ENTROPY DENSITY MAXIMIZATION WITH APPLICATION TO INCOME DISTRIBUTION

Elemary, B. R. *1, Ali, A. H. 2

*1 Department of Applied Statistics & Insurance, Faculty of Commerce, Damietta University, Egypt.
2 Msc. Student, ISSR, Cairo University, Egypt.

ABSTRACT

This paper considers the application of maximum entropy principle in estimating the income distribution when the data is in grouped format. A method of maximum entropy density estimation is used to estimate the distribution, when only limited information about intervals or ranges is available. The behavior of different forms of maximum entropy function and their relation with the conventional distributions is considered. Generalized entropy Inequality indices and Theil Index also are used. The results indicate that the Maximum Entropy formalism can provide good estimates of income distribution with high degree of accuracy.

Keywords: Maximum Entropy, Characterizing Moments, Income Distribution, Grouped Data, Goodness of Fit Measures, Inequality Measures

1. Introduction

In the recent years the applications of maximum entropy have been evolved in many fields of science like Physics, statistics, communication, and econometrics. Since complete information about the density function is not available, a parametric form is generally assumed before performing estimation. If the density function is correctly specified, the classical maximum likelihood estimation preserves efficiency and consistency. The true density is not known in almost all cases; therefore, an assumed density function could be misspecified. So the main object is to extract useful information about the unknown density from a given data by imposing some well-defined moment functions in analyzing data. By doing this it can reduce the degree of model misspecification considerably. In 1713, Bernoulli put the principle of insufficient reason, but this principle remained for many years unconvincing for lacking the justification or basis that based upon so it remained just a presumed case. In 1921, Keynes restored the formulation of this principle and called it the principle of indifference in an attempt to re-use the principle that is developed by Bernoulli but because the lack of a mathematical formula for this principle it has not been relied upon. In 1948, [1] introduced a formula to measure the amount of information in the encrypted sent messages in the field of communications and called this scale "entropy". The concept of Shannon entropy coincides with the concept of uncertainty in probability and uncertainty represents the limited and incomplete knowledge. In 1957 [2] suggested the principle of maximum entropy which is a method to choose the best estimated distribution in the case of the maximum uncertainty. In other words choosing a distribution closest to the uniform distribution and agreed with prior incomplete knowledge.

This principle states that the estimated distribution results from maximizing the entropy function under certain moments restrictions is the most likely distribution to be realized in the reality on the condition that the moments restrictions used in the experiment are the same in reality restrictions. So the maximum entropy principle is a generalization of Bernoulli and Laplace's Principle of insufficient reason. The maximum entropy principle used for performing inference about population with as less as possible assumption and this will make the problem of estimation becomes ill-posed. Mathematically this means that the number of unknown is more than the number of known. The Maximum Entropy (MaxEnt) principle is a method to estimate the underlying distribution with minimum assumptions represented by an empirical moments calculated from sample of observations, which selected from unknown distribution. The MaxEnt density provides flexible and powerful tool for density approximation and estimation since it nests a wide range of statistical distributions as special cases.

According to the maximum entropy principle, it yields the most uniform (unbiased) density estimation conditioned on the available priori knowledge. In fact, most known distributions can be regarded as the MaxEnt distribution with certain moment constraints. The computation of the MaxEnt density, however, is not an easy task because that the maximization of the entropy is usually achieved through the use of Lagrange multipliers whose numerical solution requires involved nonlinear optimization. Most existing approaches adopt the iterative Newton’s method. The Newton’s method is too computationally demanding and is sensitive to the choice of the initial values. To reduce the computational complexity, many proposed methods to compute the Lagrange multipliers have been suggested in the literature. A method by [3] was suggested to estimate a distribution when only limited information about intervals or ranges is available. An important application of this technique is the estimation of income distributions as many government agencies report summary statistics for only some ranges of the income distribution. Estimation of the exact distribution of income took a great part of attention and works of many researchers since work of Pareto 1895 until now. These works have suggested
many functional forms for probability distribution of income and this led to inventing many methods and new techniques to precisely estimate the income distribution. These different methods have competed for more than a century to prescribe the properties of income distribution. This paper considers the application of maximum entropy principle in estimating the income distribution when data is in grouped format. As well as studies the behavior of the different forms of maximum entropy function and their relation with the conventional distributions, also it evaluates the income inequality using the generalized entropy measures of inequality. The paper consists of seven sections; the current section is an introduction of the paper. Section (2) introduced the MaxEnt principle, an Algorithm to compute MaxEnt for grouped data with known interval limits, and Characterizing Moments of MaxEnt. Section (3) showed some measures of income inequality. Section (4) displayed income data models were used, and some goodnes of fit and Inequality measures. Sections (5) and (6) implied tables and graphs. Finally, section (7) was about the conclusion.

2. MAXIMUM ENTROPY

Let $x_i, i = 1, 2, ..., n$ be an i.i.d. random sample of size $n$. The MaxEnt density $f(x)$ can be obtained by maximizing Shannon’s information entropy as follow:

$$\max_{f(x)} hf(x) = -\int_R f(x) \ln f(x) dx$$

Subject to $k + 1$ known moment conditions for the entire range of the distribution.

$$\int_R f(x) dx = 1; \text{ normalization condition (natural constraint)}$$

and

$$\int_R g_i(x) f(x) dx = \xi_i; \text{ moment conditions}$$

where $i = 1, 2, ..., k$ index to the characterizing moments $\xi_i$, and their functional forms $g_i(x)$. Here $g_i(x)$ is continuous and at least twice differentiable. This optimization problem can be solved by using Lagrange’s method, which leads to a unique global maximum entropy. The solution takes the form:

$$\ell(f(x), \lambda) = -\int_R f(x) \ln f(x) dx - \lambda_0(\int_R f(x) dx - 1) - \sum_{i=1}^k \lambda_i \left( \int_R g_i(x) f(x) dx - \xi_i \right)$$

Using Lagrange’s equation (2.3) with respect to $f(x)$:

$$\frac{\partial \ell(f(x), \lambda)}{\partial f(x)} = -1 - \ln f(x) - \lambda_0 - \sum_{i=1}^k \lambda_i g_i(x) = 0$$

$$f(x) = \exp\{-\lambda_0 - \sum_{i=1}^k \lambda_i g_i(x)\}$$

(2.4)

where $f(x) \geq 0$ for all $x \in R$. $\lambda_0$ represent the normalization coefficient, which can be obtained by the formula:

$$\lambda_0 = \ln \int_R \exp\{-\sum_{i=1}^k \lambda_i g_i(x)\} dx$$

(2.5)

