On Truncated Versions of Certain Measures of Inequality and Stability

Thesis submitted to the

Cochin University of Science and Technology

in fulfillment of the requirements for

the award of the degree of

Doctor of philosophy

Under the

FACULTY OF SCIENCE

By

Mr. Abdul Sathar E.I.

DEPARTMENT OF STATISTICS

COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY

KOCHI- 682 022

December 2002

Dr. K.R.Muraleedharan Nair
M. Sc., M.Phil, Ph. D
Professor

Phone Off. 0484-575893
Phone Res. 0484-332078
Email: krm@vsnl.net

Department of Statistics
COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY
Kochi - 682 022, India

CERTIFICATE

This is to certify that the thesis entitled "**On Truncated Versions of Certain Measures of Inequality and Stability**", which is being submitted by **Mr. Abdul-Sathar E.I.**, in fulfillment of the requirements of the degree of Doctor of Philosophy, to the Cochin University of Science and Technology (CUSAT), Kochi is a record of the bonafide research work carried out by him under my guidance and supervision.

Mr. Abdul-Sathar has worked on this research problem for about three years and six months in the Department of Statistics of CUSAT. In my opinion the thesis has fulfilled all the requirements according to the regulation and has reached a standard necessary for submission. The results embodied in this thesis have not been submitted for any other degree or diploma.

Cochin-22
19.12.2002

K.R.Muraledharan Nair
(Supervising Teacher)

Contents

Chapter I

Review of Literature

1.1 Introduction

The problem of modelling income data as well as that of measurement of inequality in the income of members of a group or a society has a history of about two hundred years and has been attracting a lot of researchers in Economics, Statistics, Sociology etc. As is customary in most statistical analysis, the extend of variation in incomes is represented in terms of certain summary measures. Thus a measure of income inequality is designed to provide an index that can abridge the variations prevailing among the individuals in a group.

Although there had been many attempts to provide measures of income inequality in the nineteenth century, the first major development in this area can be attributed to the work of M.O.Lorenz in 1905. A measure of income inequality is provided through a graphical representation of incomes by plotting a curve with co-ordinates $(p, L(p))$, where $L(p)$ represents the percentage of the total income of the population accruing to the poorest p percent of the population. For different data, a comparison of inequality of income shall be accomplished from the nature of the Lorenz curve. Subsequently, Gini (1912) proposed a measure of income inequality, which is defined as twice the area between the Lorenz curve and the line of equal distribution. Although different measures of income inequality such as coefficient of variation, relative mean deviation, mean deviation, standard deviation of logarithms of incomes and some entropy indices has been suggested in literature, the Gini-index still enjoys an important role in the context of measurement of income inequality. For a detailed study on various measures of income inequality we refer to Kakwani (1980), Anand (1983) and Arnold (1987).

For statistical or administrative reasons, many surveys of income are truncated at the lower end of the income range. Since much of the data on incomes comes from income tax returns and most countries have a threshold below which no tax is levied, someone known or suggested to have a low income is much less likely to file a tax return than a person with high earnings. Hence the importance of studying inequality measures of truncated distributions is much of interest. The effect of truncation of the distribution upon the various measures of income inequality had been a theme of recent interest among researchers. Bhattacharya (1963) showed that the Lorenz curve of a left truncated distribution is independent of the point of truncation if and only if the distribution is Pareto.The right truncation case was studied by Moothathu (1986) who showed that the Lorenz curve is independent of the point of truncation if and only if the distribution is a power function distribution. Ord, Patil and Taillie (1983) examined the effects of truncation upon some derived measures of inequality and it is shown that only for the Pareto distribution are the measures invariant with respect to truncation. Dancelli (1990) has looked into the effects of the truncation upon the Zenga curve and the Zenga index and makes some numerical studies of the effect of truncation in the Dagum model type-1 distribution. Further some results connected with ordering of distributions in the context of truncation have been obtained. Ahmed (1988) studied a partial ordering for life distributions based on the mean residual life. Mailhot (1990) studied some conditions to obtain ordering of truncated distributions. Belzunce, Candel and Ruiz (1995) has looked into the problem of ordering of truncated distributions using the Lorenz and Zenga curve of concentration.

Recently concepts and ideas from Reliability theory has been extensively used to study measures of inequality. Chandra and Singpurwalla (1981) pointed out few relationships between some notions that are common to Reliability theory and Economics in the context of measuring inequality. These aspects were further investigated by Klefsjo (1984). Further Bhattacharjee (1993) stress the role of anti-aging distributions in Reliability theory as reflecting certain features of skewness and heavy tails, typical of wealth distributions.

Shannon's entropy [Shannon (1948)] has been extensively used in literature as a qualitative measure of the uncertainty associated with a random phenomenon. Further the entropy indices have been advantageously used as measures of income inequality, as pointed out in section 1.6. In the Reliability context, concepts such as failure rate or the mean residual life function comes up as a handy tool to describe the failure pattern of a component or device. Observing that highly uncertain components are inherently not reliable, recently, Ebrahimi and Pellery (1995) has used the Shannon's entropy associated with the residual life, referred to in literature as the residual entropy function, as a measure of the stability of a component or a system.

Motivated by this, the present study focuses attention on

(i) defining certain measures of income inequality for the truncated distributions and characterization of probability distributions using the functional form of these measures.

(ii) extension of some measures of inequality and stability to higher dimensions.

(iii) characterization of some bivariate models using the above concepts.

(iv) estimation of some measures of inequality using the Bayesian techniques.

1.2 Basic concepts in Reliability.

In the present section we give a brief review of the basic concepts and results in Reliability theory, which are of use in the sequel and are referred to in the text. The commonly used concepts in Reliability theory are (i) the survival function (ii) the failure rate

(iii) the mean residual life function and (iv) the vitality function. The definitions are reproduced below.

