ECONOMIC INEQUALITY: CONCEPT AND MEASUREMENT

The idea of inequality is both very simple and complex. This has enthused people with its powerful appeal to the extent that it has emerged as an important area of research, engaging several eminent scholars. The existence of various levels of society provides us with a tangible proof of inequality and it is essential for a community to know the extent of this inequality [Atkinson (1970)]. The literature is overflowing with works on inequality at various levels of aggregation. All the studies, conducted at different levels of aggregation with differing motivations and degrees of sophistication, revolve around a single theme: "How a given entity, a surrogate of the quality of human existence, is shared by the relevant units in a given time-space domain".

Of all these inequalities the study of economic inequality has assumed special significance over ages. Economic inequalities, as Tendulkar (1983, 1987) put it, are reflected in four size distributions viz wealth, earnings, income and levels of living. These four taken together, portray the distributional consequences of the operation of economic process.
The four kinds of inequalities, described above, are all inter-related. "Wealth", refers to the various types of privately-owned assets that yield income. "Earnings" refer to the economic rewards accruing to individuals in return for their contributions to various types of labour. 'Incomes' refer to the amounts received by the members of a household from all sources and shared by all its members. The size distribution of incomes bears directly on the size distribution of purchasing power, which reflects itself into the level of living partly by acquiring goods for current consumption and partly by saving for future goods and services.

But, it is not difficult to observe that major sources of "income" are deeply connected with the distribution of wealth, and earnings; which affect the levels of living. Hence the term "economic inequality" - a wide and complex term which may accommodate all the four terms into its domain, may lead us to the realm of "income inequality"; which as Tendulkar (1983) observes is the "most comprehensive size distribution in terms of the coverage of the population as well as coverage of the sources of income. A size distribution of earnings focuses on the unequal economic rewards, provided by the production processes, to those individuals who participate in it. In contrast, a size distribution of incomes seeks to focus on the unequal welfare implications of income, "earned", "unearned" and
"transferred", in so far as their disposal leads to welfare in terms of current and future consumption". Hence economic inequality, whether perceived in the welfare context or through a perception of society can be perceived through "income inequality".

In the light of the above arguments, we have focussed our attention on the "income inequality" - an integral aspect of "economic inequality". Having settled this issue of dealing with "income inequality", we are faced with the two sets of problems,

(1) The problem of defining the concept of income, deciding upon the right unit of analysis and the choice of appropriate time period.

(2) Once this problem is solved, the issue of measurement of inequality starts dominating.

With several options available and multiplicity of arguments attached to each of these, a brief discussion of these issues is necessary in order to reach at a conclusive argument.

3.1(a) The Concept of Income

The concept of "income" is a much debated one, resulting into several definitions of the term; some of which have attractive implications. For example, Sen's (1981) definition of the term as "exchange entitlement" views income as a potential yield of wealth. Roegen (1971)
asserts that income may be regarded as a means of enhancing the daily enjoyment of the individual, providing the flow of consumer goods along with more leisure time at his disposal.

Howsoever good these definitions may sound the practical applicability of these is a doubtful proposition. In both the above definitions there is quite a bit of room for imputations and consequent manoeuvres. Therefore, we have tried to adopt a simple concept of income as suggested by Murray (1959) and Simons (1969), whereby income is considered to be the money value of the net accretion to one's economic power between two points of time.

3.1(b) The Unit of Analysis

Given the peculiar organisational structure of farm organisms, a farming household is considered the natural unit of analysis as against the "individual" or the "nuclear family".

However, while measuring the incidence of income inequality due allowance will be made to remove the biases on account of variation in demographic composition of the operating households. While per capita transforms might serve a convenient route for accommodating differences in size, the use of adult equivalent scale can take care of differing age and sex composition of those units.

3.1(c) Choice of An Appropriate Time Period

The extent of "income inequality" is to be measured in some income profile for which the representative
sample is required within an appropriate time period. This time period can extend from a single day to the life time as a whole. It is argued that only life time stream truly represents the unit's economic position, as transitory fluctuations in income will disappear and a greater equality will be observed. But the obvious difficulty of playing with such data is too much to be ignored. As pointed by Iyengar (1978), "...working with life income distribution is not feasible because it involves discounting and other difficult problems such as the treatment of transfers, subsidies, public investments and taxes". Hence keeping within our limited scope we have chosen one single agricultural year (1979-80) as the appropriate time period for measuring the extent of inequality under the technological spectrum.

3.2 Statistical Measurement of Inequality.

Having chosen an appropriate measure of income, the unit of analysis and the time period, the search for suitable inequality measures forms the next logical step.

As regards the measurement, Boulding (1975) observes, "Whatever we are measuring implies complex structure, then cardinality may be an illusion. On the other hand, the measurement of equality and inequality by ordinal numbers is a problem". We accept cardinal measures as appropriate for our objective; which is strongly
recommended by Kuznets (1953) and supported by various such measurements suggested by Stark (1972), Sen (1973), Bartels (1977), Takayama (1979), Atkinson (1980) besides others.

But measurement itself is value loaded theme. It was in this context that Kolm (1976) talked of the "rightist", "leftist" and "centrist", measures of inequality which embody the economic and ethical properties of the term "inequality". The debate is wide; yet to arrive at some conclusion, the term "measures of inequality" is to be understood in some definite perspective.

The term "measure of inequality" is ambiguous in the sense that it can refer to an "index of inequality" or to an "estimated value of such an index". Our immediate concern with the term is in its former sense. The estimated value of the index will also be analysed but in the empirical analysis of this theoretical concept. While referring to various inequality indices or measures, a careful definition of the term "income inequality measure" is in order. "An income inequality measure is a condensed quantitative indication of how and to what extent a certain income distribution differs from a reference distribution in which differences have been weighted according to specific normative distributional preferences".

The definition rules out the use of purely graphical representations and implies single real number instead of a vector to be chosen as an inequality indicator.
But use of different weighting procedures and distributional preferences can yield a panorama of such indices. Formally stated an inequality measure $I$ is a real valued function $I(Y) = I(Y_1, Y_2, ..., Y_n)$ of $n$ entities, in a society consisting of $n$ units having incomes $Y_i (i = 1, ..., n)$.

The above formal definition conceals the ideological underpinning of a measure. Therefore, statistical measurement of inequality becomes more than a simple exercise in measurement.

The choice of the single best measure of income inequality might be a debatable proposition but all the measures choose the distribution characterised by equal incomes as the reference distribution. And, as an other extreme, frequently the distribution of complete inequality is choosen as the one where a single recipient receives the entire income. We can remove the first issue of the debate by agreeing on the proposition that there is no single best coefficient of inequality. Some coefficients are more suited to reflect one aspect while others might be needed to highlight a different dimension.

3.2(a) **Criteria for Indices of Inequality.**

Generally, while comparing the various measures of inequality, some conditions are laid down to judge the suitability of a measure. This tradition dates back to Dalton (1920), who puts it, "We may still lay down certain general principles, which shall serve as tests to which
various plausible measures of inequality may be submitted. These conditions though serving as a criteria for various measures are non-binding in nature". Even Kolm's (1976) definition of "leftist" and "rightist" measures of inequality stems from the fact of analysing the various properties of such measures; which explain the varying aspects of inequality.

Some of the important conditions supposed to be possessed by a good measure of inequality, are discussed below:

I. **Condition of Symmetry**

   This condition which is also known as the criterion of anonymity requires a measure to depend only on the frequency of incomes in an income profile irrespective of the order in which individuals are ranked within the distribution i.e. \( I(Y) = I(PY) \), where \( P \) is the permutation matrix of order \( n \). This criterion implies impartiality between persons (units) with the result that even if two units interchange their income positions, the measure remains unchanged.

