Modified Lorenz Curve and Its Computation

Subian Saidi, Ulfah Muharrahah, La Zakaria*, Yomi Mariska, Tiryono Ruby

Universitas Lampung, Indonesia

ABSTRACT

The Lorenz curve is generally used to find out the inequality of income distribution. Mathematically a standard form of the Lorenz curve can be modified with the aim of simplicity of its symmetric analysis and calculation of the Gini coefficient that usually accompanies it. One way to modify the shape of the Lorenz curve without losing its characteristics but is simple in the analysis of geometric shapes is through a transformation (rotation). To be efficient and effective in computing and analyzing a Lorenz curve it is necessary to consider using computer software. In this article, in addition to describing the development of the concept of using transformations (rotations) of the standard Lorenz curve in an easy-to-do form, the symmetric analysis is also described by computational techniques using Mathematica® software. From the results of the application of the development of the concept of the Lorenz curve which is carried out on a data gives a simpler picture of the computational process with relatively similar computational results.

INTRODUCTION

The Lorenz curve is a graphical representation of the proportionality of the distribution (cumulative percentage of values) (Arcagni & Porro, 2014; Todaro, 1989; Zheng et al., 2008). The application of the Lorenz curve and the Gini coefficient is common in the analysis of income distribution inequality (Gastwirth, 1971, 1972). The other side of the Lorenz curve which is relatively rarely discussed but important is its symmetry (Fajar, 2018; Sugiyarto et al., 2015). The asymmetry of the Lorenz curve is an interpretation of inequality of distribution proportionality. Several scientists have discussed the symmetrical analysis of the Lorenz curve. For example exploration of the distribution of institutional publications that use publications as a proxy (Zheng et al., 2008). In addition, the measurement of income distribution through the asymmetric analysis of the Lorenz curve based on the
Zanardi index approach has been discussed in (Tarsitano, 1988). In (Arcagni & Porro, 2014) specifically discussed the results of a review and comparison of well-known inequality curves, Lorenz curves, Bonferroni curves, and Zenga curves. For the record, the Bonferroni curve and the Zenga curve are two curves that rely on the Lorenz curve for a special discussion of cases of income distribution inequality. From the results of previous studies have not presented an asymmetrical analysis of the Lorenz curve for the large amounts of data that are commonly found in census activities. Also, curve fitting for empirical data is relatively rare to discuss. This research article discusses the asymmetry of the Lorenz curve by seeking anticipation for census empirical data, for example by creating computer programs that can read data files. Besides, a review of the asymmetry of the Lorenz curve in this study is the non-polymerization of the Lorenz curve which transformation angle of 45° in a counterclockwise direction. To read large data files, curve fitting, and the computation process used Mathematica® software from Wolfram Research Company.

METHOD

See Figure 1 along the horizontal (x-axis) line is associated with the cumulative percentage of income earners while along the vertical line (y-axis) is associated with the cumulative percentage of total income. The curved line (curve) connecting coordinates (0,0) with (1,1) is a curve known as the Lorenz curve. The curve is plotted based on the cumulative percentage of income earners (ranking from poor to rich, for example), to the cumulative percentage of total income. Along the straight line (diagonal) connecting the coordinates (0,0) with (1,1), is the income equalization line. This straight-line means that the coordinates along this line if the poorest 10% of the population will receive 10% of the total income, if there are 30% of the poorest people will receive 30% of the total income and so on. A line with an angle of 45° (diagonal line) shows a perfect (equal) income distribution while the curve shows the actual distribution of income.

![Figure 1. An illustration of the Standard Form of the Lorenz Curve](image-url)
The functional relationship on the Lorenz curve is expressed as a relationship between two variables whose values lie on the curve. In other words, the Lorenz curve represents an indication of an income distribution imbalance. The Lorenz curve in Figure 1, for example, for a point on the coordinates (62%, 20%) on the Lorenz curve (dotted line), these coordinates mean that there are 62% income earners who earn 20% of total income. As for the coordinates (86%, 60%), there are 86% (high income earners / rich people) who get 60% of total income.