The coefficients $\lambda_1, \lambda_2, ..., \lambda_k$ represent the values of marginal information content in the $i^{th}$ constraint, and if $\lambda_i$, $i = 1, 2, ..., k$ was close to zero this mean that its associated moment function does not convey any valuable information. So the object is how to find functional forms $g_i(x)$ of the characterizing moment:

$$\int_R g_i(x) f(x) dx = \xi_i, i = 1, 2, ..., k$$

That exhausts all the available information in the sample which relevant for estimating the density function. Since an analytical solution for (2.4) does not exist when $k \geq 2$, one must use a nonlinear optimization technique to solve for the MaxEnt density. One way to solve the MaxEnt problem is found by [4], is to transform the constrained optimization problem into an unconstrained optimization problem using the dual approach. Substituting Eq. (2.4) into the Lagrangian function Eq. (2.3) and
rearranging terms, the dual objective function for an unconstrained optimization problem will be:

\[ \ell(\lambda) = \lambda_0 + \sum_{i=1}^{N} \lambda_i \xi_i \]

Newton’s method is used to solve for the Lagrange multiplier \( \lambda^* = [\lambda_0, \lambda_1, \ldots, \lambda_k] \) by iteratively updating:

\[ \lambda^{(r+1)} = \lambda^{(r)} - H^{-1} \frac{\partial \ell(\lambda)}{\partial \lambda} , \quad r = 0, 1, 2, \ldots \]

Where the gradient

\[ \frac{\partial \ell(\lambda)}{\partial \lambda_i} = \xi_i - \frac{\int_{R} g_j(x) e^{-\sum_{i=1}^{k} \lambda_i g_i(x)} dx}{\int_{R} e^{-\sum_{i=1}^{k} \lambda_i g_i(x)} dx} = G_j(\lambda) , \quad i = 0, 1, \ldots, k \]

and the Hessian

\[ H_{ij} = \frac{\partial^2 \ell(\lambda)}{\partial \lambda_i \partial \lambda_j} = G_{ij}(\lambda) = G_j(\lambda)G_j(\lambda) , \quad i, j = 0, 1, \ldots, k \]

\[ G_j(\lambda) = \xi_j - \frac{\int_{R} g_j(x) e^{-\sum_{i=1}^{k} \lambda_i g_i(x)} dx}{\int_{R} e^{-\sum_{i=1}^{k} \lambda_i g_i(x)} dx} \]

Since the Hessian matrix \( H \) is convex and positive definite, there exists a unique solution. The MaxEnt estimates are shown consistent and efficient by [5]. The Newton-Raphson iterative method described above is straightforward but has at least two disadvantages: (1) it computationally takes a lot of time due to a lot of numerical integrals involved, (2) it is sensitive to the choice of the initial values \( \lambda^{(0)} \) and only works for a limited set of moment constraints. Two possible reasons suggested by [6] for these disadvantages: (a) numerical errors may build up during the updating process; (b) near-singularity of the Hessian may occur for a large range of \( \lambda \) space. The MaxEnt method is equivalent to the Maximum Likelihood ML approach where the likelihood is defined over the exponential distributions family with parameters. Hence, the log-likelihood ratio can be used to test the function specification. Given:

\[ p(x_j) = \exp \sum_{i=0}^{k} \lambda_i g_i(x_j) \]

the log-likelihood can be conveniently calculated as:

\[ L = \sum_{j=1}^{N} \ln p(x_j) = -N \sum_{i=0}^{k} \lambda_i \xi_i \]

where \( \xi_i \) is the \( i^{th} \) sample origin moment.

Since the maximized entropy subject to known moment constraints is:

\[ W = -\sum_{j=1}^{N} p(x_j) \ln p(x_j) = \sum_{i=0}^{k} \lambda_i \xi_i \]

The log-likelihood is equivalent to the maximized entropy multiplied by the number of observations. This log-likelihood ratio is asymptotically distributed as \( \chi^2 \) with two degrees of freedom. An alternative measure of information proposed by [7], distinguishability between \( p \) and \( p^* \),

\[ ID(p, p^*) = 1 - \exp(-D(p \parallel p^*)) \]

which is constrained to lie between zero and one.

(2.1) Maximum Entropy for grouped data (with known Interval Limits)
For a certain distribution, the grouped summary statistics of the conditional moments of each interval is as the form:

\[
\begin{bmatrix}
  v_{1,1} & v_{2,1} & \cdots & v_{M,1} \\
v_{1,2} & v_{2,2} & \cdots & v_{M,2} \\
  \vdots & \vdots & \ddots & \vdots \\
  v_{1,J} & v_{2,J} & \cdots & v_{M,J}
\end{bmatrix}
\]

where \( v_{m,1} \) is the share of the \( m^{th} \) interval and \( \sum_{m=1}^{M} v_{m,1} = 1 \).

The \( j^{th} \) partial moment of a distribution \( f(x) \) over the \( m^{th} \) interval is defined as:

\[
v_{m,j} = \int_{I_{m-1}}^{I_m} g_j(x)f(x)dx , \quad j = 1,2,\ldots,J
\]

where \( m \) (number of groups) = 1, ..., \( M \) and \( j \) (number of moments) = 1, ..., \( J \) and \( g_j(x) \) is continuous and at least twice differentiable, \( g_1(x) = 1 \), and \( \lambda_1 \) is the normalization coefficient. The Lagrange multipliers can be solving by iteratively updating:

\[
\lambda^{(r+1)} = \lambda^{(r)} - (G^*G)^{-1}G' b , \quad r = 0,1,\ldots
\]

With dual function:

\[
\ell(\lambda) = \ln \prod_{m=1}^{M} e^\sum_{j=2}^{J} \lambda_j g_j(x)dx + \sum_{j=2}^{J} \lambda_j v_{m,j}
\]