Let X be a non-negative random variable defined on a probability space $(\Omega, \Im, P)$ with distribution function $F(x) = P(X \leq x)$. In the Reliability context, X generally represents the lifetime of a device measured in units of time. The function

$$\begin{aligned}\bar{F}(x) &= P(X > x) \\ &= 1 - F(x) \qquad (1.2.1)\end{aligned}$$

is called the survival (Reliability) function, which indicates the probability of failure free operation of the devise up to time x. One major problem of interest in Reliability analysis is that of the determination of the functional form of the survival function.

In the bivariate case, if $X = (X_1, X_2)$ is a non-negative random vector admitting an absolutely continuous distribution function $F(x_1, x_2)$ with respect to Lebesgue measure, the survival function of X is defined as

$$\bar{F}(x_1, x_2) = P[X_1 > x_1,\ X_2 > x_2]. \qquad (1.2.2)$$

(1.2.2) represents the probability of failure free operation of a two-component system up to time $(x_1,\ x_2)$. Also we have

$$\bar{F}(x_1,\ x_2) = 1 - F_1(x_1) - F_2(x_2) + F(x_1, x_2)$$

where $F_i(x_i)$ is the distribution function of $X_i, i=1,2$. Further the density function of X is given by

$$f(x_1, x_2) = \frac{\partial^2 \bar{F}(x_1, x_2)}{\partial x_1 \partial x_2}. \qquad (1.2.3)$$

For the random vector X considered above, it is of special interest to consider the conditional distribution of X_i given $X_j > t_j, i, j = 1,2,\ i \neq j$. In a life testing experiment, if (X_1, X_2) represents the lifetimes of the components in a two component system the above conditional distribution focuses attention on the distribution of the i^{th} component subject to the condition that the other has survived up to time t_j. The survival function of X_1 given $X_2 > t_2$ takes the form

$$\bar{F}\left(t_1 \mid X_2 > t_2\right) = \frac{\bar{F}(t_1, t_2)}{\bar{F}_2(t_2)} \tag{1.2.4}$$

where $\bar{F}_i(t_i) = P(X_i > t_i)$, $i = 1,2$.

Also we have

$$\bar{F}(t_1, t_2) = \bar{F}\left(t_1 \mid X_2 > t_2\right) \bar{F}_2(t_2).$$

Differentiating with respect to t_1 we get

$$\frac{\partial \bar{F}(t_1, t_2)}{\partial t_1} = -f\left(t_1 \mid X_2 > t_2\right) \bar{F}_2(t_2)$$

so that

$$f\left(t_1 \mid X_2 > t_2\right) = \frac{-1}{\bar{F}_2(t_2)} \frac{\partial \bar{F}(t_1, t_2)}{\partial t_1}. \tag{1.2.5}$$

Failure rate

Defining the right extremity, L, of $F(x)$ by

$$L = \text{Inf}\left\{ x : F(x) = 1 \right\},$$

the failure rate $h(x)$ of X, when $F(x)$ is absolutely continuous with respect to Lebesgue measure with probability density function $f(x)$, is defined for $x < L$ by

$$\begin{aligned} h(x) &= \lim_{u \to 0^+} \frac{P\left[\, x < X < x + u \,\right]}{u} \\ &= \frac{f(x)}{\bar{F}(x)} \\ &= -\frac{d \log \bar{F}(x)}{dx}. \end{aligned} \tag{1.2.6}$$

For a random variable X defined on the entire real line, Kotz and Shanbhag (1980) defines the failure rate as the Radon-Nikodym derivative with respect to Lebesgue measure on $\{x, F(x) < 1\}$ of the hazard measure,

$$H(B) = \int_B \frac{dF}{1-F(x)}$$

for every Borel set B of $(-\infty, L)$. Further the distribution of X is uniquely determined by the failure rate through relationship

$$\bar{F}(x) = \prod_{u<x} (1-H(u)) \exp(-H_c(-\infty, c)) \qquad (1.2.7)$$

where H_c is the continuous part of H. When X is non-negative and has an absolutely continuous distribution function, (1.2.7) reduces to

$$\bar{F}(x) = \exp\left(-\int_0^x h(t)\, dt\right). \qquad (1.2.8)$$

In view of (1.2.8) $h(x)$ determines the distribution uniquely. Also the constancy of $h(x)$ is characteristic to the exponential model [Galambos and Kotz (1978)]. Further Mukherjee and Roy (1986) has established that for a non-negative random variable X in the support of the set of non-negative real numbers, a failure rate of the form

$$h(x) = \frac{1}{a\,x + b} \qquad (1.2.9)$$

is characteristic to

(i) the exponential distribution with survival function

$$\bar{F}(x) = e^{-\lambda x}, x \geq 0, \lambda > 0. \qquad (1.2.10)$$

(ii) the Pareto distribution with survival function

$$\bar{F}(x) = \left(\frac{\alpha}{x+\alpha}\right)^{\beta}, x \geq 0, \beta > 1, 0 < \alpha < \infty \qquad (1.2.11)$$

(iii) and the finite range distribution with survival function

$$\bar{F}(x) = \left(1-\frac{x}{R}\right)^{c}, 0 < x < R, c > 1 \qquad (1.2.12)$$

according as $a = 0, a > 0$ and $a < 0$.

The concept of failure rate has been extended to higher dimensions. One of the main problems encountered in generalizing a univariate concept to higher dimensions is that it cannot be done in a unique manner. Where as Basu (1971) defines the failure rate for a two dimensional random vector as a scalar, Johnson and Kotz (1975) defines the same as a vector. Assuming that (X_1, X_2) represents the lifetime of the components in a two-component system, Basu (1971) defines the failure rate as

$$a(x_1,x_2) = \frac{f(x_1,x_2)}{\bar{F}(x_1,x_2)} \text{ for } x_i > 0 \;, i=1,2. \qquad (1.2.13)$$

Basu (1971) has further shown that $a(x_1, x_2)$ is a constant independent of x_1 and x_2, if and only if (X_1, X_2) is distributed as a bivariate exponential distribution with exponential marginals. One of the main draw back of this definition is that the bivariate failure rate does not determine the distribution uniquely.