In Figure 1 the ratio between area of $A$ and area of $(A + B)$ bounded by the Lorenz curve is known as the Gini coefficient/index ($G$), which is (1)

$$G = \frac{A}{A + B} = 2A = 1 - 2B. \quad (1)$$

For certain conditions, the Gini index can be estimated accurately without first fitting the data curve (Gastwirth, 1972). With respect to the Gini index, if the income distribution is the same, then $G = 0$ or the Lorenz curve is on the diagonal line. Conversely, the income distribution becomes more equal as $G$ increases and gets closer to 1. In other words, the narrower the area, the smaller the Gini ratio, the more evenly distributed the income distribution, and vice versa. Although not discussed in detail in this article, in principle the Lorenz curve and the Gini index have a functional relationship as stated in Table 1.

### Table 1. Relationship of the Lorenz Curve and the Gini Index (Arcagni & Porro, 2014).

| Model       | Lorenz Curve $L(p)$                                                                 | Gini Index $G$ |
|-------------|------------------------------------------------------------------------------------|---------------|
| Uniform     | $p(1-\theta + \theta p)$                                                          | $\frac{\theta}{3}$ |
| Exponential | $p + (1-p) \ln (1-p)$                                                              | 0.5           |
| Pareto      | $1-(1-p)^{(\alpha-1)/\alpha}$                                                     | $\frac{1}{(2\theta - 1)}$ |
| Log-Normal  | $\Phi(\Phi^{-1}(p) - \delta)$                                                     | $2\Phi\left(\frac{\delta}{\sqrt{2}}\right) - 1$ |
| Dagum       | $B\left(p^{\frac{1}{\beta}}, \frac{1}{\beta}; 1-\frac{1}{\beta}\right)$         | $\frac{\Gamma(\beta) \Gamma(2\beta + \theta + 1)}{\Gamma(2\beta) \Gamma(\beta + \theta + 1)}$ |

In Table 1 the ratio of incomplete Beta functions is defined by:

$$B(x; a; b) = \frac{\int_{a}^{x} t^{a-1} (1-t)^{b-1} dt}{\int_{0}^{1} t^{a-1} (1-t)^{b-1} dt},$$

where $x \in [0,1], a > 0, b > 0$ and Gamma function is $\Gamma(x) = \int_{0}^{\infty} t^{x-1} e^{-t} dt.$

### Symmetry the Lorenz curve

See Figure 1, if a line connecting points in coordinates (0,1) and (1,0) is given then the Lorenz curve is now separated into two regions ($A$ and $A'$) by that line (see a dot-dashed line in Figure 2). As a result, two other regions were obtained, $B$ and $B'.$
The asymmetrical ratio of the Lorenz curve can be formulated by:

\[
l_k = \frac{\text{Area of } A}{\text{Area of } A'} \tag{4}\]

Where \( l_k \) is the asymmetrical ratio of the Lorenz curve.

The interpretation of \( l_k \) is as follows:

a. If \( l_k = 1 \), the Lorenz curve is symmetrical. This condition means that inequality does not occur. In other words, the magnitude of inequality between high-income groups and low-income groups is the same.

b. If \( l_k < 1 \), then the Lorenz curve is not symmetrical (convex top) which means the amount of inequality in high-income groups is greater than the amount of inequality in low-income groups.

c. If \( l_k > 1 \), then the Lorenz curve is not symmetrical (bottom convex) which means the amount of inequality in high-income groups is lower than the amount of inequality in low-income groups.

The numerator and denominator on the symmetry ratio of the Lorenz curve (4) calculate directly, namely:

\[
\text{Area of } A' = \int_{0}^{0.5} p - L(p) \, dp + \int_{p_1}^{p_2} ((1-p) - L(p)) \, dp \tag{5}
\]

Because \( p_1 = 0.5 \) then the equation becomes

\[
\text{Area of } A' = \int_{0}^{0.5} p - L(p) \, dp + \int_{0.5}^{p_1} ((1-p) - L(p)) \, dp
\]

or

\[
\text{Area of } A' = -0.25 + p_2 - 0.5(p_2)^2 - \int_{0.5}^{p_1} L(p) \, dp \tag{6}
\]

then,

\[
\text{Area of } A = \int_{p_1}^{p_2} p - L(p) \, dp - \int_{p_1}^{p_2} ((1-p) - L(p)) \, dp
\]

or

\[
\text{Area of } A = 0.75 - p_2 + 0.5(p_2)^2 + \int_{p_1}^{p_2} L(p) \, dp \tag{7}
\]
Gini Ratio (Gini Coefficient/Index)