Where \( v_{m,j} \) is the estimation of \( v_{m,j} \) calculated from the grouped sample by the

\[
\hat{v}_{m,j} = \hat{v}_{m,1} g_j(x_m) , \quad x_m \text{ is the class center:}
\]

\[
\hat{v}_{m,1} = \frac{1}{I_1} \int_{I_{m-1}}^{I_m} g_j(x)e^\sum_{j=2}^{J} \lambda_j g_j(x)dx , \quad i, j = 1,\ldots,J
\]

Where

\[
b_{m,j} = \frac{\partial \ell(\lambda)}{\partial \lambda_j} = \hat{v}_{m,j} - \int_{I_{m-1}}^{I_m} g_j(x)e^\sum_{j=2}^{J} \lambda_j g_j(x)dx , \quad i, j = 1,\ldots,J
\]

and

\[
f(x,\lambda) = \exp(-\sum_{j=2}^{J} \lambda_j g_j(x))
\]

Let \( G \) consists of \( M \) submatrices \( G^{(m)}(J \times J) \) stacked top of one another as:

\[
G^{(m)} = \frac{\partial^2 \ell(\lambda)}{\partial \lambda_i \partial \lambda_j} = \int_{I_{m-1}}^{I_m} g_j(x)e^\sum_{j=2}^{J} \lambda_j g_j(x)dx , \quad i, j = 1,2,\ldots,J , \quad m = 1,\ldots,M
\]

To ensure identification, the number of characterizing moments should be no larger than the number of moment constraints.

(2.2) An Algorithm to Compute MaxEnt Densities

There are many developed algorithms to approximately solve the maximum entropy to avoid the problem of integration divergence. An algorithm introduced to solve the MaxEnt by iterative Newton’s method with integration, was found by [8] as
well as by [6]. This solution converges to the solution for small set of given moments $\xi$, and even then requires the initial values for $\lambda$ to be close to the final solution. A modification and refinements to the algorithm suggested by [6] to derive a robust and efficient algorithm for the computation of $\lambda$, one may follow the procedure suggested by [9] for minimization of numerical integrals of form $\int_{\mathbb{R}} h(x,\lambda) dx$ with respect to $\lambda$.

One make use of the fact that if a standard quadrature method is used to compute the numerical integrals involved in the computation of $\lambda$, one obtains an approximation which may be written as:

$$\int_{\mathbb{R}} h(x,\lambda) dx \approx \sum_{n=0}^{N-1} w_n h(x_n,\lambda)$$

where $x_n,\ n=0,1,\ldots,N-1$ is a sequence of appropriately chosen "quadrature" points, corresponding quadrature weights. Taking account of this approximation, the constraints:

$$G_i(\lambda) = \int_{\mathbb{R}} g_i(x) f(x,\lambda) dx = \xi_i, \quad i = 0,1,\ldots,k$$

To satisfied by $\lambda$ become:

$$G_i^N(\lambda) = \sum_{n=0}^{N-1} w_n g_i(x_n) f(x_n,\lambda) = \xi_i, \quad i = 0,1,\ldots,k \quad (2.6)$$

Let $\lambda_N^*$ denote to $\lambda$-value satisfying (2.6). To compute $\lambda_N^*$, Newton’s method was applied by [10] for nonlinear equations to solve:

$$\lambda^{(r+1)} = \lambda^{(r)} - G^N(\lambda)^{-1} b^N(\lambda) \quad , \quad r = 0,1,\ldots \quad (2.7)$$

Where

$$G^N_{ij}(\lambda) = \frac{\partial G^N_i(\lambda)}{\partial \lambda_j} = \sum_{n=0}^{N-1} w_n g_i(x_n) g_j(x_n) f(x_n,\lambda) \quad , \quad i, j = 0,1,\ldots,k$$

and

$$b_i^N(\lambda) = \xi_i - G_i^N(\lambda) \quad , \quad i = 0,1,\ldots,k$$

Further, we can modify the update step (2.7) in such a way that convergence is guaranteed. This is possible because the Jacobian $G^N(\lambda)$ in the Newton-Raphson iterative step (2.7) is positive definite. The modification is to combine Newton’s method for nonlinear equation with a backtracking line search, which guarantees convergence if the backtracking algorithm is chosen appropriately. $\lambda$ is updated according to $\lambda^{(r)} = \lambda + \alpha_r G^N(\lambda)^{-1} b^N(\lambda)$ for a sequence of values $\alpha_r = 2^{-r+1}$, $r = 1,2,\ldots$, until the condition $\|b^N(\lambda^{(r)})\| \leq (1-1/4r)\|b^N(\lambda^{(r-1)})\|$ is satisfied, where $\|\|$ is the Euclidean norm.

(2.3) Characterizing Moments of Maximum Entropy

The problem of finding the MaxEnt density coefficients $\lambda_i, i = 0,1,\ldots,K$ and the form of the functional form $g_i(x)$ of the characterizing moment:

$$\int_{\mathbb{R}} g_i(x) f(x) dx = \xi_i$$

are the most crucial point in the estimation of an optimal density function for the underlying distribution for the sample data. As the maximum entropy is a nonlinear function so there is no analytical solution exists for $k \geq 2$ and it must use one of the optimization methods. On the other hand, $g_i(x)$ is a form of characterizing moment which takes different functions that are continuous and at least twice differentiable. These characterizing moments are sufficient statistics for exponential families; the entire distribution can be summarized by the characterizing moments.
There are some functions suggested by [11] that could be taken by \( g_i(x) \) with some restrictions on the parameters. These functions can be appropriate to capture the features of the empirical data histogram, which as the following forms:

\[
x^i : i = 1, 2, ..., n, \ln(1 + \frac{|x|}{r})^p, \tan^{-1}(\frac{x^2}{r^2}), \ln x, \ln(1 - x), \tan^{-1}(\frac{x}{\gamma}), \ln(1 + x^2), \arctan(x), \sin(x), \\
\cos(x), \sin(\ln(x)), \cos(\ln(x)), \frac{(x/\beta)^a}{a > 0, x^a}, \ln(r^2 + x^2), |x|^\theta.
\]

Some of these functional forms \( g_i(x) \) is used in this paper to estimate the population characterizing moment \( \xi_i = \bar{E}(g_i(x)) \) by its corresponding sample moment

\[
\xi_i = \frac{1}{n} \sum_{i=1}^{n} g_i(x_i); i = 0, 1, ..., k \text{ and } g_0(x_i) = 1.
\]

These forms are: \( \ln(1 + |x/r|^p) \) where \( p < 2 \), \( \tan^{-1}(x^2/r^2) \) capture high peakedness, \( \ln x \); \( 0 < x < \infty \), \( \ln(1 - x) \); \( 0 < x < 1 \), \( \tan^{-1}(x/\gamma) \) capture asymmetry, \( \log(1 + x^2) \) capture fat tails, \( \arctan(x) \) is an odd function: capture skewness and other deviations from the bell shape of symmetric distribution, such as that of normal or t distribution (Because its range is restricted between \( -\pi/2 \) and \( \pi/2 \) and it limits the influence of potential outliers).