A second approach to the concept of bivariate failure rate is provided by Johnson and Kotz (1975) who define it as the vector valued function

$$h(x_1,x_2) = (h_1(x_1,x_2),\; h_2(x_1,x_2)) \qquad (1.2.14)$$

where

$$h_j(x_1,x_2) = \frac{-1}{\bar{F}(x_1,x_2)} \frac{\partial \bar{F}(x_1,x_2)}{\partial x_j} \;, j=1,2. \qquad (1.2.15)$$

When the components $h_i(x_1,x_2)$ exist and are continuous in an open set containing $R_2^+ = \{ (x_1,x_2) \mid x_i > 0 \;, i=1,2 \}$, Galambos and Kotz (1978) has established that

$$\bar{F}(x_1,x_2) = \exp\left(- \int_0^{x_1} h_1(t_1,0)\, dt_1 - \int_0^{x_2} h_2(x_1,t_2)\, dt_2 \right) \qquad (1.2.16)$$

or alternatively

$$\bar{F}(x_1,x_2) = \exp\left(- \int_0^{x_2} h_1(0,t_2)\, dt_2 - \int_0^{x_1} h_2(t_1,x_2)\, dt_1 \right) \qquad (1.2.17)$$

as an extension of the one-dimensional relationship (1.2.8). Thus the vector $h(x_1, x_2)$ uniquely determines the distribution of X through (1.2.16) and (1.2.17).

Mean residual life function

The mean residual life function (MRLF) represents the average lifetime remaining for a component, which has survived up to time x. For a continuous random variable X with $E(X) < \infty$, the mean residual life function is defined as the Borel measurable function

$$r(x) = E(X - x \mid X \geq x) \qquad (1.2.18)$$

for all x such that $P(X \geq x) > 0$. If X is a random variable admitting an absolutely continuous distribution $r(x)$ can also be written as

$$r(x) = \frac{1}{\bar{F}(x)} \int_x^\infty \bar{F}(t)\, dt. \qquad (1.2.19)$$

The following relationship between the failure rate and the mean residual life function is immediate.

$$h(x) = \frac{1 + r'(x)}{r(x)}. \qquad (1.2.20)$$

Also the mean residual life function determines the distribution uniquely through the relationship

$$\bar{F}(x) = \frac{r(0)}{r(x)} \exp\left(- \int_0^x \frac{dt}{r(t)} \right) \qquad (1.2.21)$$

for every x in $(0, L)$. A set of necessary and sufficient condition for $r(x)$ to be a mean residual life function, given by Swartz (1973) is that, along with (1.2.21), the following conditions hold.

(i) $r(x) \geq 0$

(ii) $r(0) = E(X)$

(iii) $r'(x) \geq -1$

(iv) $\int_0^\infty \frac{dx}{r(x)}$ diverges.

Cox (1972) has established that the mean residual life function is a constant for the exponential distribution. Mukherjee and Roy (1986) observed that a relation of the form

$$r(x)\, h(x) = k \qquad (1.2.22)$$

where k is a constant, holds if and only if X follows the exponential distribution specified by (1.2.10) when $k = 1$, the Pareto distribution specified by (1.2.11) when $k > 1$ and the finite range distribution specified by (1.2.12) when $0 < k < 1$. The Pareto case is also discussed in Sullo and Rutherford (1977). In view of (1.2.20), Hitha (1991) has observed that a linear mean residual life function of the form

$$r(x) = a\,x + b \qquad (1.2.23)$$

is characteristic to the exponential distribution specified in (1.2.10) if $a = 0$, the Pareto distribution specified by (1.2.11) if $a > 0$, and the finite range distribution specified by (1.2.12) if $a < 0$.

As a natural extension of the mean residual life function, Buchanan and Singpurwalla (1977) defines the bivariate mean residual life function as

$$g(x_1, x_2) = \frac{\int_0^\infty\int_0^\infty P(X_1 > x_1 + t_1, X_2 > x_2 + t_2)}{\bar{F}(x_1, x_2)}, \; X_i > 0, i = 1,2. \qquad (1.2.24)$$

Although $g(x_1, x_2)$ seems to be a reasonable and direct extension, it does not share the most essential property of the univariate MRLfunction, viz, that, it should determine the corresponding distribution function uniquely.

A second definition for the bivariate mean residual life function is provided in Shanbhag and Kotz (1987) and Arnold and Zahedi (1988). Let $X = (X_1, X_2)$ be a random vector defined on R_2^+ with joint distribution function $F(x_1, x_2)$ and $L = (L_1, L_2)$ be a vector of extended real numbers such that $L_i = \inf\{x \mid F_i(x_i) = 1\}$ where $F_i(x_i)$ is the distribution function of $X_i, i=1,2$. Further let $E(X_i) < \infty, i=1,2$.

The vector valued Borel measurable function $r(x_1,x_2)$ on R_2 defined by

$$\begin{aligned} r(x_1,x_2) &= E(X-x \mid X \geq x) \\ &= (r_1(x_1,x_2)\,,\ r_2(x_1,x_2)) \end{aligned} \tag{1.2.25}$$

for all $x = (x_1,x_2) \in R_2^{+}, x_i < L_i$, $i=1,2$, such that $P(X>x)>0$ and $X \geq x$ implies $X_i \geq x_i, i=1,2$ is called the bivariate mean residual life function (BVMRLF). When (X_1, X_2) is continuous and non-negative the components of the BVMRLF are given by

$$\begin{aligned} r_1(x_1,x_2) &= E(X_1-x_1 \mid X \geq x) \\ &= \frac{1}{\bar{F}(x_1,x_2)} \int_{x_1}^{\infty} \bar{F}(t,x_2)\,dt \end{aligned} \tag{1.2.26}$$

and

$$\begin{aligned} r_2(x_1,x_2) &= E(X_2-x_2 \mid X \geq x) \\ &= \frac{1}{\bar{F}(x_1,x_2)} \int_{x_2}^{\infty} \bar{F}(x_1,t)\,dt. \end{aligned} \tag{1.2.27}$$

