Maximum Second Order Entropy Lorenz Curve

M. Yaghoobi Avval Riabi*  G. R. Mohtashami Borzadaran †
M. Nezhadabdolah †

* Islamic Azad University, Gonabad Branch, Gonabad, Iran
E-mail: Corresponding author: m.yaghoobiavl@yahoo.com.

† Ferdowsi University of Mashhad, Mashhad, Iran, Islamic Azad University, Mashhad Branch, Mashhad, Iran
E-mails: grmohtashami@um.ac.ir, m.nezhadabdolah@yahoo.com

Abstract
In this work, we derive a family of maximum second order entropy distributions provided that the mean and the Gini index are known as the optional side conditions. Then via these distributions, families of Lorenz curves are obtained which are compatible with the optional side conditions. Also some results are derived via maximization of the Second order entropy in view of income distributions.

Keywords and phrases: Gini Index, Lorenz Curve, Maximum Entropy, Income Inequality

1. Introduction

Suppose a probability density function of a random variable is at hand, thus the distribution is completely known. But in many cases, the explicit density is hidden and must be estimated. The classical procedure consists of fitting an analytical function on observations. Another more logical approach is to apply the maximum entropy that was proposed by Jaynes (1957). This technique allows to choose among all the possible probability distributions the most suitable one with respect to the available knowledge. The maximum entropy principle state that, when given some information about a random variable, the least biased probability distribution is obtained by maximizing entropy subject to the given constraints. Lorenz curve is one of the most important tools for analyzing the income distributions, proposed by Max Lorenz (1905). The Lorenz curve relates the cumulative proportion of income to the cumulative proportion of population, when population is arranged according to increasing level of income. The Gini index, that is widely used in the study of the inequality of income distributions, was proposed by Corrado Gini (1912). The Gini index measures the ratio of the area, between the Lorenz curve and the equality line.

Holm (1993), derived a family of maximum Shannon entropy density functions under side conditions on the mean and Gini index. Ryu (2008) determined the functional form of the share function (as a density function) via maximum Shannon entropy method under side conditions on the Bonferroni index. Over the past sixty years, after Shannon (1948) introduced his measure of entropy, various forms of the entropy suggested. One of them is
the Second order entropy that was introduced by Kemp (1975).


There are two main objective of this paper. Firstly, we derive a family of maximum second order entropy distributions under the conditions on the mean and the Gini index. Secondly, as a consequence of the first, some family of Lorenz curves are found.

2. Preliminaries

In this section, some basic notions such as, Lorenz curve, Gini index that are necessary in the main results are reviewed.

2.1. Second Order Entropy. The second order entropy of a continuous random variable $X$, where taking its values in $R$, with probability density function $f_x(x)$, is defined by

$$H_s(f) = \left[1 - \int_R f^2(x)dx \right].$$

provided that, the integral exists.

We can have a similar version for discrete cases.

2.2. Lorenz Curve. Lorenz curve as an important tools for analyzing the income distributions introduced by Max Lorenz (1905). He proposed a simple graphical, means to summarize the inequality of wealth. Graphically, the Lorenz curve gives the proportional of the total societal income accruing to the lowest earning proportion of income earners.

Definition 1. Let $X$ be a non-negative random variable with probability density function $f_X(x)$ and distribution function $F_X(x)$. The Lorenz curve of $X$ denoted by $L_X(p)$ and defined as:

$$L_X(p) = \frac{1}{\mu} \int_0^p F_X^{-1}(y)dy, \quad 0 \leq p \leq 1,$$

where, $F_X^{-1}(y) = \inf\{x : F_X(x) \geq y\}$ is the inverse function of $F_X(x)$ and $\mu$ is the mean of $X$. From Definition 1, we can show that, $L$ is a continuous function on $[0, 1]$, with $L(0) = 0$ and $L(1) = 1$. Also $L$ is an increasing and convex function of $p$.