The Gini Ratio (RG) or commonly referred to as the Gini coefficient/index whose formula is stated in equation (1) is a measurement or medium that is relatively easy to use to measure the degree of inequality in income distribution. RG is obtained by calculating the ratio that lies between the diagonal lines of the Lorenz Curve divided by the area of the half rectangle where the Lorenz curve is located. RG which is stated in equation (1), can be formulated using the concept of integration, i.e

\[ G = 1 - 2 \int_0^1 L(p) dp \]  

(8)

where \( L(p) \) is the Lorenz curve equation (Arcagni & Porro, 2014).

The Gini (G) ratio for n cumulative distribution data can also be known through the following equation (Zheng et al, 2008):

\[ G = 1 - \frac{1}{2} \sum_{i=0}^{n-1} (x_{i+1} - x_i)(y_i + y_{i+1}) \]

or

\[ G = \sum_{i=1}^{n-1} (x_i y_{i+1} - x_{i+1} y_i) \]

(9)

The income inequality criteria based on the Gini ratio/coefficient are:

If the value of \( G = 0 \), it means that the income distribution is perfectly even.

If the value of \( 0 < G < 0.4 \) then it means that the level of income distribution inequality is low.

If the value is \( 0.4 < G < 1 \), this means that the level of income distribution is moderate.

If the value of \( G = 1 \) then it means that the income distribution is perfectly uneven.

RESULTS AND DISCUSSION

Modification of the Lorenz Curve

Look at Figure 1. Assume the Lorenz curve is a curve produced by a degree 3 polynomial function that is defined by

\[ y(x) = ax - bx^2 + cx^3 \]  

(10)

with \( a, b, c \in \text{real number} \) and \( x, y \in [0,1] \).

Coefficient \( a, b, \) and \( c \) in equation (10) is a real constant that corresponds to the income distribution data. In addition to the Lorenz curve, in Figure 2 there are also the following equations

\[ g_1 : y(x) = x \]
\[ g_2 : y(x) = 1 - x \]

(11)

with \( x, y \in [0,1] \).

Then, assume a polynomial function in equation (10) and equation (11) is the parameter equation notated in the following form.

\[
\begin{align*}
  \begin{pmatrix}
    x(t) \\
    y(t)
  \end{pmatrix} &= \begin{pmatrix}
    t \\
    at^3 - bt^2 + ct
  \end{pmatrix}, \\
  \begin{pmatrix}
    x(t) \\
    y(t)
  \end{pmatrix} &= \begin{pmatrix}
    t \\
    1 - t
  \end{pmatrix};
\end{align*}
\]

(12)

with \( t \in [0,1] \); \( a, b, c \in \text{real number} \).

Look at the parameter equations in (12). By doing a transformation (rotation) with a fixed point of rotation at \( \alpha \) and \( \theta \) as the angle of rotation we will get new parameter equations namely

\[
\begin{align*}
  x'(t) &= (\alpha - \sin(\theta)(-\alpha - \alpha t + \alpha t - \alpha)) + \alpha + \cos(\theta)(t - \alpha) \\
  y'(t) &= (\alpha - \sin(\theta)(\alpha - \alpha t + \cos(\theta)(t - \alpha)) + \alpha + \sin(\theta)(t - \alpha)) \cos(\theta)(\alpha - \alpha t + \alpha t - \alpha) \\
  x'(t) &= (\alpha - \sin(\theta)(\alpha - \alpha t + \alpha t - \alpha)) + \alpha + \cos(\theta)(t - \alpha) \\
  y'(t) &= (\alpha - \sin(\theta)(\alpha - \alpha t + \alpha t - \alpha)) + \alpha + \sin(\theta)(t - \alpha) \cos(\theta)(\alpha - \alpha t + \alpha t - \alpha)
\end{align*}
\]

(13)

(14)

(15)

with a "prime" sign is a new function of the results of rotation.