It is established that $r(x_1,x_2)$ determine the distribution of X uniquely. The unique representation of the survival function in terms of $r(x_1,x_2)$ is provided in Nair and Nair (1988) as

$$\bar{F}(x_1,x_2) = \frac{r_1(0,0)\ r_2(x_1,0)}{r_1(x_1,0)\ r_2(x_1,x_2)} \exp\left[-\int_0^{x_1} \frac{dt}{r_1(t,0)} - \int_0^{x_2} \frac{dt}{r_2(x_1,t)}\right] \tag{1.2.28}$$

or alternatively

$$\bar{F}(x_1,x_2) = \frac{r_2(0,0)\ r_1(0,x_2)}{r_2(0,x_2)\ r_1(x_1,x_2)} \exp\left[-\int_0^{x_2} \frac{dt}{r_2(0,t)} - \int_0^{x_1} \frac{dt}{r_1(t,x_2)}\right]. \tag{1.2.29}$$

The BVMRL function in (1.2.25) and the bivariate failure rate in (1.2.14) are connected through the relationship

$$h_i(x_1,x_2) = \frac{1 + \dfrac{\partial r_i(x_1,x_2)}{\partial x_i}}{r_i(x_1,x_2)}, i = 1,2. \tag{1.2.30}$$

A necessary and sufficient condition for a vector valued function $r(x_1,x_2)$ to be BVMRLF are

(i) $r_i(x_1,x_2) \geq 0$

(ii) $r_i(0,0) = E(X_i)$

(iii) $\dfrac{\partial r_i(x_1,x_2)}{\partial x_i} \geq -1 \;, i=1,2$

(iv) $\int_0^\infty \dfrac{dy}{r_1(y,x_2)}$ and $\int_0^\infty \dfrac{dy}{r_2(x_1,y)}$ diverge.

The result of Mukherjee and Roy (1986) has been generalized by Roy (1989). It is established that a relationship of the form

$$h_i(x_1,x_2)\; r_i(x_1,x_2) = c \;, i=1,2 \qquad (1.2.31)$$

is characteristic to

(i) the Gumbels bivariate exponential distribution with survival function

$$\bar{F}(t_1,t_2) = e^{-\lambda_1 t_1 - \lambda_2 t_2 - \theta t_1 t_2} \;, \lambda_1,\lambda_2 > 0 \;, t_1,t_2 > 0 \;, 0 \leq \theta \leq \lambda_1\lambda_2 \qquad (1.2.32)$$

if $c = 1$

(ii) the bivariate Pareto type-II distribution with survival function

$$\bar{F}(t_1,t_2) = (1 + a_1 t_1 + a_2 t_2 + b t_1 t_2)^{-c} \;, t_1,t_2 > 0 \;, a_1,a_2,c > 0 \;, 0 \leq b \leq (c+1)\, a_1 a_2 \qquad (1.2.33)$$

if $c > 1$ and

(iii) the bivariate finite-range distribution with survival function

$$\bar{F}(t_1,t_2) = (1 - p_1 t_1 - p_2 t_2 + q t_1 t_2)^{d} \;, 0 < t_1 < \frac{1}{p_1}, 0 < t_2 < \frac{1-p_1 t_1}{p_2 - q t_1} \;, p_1, p_2 > 0 \;, 1-d \leq \frac{q}{p_1 p_2} \leq 1, d > 0 \qquad (1.2.34)$$

if $c < 1$.

Sankaran (1992) has proved that a relationship of the form

$$r_i(x_1,x_2) = A x_i + B_i(x_j) \;, i,j=1,2 \;, i \neq j \qquad (1.2.35)$$

where, $B_i(x_j)>0$ for all $x_j>0$ holds if and only if X is distributed as (1.2.32) when $A=0$, the Pareto distribution specified by (1.2.33) when $A>0$ and the finite range distribution specified by (1.2.34) when $A<0$.

Vitality function

The concept of vitality function was introduced by Kupka and Loo (1989). For a non-negative random variable X admitting an absolutely continuous distribution function, the vitality function is defined as the B-measurable function defined on the real line given by

$$\begin{aligned} m(x) &= E(X \mid X \geq x) \\ &= \frac{1}{\bar{F}(x)} \int_x^{\infty} t\, dF(t). \end{aligned} \qquad (1.2.36)$$

The vitality function satisfies the following properties.

(i) $m(x)$ is non-decreasing and right continuous on $(-\infty, L)$

(ii) $m(x) \geq x$ for all $x<L$

(iii) $\lim_{x\to L^-} m(x) = L$

(iv) $\lim_{x\to -\infty} m(x) = E(X)$

Moreover, $m(x)$ is related to $r(x)$ through the relationships

$$m(x) = x + r(x) \qquad (1.2.37)$$

and

$$m'(x) = r(x)\, h(x). \qquad (1.2.38)$$

In the bivariate case, let $X = (X_1, X_2)$ be a random vector in the support of $\{ (x_1,x_2) \mid a_i \leq x_i \leq b_i, i=1,2 \}$ for $a_i \geq -\infty$ and $b_i \leq +\infty$ with survival function $\bar{F}(x_1,x_2)$. For values of $x_i < b_i$ such that $P(X \geq x) > 0$ and $X_i^+ = \max(0, X_i)$ satisfying $E(X_i^+) < \infty$, Sankaran and Nair (1991) defines the bivariate vitality function as the vector

$$m(x_1,x_2) = (m_1(x_1,x_2), m_2(x_1,x_2)) \qquad (1.2.39)$$

where

$$m_i(x_1,x_2) = E\,(X_i \mid X_1 \ge x_1\,, X_2 \ge x_2)\ ,i=1,2. \qquad (1.2.40)$$