The MaxEnt Density is a form of an exponential family distribution expressed by Lagrange multiplier coefficients rather than the original parameters. So, the relation between traditional distributions and MaxEnt density as follow:

\[
\text{MaxEnt density} = e^{\ln(\text{traditional exponential family distributions})}
\]

\[
\ln(\text{Traditional exponential family distributions}) = \ln(\text{MaxEnt density})
\]

Traditional exponential family distributions \( = e^{\ln(\text{MaxEnt density})} \)

The above transformation admit us to estimate the parameters of the traditional distribution in the form of Lagrange multipliers by many methods of optimization like Newton-Raphson iterative method, nonlinear least squares, though they require cumbersome calculation, but they guarantee an optimal solution for the parameters.

3. Income Distribution

In economics, income distribution is how a nation’s total gross domestic product is distributed amongst its population. Income distribution has always been a central concern of economic theory and economic policy, which tells us how income is divided between different groups of individuals. When dealing with the economical phenomenon like income and its distribution among the individuals in the community categories we must make a non-doubtable decision and ultimately guarantee the welfare and equality and eliminate the poverty. Estimating the exact distribution of income took a great part of attention of many researchers and this led to invent many methods and new techniques to precisely estimate the income distribution. These different methods have competed for more than a century, by using conventional methods of making parametric inference and measure the distribution fitting which is best representing the data set generated by some random process. Recently, the MaxEnt probability density principle was invented as a criterion used in information theory to measure the expected value of the information contained in a particular message and then greatly contributed in the statistical mechanics and it’s important in predicing the least biased estimation of the underlying distribution.

3.1 Measures of Income Inequality

There are many income inequality indices which have been suggested in the literature to detect the discrepancy of income among individuals. The most common index used for this purpose is the Gini coefficient which depends on the Lorenz curve to interpret how the distribution of income is unequal distributed. Although this index is commonly used in distributional analysis of all kinds, it is one of many measures available, and it incorporates particular assumptions about how income differences at different points along the income range are summarized. In other forms of distribution analysis, researchers as [12] commonly use Generalized Entropy (GE) and indices to assess inequality trends and differences. These one-parameter families’ indices have the advantage that variations in inequality aversion are straightforwardly incorporated.

3.1.1 Lorenz Curve

The Lorenz curve depicts the relationship between the percent of income received by different percentages of a given population. It is a graph shows for the bottom \( x\% \) of household, what percentage \( y\% \) of the total income they have. The percentage of households is plotted on the x-axis, the percentage of income on the y-axis. The graph of Lorenz curve is represented in Figure1. The Lorenz curve can be formally defined by:

\[
L(p) = \mu \int_{0}^{p} F^{-1}(t)dt \in [0, 1]
\]

Where \( F^{-1} \) is the quintile function and it is defined as:

\[
F^{-1}(t) = \sup\{x \mid F(x) \leq t\} \in [0, 1]
\]

3.1.2 Gini Coefficient

The Gini Coefficient can be interpreted as twice the area \( A \) of concentration between the Lorenz curve and the 45 degree line of perfect equality. The higher the coefficient, the more unequal the distribution is. Gini coefficient; [13] is defined as:
\[
G = 2 \int_0^1 [u - L(u)] \, du = 1 - 2 \int_0^1 L(u) \, du
\]

3.1.3 Generalized Entropy (GE)-Based Inequality

Building on Shannon’s work, a number of generalized information measures were developed. Though none of these measures exhibits the exact properties of the Shannon’s entropy, these measures are used often in econometrics and provide a basis for defining information theory estimators. These generalized information measures are all indexed by a single parameter \( \alpha \) and can be expressed as:

\[
GE(\alpha) = \frac{1}{N\alpha(\alpha - 1)} \sum_{i=1}^{N} \left[ (x_i / \bar{x})^\alpha - 1 \right], \text{ for real values of } \alpha.
\]

Generalized Entropy-based inequality (GE) measures are related to the concept of information theory; they have no direct interpretation in terms of the Lorenz curve. This family of measures includes popular inequality indices such as Theil’s \( T \) [14] and Theil’s \( L \) [15] measures, as well as half the square of the coefficient of variation. While this group of measures may be criticized for lacking an intuitive appeal they possess decomposability characteristics that make them particularly useful for analyzing inequality, [16]. Measure is included with GE measures in this analysis by [12]. Also [16] showed that the measure is closely related to the family of GE measures.

Consider the distribution of income as it accrues unevenly across a population of \( j \) individuals. Also, assume that all incomes are non-negative, and denote the income of the \( k^{th} \) individual to be \( X_k \). If the population is arbitrarily partitioned such that it has \( n \) groups, the income share and population share of the \( i^{th} \) group are denoted \( q_i \) and \( p_i \) respectively. The income share of group \( i \) may be calculated as the total income accruing to persons within income group \( i \), divided by the total income of the population. Similarly the population share of group \( i \) is the proportion of total population contained within that group. For income group \( i \) that contains \( b_i \) individuals, the income and population shares may be calculated as:

\[
q_i = \frac{\sum_{k: x_k \in X_i} X_k}{\sum_{k=1}^{j} X_k}, \quad p_i = \frac{b_i}{j}
\]

Clearly

\[
\sum_{i=1}^{n} q_i = 1, \quad \sum_{i=1}^{n} p_i = 1
\]

Generalized Entropy inequality measures are easily calculated from income and population share data. These measures take the general form:

\[
GE(\alpha) = \frac{1}{\alpha^2 - \alpha} \left[ \frac{1}{n} \sum_{i=1}^{n} \left( \frac{q_i}{p_i} \right)^\alpha - 1 \right]
\]