II. **Pigou-Dalton Condition**

   According to this criterion if a distribution is modified by altering two incomes only so as to leave their total unaltered, then the index concerned must be increased, unchanged or decreased, according as the absolute difference between the two incomes is increased,unchanged or decreased.
Hence, if a strictly positive transfer of income, \( h \), takes place from a unit with income \( Y_p \) to a unit with lower income \( Y_r \) such that

\[ Y_p + h < Y_r - h, \quad h > 0, \quad p < r \]  

then according to this criterion

\[ I(Y_1, Y_2, \ldots, Y_p, \ldots, Y_r, \ldots, Y_n) > I(Y_1, Y_2, \ldots, Y_p + h, \ldots, Y_r - h, \ldots, Y_n) \]  

(3.2)

This condition implies that the measure of inequality should be strictly Schur-Convex.9

III. Condition of Scale Invariance

According to this condition the inequality index should remain unaffected if we keep the proportionate distribution of income along the income scale unaltered even if we increase or decrease the total amount of income. So, according to this condition;

\[ I(Y) = I(\lambda Y) \text{ for } \lambda > 0 \]  

(3.3)

i.e. a measure should be independent of the scale in which incomes are measured. There is sharp disagreement over this condition. While Sen (1973) feels that inequality should decrease when all incomes are raised proportionately, Kolm (1976) argues that such a move amounts to an increase in inequality.
IV. Condition of Sign and Size

According to this criterion, a measure should be zero at equality, positive in case of inequality and unity in case of extreme inequality i.e.

\[ I(\bar{Y}) = 0; \quad I(\lambda \bar{Y}) = 0, \quad \lambda > 0 \]  
(3.4)

\[ I(Y) > 0; \quad Y_i \neq \bar{Y} \text{ for some } i \]
\[ \quad \text{and } i = 1, 2, \ldots, n \]  
(3.5)

\[ I(Y) = 1; \quad \text{where } Y_i = 0 \text{ for every } i \neq k \]
\[ \quad \text{and } Y_k = n\bar{Y} \]  
(3.6)

V. Familiarity and Convenience of Computation

Meeting the other requirements mentioned so far, an index applied for practical purposes must be easy to compute and capable of being estimated from the statistics in a readily available form.

Summing up these conditions, we may again emphasize the non-binding nature of these conditions. Depending on the purpose, some conditions may be picked up and others may be left. These conditions are desirable but not necessary and we may not come across a single index possessing all these properties simultaneously. Mentioning these conditions only facilitates one's decision to choose a particular measure for a particular purpose.

3.2(b) Some Important Measures of Inequality

After discussing the criteria for good measures of inequality, some important measures of inequality are discussed below, along with some of their properties. Though
we don't intend to adopt all these measures for studying the extent of inequality in our data profile related to the rural Punjab, yet a brief discussion about those measures would help us choose a particular index for our study.

The following indices have been picked up for the purpose,

(i) The variance and the co-efficient of variation,
(ii) The relative mean deviation,
(iii) The standard deviation of logarithms,
(iv) The relative mean difference and related measures—Gini's co-efficient of concentration,
(v) The extreme mean disparity ratio,
(vi) Theil's entropy measure,
(vii) Atkinson's measure.

The first six measures listed above can be categorised as "Positive Measures", while the last measure falls under the category of "Normative Measures". Strictly speaking this distinction is superficial as Sen (1973) observes "In some ways the positive measures .....can also be viewed as normative measures with specific assumptions about social welfare function".

In order to define the statistical indices of inequality, we consider the income as a random variable \( Y \) with specific values \( Y_i, \ i=1, \ldots, n \), where \( n \) is the total number of recipients studied. The definition of indices of inequality may be based on two different approaches, i.e.
one that assumes income to be a continuous random variable and another starting from the assumption of a discrete random variable.

In the discussion that follows, both these approaches have been adopted.

(i) The Variance and the Co-efficient of Variation

This index of inequality takes into account the difference between all the possible variate values.

If $Y$ is a continuous random variable, variance is defined as

$$\sigma^2 = E[(Y-\mu)^2] \quad (3.7)$$

where, $\mu = E[Y] = \text{Mean Income}$

If $Y$ is assumed to be discrete random variable, variance is defined as

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \quad (3.8)$$

where, $\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i = \text{Mean Income}$.

Unlike range, variance takes into account the total distribution. Variance is homogenous of degree 2; if homogeneity of degree 1 is preferred, one should use standard deviation [Bartels (1977)].

In case of perfect equality, the variance is zero, while situation of complete inequality gives a value equal to $(n-1)\bar{Y}^2$ (in the discrete case) and $(n-1)\mu^2$ (in continuous case). The measure is widely used in literature.

As Dalton (1920) pointed out, variance is sensitive to transfer of income, but it possesses increased
sensitivity to changes at higher incomes. One may use a different reference value instead of the mean if this sensitivity is regarded as a drawback. Bartels (1977) suggests a new measure on this line when the maximum income value is taken as a reference value. Another disadvantage using mean income level is pointed by Sen (1973), "One distribution may show much greater relative variation than another and still end up having a lower variance, if the mean income level around which the variations take place is smaller than the other distribution".

Another measure which concentrates on relative variation and is free from the above drawback, is the "coefficient of variation"; defined as follows:

$$C.V. = \frac{\text{Standard Deviation}}{\text{Mean}}$$  \hspace{1cm} (3.9)

This measure takes a value, 0, when all the incomes are equal, a case of perfect equality, and is, $\sqrt{n-1}$ in the event of all the income going to a single unit.

This measure is sensitive to transfers of income at all levels but the effect of transfer is the same, independent of the income level at which it is made.\(^\text{13}\)

Both variance and coefficient of variation are strictly Schur-Convex.

(ii) \textbf{The Relative Mean Deviation}

In this measure the deviations from mean are not squared, leading to heavy weight at large deviations from
mean; rather in this measure all deviations from mean are judged to be equally important. The measure is defined as

\[
R.M.D. = \frac{M.D.}{\mu}
\]  
(3.10)

where,  

\[
M.D. = \text{Mean Deviation of Y about } \mu = E(|Y-\mu|)
\]  
(3.11)

and  

\[
\mu = \text{First order moment about the origin}
\]  
(3.12)

For the discrete random variable case we drive

\[
R.M.D. = \frac{1}{n} \sum_{i=1}^{n} |Y_i-Y| 
\]  
(3.13)

The relative mean deviation takes values\(^{14}\) on the interval 

\([0, 2(n-1)/n]\)

For a large \(n\), the upper limit approximates the value 2. That is why the associated inequality measure is generally taken as \(1/2 \ R.M.D.\), which is \([0,1]\), and is known as Pietra ratio. An extensively used measure in literature\(^{15}\), this measure is also known by other names viz. "maximum equalisation percentage" and "Kuznets' ratio". Mehran (1976) has shown that "The relative mean deviation can be viewed as the maximum distance between the Lorenz curve and the line of perfect equality".

The measure uses a distribution of complete equality as a reference distribution and absolute differences from the reference income are weighted equally. One could also use different convenient reference income
distributions and different weighting schemes depending on the specific reasons.