An Illustration:

Suppose \( a = 1.7, b = 0.9, \) and \( c = 0.2. \) if the center of rotation is at \( (0,0) \) with a rotation angle of \( -\pi/4 \), then the parameter equation (12) becomes
Desimal, 3 (2), 2020 - 102
Subian Saidi, Ulfah Muharramah, La Zakaria, Yomi Mariska, Tiryono Ruby

If the parameter equation is plotted (16) - (18) for the value $t \in [0,1]$, then the curves are obtained as shown in Figure 3.

$$\begin{align*}
\left( x'(t) \right) &= \begin{pmatrix} 1.7t^3 - 0.9t^2 + 0.2t \over \sqrt{2} + t \over \sqrt{2} \\ 1.7t^3 - 0.9t^2 + 0.2t \over \sqrt{2} - t \over \sqrt{2} \end{pmatrix} \\
\left( y'(t) \right) &= \begin{pmatrix} t \sqrt{2} \\ 0 \end{pmatrix}
\end{align*}$$

If the parameter equation is plotted (16) - (18) for the value $t \in [0,1]$, then the curves are obtained as shown in Figure 3.

$$\begin{align*}
\left( x'(t) \right) &= \begin{pmatrix} 1-t \over \sqrt{2} + t \over \sqrt{2} \\ 1-t \over \sqrt{2} - t \over \sqrt{2} \end{pmatrix} \\
\left( y'(t) \right) &= \begin{pmatrix} a \sqrt{2} - bt^2 + ct \\ b \sqrt{2} - bt^2 + ct \end{pmatrix}
\end{align*}$$

Figure 3. Lorenz curve before transformation (rotation) obtained from the parameter equation $x(t) = t$ and $y(t) = at^3 - bt^2 + ct$ with $t \in [0,1]$, $a = 1.7$, $b = 0.9$, and $c = 0.2$.

Meanwhile from the parameter equation (16) - (18) the curve is obtained as shown in Figure 4.
Determination of the area formed as a result of the existence of asymmetrical lines on the Lorenz curve will be simpler to do if in the geometric form as in Figure 4 compared to the geometry as in Figure 3.

**Case illustration: Computation of a modified Lorenz curve**

This case illustration is based on the example provided by Bellù and Liberati (Bellu & Liberati, 2005). Look at the results of a survey on population income in one year of evaluation. For example, a population consisting of 100 individuals obtained data as given in Table 2.

| Individu | Income (US $/Year) |
|----------|---------------------|
| 1        | 2,4170              |
| 2        | 7,8000              |
| 3        | 8,4890              |
| 4        | 10,072              |
| 5        | 12,957              |

Table 2 means that each person has (1/100) of total income, the graph of the distribution of income that underlies the equalization line. Suppose income distribution is made from poor to rich, this means that poor individuals have less income than richer individuals. Therefore, from Table 2 obtained information that individual 1 has the US $ 2,417 / year (the poorest), while individual 5 has the US $ 12,957 / year (the richest).

For efficiency and effectiveness in handling geometric shapes, computational symmetry of the Lorenz curve, and determination of the Gini index the Mathematica® program will be used with the algorithm description as follows:

- **Algorithm (Description)**
  a. Read data \((X_i, Y_i)\).
  b. Compute the cumulative frequency of the data \(X_{(Cum,i)}, Y_{(Cum,i)}\).
c. Determine the best Nonlinear Model for data \( (X_{\text{kum},i}, Y_{\text{kum},i}) \).
d. Determine the geometric shape of the Lorenz curve based on data \( (X_{\text{kum},i}, Y_{\text{kum},i}) \).
e. Compute transformation (rotation) at the origin with an angle of \(-45^\circ\) for data \( (X_{\text{kum},i}, Y_{\text{kum},i}) \) and the best Nonlinear Model obtained from c.
f. Computing Symmetry Lorenz curves based on output data from e.

- **Algorithm Implementation**

  For the implementation of the algorithm, the Mathematica® software is used. The intended implementation is easy to do (Wellin, P.R., Richard J. Gaylord, R.J. & Samuel N. Kamin, 2005) for programming techniques. The output of the Mathematica® program listing created based on the algorithm given is the function graphs as shown in Figure 5.