Where \( \alpha \) is a non-negative sensitivity parameter, which dictates the emphasis of inequality measure places on higher and lower ends of the distribution. Low values for \( \alpha \) will place extra emphasis on the lower end of the distribution, while higher values will place emphasis on the higher incomes. Taking the limits of this equation as \( \alpha \rightarrow 0,1 \) gives Theil’s \( L \) and \( T \) inequality measures respectively:

\[
L = \sum_{i=1}^{n} p_i \ln \left( \frac{p_i}{q_i} \right)
\]

\[
T = \sum_{i=1}^{n} q_i \ln \left( \frac{q_i}{p_i} \right)
\]

Atkinson’s measure is given as:

\[
A = 1 - \left[ \frac{1}{n} \sum_{i=1}^{n} \left( \frac{p_i}{q_i} \right)^{1-\varepsilon} \right]^{1-\varepsilon}
\]

Where \( \varepsilon \) is an inequality aversion parameter and lies between zero and infinity. The higher the value that \( \varepsilon \) takes, the more harmful inequality is to the considered society. [16] shows that this measure is ordinaly equivalent to GE measures when \( \alpha = 1 - \varepsilon \).

4. Income Data

The maximum entropy principle is used to approximate the size distribution of household income for three countries Egypt, Iraq and United States. For Egypt, household income data which include three types; earning, total income and net disposable income are got from Central Agency for Public Mobilization and Statistics, the total money income is used as a measure for income in (2009). The sample data is corresponding to 21,300 incomes of households, which were combined into 19 income groups. For Iraq, household’s income data are got from Central Statistical Organization (CSO), the total money income is used as a measure for income in (2007). The sample data is corresponding to 12,454 incomes of households in raw data, which is transformed to grouped data combined from 15 income groups using the conventional method. For United States, household’s income data was obtained from U.S.A. Census Bureau, which divided the households to many races. The selected races for this study are: Asian, Black, and Hispanic household’s data income in (2008), where the data were in grouped format. The data correspond to three samples as: 4,573 for Asian, 14,595 for Black and 13,425 for Hispanic, combined into 43 income groups. (it can back to Author for getting real income data)

4.1 Income data models

The main problem of maximum entropy densities is concentrated in the selection of appropriate characterizing moments \( g_r(x) \) that can capture the features of the empirical
data distribution, and in the estimation of entropy coefficients $\lambda$. To deal with this problem this paper used the curve fitting algorithm to find the appropriate model for the empirical data by adding functional form to the exponent sequentially and solve the nonlinear least square to estimate the model with least sum squares error and greatest determination coefficient using the Matlab curve fitting tool as follow:

The Egypt model:

$$f(x) = e^{-a - bx - cx \sin(x/100) - dx \sin(x/50) - f \sin(x/27) - g \sin(x/29) - h \sin(x/28) - r \sin(x/28) - t \sin(x/6/100)}$$

$1 \leq x \leq 100000$

The Iraq model:

$$f(x) = e^{-a - bx - c x \sin(x/100) - d x \sin(x/50) - f \sin(x/27) - g \sin(x/29) - h \sin(x/28) - r \sin(x/28) - t \sin(x/6/100)}$$

$1 \leq x \leq 21991.67$

The U.S.A. Black model:

$$f(x) = e^{-a - bx - c x \sin(x/100) - d x \sin(x/50) - f \sin(x/27) - g \sin(x/29) - h \sin(x/28) - r \sin(x/28) - t \sin(x/6/100)}$$

$1 \leq x \leq 300000$

The U.S.A. Hispanic model:

$$f(x) = e^{-a - bx - c x \sin(x/100) - d x \sin(x/50) - f \sin(x/27) - g \sin(x/29) - h \sin(x/28) - r \sin(x/28) - t \sin(x/6/100)}$$

$1 \leq x \leq 300000$

The U.S.A. Asian model:

$$f(x) = e^{-a - bx - c x \sin(x/100) - d x \sin(x/50) - f \sin(x/27) - g \sin(x/29) - h \sin(x/28) - r \sin(x/28) - t \sin(x/6/100)}$$

$1 \leq x \leq 300000$

Although these models will not warrant the representation of the data in the means of moments, it will give a good starting point for the second stage by equating the calculated sample moment with the selected model moments. The best suggested forms of characterizing moments to capture the curve of the used data and to satisfy the least sum squares error and greatest determination coefficient are found to be the fractional exponent of the variable $x^r$, $0 \leq r \leq 1$. That can reveal many feature of the data curve and the trigonometric function of the fractional exponent of variable $\sin(x^r)$, $0 \leq r \leq 1$. Which is a periodic function that can employ to capture skewness and other deviations from the bell shape of symmetric distribution and the multimodal form of the distribution, the fractional exponent of the variable with trigonometric function can control its periodicity. Generally, using periodic functions like $\sin(x)$ and $\cos(x)$ requires rescaling the data to be within $[-\pi, \pi]$. Sine and cosine terms are bounded in $[-1,1]$, thus these two moments always exist and are not sensitive to outliers. Moreover, sine and cosine can be expressed as infinite power series.

$$\sin(x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \ldots$$

$$\cos(x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \ldots$$

By turning to our suggested models and run Algorithm described in sec. (2.2) with some modifications by using a simplified Simpson method for integral approximation which is equivalent to Trapezoidal method it found the best suggested models for data income.

The best suggested model for Egypt income data is found to be:

$$f(x) = e^{-a_1 - \lambda_2 x - \lambda_3 x^{0.6} - \lambda_4 x^{0.64} - \lambda_5 x^{0.7} - \lambda_6 \sin(x/27) - \lambda_7 \sin(x/29) - \lambda_8 \sin(x/28) - \lambda_9 \sin(x/0.28) - \lambda_{10} \sin(x/0.16/100)}$$

$1 \leq x \leq 100000$

The best suggested model for Iraq income data is found to be:
The best suggested model for U.S.A. Black income data is found to be:

\[ f(x) = e^{-\lambda_1 - \lambda_2 x - \lambda_3 50 x^{0.26} - \lambda_4 30 x^{0.8} - \lambda_5 0.9 - \lambda_6 \sin(x^{0.11}) - \lambda_7 \sin(x^{0.6}/100) - \lambda_8 \sin(x^{0.6}/50) - \lambda_9 2^{*}\log(x)} \quad 1 \leq x \leq 300000 \]