(ii) The Standard Deviation of Logarithms

This measure which attaches greater importance to income transfers at the lower end rather than to transfers at high income levels is defined as

\[ S.D.\text{logs} = \log[\text{E}(Y-u)^2]^{\frac{1}{2}} \]  \hspace{1cm} (3.14)

in case of continuous random variable.

and \[ S.D.\text{logs} = \left[ \frac{1}{n} \sum_{i=1}^{n} \left( \log \frac{Y_i}{\bar{Y}} \right)^2 \right]^{\frac{1}{2}} \] \hspace{1cm} (3.15)

in case \( Y \) is discrete.

In case of complete equality, the measure takes a value zero but the upper limit is not easy to obtain because of the involvement of logarithm of zero. It may be pointed out that in standard statistical literature, the deviation is taken from geometric mean rather than the arithmetic mean, though the use of arithmetic mean is also not uncommon.

It is observed that at very high levels of income the measure decreases instead of increasing, when there is a transfer of incomes from a relatively low-income unit to a high income unit. But since income levels, as they get higher and higher, show severe contraction, this drawback is not very serious.
The measure mean difference takes into account the absolute differences between all possible pairs of values of variable \( Y \). Let \( F(Y) \) denote the Cumulative Distribution Function of continuous r.v \( Y \), then mean difference is defined as

\[
\Delta = \mathbb{E}[|Y_1 - Y_2|] \quad (3.16)
\]

In the discrete case, mean difference is defined as

\[
\Delta = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1 \neq i}^{n} |Y_i - Y_j| \quad (3.17)
\]

with repetition and

\[
\Delta = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1 \neq i}^{n} |Y_i - Y_j| \quad (3.18)
\]

without repetition.

It can also be looked upon as a particular case of Bartle's (1977) "generalised weighted difference indicator". Two other weighting schemes are also suggested by Bartels (1977), the "income weighted mean difference" and "rank-order weighted mean difference".

The special measure has the standardised version \( \Delta/\mu \) or \( \Delta/Y \) (as the case may be continuous or discrete) known as relative mean difference. But the commonly used normalised version is \( \Delta/2\mu \) (it is \( \Delta/2Y \) in discrete case), popularly known as "Gini coefficient of concentration".
Thus Gini coefficient of concentration is defined as

\[ G = \frac{E[|Y_1-Y_2|]}{2\mu} \] (3.19)

where \( \mu = E(Y) \) & \( Y \) is a continuous random variable.

While in discrete case,

\[ G = \frac{1}{2n(n-1)Y} \sum_{i<j} |Y_i-Y_j| \] (3.20)

with repetition and

\[ G = \frac{1}{2n^2 Y} \sum_{i<j} |Y_i-Y_j| \] (3.21)

without repetition.

The measure lies between 0 and 1 with some exceptions \(^{21}\) and can be related to a widely-used graphical representation of an income distribution viz. "The Lorenz Curve" \(^{22}\).

The measure has a long tradition in the study of income distribution and is one of the most popular measures of income inequality.

We discuss in some detail its properties and some modifications applied to this index in the following sub-section.

3.2b-IV(a) Gini Index and its Modifications

Ever since its advent in 1912, the Gini ratio along with the accompanying Lorenz curve remained the subject matter of intellectual probe by researchers both in the field of theory and empirical study on income distribution. Dalton (1920) discussed this measure in his study on income inequality and appreciated its being "perfectly sensitive to
transfers" of income. As defined in the previous section, "Gini mean difference" is the arithmetic average of the absolute difference between all possible pairs of income. Dividing it by the arithmetic mean, we obtain a measure of relative deprivation called the "Gini ratio" or "Gini coefficient of concentration".

Gini mean difference derives its appeal mainly from the fact that it can be shown to be related to a widely used graphical representation of an income distribution, viz., the Lorenz curve.

The Lorenz curve due to Lorenz (1909) is a graphical representation of an income distribution, whereby the cumulative proportion of the population arranged from the poorest to the richest is depicted on the horizontal axis and the proportion of income enjoyed by the bottom Y% of the population is shown on the vertical axis.

Obviously 0% of the population enjoys 0% of the income and 100% of the population enjoys all the income. Hence Lorenz curve runs from one corner of the unit square to the diametrically opposite corner.

Let Y be a random variable denoting income of a set of individuals (units).

Let \[ F(y) = P[Y \leq y] \] (3.22)
be the c.d.f of Y assumed to be continuous and differentiable to at least second order.
Then \( F_1(y) = \frac{1}{\mu} \int_0^y udF(u) \) (3.23) is the proportion of income received by units having income less than or equal to \( y \), when all incomes are assumed positive and the mean and variance of \( Y \) i.e. \( E(Y) = \mu \) and \( \text{var}(Y) = \sigma^2 \), exist and are finite.

A plot of \( F(y) \) on the horizontal axis and \( F_1(y) \) on the vertical axis according to increasing \( y \) is the Lorenz curve. Here \( F_1(y) \) is known as the Lorenz curve ordinate and \([F(y), F_1(y), 0 < y < \infty] \) is the equation of the Lorenz curve.

**Fig.3.1 LORENZ CURVE**

Since \([dF_1(y)/dF(y)] = y/\mu \geq 0\), this implies that the Lorenz curve is increasing monotonically. As \( y \) approaches zero, \([dF_1(y)/dF(y)] \) also approaches zero and for \( y \) approaching \( \infty \), \([dF_1(y)/dF(y)] \) approaches \( \infty \). Hence Lorenz curve tends to flatten out on both the extremes, horizontal near the origin and vertical at the farthest end.
Moreover, \( \frac{d^2 F_1(y)}{d F(y)} = \frac{1}{\mu f(y)} > 0 \) implying thereby that Lorenz curve is convex towards the population axis and lies below the egalitarian line. (If the total income is uniformly distributed in the population, the Lorenz curve is a straight line, called egalitarian line). The equation of the egalitarian line is given by \( F_1(y) = F(y) \). The distance between the Lorenz curve and the egalitarian line (the line of absolute equality) known as the "concentration surface" is the measure of the extent of inequality present in the distribution.

An obvious measure related to this graphical presentation is the ratio of concentration surface to the triangular area underneath the diagonal line (Ref. fig.3.1), while the later is \( 1/2 \). Consequently, the ratio equals "twice the concentration surface", which is also known as the "Gini ratio" or "Gini ratio of concentration" or "Gini coefficient of concentration". Further, it is easy to prove that "twice the concentration surface" equals \( \Delta/2 \mu \) which is another form of Gini ratio used in practice.

Following Sen (1973) the Gini coefficient can be rewritten in several ways:

\[
G = \frac{1}{2n^2Y} \sum_{i<j} (Y_i - Y_j) \quad (3.24)
\]

\[
= 1 - \frac{1}{n^2Y} \sum_{i<j} \min(Y_i, Y_j) \quad (3.25)
\]

\[
= 1 + \frac{1}{n} - \frac{2}{n^2Y} \sum_{i<j} Y_i \quad (3.26)
\]

where, \( Y_1 \geq Y_2 \geq \ldots \geq Y_n \).
Still another version of Gini ratio based on order statistics is given as [Glasser (1962)]

\[
G = \frac{1}{n^2} \sum_{i=1}^{n} (2i-n-1)Y_i
\]  

(3.27)

where, \(Y_i\) is the ith order statistic, which suggested that Gini ratio is a weighted sum of relative income shares with weights heavier at both extremes. Because of this property the measure is very sensitive to income transfers around the middle of the profile than at the tails. This is taken as the drawback of the measure; but this property may be considered attractive if income policy is most concerned with attempts to change the position of low and high income recipients.