There are various ways for the construction of the parametric families of Lorenz curve. The first way consists of using from parametric family distribution and substituted in (2) (see, for example Gastwirth (1971)). Note that the explicit distribution in many cases hidden and must be estimated (by classical method or maximum entropy or other method).

Another way consists of selecting parametric families of curves that, satisfying in the Lorenz curves’s properties. Some of the most common parametric Lorenz curves are as follows:

- The Kakwani and Podder specification
The pioneer model was proposed by Kakwani and Podder (1973). The Kakwani - Podder Lorenz curve is specified as:

\[ L(p) = p^\alpha e^{-\beta(1-p)}, \quad 0 \leq p \leq 1, \quad 1 < \alpha < 2, \quad \beta > 0. \]

The Kakwani and Podder (1976) also proposed a new parametric model based on a geometric motivation as:

\[ L(p) = \gamma p^\alpha (\sqrt{2} - p)^\beta, \quad 0 \leq \alpha \leq 1, \quad 0 < \beta \leq 1, \quad \gamma \geq 0. \]

**• The Gupta specification**

Gupta (1984) provides another parametric Lorenz curve as following form:

\[ L(p) = p^{\alpha} p^{1-\alpha}, \quad 0 \leq p \leq 1, \quad \alpha > 0. \]

**• The Rasche specification**

Rasche et al. (1980) suggested the following form for Lorenz curve,

\[ L(p) = [1 - (1 - p)^\alpha]^{\frac{1}{\beta}}, \quad 0 \leq p \leq 1, \quad 0 < \beta, \quad 0 < \alpha \leq 1. \]

If \( \beta = 1 \) and \( \alpha < 1 \), we obtain the Lorenz curve corresponding to the classical Pareto distribution with the form:

\[ L(p) = 1 - (1 - p)^\alpha, \quad 0 \leq p \leq 1, \quad 0 < \alpha \leq 1. \]

**• The Chotikopanich specification**

The Chotikopanich Lorenz curve is a functional form that proposed by Chotikopanich (1993). This model is as:

\[ L(p) = \frac{e^{kp} - 1}{e^k - 1}, \quad 0 \leq p \leq 1, \quad k > 0. \]

**• The Pake specification**

The family proposed by Pakes (1981) has a Lorenz curve of the form:

\[ L(p) = \frac{\int_0^p x^{\alpha-1}(1-x)^{\beta-1}dx}{B(\alpha, \beta)}, \quad 0 \leq p \leq 1, \]

where \( \alpha \geq 1, \quad 0 < \beta \leq 1 \) and \( B \) is the beta function.


2.3. **Gini Index.** The Gini index was developed by Corrado Gini (1912), it is strictly connected with representation of income inequality due to Lorenz curve. The Gini index is given as twice the area between the Lorenz curve and the equality line.

**Definition 2.** Let \( X \) be non-negative random variable with Lorenz curve \( L(p) \). The Gini index denoted by \( G \) and defined as:

\[ G = 2 \int_0^1 (p - L(p))dp = 1 - 2 \int_0^1 L(p)dp, \quad (2.3) \]
on noting that, $0 \leq G \leq 1$.

By setting $L(p)$ from (2) in (3), after some calculation the following useful form is obtained:

$$G = \frac{1}{\mu} \int_0^1 (2p - 1)F^{-1}(p)dp.$$  \hspace{1cm} (2.4)

Note that, the mean income can be written as:

$$\mu = \int_0^1 F^{-1}(p)dp.$$  \hspace{1cm} (2.5)

For more details see Gastwirth (1972) and Nembua (2006).

Usually the Gini index is calculated from the Lorenz curve, but we shall do the reverse.

3. Main Result

Let $X$ be a random variable indicating the level of income, where taking its values in $[x_0, x_1]$, also, suppose $f_X(x)$ and $F_X(x)$ be the probability density function and the distribution function of $X$ respectively. Second order entropy of distribution is defined as:

$$H_s(f) = \left[ 1 - \int_{x_0}^{x_1} f^2(x)dx \right],$$  \hspace{1cm} (3.1)

where on choosing $F(x) = p$, we obtain,

$$H_s(f) = \left[ 1 - \int_0^1 \left( \frac{dF^{-1}}{dp} \right)^{-1}dp \right],$$  \hspace{1cm} (3.2)

such that, $F^{-1}$ is the inverse function of $F$.