The best suggested model for U.S.A. Hispanic income data is found to be:

\[ f(x) = e^{-\lambda_1 - \lambda_2 x - \lambda_3 50 x^{0.29} - \lambda_4 30 x^{0.8} - \lambda_5 0.9 - \lambda_6 \sin(x^{0.32}) - \lambda_7 \sin(x^{0.6}/100) - \lambda_8 \sin(x^{0.6}/50) - \lambda_9 2^{*}\log(x)} \quad 1 \leq x \leq 300000 \]

The best suggested model for U.S.A. Asian income data is found to be:

\[ f(x) = e^{-\lambda_1 - \lambda_2 x - \lambda_3 50 x^{0.09} - \lambda_4 30 x^{0.8} - \lambda_5 0.9 - \lambda_6 \sin(x^{0.05}) - \lambda_7 \sin(x^{0.6}/100) - \lambda_8 \sin(x^{0.6}/50) - \lambda_9 2^{*}\log(x)} \quad 1 \leq x \leq 300000 \]

The use of fractional exponent can extract useful information from the data that eventually can build excellent representative model. A vector of zeros was used as initial values for \( \hat{\lambda} \)'s. The simplified Simpson method using equal space increment was used to approximate the integration to find the MaxEnt density coefficients \( \hat{\lambda} \)'s using Matlab code, (with increment=0.5). The number of maximum iterations is restricted to 100 iteration, to warrant obtaining the solution before the Hessian become singular, and the tolerance value is 0.000001. In these models, the fractional exponent is used to capture the variability in the lower values of the empirical distribution with the trigonometric function. Which the empirical distributions have a multimodal form and as the trigonometric function is highly periodic we decrease it periodicity by the fractional exponent of the variable.

4.2 Goodness-of-fit measures

The measures of goodness-of-fit used in this paper are sum of squared errors (SSE), sum of absolute errors (SAE), and chi-square (\( \chi^2 \)) test, which are described for grouped data. These measures are defined as follow:

\[
SSE = \sum_{i=1}^{G} \left( \frac{n_i}{N} - p_i(\hat{\theta}) \right)^2
\]

\[
SAE = \sum_{i=1}^{G} \left| \frac{n_i}{N} - p_i(\hat{\theta}) \right|
\]

\[
\chi^2 = N \sum_{i=1}^{G} \left( \frac{n_i}{N} - p_i(\hat{\theta}) \right)^2 / p_i(\hat{\theta})
\]

Where

\[
p_i(\hat{\theta}) = \int_{x_i}^{x_{i-1}} f(x) \, dx
\]

\( x_i, x_{i-1} \) represent the upper and lower limits of the group \( i \). \( n_i / N \) are the observed relative frequencies. The null hypothesis is being to examine will be:

\( H_0 \) : \( x \) is distributed as the suggested model.

against \( H_1 \) : \( x \) is not distributed as the suggested model

These measures compare the estimated and observed cell frequencies, \( p_i(\hat{\theta}) \) and \( n_i / N \), respectively. The results of fitting the MaxEnt density to the data are reported in Sec. (5)

4.3 Inequality Measures

The used measures of inequality in this paper is first the Gini Coefficient which is calculated numerically (because it is impossible to derive the inverse of the suggested model) according to the following steps.

1. Calculate the Lorenz curve numerically as follow:
\[
L(u) = \mu^{-1} \int_0^p F^{-1}(t) \, dt \quad 0 \leq p \leq 1 \tag{4.1}
\]

Where:

\[
t = F(x), \quad x = F^{-1}(t),
\]

\[
p = F^{-1}(t) = x, \quad dt = f(x) \, dx
\]

By substituting in (4.1):

\[
L(u) = \mu^{-1} \int_0^x x f(x) \, dx \quad 0 \leq x \leq \infty
\]

Generate numbers from \( \min(x) \) to \( \max(x) \) for \( x \), with an optional equal increment to obtain a vector of cumulative Lorenz curve as follow:

\[
L(u_i) = \mu^{-1} \sum_{i=1}^n x_i \, f(x_i) \, w_i \quad i = 1, 2, \ldots, n
\]

Where \( w_i \), \( i = 1, 2, \ldots, n \) represent the increment (weight).

2. Using the calculated Lorenz curve vector to calculate the Gini Coefficient as follow:

\[
\cdot G = 1 - 2 \int_0^1 L(u) \, du \quad 0 \leq u \leq 1
\]

\[
G = 1 - 2 \sum_{i=1}^n L(u_i) \, du_i
\]

\[
u_i = F(x_i)
\]

\[
du_i = F(x_i) - F(x_{i-1}) \quad i = 2, 3, \ldots, n
\]

Where \( du_1 = F(x_1) \).

The second measures of inequality are the generalized entropy indices of inequality, and calculated by the following equations:

The generalized entropy index of inequality, \([14]\):

\[
I_{\alpha} = \int_0^\infty \frac{1}{\alpha(1-\alpha)} \left[ \frac{x}{\mu} \right]^\alpha - 1 \, dF(x)
\]

Where \( I_{\alpha} = -\alpha^{-1} h_{\alpha}(s(q)) \) and \( \alpha \) represent the sensitivity index. If \( \alpha = 1 \) we get Theil’s L \([14]\) inequality index:

\[
I_1 = \int_0^\infty x \log \frac{x}{\mu} \, dF(x)
\]

Where \( I_1 = -h_{1}(s(q)) \). If \( \alpha = 0 \) we get Theil’s L \([15]\) second index:

\[
I_0 = \int_0^\infty \log \frac{x}{\mu} \, dF(x)
\]

Also known as mean logarithmic deviation (MLD) index:

\[
\int_0^\infty \log(\mu) - \log(x) dF(x)
\]

If \( \alpha = 1 - \varepsilon \) for \( \varepsilon < 1 \) gives \([12]\) inequality aversion:

\[
I_A^\varepsilon = 1 - \frac{1}{\mu(F)} \left[ \int_0^\infty x^{1-\varepsilon} dF(x) \right]^{\frac{1}{1-\varepsilon}}
\]

The inequality indices reported in Sec. (5).

5. Tables
This section implies first some tables which content the estimated nonlinear least squares coefficients and goodness of fit measures for data income in three counties: Egypt, Iraq, and U.S.A: Black, Hispanic and Asian. Also, tables content Estimation of MaxEnt Density Coefficients and Inequality measures were found.