The Gini ratio will yield biased estimates in case of grouped data, since within - class inequality is ignored. But it is possible to derive lower and upper bounds for the real value of Gini index, as suggested by Gastwirth(1972); if both class bounds and average within class incomes are known. But when class bounds and/or within class average incomes are not known, a derivation of lower and upper bounds can still be found in Mehran (1975b). Still the concentration ratio is very much biased particularly is highly positively skewed distributions.

Gini ratio has been interpreted differently by different researchers. Mehran (1976) has shown Gini ratio and the relative mean deviation to be the particular cases
of a general class of linear measures of income inequality. It is shown that only Gini ratio as a linear measure satisfies "Pigou-Dalton transfer principle" but fails to satisfy a stronger transfer principle than Pigou - Dalton's viz. "Diminising Transfer Principle." 26.

Another significant point to be mentioned with regard to Gini ratio is the fact that Gini ratio can be decomposed into "within" and "between" groups inequality to analyse the overall inequality. As Mehran (1975a) suggests

\[
Gini \text{ ratio} = \frac{\Delta}{2\mu} = \frac{\sum_{g=1}^{G} \frac{N_g Y_g}{NY} G_g}{2\mu} + \sum_{g=1}^{G} \frac{\sum_{g' \neq g} \frac{N_g Y_g}{NY} G_{g,g'}}{2\mu}
\]

where \( G \) is the Gini ratio obtained for group \( g \)

\( G_g \) = Gini ratio obtained for group \( g \)

and \( G_{g,g'} \) = Gini ratio obtained by taking all pairs of incomes, one in group \( g \) and other in group \( g' \). Bhattacharya and Mahalanobis (1967) suggested a decomposition of "Gini's mean difference"

\[
\Delta = \text{E}[|Y^{(1)} - Y^{(2)}|]
\]

where \( Y^{(1)} \) and \( Y^{(2)} \) are two independent observations from
the given population, as

$$\Delta = \sum_{i=1}^{k} p_i \Delta_i + \sum_i p_i p_j (E|Y_i^{(1)} - Y_j^{(2)}|)$$

(3.30)

where $\Delta_i = \text{Gini mean difference for the group distribution}$

$F_i(Y)$ and $Y_i$ and $Y_j$ are assumed to be values associated with

non-negative variate $Y$ with probabilities $p_1, p_2, \ldots, p_k$

such that

$$\sum_{i=1}^{k} p_i = 1$$

(3.31)

Pyatt's (1976) interpretation of Gini coefficient

is in terms of the expected value of a game in which each

individual is able to compare himself with some other drawn

at random from the total population. Pyatt shows that Gini

ratio can be expressed in the form

$$G = \frac{\frac{1}{n^2} \sum_{i} \sum_{j} \max(0, Y_i - Y_j)}{\frac{1}{n} \sum_{i} Y_i}$$

(3.32)

where,

$$\frac{1}{n} \sum_{j=1}^{n} \max(0, Y_i - Y_j) > 0 \text{  for all } i$$

(3.33)

is nothing but the expected gain for individual $i$, in game

theoretical sense.

Hence Gini ratio as defined in (3.32) is the

average gain to be expected from the option of being some

one else in the population, divided by the average income.

Based on this notion of average gain as defined above, a
decomposition of Gini coefficient in terms of conditional expectations of the value of the game is established. The decomposition separates out three components, within-group Gini ratios, inequality between group means and a term related to overlapping among groups. This same disaggregation has been derived previously by Bhattacharya and Mahalanobis (1967) using a more conventional approach in terms of absolute difference, while Pyatt's approach is new and unconventional; which is shown to have particular relevance to the studies of migration and discrimination.

Paglin (1975) attacked the Lorenz-Gini methodology by bringing in the concept of "age-Gini-ratio" to have true estimates of inequality. He argued in favour of changing the concept of perfect equality inherited in the usual "Lorenz-Gini methodology" and suggested a new P-reference line of equality as against the $45^\circ$ line used so far. He suggested that equality be interpreted as equal life-time earning; because every family at a given stage of its life-circle would have the same income. Therefore, the Lorenz-Gini ratio should be corrected to account for the "age-Gini ratio" to estimate the true inequality.

Accordingly,

$$\text{Paglin Gini ratio} = \text{The Lorenz Gini ratio minus the age-Gini ratio.}$$

The Paglin Gini is simply

$$\text{P.G.} = \frac{(\Delta_1 - \Delta_2')}{2\mu}$$

where, $\Delta_1 = \text{Gini coefficient of mean difference}$
and \( \Delta' \) = the mean difference of the P curve distribution

\( \mu \) = mean income of the population.

**Fig. 3.2 P CURVE FOR DISTRIBUTION**

We however feel that such revisions cannot be put to use in the current context of farm-families, as the age-Gini ratio itself contains non-age sources of inequality and thus cannot be, and should not be entirely subtracted from the Lorenz-Gini ratio.

3.2(b)-IV(b) Gini Index and Negative Incomes

The preceding section concentrates on the development of the measure Gini index, insinuating the various properties which can be attached to this widely used measure of inequality. But an important point which must draw proper attention in our study is the existence of negative entries in the income profile, called negative income, an aspect too important to be ignored while studying the measurement of deprivation. The adaptation of Gini ratio
to such an income profile is both interesting and important, because of the lack of attention being paid to it in the various studies scattered throughout the literature. Since existence of negative incomes (negative entries) in any distribution profile is a real life situation; particularly in "farm business income"\textsuperscript{27}, the application of Gini ratio as an appropriate index of inequality gets added significance, and demands our immediate attention. Accordingly, the present section is devoted to this aspect of the index.

Though some of the indices of inequality, viz. mean, and variance do admit of both negative and positive entries and their sampling distributions are also available in literature\textsuperscript{28}; yet the usual application of Gini ratio to these cases when negative entries are also present, is curious. Either the negative entries are ignored (considered to be zero) or even if they are considered; the usual definition of Gini ratio becomes inappropriate for the purpose [Ref. Chen et al.(1982)].

A remedy for these cases on the lines of Chen et al.(op cit.) is suggested here and the popularity of this measure gets enhanced, thanks to the inclusion of this new dimension and improved results. Once it is possible to consider both the positive and negative entries in a distribution profile, through the application of Gini ratio; it is also possible to compare two distribution profiles -
one involving negative entries and other having only the positive entries.

Viewed from this angle, the choice of Gini ratio as an appropriate index of inequality has an added advantage and opens new avenues for research on various aspects of inequality.

With this background we now shift to the definition of Gini ratio as applied in our study. Section 3.2b-IV(c) carries out the definitions and other preliminaries required for the purpose.

3.2b-IV(c) Certain Definitions and Preliminaries

As seen earlier, the Gini ratio is defined as

\[ G = \frac{\Delta}{2\mu} \]

where, \( \Delta = \text{popn mean difference} \)
and \( \mu = \text{popn mean} \)

But in case negative entries are present, some modification is necessary in the usual definition of Gini ratio as described below.

Let \( Y_1, Y_2, \ldots, Y_n \) be a random sample of size \( n \), drawn from an unknown income distribution \( F(Y) \). It is assumed that the distribution has a finite mean \( \mu \) and has a differentiable density. Though some of the incomes \( Y_i's \) might be negative but the total income must be positive i.e.

\[ \sum_{i=1}^{n} Y_i > 0 \text{ and hence } \bar{Y} > 0 \]  \hspace{1cm} (3.35)
It might be added that the case of $\sum_{i=1}^{n} Y_i < 0$ might throw further theoretical possibilities but this makes an economic absurdity. Hence we restrict ourselves to the realistic case of a viable economy producing positive income in the aggregate.