In a two-component system, where the life lengths of the components are X_1 and X_2 (which are non-negative), $m_1(x_1,x_2)$ measures the expected age at failure of the first component as the sum of the present age x_1 and the average lifetime remaining to it, given the survival of the second at age x_2. A similar interpretation can be given to $m_2(x_1,x_2)$. Also we have

$$m_1(x_1,x_2) = x_1 + \frac{1}{\overline{F}(x_1,x_2)} \int_{x_1}^{b_1} \overline{F}(t_1,x_2)\, dt_1 \qquad (1.2.41)$$

and

$$m_2(x_1,x_2) = x_2 + \frac{1}{\overline{F}(x_1,x_2)} \int_{x_2}^{b_2} \overline{F}(x_1,t_2)\, dt_2 \qquad (1.2.42)$$

The following relationship is immediate

$$m_i(x_1,x_2) = x_i + r_i(x_1,x_2)\ ,i=1,2. \qquad (1.2.43)$$

In view of (1.2.43) and (1.2.28), $\overline{F}(x_1,x_2)$ is uniquely determined from the bivariate vitality function. Also, the bivariate failure rate $h(x_1,x_2)$ given in (1.2.14) is related to $m(x_1,x_2)$ through the relationship

$$h_i(x_1,x_2) = \frac{\frac{\partial}{\partial x_i}\, m_i(x_1,x_2)}{\left(m_i(x_1,x_2)-x_i\right)}\ ,\ i=1,2. \qquad (1.2.44)$$

or

$$\frac{\partial}{\partial x_i}\, m_i(x_1,x_2) = h_i(x_1,x_2)\, r_i(x_1,x_2)\ ,i=1,2. \qquad (1.2.45)$$

1.3 The Lorenz Curve

To compare the distribution of income of a country at different periods of time or of different countries at the same time, Lorenz (1905) introduced an approach, later termed as the Lorenz curve, which simultaneously takes into account the changes in income and population.

Let X be a non-negative random variable admitting an absolutely continuous distribution function $F(x)$, with finite mean μ. The Lorenz curve $L(p)$ of X is defined in terms of two parametric equations in x [Kendall and Stuart (1977)] namely

$$p = F(x) = \int_0^x f(t)\, dt$$

and

$$L(p) = F_1(x) = \frac{1}{\mu} \int_0^x t\, f(t)\, dt. \qquad (1.3.1)$$

$L(p)$ determined by (1.3.1) is called 'the standard Lorenz curve'.

$F(x)$ can be interpreted as the proportion of individuals having income less than or equal to x. $F_1(x)$ can be viewed as the proportional share of the total income of individuals having an income less than or equal to x. It follows from (1.3.1) that the Lorenz curve is the first moment distribution function of $F(x)$. It may be noticed that both F(x) and $F_1(x)$ lies between zero and one and the Lorenz curve being the plot of the points $(F(x), F_1(x))$ is represented in the unit square. $L(p)$ can be interpreted as the proportion of the total wealth owned by the poorest p^{th} fraction of the population. The Lorenz curve defined by (1.3.1) satisfies the following properties.

(i) $L(0) = 0, L(1) = 1$, $L(p)$ is continuous and strictly increasing on $(0,1)$, as

$$L'(p) = \frac{1}{\mu}\, x > 0.$$

(ii) $L(p)$ is twice differentiable and is strictly convex on $(0,1)$ as

$$L''(p) = \frac{1}{\mu\, f(x)} > 0.$$

Gastwirth (1971) gave a general definition of the Lorenz curve. For any non-negative random variable X with distribution function $F(x)$ and a finite mean μ, the Lorenz curve $L(p)$ is defined as

$$L(p) = \frac{1}{\mu}\int_0^p F^{-1}(t)\, dt\,,\ 0 \le p \le 1 \qquad (1.3.2)$$

where $F^{-1}(t) = \inf_x \{x : F(x) \ge t\}$ is the left continuous inverse of $F(x)$ (also known as the quantile function).

Thompson (1976) has proved the following properties for the Lorenz curve defined by (1.3.2).

(i) $L(p)$ is continuous, has a left derivative and is convex on $[0,1]$

(ii) $L(p) \le p$ and equality holds if and only if F places all its probability mass at one point.

(iii) Given a convex, non-decreasing function $g(p)$ on $[0,1]$, which satisfies $g(0)=0$, and $g(1)=1$, there is a distribution function for which $g(p)$ is the Lorenz curve.

(iv) $\dfrac{E|X-\mu|}{2\mu} = F(\mu) - L(\mu) = \max_y \left[F(y) - L(y)\right]$

(v) $\dfrac{E|X-m|}{2\mu} = F(m) - L(m) = \dfrac{1}{2} - L(m)$, where m is the median of income.

Kakwani and Podder (1976) introduced a new co-ordinate system for the Lorenz curve. Consider the standard Lorenz curve and let P be a point on this Lorenz curve with co-ordinates (F, F_1). Now define

$$\eta = \frac{(F - F_1)}{\sqrt{2}}$$

and

$$\pi = \frac{(F + F_1)}{\sqrt{2}}.$$

Then η is the length of the perpendicular line on the egalitarian line from P and π is the distance from origin (0,0) to the foot of the above perpendicular line on the egalitarian line. With this co-ordinate system, they considered the following Lorenz curve

$$\eta = a\pi^{\alpha}\left(\sqrt{2} - \pi\right)^{\beta} \tag{1.3.3}$$

where $a > 0, 0 \le \alpha \le 1, 0 \le \beta \le 1$ are parameters.