### Table 1: Estimation of Nonlinear least squares Coefficients

<table>
<thead>
<tr>
<th>Country</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>r</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Egypt</td>
<td>50.36</td>
<td>0.00924</td>
<td>8.356</td>
<td>-16.52</td>
<td>-2.294</td>
<td>-0.2215</td>
<td>1.512</td>
<td>13.36</td>
<td>4.324</td>
<td>15.16</td>
</tr>
<tr>
<td>Iraq</td>
<td>1577</td>
<td>1.445</td>
<td>-1628</td>
<td>11.78</td>
<td>-8.175</td>
<td>-0.2369</td>
<td>-19.26</td>
<td>-71.37</td>
<td>87.74</td>
<td>-</td>
</tr>
<tr>
<td>U.S.A. Black</td>
<td>-3.897</td>
<td>0.002182</td>
<td>-0.0694</td>
<td>0.00148</td>
<td>-0.0193</td>
<td>-2.542</td>
<td>-0.031</td>
<td>-0.0115</td>
<td>2.25</td>
<td>-</td>
</tr>
<tr>
<td>U.S.A. Hispanic</td>
<td>-3.881</td>
<td>0.002125</td>
<td>-0.072</td>
<td>0.00147</td>
<td>-0.019</td>
<td>-2.461</td>
<td>-0.0119</td>
<td>-0.0255</td>
<td>2.35</td>
<td>-</td>
</tr>
<tr>
<td>U.S.A. Asian</td>
<td>1.509</td>
<td>0.002566</td>
<td>0.2052</td>
<td>0.00104</td>
<td>-0.0179</td>
<td>1.382</td>
<td>-0.5109</td>
<td>-1.385</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

### Table 2: Goodness of fit Measures

<table>
<thead>
<tr>
<th>Country</th>
<th>SSE</th>
<th>R²</th>
<th>$R^2$</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Egypt</td>
<td>0.001089</td>
<td>0.9765</td>
<td>0.953</td>
<td>0.011</td>
</tr>
<tr>
<td>Iraq</td>
<td>0.0003373</td>
<td>0.9925</td>
<td>0.9824</td>
<td>0.007498</td>
</tr>
<tr>
<td>U.S.A. Black</td>
<td>0.005318</td>
<td>0.4403</td>
<td>0.3086</td>
<td>0.01251</td>
</tr>
<tr>
<td>U.S.A. Hispanic</td>
<td>0.006125</td>
<td>0.2979</td>
<td>0.1327</td>
<td>0.01342</td>
</tr>
<tr>
<td>U.S.A. Asian</td>
<td>0.006999</td>
<td>0.7823</td>
<td>0.7388</td>
<td>0.01413</td>
</tr>
</tbody>
</table>

### Table 3: Estimation of MaxEnt Density Coefficients

#### Egypt

<table>
<thead>
<tr>
<th>Mean’</th>
<th>SSE</th>
<th>SAE</th>
<th>Chi-square</th>
<th>MaxEnt Density Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>23002.59</td>
<td>9.82×10^5</td>
<td>0.026</td>
<td>364.83</td>
<td>$\lambda_1 \quad \lambda_2 \quad \lambda_3 \quad \lambda_4 \quad \lambda_5 \quad \lambda_6 \quad \lambda_7 \quad \lambda_8 \quad \lambda_9 \quad \lambda_{10}$</td>
</tr>
<tr>
<td>Sample Mean = 23305.094</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 4: Estimation of MaxEnt Density Coefficients

#### Iraq

<table>
<thead>
<tr>
<th>Mean’</th>
<th>SSE</th>
<th>SAE</th>
<th>Chi-square</th>
<th>MaxEnt Density Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>594.878</td>
<td>0.0042</td>
<td>0.1352</td>
<td>5761.44</td>
<td>$\lambda_1 \quad \lambda_2 \quad \lambda_3 \quad \lambda_4 \quad \lambda_5 \quad \lambda_6 \quad \lambda_7 \quad \lambda_8 \quad \lambda_9 \quad \lambda_{10}$</td>
</tr>
<tr>
<td>Sample Mean = 595.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 5: Estimation of MaxEnt Density Coefficients

#### U.S.A. Black

<table>
<thead>
<tr>
<th>Mean’</th>
<th>SSE</th>
<th>SAE</th>
<th>Chi-square</th>
<th>MaxEnt Density Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>52206.65</td>
<td>0.0027</td>
<td>0.2717</td>
<td>5018.59</td>
<td>$\lambda_1 \quad \lambda_2 \quad \lambda_3 \quad \lambda_4 \quad \lambda_5 \quad \lambda_6 \quad \lambda_7 \quad \lambda_8 \quad \lambda_9 \quad \lambda_{10}$</td>
</tr>
<tr>
<td>Sample Mean = 52425.167</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 6: Estimation of MaxEnt Density Coefficients

#### U.S.A. Hispanic

<table>
<thead>
<tr>
<th>Mean’</th>
<th>SSE</th>
<th>SAE</th>
<th>Chi-square</th>
<th>MaxEnt Density Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>57950.69</td>
<td>0.00295</td>
<td>0.273</td>
<td>4694.8</td>
<td>$\lambda_1 \quad \lambda_2 \quad \lambda_3 \quad \lambda_4 \quad \lambda_5 \quad \lambda_6 \quad \lambda_7 \quad \lambda_8 \quad \lambda_9 \quad \lambda_{10}$</td>
</tr>
<tr>
<td>Sample Mean = 58245.358</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 7: Estimation of MaxEnt Density Coefficients

#### U.S.A. Asian

<table>
<thead>
<tr>
<th>Mean’</th>
<th>SSE</th>
<th>SAE</th>
<th>Chi-square</th>
<th>MaxEnt Density Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1 \quad \lambda_2 \quad \lambda_3 \quad \lambda_4 \quad \lambda_5 \quad \lambda_6 \quad \lambda_7 \quad \lambda_8$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 8: Inequality measures - Egypt