Let the income be ordered as

$$Y_1 \leq Y_2 \leq \cdots \leq Y_n$$  \hspace{1cm} (3.36)

Then, the Gini coefficient is estimated as

$$\hat{G} = \frac{\hat{\Delta}}{2Y}$$

where,

$$\hat{\Delta} = \frac{2}{n(n-1)} \sum_{i}^{n} \sum_{i < j}^{n} (Y_j - Y_i)$$

and

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

Define $y_i = \frac{Y_i}{n\bar{Y}}$, \hspace{1cm} i = 1 \ldots \ldots n  \hspace{1cm} (3.37)

where $y_i$ is the income share of the ith unit.

Let us so define $k$ such that

$$\sum_{i=1}^{k} y_i = 0$$  \hspace{1cm} (3.38)

Following Sen (1973) and using these two definitions viz. (3.37) and (3.38), the expression for $\hat{G}$ can be written as follows [Ref. Chen et al.(1982)]
\[
\hat{G} = \frac{1}{n-1} \sum_{i=1}^{n} \frac{1}{\varepsilon} \sum_{j<i}^{n} (y_i - y_j)
\]

\[
= \frac{1}{n-1} \left[ \sum_{i=1}^{n} y_i (2i-(n+1)) \right]
\]

\[
= \frac{1}{n-1} \left[ \sum_{i=1}^{n} y_i (n-(1+2(n-i))) \right]
\]

\[
= 1 + \frac{1}{n-1} - \frac{1}{n-1} \sum_{i=1}^{n} y_i (1+2(n-i))
\]

\[
= 1 + \frac{1}{n-1} + \frac{2}{n-1} \left( \sum_{i=1}^{k} i y_i \right) - \frac{1}{n-1} \sum_{i=k+1}^{n} y_i (1+2(n-i))
\]

\[\text{(3.39)}\]

As shown by Chen et al. (1982), such a formulation brings home the fact that \(\hat{G}\) can take values greater than one, when negative entries are present in the distribution.

Let us consider the case of extreme inequality, a situation when one family, say nth, earns all the income and the other \((n-1)\) families together earn nothing at all.

Now \(K = n-1\), \(Y_n = 1\), \(\sum_{i=1}^{n-1} y_i = 0\); and \(\hat{G}\) reduces to

\[
\hat{G} = 1 + \frac{2}{n-1} \sum_{i=1}^{n-1} iy_i
\]

\[\text{(3.40)}\]

If all incomes were non-negative, then in the case of extreme inequality \(y_i = 0\) for all \(i < n\) and \(G\) equals one.

However, in the presence of negative income, we will have

\[y_i < 0\] for all \(i < n\) and the term
Consequently $G$ exceeds one.

In general, $G > 1$ depending upon

$$\frac{1}{n-1} + \frac{2}{n-1} \left( \sum_{i=1}^{k} y_i \right) > \frac{1}{n-1} \sum_{i=k+1}^{n} y_i(1+2(n-i))$$

(3.42)

Such a discussion suggests that $\hat{G}$ may overestimate inequality, in case negative entries are present in the distribution profile. To avoid this situation, some adjustments or modifications are needed in the definition of Gini ratio when negative entries are also included in the data profile.

As proposed by Chen et al. (1982), a normalised Gini ratio $\hat{G}^*$ is defined as follows:

$$\hat{G}^* = \frac{1 + \frac{1}{n-1} + \frac{2}{n-1} \left( \sum_{i=1}^{k} y_i \right) - \frac{1}{n-1} \sum_{i=k+1}^{n} y_i(1+2(n-i))}{1 + \frac{2}{n-1} \sum_{i=1}^{k} y_i}$$

(3.43)

It can be seen [Ref. Chen et al. (1982)] that $\hat{G}^*$ retains all the basic properties of conventional Gini ratio $\hat{G}$ and lies between 0 and 1. However, $\hat{G}^*$ will be equal to $\hat{G}$ if incomes are non-negative only.

But the way $k$ has been defined in equation (3.38) its validity is limited. This, in turn, leads to the limited application of $\hat{G}^*$ to the empirical data. In fact, in most
empirical works we shall be facing situations wherein there exists a value of $k$, say $k'$, such that

$$\sum_{i=1}^{k'} y_i < 0 \text{ and } \sum_{i=1}^{k'+1} y_i > 0 \quad (3.44)$$

Thus there is a need to remodify $\hat{G}^*$ to get another version of normalised Gini ratio, say, $\hat{G}^{**}$.

In such case, following Chen et al. (1982), Berrebi & Silber (1985) and Chen et al. (1985), the entity

$$\frac{2}{n-1} \sum_{i=1}^{k'} i y_i \text{ is written as}$$

$$\frac{2}{n-1} \sum_{i=1}^{k'} i y_i + \frac{1}{n-1} \sum_{i=1}^{k'} y_i \left( \sum_{i=1}^{k'} y_i \right) - (1+2k') \quad (3.45)$$

and the normalised Gini ratio takes the following form:

$$\hat{G}^{**} = \frac{A}{B} \quad (3.46)$$

where,

$$A = 1 + \frac{1}{n-1} + \frac{2}{n-1} \sum_{i=1}^{k'} i y_i + \frac{1}{n-1} \sum_{i=1}^{k'} y_i \left[ -(1+2k') \right]$$

$$B = 1 + \frac{2}{n-1} \sum_{i=1}^{k'} i y_i + \frac{1}{n-1} \sum_{i=1}^{k'} y_i \left[ \sum_{i=1}^{k'} y_i \right] - (1+2k')$$

It can be seen easily that $\hat{G}^{**}$ also lies between 0 & 1, and in the case of positive incomes only, both $\hat{G}^{**}$ and $\hat{G}$ are the same.
Like $\hat{G}$ in case of positive income, the new versions of Gini ratio, viz., $\hat{G}^*$ and $\hat{G}^{**}$ can also be represented graphically with the help of associated Lorenz curves of income [ref. Chen et al. (1982)].

Let us suppose that we have a random sample of $n$ families, where $k$ is the number defined in the above manner, i.e.

$$\sum_{i=1}^{k} y_i = 0$$

The associated Lorenz curve is shown in fig. 3.3.

Fig. 3.3 LORENZ CURVE WITH NEGATIVE INCOME (Case I)
In this case, area under the horizontal line i.e. $A$ is

$$A = - \frac{n}{n-1} \left[ \sum_{i=1}^{k} \frac{y_i(1+2(k-i))}{2n} \right]$$

and the area under the Lorenz curve, $C$, by the same token is

$$C = \frac{n}{n-1} \left[ \sum_{i=k+1}^{n} \frac{y_i(1+2(n-i))}{2n} \right]$$

Hence, in this case

$$\hat{G}^* = \frac{1+\frac{1}{n-1} + 2(A-C)}{1+2A}$$

(Ref. to fig.3.3)

But, in case we use the other version of $\hat{G}$ i.e. $\hat{G}^{**}$, then $\hat{G}^{**}$ can be represented graphically as in fig.3.4, so that

$$\hat{G}^{**} = \frac{1+\frac{1}{n-1} + 2(A-C)}{1+2A}$$

where,

$$A = \frac{n}{n-1} \left[ \frac{1}{2ny} \sum_{k'+1}^{k'} \frac{(y_i)^2}{k'+1} - \frac{1}{2n} \sum_{i=1}^{k'} \frac{y_i(1+2(k-i))}{2n} \right]$$

and

$$C = \frac{n}{n-1} \left[ \frac{1}{2n} \sum_{i=k'+1}^{n} \frac{y_i(1+2(n-i))}{y_i(k'+1)} \right]$$

where, $k'$ is so defined that
\[
\sum_{i=1}^{k'} y_i < 0 \text{ but } \sum_{i=1}^{k'+1} y_i > 0
\]

In both the above cases, if no negative income is present, then \( A = 0 \), and

\[
\hat{G} = \hat{G}^* = \hat{G}^{**} = 1 + \frac{1}{n-1} - 2C
\]

\[
= \frac{n}{n-1} - 2C
\]

\[
= 1 + \frac{1}{n-1} - 2C
\]

(3.51)

Fig. 3.4 LORENZ CURVE WITH NEGATIVE INCOME (Case II)
After discussing the properties of Gini ratio, let us discuss in brief the remaining indices of inequality; referred to in section 3.2(b).