![Figure 5. Lorenz curves sourced from the data in Table 3. Before rotating (left) and after rotating towards the origin (0.0) and rotating angle of \(-\pi/4\) (right).](image)

**Symmetrical computation of modified Lorenz and Lorenz curves.**

This section shows the modified computing results of the Lorenz curve. For efficient and effective computing, a list of programs in the Mathematica® format created based on previous algorithms can be requested from the author via e-mail. The formula that is involved in the computation carried out is the formula given in equation (4), equation (8), and equation (9) for standard Lorenz curves and modified Lorenz curves (equation (16), equation (17), and equation (18).
Table 3. The proportion of income and individual as well as cumulative each proportion.

| Income Distribution | The proportion of income owned by each individual and the proportion of the population that corresponds with it | The cumulative proportion of income owned by each individual and the cumulative proportion of the population that corresponds with it |
|---------------------|-------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| Individual          | Income                                                                                            |                                                                                                 |
| 1                   | 2,417                                                                                           | 0.0579                                                                                           |
| 2                   | 7,800                                                                                           | 0.1869                                                                                           |
| 3                   | 8,489                                                                                           | 0.2034                                                                                           |
| 4                   | 10,072                                                                                          | 0.2413                                                                                           |
| 5                   | 12,957                                                                                          | 0.3105                                                                                           |
| Total               | 41,735                                                                                          | 1.0000                                                                                           |

Using the data in column 6 as the value for the independent variable x and the data in column 5 as the dependent variable y can be determined according to the regression form, namely:

\[ f_{\text{regression}} = 0.16823062x + 1.13283177x^2 - 0.30588678x^3 \]  \hfill (19)

To get the coordinates of the symmetry line, complete the following equation:

\[ f_{\text{regression}} + (x - 1) = 0 \]

that is

\[ (x^*, y^*) = (0.58047096, 0.41952904). \]  \hfill (20)

Therefore it can be determined:

\[ \text{Area of } A = \int_0^{0.5} ((1 - x) - f_{\text{regression}}) \, dx + \int_0^{0.5} (x - f_{\text{regression}}) \, dx = 0.06848141 \]

and

\[ \text{Area of } A' = \int_{0.5}^{1} (x - f_{\text{regression}}) \, dx + \int_{0.5}^{1} (x - (1 - x)) \, dx = 0.04626439 \]

as a result

\[ \frac{l_k}{\text{Area of } A} = \frac{0.06848141}{0.04626439} = 1.48021872 \]  \hfill (21)

\[ \frac{l_k}{\text{Area of } A'} = \frac{0.06752027}{0.04501061} = 1.50009661 \]  \hfill (23)

Because \( l_k = 1.48021872 \) and \( l_k = 1.50009661 \) more than 1, then the Lorenz curve is claimed to be asymmetrical (bottom convex) which means the amount
of inequality in high-income groups is lower than the amount of inequality in low-income groups.

Gini Ratio Computation
To get the Gini ratio value from the case discussed in the previous section, you can use the formula in equation (8) or equation (9). In this section, we will review both the Lorenz curve before and after the transformation.

Look at the data in column 5 and column 6 in Table 3. For the transformation data shown in Table 4, from the data in Table 4, we can determine the value of the Gini ratio before rotation. If calculated using the formula in equation (9) or using the function in equation (19) for the Gini ratio formula based on equation (8), we get:

\[
G = 1 - 2 \int_0^1 f_{\text{regression}} \, dx = 0.22949. \tag{24}
\]

or

\[
G = \sum_{i=1}^4 (x_i y_{i+1} - x_{i+1} y_i) = 0.22384. \tag{25}
\]

Meanwhile for the data in Table 4 column-5 and column-6 can be determined the value of the Gini coefficient after rotation is calculated based on equation (8), namely:

\[
G = \frac{- \int f_{\text{rotasi}} \, dx}{0.25 \left( \sqrt{2} \right)^2} = 0.22506. \tag{26}
\]

or

\[
G = \sum_{i=1}^4 (x_i' y_{i+1}' - x_{i+1}' y_i') = 0.22384. \tag{27}
\]