<table>
<thead>
<tr>
<th>Gini Coefficient</th>
<th>Theil’s $T$ Index $\alpha = 1$</th>
<th>Theil’s $L$ Index $\alpha = 0$</th>
<th>Atkinson Index $\varepsilon$</th>
<th>Generalized Entropy Index $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3210662</td>
<td>0.176014</td>
<td>0.16897</td>
<td>0.0174</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0827</td>
<td>0.1129</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1129</td>
<td>0.142</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.142</td>
<td>0.1684</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1684</td>
<td>0.168</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.168</td>
<td>0.1709</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1709</td>
<td>0.17399</td>
</tr>
</tbody>
</table>

### Table 9: Inequality measures - Iraq

<table>
<thead>
<tr>
<th>Gini Coefficient</th>
<th>Theil’s $T$ Index $\alpha = 1$</th>
<th>Theil’s $L$ Index $\alpha = 0$</th>
<th>Atkinson Index $\varepsilon$</th>
<th>Generalized Entropy Index $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4887</td>
<td>0.48075</td>
<td>0.39979</td>
<td>0.0466</td>
<td>0.1289</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.19998</td>
<td>0.262</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.262</td>
<td>0.325</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.325</td>
<td>0.3995</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.3995</td>
<td>0.403</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.403</td>
<td>0.412</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.412</td>
<td>0.4265</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.4265</td>
<td>0.4359</td>
</tr>
</tbody>
</table>

### Table 10: Inequality measures - U.S.A. Black

<table>
<thead>
<tr>
<th>Gini Coefficient</th>
<th>Theil’s $T$ Index $\alpha = 1$</th>
<th>Theil’s $L$ Index $\alpha = 0$</th>
<th>Atkinson Index $\varepsilon$</th>
<th>Generalized Entropy Index $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.51</td>
<td>0.458508</td>
<td>0.538711</td>
<td>0.04</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.219</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.3</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.38</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.52</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.48</td>
<td>0.465</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.465</td>
<td>0.456</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.456</td>
<td>0.456</td>
</tr>
</tbody>
</table>

### Table 10: Inequality measures - U.S.A. Hispanic

<table>
<thead>
<tr>
<th>Gini Coefficient</th>
<th>Theil’s $T$ Index $\alpha = 1$</th>
<th>Theil’s $L$ Index $\alpha = 0$</th>
<th>Atkinson Index $\varepsilon$</th>
<th>Generalized Entropy Index $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4917</td>
<td>0.422234</td>
<td>0.5083</td>
<td>0.042</td>
<td>0.124</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2</td>
<td>0.282</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.282</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.36</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.48</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.45</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.43</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.42</td>
<td>0.42</td>
</tr>
</tbody>
</table>

### Table 10: Inequality measures - U.S.A. Asian

<table>
<thead>
<tr>
<th>Gini Coefficient</th>
<th>Theil’s $T$ Index $\alpha = 1$</th>
<th>Theil’s $L$ Index $\alpha = 0$</th>
<th>Atkinson Index $\varepsilon$</th>
<th>Generalized Entropy Index $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.48025</td>
<td>0.3943</td>
<td>0.65428</td>
<td>0.040</td>
<td>0.125</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2</td>
<td>0.315</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.315</td>
<td>0.423</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.423</td>
<td>0.594534</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.594534</td>
<td>0.51145</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.51145</td>
<td>0.4589</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.4589</td>
<td>0.4247</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.4247</td>
<td>0.4023</td>
</tr>
</tbody>
</table>

### 6. Graphs

![Lorenz Curve](image-url)
The following graphs showed the Nonlinear least squares curves for all applied Data Income:

**Figure 2**
Nonlinear least squares curve (Egypt)
Levenberg Marquadt Algorithm

**Figure 3**
Nonlinear least squares curve (Iraq)
Gauss Newton Algorithm

**Figure 4**
Nonlinear least squares curve (U.S.A. Black)
Levenberg Marquadt Algorithm

**Figure 5**
Nonlinear least squares curve (U.S.A. Hispanic)
Levenberg Marquadt Algorithm

**Figure 6**
Nonlinear least squares curve (U.S.A. Asian)
Levenberg Marquadt Algorithm

The following Figures report the histograms of the income sample and the estimated MaxEnt densities, the fitted densities closely resemble the shape of the histogram of the sample.
Figure 7. Egypt Data Income

Figure 8. Iraq Data Income

Figure 9. U.S.A Data Income (Black)
7. CONCLUSION

This paper have considered the application of the Principle of Maximum Entropy to estimate the probability distribution of income for grouped data, given the groups and the frequency of each group which contribute as constraints matrix on income distribution. Since entropy measures the amount of missing information, the MaxEnt is the least committal with respect to this missing information. From the viewpoint of statistical inference, there is no reason to prefer any other distribution of the specific form. The results obtained in this paper indicate that the MaxEnt formalism can provide good estimates of income distribution with high degree of accuracy, but this degree of accuracy depends on many factors. The number of groups of the data is in grouped format, as the number of groups is large (e.g. U.S.A. data) this will give more information but will require more conditional moments to be entered in the exponent of the MaxEnt function. This will lead to the presence of correlation problem between these conditional moments, which make the Hessian matrix approximately ill-conditioned and prevent the iterative optimization method to converge to the optimal solution. The suggested treatment for this problem is to eliminate the correlation by choosing conditional moments uncorrelated with each other, or an additional step is introduced in the iterations to deal with this problem. The Newton step is re-defined as:

\[ \lambda^{(r)} = \lambda - \alpha_r H^{-1} b \]

where \( \alpha_r \) is called the step length. The step length is determined by a local optimization of the function, called a line search that is given the direction and the starting point this will guarantee a faster convergence before the Hessian matrix arrives its singularity and improve the estimated coefficients. Even though there are rules for choosing functional form of the conditional moments, but these rules do not cover all the existing functions that could be more convenient and flexible to adapt with many forms of empirical distributions which a case to be more investigated in the future. As the accuracy of integral approximation improves the estimation of the maximum entropy model, choosing appropriate method to approximate the integral could give a better result; the most reliable method is Gauss-Kronrod quadrature method. Generalized entropy Inequality indices and Theil Index indicate that the most of inequality concentrated in the lower end of the distribution and begin to decrease slightly as the sensitivity parameter \( \alpha \) increase for U.S.A. while the most of inequality concentrated in the upper end of the distribution and begin to decrease slightly as the sensitivity parameter \( \alpha \) decrease for Egypt and Iraq.
References


