}^{N}(2z_1+2z_2+\ldots+2z_{k-1}+z_k)$$

http://www3.nccu.edu.tw/~jthuang/Gini.pdf:

Advantages as a measure of inequality

- The Gini coefficient's main advantage is that it is a measure of inequality by means of a ratio analysis, rather than a variable unrepresentative of most of the population, such as per capita income or gross domestic product.
- It can be used to compare income distributions across different population sectors as well as countries, for example the Gini coefficient for urban areas differs from that of rural areas in many countries (though the United States' urban and rural Gini coefficients are nearly identical).
- It is sufficiently simple that it can be compared across countries and be easily interpreted. GDP statistics are often criticised as they do not represent changes for the whole population; the Gini coefficient demonstrates how income has changed for poor and rich. If the Gini coefficient is rising as well as GDP, poverty may not be improving for the majority of the population.
- The Gini coefficient can be used to indicate how the distribution of income has changed within a country over a period of time, thus it is possible to see if inequality is increasing or decreasing.

The Gini coefficient satisfies four important principles:

- Anonymity: it does not matter who the high and low earners are.
- Scale independence: the Gini coefficient does not consider the size of the economy, the way it is measured, or whether it is a rich or poor country on average.
- Population independence: it does not matter how large the population of the country is.
- Transfer principle: if income (less than the difference), is transferred from a rich person to a poor person the resulting distribution is more equal.

Disadvantages as a measure of inequality

-

The Gini coefficient measured for a large economically diverse country will generally result in a much higher coefficient than each of its regions has individually. For this reason the scores calculated for individual countries within the EU are difficult to compare with the score of the entire US.

•

Comparing income distributions among countries may be difficult because benefits systems may differ. For example, some countries give benefits in the form of money while others give food stamps, which may not be counted as income in the Lorenz curve and therefore not taken into account in the Gini coefficient.

•

The measure will give different results when applied to individuals instead of households. When different populations are not measured with consistent definitions, comparison is not meaningful.

•

The Lorenz curve may understate the actual amount of inequality if richer households are able to use income more efficiently than lower income households. From another point of view, measured inequality may be the result of more or less efficient use of household incomes.

•

As for all statistics, there will be systematic and random errors in the data. The meaning of the Gini coefficient decreases as the data become less accurate. Also, countries may collect data differently, making it difficult to compare statistics between countries.

•

Economies with similar incomes and Gini coefficients can still have very different income distributions. This is because the Lorenz curves can have different shapes and yet still yield the same Gini coefficient. As an extreme example, an economy where half the households have no income, and the other half share income equally has a Gini coefficient of ½; but an conomy with complete income equality, except for one wealthy household that has half the total income, also has a Gini coefficient of ½.

•

Too often only the Gini coefficient is quoted without describing the proportions of the quantiles used for measurement. As with other inequality coefficients, the Gini coefficient is influenced by the granularity of the measurements. For example, five 20% quantiles (low granularity) will yield a lower Gini coefficient than twenty 5% quantiles (high granularity) taken from the same distribution.

As one result of this criticism, additionally to or in competition with the Gini coefficient entropy measures are frequently used (e.g. the Atkinson and Theil indices).
These measures attempt to compare the distribution of resources by intelligent players in the market with a maximum entropy random distribution, which would occur if these players acted like non-intelligent particles in a closed system following the laws of statistical physics.

h_j is the share of
With these sequence we can describe the distribution of the wealth (among as follows:
A pair (h_j, s_j) says that the poorest *hj* where

The distribution of the wealth can be described via the points $(h_j, s_j)_{j=1..N}$, a point (h_j, s_j) says that the poorest *hj* where
Let h_j be the j-th share

Let's interpret a point (h_j, s_j) as follows:

Their elements
Any element (h_j) is interpreted as the share
They determine points (h_j, s_j) referred to as Lorenz points, and the polygonal line
Given a distribution
$(h_k, s_k)_{k=1..N}$, where

Gini Coefficient

Authors: Dr. Brian Slack and Dr. Jean-Paul Rodrigue

1. Definition

2. Index of Dissimilarity (ID)

The **dissimilarity index** is the summation of vertical deviations between the Lorenz curve and the line of perfect equality, also known as the summation of Lorenz differences. The closer the ID is to 1 (or 100 if percentages are used instead of fractions), the more dissimilar the distribution is to the line of perfect equality.

$$ID = 0.5\sum_{i=1}^{N}|X_i - Y_i|$$

where X and Y are percentages (or fractions) of the total number of elements and their respective values (traffic being the most common). N is the number of elements (observations). For instance, the following considers the distribution of traffic among 5 terminals:

Terminal	Traffic	X	Y	|X-Y|
A	25,000	0.2	0.438	0.238
B	18,000	0.2	0.316	0.116
C	9,000	0.2	0.158	0.042
D	3,000	0.2	0.053	0.147
E	2,000	0.2	0.035	0.165
Total	57,000	1.0	1.0	0.708

Terminal B, with a traffic of 18,000 accounts for 0.2 (or 20%; X) of all terminals and 0.316 (or 31.6%; Y) of all traffic. The index of dissimilarity of this distribution is 0.354 (0.708 * 0.5), which indicates an average level of concentration. A more complex example is provided here.

3. Gini's Coefficient (G)

The Gini Coefficient represents the area of concentration between the Lorenz curve and the line of perfect equality as it expresses a proportion of the area enclosed by the triangle defined by the line of perfect equality and the line of perfect inequality. The closer the coefficient is to 1, the more unequal the distribution.

$$G = 1 - \sum_{i=0}^{N} \left(\sigma Y_{i-1} + \sigma Y_i\right)\left(\sigma X_{i-1} - \sigma X_i\right)$$

Where σX and σY are cumulative percentages of Xs and Ys (in fractions) and N is the number of elements (observations). Using the same example as above, the following table demonstrates the calculation of the Gini coefficient:

Terminal	Traffic	X	Y	σX	σY	$\sigma X_{i-1} - \sigma X_i$ (B)	$\sigma Y_{i-1} + \sigma Y_i$ (A)	A*B
A	25,000	0.2	0.438	0.2	0.438	0.2	0.438	0.088
B	18,000	0.2	0.316	0.4	0.754	0.2	1.192	0.238
C	9,000	0.2	0.158	0.6	0.912	0.2	1.666	0.333
D	3,000	0.2	0.053	0.8	0.965	0.2	1.877	0.375
E	2,000	0.2	0.035	1.0	1.000	0.2	1.965	0.393
Total	57,000	1.0	1.000					1.427

The Gini coefficient for this distribution is 0.427 (|1-1.427|).