We want to find the maximum of function described in (13) subject to known the mean and the Gini index.

On the other hand, via integration by part it can easy to show that, the relations (4) and (3) are equal to,

$$\int_0^1 (1 - p) \frac{dF^{-1}}{dp}dp = \mu - x_0,$$  \hspace{1cm} (3.3)

$$\int_0^1 (1 - p)p \frac{dF^{-1}}{dp}dp = G\mu,$$  \hspace{1cm} (3.4)

respectively.

Now, we consider the following optimization problem,

$$\text{Max } H_s(f) = \left[ 1 - \int_0^1 \left( \frac{dF^{-1}}{dp} \right)^{-1}dp \right],$$

such that,

$$\int_0^1 (1 - p) \frac{dF^{-1}}{dp}dp = \mu - x_0,$$

$$\int_0^1 (1 - p)p \frac{dF^{-1}}{dp}dp = G\mu.$$  

By using the lagrangian method we obtain:

$$\frac{dF^{-1}}{dp} = \frac{1}{[\lambda_1(1 - p) + \lambda_2p(1 - p)]^{1/2}},$$  \hspace{1cm} (3.5)
in which, \( \lambda_1 \) and \( \lambda_2 \) are the undetermined Lagrange multipliers arising from the condition (15) and (16), respectively. Various values of \( \lambda_1 \) and \( \lambda_2 \) leads to familiar distributions, where, some of them discuss in next subsections.

3.1. **Maximum Second Order Entropy Lorenz Curve.**

A family of Lorenz curves that agree with the conditions will be generated from (11). For all members of this family \( L(0) = 0, L(1) = 1. \) the different cases are classified below.

3.1.1. \( \lambda_1 \neq 0 \) and \( \lambda_2 = 0. \)

In this case, equation (17) reduces to

\[
\frac{dF^{-1}}{dp} = \frac{(1-p)^{1/2}}{\sqrt{\lambda_1}}.
\]

An integration of \( \frac{dF^{-1}}{dp} \) over \([0, p]\), we obtain

\[
F^{-1}(p) = \frac{2}{\sqrt{\lambda_1}}\left\{1 - (1-p)^{1/2}\right\} + x_0. \quad (3.6)
\]

Now by using of (3) and (4) we derive:

\[
x_0 = \left\{1 - \left(\frac{5}{2}\right)G\right\} \mu,
\]

and so,

\[
F^{-1}(p) = \mu\left[1 + 5G\left[1 - \left(\frac{15}{2}\right)(1-p)^{1/2}\right]\right]. \quad (3.7)
\]

By some mathematical calculations we derive:

\[
F(x) = 1 - \left\{\frac{2}{(15 + \frac{x - \mu}{G\mu\frac{15}{2}})^2}\right\} \quad (3.8)
\]

Finally substituting \( F^{-1} \) into (3), gives the one parametric model:

\[
L^*_s(p) = p + 5G\left\{p - 1 + (1-p)^{1/2}\right\}, \quad (3.9)
\]

- A predicted minimum income \( x_0 = 0, \) the standard low bound of many income distributions is obtained for \( G = \frac{5}{2}. \)

3.1.2. \( \lambda_1 = 0 \) and \( \lambda_2 \neq 0. \)