(v) The Extreme Mean Disparity Ratio

The measure is suggested by Iyengar and Vani (1986), and is based on the notion of, "extreme disparity ratio".

For any distribution profile, the "extreme disparity ratio" is defined to be the ratio of two extreme observations which is \( \frac{\text{maximum}}{\text{minimum}} \). Let \( F(Y) \) denote the distribution function for the r.v. \( Y \) denoting the income of \( n \) individuals.

Let

\[
Y^{(1)} \leq Y^{(2)} \leq \ldots \leq Y^{(n)}
\]  

(3.52)

Then "extreme disparity ratio" is

\[
\frac{Y^{(n)}}{Y^{(1)}}
\]

(3.53)

By definition, the extreme disparity ratio itself can be treated as a measure of inequality.

Clearly, the measure lies between 0 and 1. In case of perfect equality it will take value zero, while in case of perfect inequality it will be 1. As is obvious, the measure does not take into consideration the entire distribution. Hence, it cannot be treated as a very reliable measure of inequality. Still it is a direct measure which gives a fairly
But Iyenger & Vani (1986) suggest that in case the populations are large, it may not always be easy to estimate it. Therefore, they suggested a close lower bound to this ratio, defined in terms of extreme fractile means; which provides a lower approximation to the ratio of the extreme observations.

Let $F_1, F_2, \ldots, F_g$ be the $g$ fractile groups (subpopulations of $F$) and $\mu_1, \mu_2, \ldots, \mu_g$ be the respective fractile means as defined below:

$$P(X \in F_i) = \frac{1}{g}$$

and

$$E(X|X \in F_i) = \mu_i \quad \text{for } i = 1, \ldots, g \quad (3.54)$$

The extreme mean disparity index $R$ for the distribution may be defined as the ratio

$$R = \frac{\mu_g}{\mu_1} \quad (3.55)$$

In case, $R$ is to be estimated from sample observations it is defined as

$$\hat{R} = \frac{\hat{\mu}_g}{\hat{\mu}_1} \quad (3.56)$$

where, $\hat{\mu}_g$ and $\hat{\mu}_1$ are the sample estimates of $\mu_g$ and $\mu_1$ respectively.

The index, as defined above, can be modified further to yield better approximations to "extreme disparity ratio". This will depend on the number of fractile groups i.e $g$. An increase in the value of $g$ will result in better
approximation of the index R. But at the same time it will cause the variance of the estimated fractile means \( \mu_g \) and \( \mu_1 \) to increase because of its resulting in smaller number of observations in each fractile group. Therefore, one has to choose an optimal \( g \) to get a stable estimate of \( R \).

The mean disparity ratio, like the Lorenz curve or Gini ratio, is a statistical measure. In fact, it is the ratio of the consumption share of the highest fractile to that of the lowest fractile. Both these shares could be obtained from the empirical Lorenz curves. But in case the basic data is not available in the form of fixed fractile distributions, the available fixed-class interval data has to be interpolated or a suitable model can be found which may fit the data well.

(vi) **Theil's Entropy Measure**

A commonly used measure of inequality is based on the notion of entropy. In the context of thermodynamics the term is a measure of disorder, while in information theory, entropy is interpreted as representing expected information.

Theil (1967) employed this notion of entropy in information theory to define his measure of inequality.

Let \( y_i^* \) be the fraction of total income going to the \( i \)th unit, with the condition that \( y_i^* > 0 \).

The entropy of the income shares \( y_1^*, y_2^*, \ldots, y_n^* \) is defined as

\[
H = \sum_{i=1}^{n} y_i^* \log \frac{1}{y_i^*} \quad (3.57)
\]
which is a strictly Schur-concave function [Marshall & Olkin (1979). It is easy to verify that $H$ takes maximum value as \( \log n \) (in case of perfect equality) while the minimum value of $H$ is 0 (in case of perfect inequality).

Theil's inequality measure is obtained as

$$T = \log n - H = \sum_{i=1}^{n} y_i^* \log n y_i^*$$

while the normalised version is given by

$$T^* = \sum_{i=1}^{n} \frac{y_i^* \log n y_i^*}{\log n}$$

$T^*$ lies between 0 and 1 in case of complete equality and complete inequality respectively.

The measure is most sensitive to changes at low and high income levels and easy decomposability into between and within set inequalities adds its attractiveness as a measure of inequality. For this reason, the measure received most attention in the past and can't be dismissed as an arbitrary formula.

(vii) Atkinson's Measure

The normative measures were first introduced by Dalton (1920) and Atkinson's measure is an improvement of the measure given by the former.

Let $U_i(Y)$ be the utility function of $i$th individual which is positive, increasing and concave. Let social welfare function be the sum of individual utilities flowing from an income profile, under the condition that
each individual has the same utility function. The social welfare function would be additive, symmetric and separable.

Thus, Dalton's measure of inequality is given by

$$D = 1 - \frac{\sum_{i=1}^{n} u(Y_i)}{nu(\bar{Y})}$$

(3.60)

which amounts to the proportional loss of welfare resulting from being in a state of inequality.

But as Atkinson (1970) suggested this measure is not invariant with respect to positive linear transformations of utility function and proposed a new measure.

Atkinson's measure is derived from the concept of "equally distributed equivalent income" called $Y_{EDE}$ which defines that level of income per head which if equally distributed would give the same level of social welfare as the present distribution i.e.

$$Y_{EDE} = \frac{1}{n} \sum_{i=1}^{n} u(Y_i)$$

Then Atkinson's measure is defined as

$$A = 1 - \frac{Y_{EDE}}{\bar{Y}}$$

The measure lies between zero and one and is Schur-convex function [Ref. Marshall & Olkin (1979)]. Though the measure has intuitive appeal, yet some drawback of this measure with regard to the assumption of utility function can't be ignored. Still the measure is extensively used in the
field of income distribution.

All these measures are good and appropriate depending upon the purpose for which these are applied. All these measures can be the true reflectors of inequality; if the question regarding what is expected of inequality and what inequality means is understood properly. That is why it is customary to report the results on inequality using several indices at the same time. However, for our study we intend to rely only on the Lorenz-Gini ratio by virtue of the measure possessing a relative edge over other indices of inequality, as discussed below.

3.2(c) Comparison of Gini Ratio with other Measures of Inequality

1) Apparently one appeal of Gini ratio is inherent in the fact that it is a very direct measure of income difference, taking note of differences between every pair of incomes. This quality is found to be lacking in other measures of inequality (mentioned above) which involves some reference value (e.g. mean or median) with which the values of the random variable Y are compared\(^3\)\(^4\).

2) As this measure does not involve the arbitrary squaring procedure as in the case of variance or coefficient of variation, it does not involve the drawback of attaching more weights to high income levels. As Yule & Kendall (1958) have shown, there is a strong link between the standard deviation and the mean difference; hence concentration ratio
will suffer from the disadvantage of any measure involving standard deviation, which is highly biased towards the dispersion at the upper tail of the distribution\textsuperscript{35}.