Many authors have extended the concept of Lorenz curve to higher dimensions. Taguchi (1972a) defined the 'concentration surface' of a two dimensional random vector (X, Y) having a continuous density function $f(x,y)$ and having non-zero finite mean values μ_x and μ_y for X and Y respectively, by the following implicit function

$$L(p_1, p_2, p_3) = 0 \tag{1.3.4}$$

where

$$p_1 = \int_{-\infty}^{y}\int_{-\infty}^{x} f(u,v)\, du\, dv$$

$$p_2 = \frac{1}{\mu_x}\int_{-\infty}^{y}\int_{-\infty}^{x} u\, f(u,v)\, du\, dv$$

and

$$p_3 = \frac{1}{\mu_y}\int_{-\infty}^{y}\int_{-\infty}^{x} v\, f(u,v)\, du\, dv. \tag{1.3.5}$$

He proved that the transformations (1.3.5) provides a one-to-one correspondence between (x, y) and (p_1, p_2, p_3). Hence the concentration surface defined by (1.3.4) can always be expressed as a single-valued explicit function

$$p_3 = L(p_1, p_2) \tag{1.3.6}$$

Taguchi (1972b) extended the notion of concentration surface to complete surface, which he called as the Lorenz manifold. Arnold

(1987) introduced the following definition, which is much easier to handle. The Lorenz-Arnold surface of F is the graph of the function

$$L(F,s,t) = \frac{\int_0^{\xi}\int_0^{\eta} x_1 x_2 \, dF(x_1,x_2)}{\int_0^{\infty}\int_0^{\infty} x_1 x_2 \, dF(x_1,x_2)}, \qquad (1.3.7)$$

where

$$s = \int_0^{\xi} dF^1(x_1), \; t = \int_0^{\eta} dF^2(x_2), \; 0 \le s,t \le 1,$$

F^1 and F^2 being the marginals of F.

The drawback of above definitions is that neither Arnold's nor Taguchi's definition has an economic interpretation. Koshevoy and Mosler (1996) have provided an extension of the usual Lorenz curve of the univariate distribution to the multivariate case, which does have an economic interpretation. For a given probability distribution in non-negative d space, $d \ge 1$, they define and investigate the Lorenz zonoid and the Lorenz surface, which are sets in $(d+1)$space. The surface equals the usual Lorenz curve when $d=1$. They interpreted the Lorenz surface as the endowments of economic units in d commodities.

1.4 The Gini-index

For a non-negative random variable with distribution function $F(x)$ and a finite mean μ, the Gini-index [Gini, (1912)] is defined in terms of mean difference as

$$G = \frac{1}{2\mu} \int\int |x-y| \, dF(x) \, dF(y). \qquad (1.4.1)$$

As a function of Lorenz curve it can also be defined as [Frosin, (1988)] twice the area between the Lorenz curve and the diagonal segment joining the points $(0,0)$ and $(1,1)$. That is

$$G = 1-2\int_0^{\infty} F_1(x) \, dF(x) \qquad (1.4.2)$$

or

$$G = 1 - 2 \int_{0}^{1} L(p)\, dp. \qquad (1.4.3)$$

The line segment joining the points (0,0) and (1,1) is known as line of equal distribution or egalitarian line. The value of G lies between 0 and 1,with $G=0$ representing perfect equality and $G=1$ representing perfect inequality. The Gini-index is also referred to literature under the names, coefficient of concentration, Lorenz concentration ratio, and the Gini-coefficient.

Chakrabarthy (1982) points out that the analysis and criticism of Gini-index and Lorenz curve constitute a major part of the growing literature on inequality, its measurement and interpretation and stated that Lorenz curve and Gini-index have remained the most popular and powerful tool in the analysis of size distribution of income, both empirical and theoretical.

Based on his axiomatic approach, Takayama (1979) recommended the Gini coefficient of the income distribution censored at the poverty line as a proper measure of poverty. For a detailed discussion of poverty indices based on Gini-index, we refer to Sen (1976), Foster (1984) and Sen (1986).

The Lorenz curve and the Gini-index find applications in several branches of learning. They have been extensively used in the study of inequality of distributions. For, example, they have been used in connection with studies of distribution of income by Kakwani and Podder (1976), Gastwirth (1972), and regional disparities in the house hold consumption in India by Bhattacharya and Mahalanobis (1967), and Chatterjee and Bhattacharya (1974), concentration of domestic manufacturing establishment output by Enhorn (1962), business

Recently, in connection with their study on ordering and asymptotic properties of residual income distribution, Belzunce, Candel and Ruiz (1998) introduce a measure of income gap ratio among the rich, defined by

$$\beta^*(t) = 1 - \frac{t}{E(X \mid X > t)}$$

$$= 1 - \frac{t}{m(t)} \tag{1.4.6}$$

1.5 Total time on test transform

For the random variable X considered in section 1.4, the total time on test (TTT) transform $H^{-1}(t)$ corresponding to F is defined by the relation

$$H^{-1}(t) = \int_0^{F^{-1}(t)} \bar{F}(u)\, du. \tag{1.5.1}$$

The scaled TTT transform [Barlow and Campo (1975)] is defined as

$$\phi(t) = \frac{1}{\mu} \int_0^{F^{-1}(t)} \bar{F}(u)\, du. \tag{1.5.2}$$

where

$$\mu = \int_0^{\infty} \bar{F}(u)\, du \text{ is the expectation of } X$$

The TTT transform determines the distribution through the relation

$$F^{-1}(t) = \int_0^{t} \frac{\frac{d}{du} H^{-1}(u)}{(1-u)}\, du. \tag{1.5.3}$$

In view of (1.5.3), properties of F may be studied and verified through that of $H^{-1}(t)$ or $\phi(t)$. This aspect was studied by Barlow, Bartholoma, Bremmer and Brunk (1972) and subsequently by Barlow and Campo (1975), Barlow (1979), Klefsjo (1982), Suresh (1987), and Deshpande and Suresh (1990).