Equation (17) with \( \lambda_1 = 0, \lambda_2 \neq 0 \) implies a density quantile function symmetric with respect to \( p = 0.5. \) By integration of \( \frac{dF^{-1}}{dp} \) over \([0, p]\), we obtain:

\[
F^{-1}(p) = \frac{1}{\sqrt{\lambda_2}} \int_0^p t^{-\frac{1}{2}}(1-t)^{-\frac{1}{2}} dt + x_0, \quad (3.10)
\]

considering expansion of \( (1-t)^{-\frac{1}{2}} = \sum_{j=0}^{\infty} \left(-\frac{1}{2}\right)^j (-1)^j t^j, \) lead us to:

\[
F^{-1}(p) = \frac{1}{\sqrt{\lambda_2}} \sum_{j=0}^{\infty} \left(-\frac{1}{2}\right)^j (-1)^j \frac{p^{j+\frac{1}{2}}}{j + \frac{1}{2}} + x_0. \quad (3.11)
\]
Now by using of (15) and (16) we derive:
\[ x_0 = \mu \{ 1 - G \frac{B(\frac{1}{2} + \frac{1}{2})}{B(\frac{1}{2}, \frac{1}{2})} \}. \]

Finally, by substitution of \( F^{-1} \) into (2) we obtain:
\[ L^*_s(p) = p + G \{ -4p + \sum_{j=0}^{\infty} \frac{(-1)^j (1 - \frac{1}{2})^j}{B(\frac{3}{2}, \frac{3}{2})} \}, \]
(3.12)
in which, \( B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt \)

3.1.3. \( \lambda_1 \neq 0 \) and \( \lambda_2 \neq 0 \).

Equation (17) with \( \lambda_1 \neq 0 \), \( \lambda_2 \neq 0 \) represents the following complex model,
\[ \frac{dF^{-1}}{dp} = \frac{1}{|\lambda_1 (1-p) + \lambda_2 p(1-p)]^{\frac{1}{2}}} = \frac{1}{\sqrt{\lambda_1}} (1-p)^{-\frac{1}{2}} (1+ap)^{-\frac{1}{2}}, \]
(3.13)
where \( a = \frac{4}{\lambda_1 + \lambda_2} \).

With similar process, integration over \([0 \ p] \) gives
\[ F^{-1}(p) = \frac{1}{\sqrt{\lambda_1}} \int_0^p \frac{1}{(1-t)^{-\frac{1}{2}} (1+at)^{-\frac{1}{2}}} dt + x_0. \]
(3.14)

we know that, \( (1+x)^x = \sum_{m=0}^{\infty} \frac{A(s, m)x^m}{\Gamma(s+1)} \)
and \( (1-x)^x = \sum_{m=0}^{\infty} (-1)^mA(s, m)x^m \) in which, \( A(s, m) = \frac{\Gamma(s+1-m)}{\Gamma(s+1)(m+1)} \).

By using the above expansion for \( (1+at)^{-\frac{1}{2}} \) and \( (1-t)^{-\frac{1}{2}} \) and some calculation, we derive:
\[ F^{-1}(p) = \frac{1}{\sqrt{\lambda_1}} \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} (-1)^n a^m A(-\frac{1}{2}, n) A(-\frac{1}{2}, m) \frac{b^{m+n+1}}{m+n+1} + x_0, \ a < 1. \]
(3.15)

The familiar procedure results in the two-parameter model:
\[ L^*_s(p) = p - G \sum_{m=0}^{\infty} A(-\frac{1}{2}, m) a^m B(m+1, 2-\frac{1}{2}) \sum_{m=0}^{\infty} A(-\frac{1}{2}, m) a^m B(m+2, 2-\frac{1}{2}) \]
\[ + G \sum_{m=0}^{\infty} \frac{\sum_{n=0}^{\infty} (-1)^n a^m A(-\frac{1}{2}, n) A(-\frac{1}{2}, m) \frac{b^{m+n+2}}{(m+n+1)(m+n+2)}}{\sum_{n=0}^{\infty} A(-\frac{1}{2}, m) a^m B(m+2, 2-\frac{1}{2})}. \]
(3.16)

3.2. An application and comparisons.

For using a specific model it must pass the test of empirical data. The maximum Tsallis entropy model (15) will be evaluated by examining the residual vector \( R_i = L(p_i) - L_i \) for an actual data set \{\( p_i \ L_i \) | \( i = 1, 2, ..., 9 \)\} given in Table 2 of Villasenor and Arnold (1989) with an actual value \( G = 0.3196 \).