But viewed from the practical angle, contrasting with other measures, this is not a serious mistake and can be dismissed as having no more offensive quality as other measures possess.

3) A glance at the limiting value of all the measures considered above suggests that Gini ratio is invariant to the attribute "size of a distribution", while other measures do change their values dependent upon the size of the distribution. This perhaps is one of the most important reasons for the wide applicability of this measure. Gini ratio will be either 0 or 1 in case of perfect equality or perfect inequality. Whatever the sample size may be the estimate of Gini ratio is not affected in these cases. Most of the other indices of inequality mentioned in section 3.2(b) viz. variance, coefficient of variation, relative mean deviation take the values $\sqrt{(n-1)\overline{Y}^2}$, $\sqrt{n-1}$, $2(n-1)/n$ respectively in case of perfect inequality and hence are dependent upon the sample sizes for calculating the limiting value of these indices. Gini index thus appeals more to the user as being independent of the number of units involved and consequently, appearing as a sort of standardised measure. No doubt the index "extreme mean disparity ratio" is also independent of the number of units involved in the
distribution, as it also lies between 0 and 1, in case of both perfect equality and perfect inequality. But, this measure has the major drawback of being dependent on the two extreme mean distributions and ignores the distributions in the middle. Hence, the measure cannot be preferred to Gini ratio even in this connection.

4) All the measures of income inequality, viz., the variance, the coefficient of variation, the standard deviation of logarithms and 'Gini coefficient' are sensitive to transfers of income at all income levels, though mean deviation is completely insensitive to transfers between people on the same side of mean. But on examining the relative sensitivity of these measures at different income levels, Gini coefficient seems to be best suited for the purpose. James L. McCabe (1974) pointed out that "The Kuznets Index" is more sensitive to concentration at the extreme ends of the distribution than is the "Gini coefficient". The above statement was attacked by Kondor (1975) who asserted that the "Gini coefficient" is more sensitive than the Kuznets index to extreme riches or extreme poverty - not the other way round, as stated by McCabe (op.cit.).

5) Perhaps the most popular reason for the choice of Gini index as an index of inequality is its graphical representation in terms of Lorenz curve, a curve which itself is regarded as a most popular measure of inequality.
The relationship between the two is already established in section 3.2b-IV(a), and it has become customary to talk of Gini ratio and the associated Lorenz curve. It is well established that a higher Lorenz curve (one close to the line of absolute equality) will be associated with a smaller value of Gini ratio as compared with another Lorenz curve lying below the previous one. This feature of unique ranking of distribution functions via non-intersecting Lorenz curves has also welfare interpretations as shown by Shorrocks (1983). In case, the two Lorenz curves intersect, ambiguity surrounds the inferences to be drawn about the underlying distributions, though Shorrocks (op.cit.) tries to present the ranking of distributions (in terms of inequality) through the welfare interpretations even in this case.

As mentioned earlier, we could have operated with any one or all of these measures because of their widespread use, but we opted in favour of the Gini-Lorenz ratio. It was not an easy task to choose a single index of inequality, as the ultimate choice will always, to a certain extent, be arbitrary; since no established analytical framework implicitly defines a unique concept of inequality which must be employed in a given situation. But the following features of the Gini index can be advanced as a rationale for our choice.
3.2(d) The Choice of Gini Ratio - Its Justification

Besides the desirable properties of this measure [Ref. section 3.2(c)] its familiarity and frequency of use in the study of income data has helped us in deciding in the favour of this measure. Moreover, as the empirical data pertaining to our study consists of the cultivating households, the pictorial representation of this data in terms of Lorenz curve is an added attraction for adopting this measure.

Further, another reason for our choice of this index stems from the fact that variance, coefficient of variation or the S.D. of logarithms are the well known statistical measures; whose moments and distributions have been worked out both in the parametric and nonparametric cases. There are variety of test-statistics based on these measures in both parametric and nonparametric cases. In the parametric case, the usual t-test, z-test and F-test are the standard tests available, while in case of nonparametric inference, Kolmogorov-Smirnov test, median test, sign-test, Mood test etc. are some tests relating to testing the significance of these measures. For studying income inequality, the distribution of such type of measure is discussed by many researchers such as Hart and Prais (1956), Ullah and Tiwari (1972), Motthathu (1984). But Gini index, though quite popular measure, seems to have drawn very little attention of the researchers regarding its
distribution and associated statistical inference.

The measure has been used mostly as a descriptive device, and a little attempt has been made to find the tests of significance associated with it. Some development in this field has been made mostly in the parametric case; while in the nonparametric case not much work has been carried out.

In the parametric set up, when the form of the underlying distribution is known; or assumed to be known, the asymptotic sampling distribution of Gini ratio is derived, that too in the case when the underlying distribution is positive only. Hart and Prais (1956) pointed out that if the underlying distribution is assumed to be log-normal, the Gini measure of concentration depends only upon the variance of the distribution [Aitchison and Brown (1954)]. By finding the standard error of the variance of the distribution the sampling distribution of Gini ratio could be found out. Iyengar (1960) showed that the maximum likelihood estimator (M.L.E.) of the Gini Index of the lognormal distribution is asymptotically normal. Ullah and Tewari (1972) obtained asymptotic expression for the mathematical expectation and variance of exact sampling distribution of the maximum likelihood estimator of concentration (or Lorenz ratio) when the parent distribution is lognormal. Moothathu (1985) derived the exact sampling distribution and moments of the M.L.E. of the Lorenz curve.
and Gini index of the exponential and Pareto distributions.

All these studies mentioned above explore only one aspect of the problem when it is assumed that the underlying distribution must correspond to some famous distributions such as lognormal or Pareto.

The validity of these assumptions has been a questionable proposition. If lognormal distribution is undertaken, it is to be argued whether the two-parameter lognormal is to be taken into consideration or three-parameter lognormal will be better suited for the purpose. In addition, the existence of negative entries in the distribution profile is completely ignored. So, how would one tackle the case of mixed income profiles (i.e. containing both negative and positive incomes) in the context of these favourites? Therefore, availability of nonparametric tests of significance in this generalised case holds lots of promise.

In the context of nonparametric distribution, Charles and Beach (1983) derived the results for sampling distribution of estimated Lorenz-curves and associated tests of significance. As suggested by them, the asymptotic distribution of Gini ratio can also be derived from the above results. Ramakrishnan (1984) found the asymptotic distribution of Gini ratio to be normal, in the nonparametric set-up. A mention may be made of the works of Blyn (1983) and Jain (1984). These authors based their
inference on Gini ratios in a roundabout manner by appealing to certain trends in the population. But the above results are applicable only to the positive distribution; as Gini ratio in these results is defined only for the case when no negative entry is considered.

Hence, even at the present state of the art, the problem remains: How to test statistically the significance of difference related to Gini ratio in case of single sample or two independent sample profiles, with the assumption that the underlying distribution is continuous but unknown; involving both the positive as well as negative entries?

Having made our choice in favour of the Lorenz-Gini ratio, further familiarity with this index can be gained by a comprehension of the welfare premises of this index.

3.2(e) Welfare Interpretation of Gini Ratio

The welfare aspects of this measure are interesting to visualise. The definition of Gini ratio implies a social welfare function which is just a rank order weighted sum of different people's income shares. So, the actual weights would depend upon the way the population is distributed over income sizes. Hence Gini coefficient depends not only on the size of the income levels but on the number of persons between them.