Table 4. Data development from Table 3 after being transformed with regard to the proportion of income and individual as well as the cumulative of each proportion.

| Income Distribution | The cumulative proportion of income owned by each individual and the cumulative proportion of the population that corresponds with it. (Before transformation) | The cumulative proportion of income owned by each individual and the cumulative proportion of the population that corresponds with it (After transformation) |
|---------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| Individual         | Cumulative Proportion of income owned by each individual | Cumulative Proportion of income owned by each individual in the Total Population | Cumulative Proportion of income owned by each individual | Cumulative Proportion of each Individual in the Total Population |
| 1                   | 2,417 | 0.0579 | 0.2000 | -0.1005 | 0.1824 |
| 2                   | 7,800 | 0.2448 | 0.4000 | -0.1097 | 0.4560 |
| 3                   | 8,489 | 0.4482 | 0.6000 | -0.1073 | 0.7412 |
| 4                   | 10,072 | 0.6895 | 0.8000 | -0.0781 | 1.0533 |
| 5                   | 12,957 | 1.0000 | 1.0000 | 0 | 1.4142 |
| Total               | 41,735 | 2.4405 | 3.0000 | -0.3956 | 3.8471 |

From the computational results carried out which are shown in equation (24-27) and based on the income inequality criteria measured using the Gini ratio/coefficient for the value \( 0 < G \approx 0.22 < 0.4 \) is interpreted as a low level of inequality. In other words, the level of income distribution in the community measured is relatively evenly distributed.

CONCLUSIONS AND SUGGESTIONS
This conceptual research has provided an alternative in dealing with the Lorenz curve concept. Modifications made through transformation have been able to reduce the process of determining coordinates as the basis for computing the Lorenz curve.
This article has not yet discussed all the forms of the Lorenz curve and the corresponding Gini index as given in Table 1. Therefore research in this field can be followed up to find out the formula of the Lorenz curve in other forms and the corresponding Gini index after transforming (rotation).

ACKNOWLEDGMENT

The author would like to thank the reviewer and University of Lampung for facilitating the research on which this article was based.

REFERENCES

Arcagni, A., & Porro, F. (2014). The Graphical Representation of Inequality. Revista Colombiana de Estadística, 37(2Spe), 419. https://doi.org/10.15446/rce.v37n2spe47947

Bellu, L. G., & Liberati, P. (2005). Charting Income Inequality. The Lorenz Curve. EASYPol Module 000, 17.

Fajar, M. (2018). Sebuah Ukuran Baru Ketidaksimetrisan Kurva Lorenz (A New Measure of Asymmetry of Lorenz Curve ). August. https://doi.org/10.13140/RG.2.2.28204.16003

Gastwirth, J. L. (1971). A General Definition of the Lorenz Curve. Econometrica, 39(6), 1037. https://doi.org/10.2307/1909675

Gastwirth, J. L. (1972). The Estimation of the Lorenz Curve and Gini Index. The Review of Economics and Statistics, 54(3), 306. https://doi.org/10.2307/1937992

Sugiyarto, Mulyo, J. H., & Seleky, R. N. (2015). Kemiskinan and Ketimpangan Pendapatan Rumah Tangga di Kabupaten Bojonegoro. Jurnal Agro Ekonomi, 26(2), 115–120.

Tarsitano, A. (1988). Measuring the asymmetry of the Lorenz curve. Ricerche Economiche, 42(3), 507–519.

Todaro, M. P. (1989). Economic Development in the Third World.

Wellin, P.R., Richard J. Gaylord, R.J. & Samuel N. Kamin, S. N. (2005). An introduction to programming with Mathematica®.

Zheng, M., Junpeng, Y., Cheng, S., Zhiyu, H., Zhenglu, Y., Yuntao, P., & Yishan, W. (2008). Using Lorenz Curve and Gini Coefficient to Reflect the Inequality Degree of S&T Publications: An Examination of the Institutional Distribution of Publications in China and other Countries. 1–9. http://www.collnet.de/Berlin-2008/MaZhengWIS2008ulc.pdf