The scaled total time on test transform (1.5.2) is similar to the Lorenz curve $L(p)$ defined in (1.3.2), in many respects. Its shape is like that of Lorenz curve, but it is concave rather than convex. Chandra and Singpurwalla (1981) have mentioned certain relationships between the Lorenz curve and the Gini-index using the TTT transform. They noted that Lorenz curve and TTT transform are connected by the relation

$$L(p) = \frac{-1}{\mu}(1-p)\,F^{-1}(p) + \phi(p), \text{ for } 0 \le p \le 1 \quad (1.5.4)$$

Also they define the cumulative total time on test transform as

$$V = \frac{1}{\mu}\int_0^1 H^{-1}(u)\,du \quad (1.5.5)$$

and they showed that the Gini-index G is related to TTT transform by

$$G = 1 - V \quad (1.5.6)$$

Further the above relation was used to derive a test for exponentiality based on the Gini-index, identical to the one based on the total time on test transform.

Pham and Turkkan (1994) has listed the following properties of the TTT-curve, which are analogous to that of Lorenz curve.

(i) $\phi(t)$ strictly increases within the unit square, with $\phi(0) = 0$ and $\phi(1) = 1$. Moreover

$$\phi(\,F(\mu)\,) = 1 - \frac{E|X-\mu|}{2\mu}$$

and

$$\phi(m) = \phi\left(\frac{1}{2}\right) = \frac{1}{2} + (\,m - E|X-m|\,)\,\frac{1}{2\mu}.$$

(ii) In the unit square, the area between the TTT-curve and the Lorenz curve is equal to the area below the Lorenz curve. The area above the TTT-curve is coincide with that of the Gini-index.

(iii) When $F^{-1}(p)$ is continuous, $L(p)$ and $\phi(p)$ are related by

$$L(p) = (1-p)\int_0^p \frac{\phi(t)}{(1-t)^2}\,dt\,, 0 \le p \le 1$$

Pham and Turkkan (1994) also listed some applications of Lorenz curve and TTT-curve in the Reliability context.

(i) G is the area above the TTT-curve in the unit square. Hence $0 \le G \le 1$, with the extreme values corresponding respectively to the most IFR and most DFR distributions.

(ii) $G=0$ implies (a) Lorenz curve coincide with the diagonal and (b) TTT-curve is the upper side of the unit square $(\phi(F) = 1,\ 0 < F \le 1,\ \phi(0) = 0)$. The corresponding distribution is degenerate, concentrated at μ. In economic terms this corresponds to the situation where each element of the population receives the same income μ.

(iii) G=1 implies that, both the Lorenz curve and TTT-curve are on the lower side of the unit square, $L(F) = \phi(F) = 0,\ 0 \le F \le 1$ and $L(1) = \phi(1) = 1$. $F(x)$ is then the limit Pareto distribution. In this situation every element of the population receives no income, except one, which receives the total.

1.6 The entropy measure

The Shannon entropy [Shannon (1948)] has been extensively used as a measure of income inequality. If there are N individuals in a society, there are N non-negative amounts of individual income, which adds up to the total income. Each of the individual earns non-negative fractions $y_1, y_2, \ldots, y_N$ of total income where y_i's are non-negative numbers which add up to one. When there is equality of income

$y_1 = y_2 = \dots = y_N = \frac{1}{N}$ and in the case of complete inequality $y_i = 1$ for some i and zero for each $i \neq j$. The quantity

$$H(y) = \sum_{i=1}^{N} y_i \log\left(\frac{1}{y_i}\right) \qquad (1.6.1)$$

is the entropy of income shares. A measure of income inequality is defined as

$$\log N - H(y) = \sum_{i=1}^{N} y_i \log(Ny_i) \qquad (1.6.2)$$

where $\log N$ is the maximum value that $H(y)$ can attain. Perfect equality is achieved when there is maximum entropy.

Let X be a non-negative random variable with distribution function $F(x)$ and with a finite mean μ, Theil (1967) used the quantity

$$\begin{aligned} R_F &= \frac{1}{\mu} \int_0^\infty x \log\frac{x}{\mu}\, f(x)\, dx \\ &= E\left(\frac{X}{\mu} \log\frac{X}{\mu}\right) \qquad (1.6.3) \end{aligned}$$

as a reasonable measure of income inequality.

Recently, Ebrahimi (1996) defines the residual entropy function as the Shannon's entropy associated with the residual life distribution, that is, the Shannon's entropy associated with the random variable $(X-t)$ truncated at $t>0$. This has the form

$$H(f,t) = -\int_t^\infty \frac{f(x)}{\bar{F}(t)} \log \frac{f(x)}{\bar{F}(t)}\, dx. \qquad (1.6.4)$$

(1.6.4) can also be written as

$$H(f,t) = \log\bar{F}(t) - \frac{1}{\bar{F}(t)} \int_t^\infty f(x) \log f(x)\, dx. \qquad (1.6.5)$$

The residual entropy function can be expressed in terms of the failure rate encountered in section 1.2, through the relation

$$H(f,t) = 1 - \frac{1}{\bar{F}(t)} \int_t^\infty f(x) \log h(x)\, dx \qquad (1.6.6)$$

$H(f,t)$ measures the expected uncertainty contained in the conditional density of $(X-t)$ given $X>t$ about the predictability of remaining lifetime of the component. It may be noticed that $-\infty \le H(f,t) \le \infty$ and $H(f,0)$ reduces to Shannon's entropy defined over $(0,\infty)$. It is established that $H(f,t)$ determines the distribution uniquely.

Bhattacharjee (1993) stress the importance of considering the random variable $Y=X-t \mid X>t$, in the context of income distributions. He interpreted it as, for any threshold t, the residual holding of the amount of wealth in excess of a threshold t among those who own at least as much. The residual entropy function in the discrete time domain is studied by Rajesh and Nair (1998). Further, characterization results associated with the geometric distribution using the functional form of the residual entropy function are also obtained.