In Table 1, appear the actual values \( L \) and estimated values via Kawkani-Podder \( L_{KP} \), with \( \alpha = 1.462 \) and \( \beta = 0.501 \), Pakes \( L_p \), with \( \alpha = 1.33 \) and \( \beta = 0.727 \) and classic Pareto \( L_{CL} \), with \( \alpha = 0.583 \) that have been obtained by Villasenor and Arnold (1989), also, the obtained values from maximum Second order entropy model \( L^*_s \). It is obvious that the last column in
<table>
<thead>
<tr>
<th>p</th>
<th>(L)</th>
<th>(\hat{L}_{KP})</th>
<th>(\hat{L}_P)</th>
<th>(\hat{L}_{CL})</th>
<th>(\hat{L}_{1.3})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.0213</td>
<td>0.02198</td>
<td>0.0331</td>
<td>0.05961</td>
<td>0.03321</td>
</tr>
<tr>
<td>0.2</td>
<td>0.0657</td>
<td>0.06368</td>
<td>0.0843</td>
<td>0.12206</td>
<td>0.07595</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1273</td>
<td>0.12112</td>
<td>0.1470</td>
<td>0.18785</td>
<td>0.12912</td>
</tr>
<tr>
<td>0.4</td>
<td>0.2001</td>
<td>0.19303</td>
<td>0.2200</td>
<td>0.25769</td>
<td>0.19387</td>
</tr>
<tr>
<td>0.5</td>
<td>0.2833</td>
<td>0.28255</td>
<td>0.3025</td>
<td>0.32259</td>
<td>0.27166</td>
</tr>
<tr>
<td>0.6</td>
<td>0.3781</td>
<td>0.38781</td>
<td>0.3958</td>
<td>0.41405</td>
<td>0.36447</td>
</tr>
<tr>
<td>0.7</td>
<td>0.4867</td>
<td>0.51080</td>
<td>0.5013</td>
<td>0.50458</td>
<td>0.47517</td>
</tr>
<tr>
<td>0.8</td>
<td>0.6119</td>
<td>0.65283</td>
<td>0.6225</td>
<td>0.60893</td>
<td>0.60831</td>
</tr>
<tr>
<td>0.9</td>
<td>0.7624</td>
<td>0.81535</td>
<td>0.7683</td>
<td>0.73900</td>
<td>0.77265</td>
</tr>
<tr>
<td></td>
<td>SSE</td>
<td>0.00523</td>
<td>0.00231</td>
<td>0.01622</td>
<td>0.00086</td>
</tr>
<tr>
<td></td>
<td>SAE</td>
<td>0.14349</td>
<td>0.138</td>
<td>0.3423</td>
<td>0.0808</td>
</tr>
</tbody>
</table>

**Table 1**

\(L\): actual value  \(\hat{L}_{KP}\): Kakwani - Podder model  \(\hat{L}_P\): peak model  \(\hat{L}_{CL}\): classical Pareto model  \(\hat{L}_{1.3}\): Max Second Order Ent

Table 2 is our work and others are derived from Villasenor and Arnold (1989).

We observe in Table 1, on the basis of the sum of square errors (SSE) and the sum of absolute errors (SAE)\(^1\) (as the goodness of fit criterions), \(\hat{L}_{1.3}\) is improved upon by \(\hat{L}_{KP}\) and \(\hat{L}_P\) and \(\hat{L}_{CL}\).

4. Conclusion

In this work, the maximum second order entropy density function of an income distribution was found. It provided that the mean and the Gini index are known and as a consequence of this, some models for Lorenz curve are obtained. Finally, by using of an actual data, we compared maximum Tsallis entropy Lorenz curve with some parametric Lorenz curves, in view of data in Villasenor and Arnold (1989).

References


Lorenz, M. O. Methods of measuring the concentration of wealth. Publication of the American Statistical Association, 1905.


Pakes, A. G. On Income Distributions and Their Lorenz Curves, Tech. rep., Department of Mathematics, University of Western Australia, 1981.