Since G is a linear function of income, it does not imply a strictly concave group welfare function. In the
light of this view it is observed that Gini ratio cannot order distributions in the same way as given by any additive group welfare, strictly concave, and differentiable individual utility functions. But Sen (1973) defends this attack by suggesting that implied group welfare function may not be strictly concave but it is concave and any transfer from the poor to the rich or vice versa is strictly recorded in the Gini coefficient in an appropriate direction.

A systematic debate regarding the welfare interpretation of this measure is found in Kakwani (1980). Though much debate is on, against and in favour of this measure, still its popularity remains undiminished till to-date.

The phenomenon of economic inequality becomes much more disturbing when coupled with high incidence of absolute deprivation. In fact, inequality and poverty are but two facets of the problem—"unjust" distribution. If inequality enhances the sense of being left-out, poverty questions the very basis of existence of the society. So, it is not merely customary to consider the two problems together but a logical compulsion for the students of the society.

Accordingly, the next chapter deals with the phenomenon of poverty, its nature and measurement.
Notes and References

(Chapter 3)


5. Such a procedure is necessarily called for if one does not wish to ignore the internal structure of farm families. When the emphasis is on consumption needs, "Adult Equivalent Scales" developed by ICMR, New Delhi, may be referred to wherein adult male = 1.0, adult female = 0.9 and child (below 16 years) = 0.6. On the other hand, if the working force aspect is being considered, it is the standardisation of male workers that is required. The prevalent practice is to consider adult male = 1.0, adult female and child working on the farm = .5.

7. The view is expressed by Bartels (1977), which supports the idea of inequality expressed by Atkinson (1975, pp 3).

8. Various terms and conditions for measures of inequality have been discussed by many authors e.g. ref. to Dalton (1920), Atkinson (1970), Champernowne (1974), Kolm (1976), Bartels (1977). Kakwani (1980) points to the non-binding nature of these conditions.

9. If a function $\phi$ is to be used as a measure of inequality then it should satisfy
   (i) $x < y \Rightarrow \phi(x) \leq \phi(y)$
   i.e. $\phi$ should be Schur-convex.
   Even more, $\phi$ should satisfy.
   (i') $x < y$ and $x$ is not a permutation of $y$
   $\Rightarrow \phi(x) < \phi(y)$,
   i.e. should be strictly Schur-convex.

The above definition of Schur-convex functions is due to Marshall & Olkin (1979); who also brings forth the relationship between various measures of income inequality and Schur-convex functions.

10. It is customary to divide the measures of inequality into two broad categories - Normative and Positive Measures. For details one may refer to Sen (1973).
11. Since \( \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \)

\[ = \sigma_d^2 \] (in discrete case)

This implies a homogeneity of order 2.

12. For discrete case, we can see that when \( Y_i = \bar{Y} \) for all \( i \)

\( \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (\bar{Y} - \bar{Y})^2 = 0 \)

when \( Y_i = 0 \) for all \( i \neq k \) and \( Y_k = n \bar{Y} \).

\[ \sigma^2 = \frac{1}{n} \left[ (0 - \bar{Y})^2 + (0 - \bar{Y})^2 + \ldots + (n\bar{Y} - \bar{Y})^2 \right] \]

\[ = \frac{n\bar{Y}^2(n-1)}{n} \]

\[ = (n-1)\bar{Y}^2 \]

13. Kakwani (1980) in this context, points out the drawback of this measure in case one wishes to attach more weight to transfers at the lower end of the distribution than at the top.

14. When \( Y_i = 0 \) for \( i \neq k \) and \( Y_k = n\bar{Y} \),

\[ R.M.D. = \frac{1}{n\bar{Y}} [(n-1)\bar{Y} + (n-1)\bar{Y}] \]

\[ = \frac{2(n-1)}{n} . \]

15. It was called "maximum equalisation percentage" by the UN Economic Commission for Europe, 1957. Also refer to Rosenbluth (1951), Schutz (1951), Kuznets (1988), Mehran (1976).
19. Bartels's "Generalised weighted difference indicator", is defined as

\[ GWD = \left[ \sum_{i,j} |Y_i - Y_j|^P w(Y_i, Y_j) \right]^{1/P} , \quad P \geq 1 \]

where \( w(Y_i, Y_j) \) is a weight function representing weights associated with the pairs of incomes compared. In case the weight function is one i.e. \( w(Y_i, Y_j) = 1 \) and \( P = 1 \), 
GWD reduces to mean difference.

20. Gini proposed this measure in 1912, vide variabilita's mutabilita, Bologna, 1912. The reason for mentioning the Gini coefficient of concentration is that it can be demonstrated that "twice the concentration surface" equals Gini ratio, where "concentration surface" is the area between the Lorenz curve and the line of absolute equality. For proof ref. Kendal & Stuart (1958), pp 49.


22. As Dalton (1920) puts it, "If a single measure is to be used, the relative mean difference is,
perhaps, slightly preferable, owing to the graphical convenience of the Lorenz curve.

23. The proof was given first, apparently, by Professor Gini. Compare with Kendall and Stuart (1958), pp 49.


25. Atkinson (1970) does point out towards this drawback saying, "It is not clear that such a weighting would necessarily accord with social values."

26. According to this principle, "a small positive transfer from a richer to a poorer individual, with a given proportion of the population in between them, decreases the inequality and the decrease is larger the poorer the recipient". Ref. Mehran (1976).


28. One may refer to Hart and Prais (1956), Ullah & Tiwari (1972), Moothathu (1985) to name a few.

29. In the case of equally distributed incomes $y_i = 1/n$ for all $i$ and $G^*$ reduces to $G$, taking a value zero. And, in the event of extreme inequality
\[
G^* = \frac{1 + \frac{1}{n-1} + \frac{2}{n-1} \sum_{i=1}^{n-1} iy_i - \frac{1}{n-1}}{1 + \frac{2}{n-1} \sum_{i=1}^{n-1} iy_i}
\]
which equals one.


31. This property is caused by the additivity condition underlying the definition of this class of information measures. Ref. Theil (op. cit) and Bartels (1977).

32. Sen (1973) criticises this measure as being based on an arbitrary formula; but Bartels (1977) has suggested a number of entropy measures due to wide applicability of entropy as a measure of inequality e.g. Renyi's (1965) \( \alpha \)-order entropy measure.


34. For details one may refer to Bartels (1977), Sen Sen (1973).

35. Stark (1972).

36. Schutz (1951) argued that the relative mean deviation is preferable to the Gini coefficient. He pointed out that the slope of the Lorenz curve may be infinitely varied without any change in Gini coefficient.

It is defined as

\[ k = \frac{1}{n} \left[ \sum_{i=1}^{k} \left( \frac{y_i}{M-1} \right) - \sum_{i=k+1}^{n} \left( \frac{y_i}{M-1} \right) \right] \]

for \( y_1 \geq y_2 \geq \ldots \geq y_n \) and \( y_{k+1} \leq M \leq y_k \)

where \( y_i \) = total income of \( i \)th group

\( M \) = Mean income for all groups.

\( n \) = Number of income groups with equal number of persons.

The same view is held by many others e.g. refer Stark (1972), Sen (1973), Farbman (1975), Bartels (1977).

Atkinson (1970) also points out the aspect of unique ranking of two distributions of income in case their Lorenz curves don't intersect. Some empirical results are also quoted by him to elaborate his point.

In general, the practice is to comment upon the inequality by looking at the size of the index only and by looking at the shape of the corresponding Lorenz curve. A higher index is supposed to contain more inequality and smaller one is associated with small inequality.