1.7 Geometric vitality function

Let X be a non-negative variable admitting an absolutely continuous distribution function, $F(x)$, with respect to Lebesgue measure on $(0,L)$, where

$$L = \inf\{x : F(x) = 1\}.$$

with $E(X)<\infty$, Nair and Rajesh (2000) defines the geometric vitality function $G(t)$, for $t>0$ as

$$\log G(t) = E(\log X \mid X>t)$$

$$= \frac{1}{\bar{F}(t)} \int_t^\infty \log x\, f(x)\, dx. \qquad (1.7.1)$$

In the Reliability context, if X represents the life length of a component, $G(t)$ represents the geometric mean of lifetime of the components which has survived up to time t. (1.7.1) can also be written as

$$\log\left(\frac{G(t)}{t}\right) = \frac{1}{\bar{F}(t)}\int_t^\infty \frac{\bar{F}(x)}{x}\,dx. \tag{1.7.2}$$

The following properties of geometric vitality function have been established.

(i) $\log G(t)$ is non-decreasing

(ii) $\lim_{t\to 0} \log G(t) = E(\log X)$

(iii) $m(t) \geq \log G(t)$, for all $t>0$

(iv) If $h(t) = \frac{f(t)}{\bar{F}(t)}$ is the failure rate of X then

$$h(t) = \frac{\frac{d}{dt}\log G(t)}{\log\frac{G(t)}{t}}. \tag{1.7.3}$$

It is further established that geometric vitality function determines the distribution uniquely.

The utility of the geometric mean to obtain summary measures of income inequality is evident form the works of Ord, Patil and Taillie (1983). If the random variable X represents the income of people in a locality, the geometric vitality function, being the geometric mean of the income of people whose income greater than a threshold t can be reasonably be taken as a summary measure of income inequality.

1.8 Some inference problems

Moothathu (1985a, 1985b, 1985c and 1989a) has obtained the maximum likelihood estimators (MLE) of the Lorenz curve and the Gini-index for the exponential, Pareto and lognormal distributions in the classical framework. He showed that each of these MLEs is strongly consistent, converges in the r^{th} mean and has obtained their exact distributions. Moothathu (1989b) has obtained the uniformly minimum variance unbiased estimator for the Gini-index of the lognormal distribution along with its variance. Further the best estimate for the Lorenz curve, and the Gini-index of the Pareto distribution, along with its variance have been obtained.

The Bayesian approach to estimation in specific distributions assumes the existence of a joint probability measure on $(\Theta \times X)$, where $\Theta \in \mathbb{R}_k$ is the parametric space corresponding to a vector of parameters $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$ and X is the sample space. The joint measure is determined through a prior measure on Θ and the conditional measure on X for a given θ in Θ which in turn provides the posterior measure on Θ for a specified x in X along with a marginal measure on X. In this formulation the posterior density function of θ can be obtained through Bayes theorem as [Raiffa and Schlaifer (1961)]

$$f(\theta | \underline{x}) = \phi(\theta)\, l(\underline{x} | \theta)\, C(\underline{x}) \qquad (1.8.1)$$

where $\phi(\theta)$ is the prior density and $C(\underline{x})$ is a normalizing constant independent of θ given by

$$\int_{\Theta} f(\theta | \underline{x})\, d\theta = C(\underline{x}) \int_{\Theta} \phi(\theta)\, l(\underline{x} | \theta)\, d\theta = 1 \qquad (1.8.2)$$

For mathematical tractability it is common to use the conjugate prior to arrive at the desired posterior distribution. In finding point estimate of θ we employ either the mode of (1.8.1) or make use of the quadratic loss function

$$L\left(\hat{\theta}(\underline{x}) - \theta\right) = \left(\hat{\theta}(\underline{x}) - \theta\right)^2 \qquad (1.8.3)$$

to prescribe the estimate as one that minimizes

$$E\left(L\left(\hat{\theta}(\underline{x}) - \theta\right)\right) = \int_{\Theta} \left(\hat{\theta}(\underline{x}) - \theta\right)^2 f(\theta | \underline{x})\, d\theta \qquad (1.8.4)$$

or

$$\hat{\theta}(\underline{x}) = E(\theta | \underline{x}) \qquad (1.8.5)$$

The expected loss, resulting from the use of (1.8.5) as the estimator of θ, is the posterior variance of θ. Since (1.8.5) is calculated for a specific sample point $\underline{x}$, some times it is of advantage to look at the Bayes risk

$$R\left(\hat{\theta},\theta\right) = \int_{\Theta}\int_{X} L\left(\hat{\theta},\theta\right) l(\underline{x}|\theta)\, \phi(\theta)\, d\underline{x}\, d\theta \qquad (1.8.6)$$

1.9 Present study

The present work is organized into six chapters. After the present introductory chapter, which focuses attention on a brief review of the basic concepts, in chapter 2 we define certain measures of income inequality for the truncated distributions and study the effect of truncation upon these measures. It is shown that the Pareto distribution is the only distribution for which these measures are unaffected by truncation. Characterization results in respect of some specific models such as exponential, Pareto and finite range based on the functional form of these measures are also discussed.

Considering the importance of the study of disparity of a population with respect to more than one attribute, in chapter 3, we extend the Gini-index to the bivariate setup. Although several extensions of Gini-index are available in the literature, they are not mathematically tractable from the point of view of characterization of probability distributions. In the present chapter we provide a definition for the Gini-index in higher dimensions, similar to that of the definition of the vector valued failure rate reviewed in section 1.2. Characterization problems associated with certain bivariate models such as the Gumbel's bivariate exponential, bivariate Pareto and bivariate finite range based on the form of the bivariate Gini-index are also investigated.

An important measure, used in Reliability theory, to measure the stability of the component is the residual entropy function. This concept can advantageously used as a measure of inequality of truncated distributions. In chapter 4 we extend this concept to the bivariate setup and provide characterization results for some bivariate models using the same.

The geometric mean comes up as a handy tool in the measurement of income inequality. The geometric vitality function being the geometric mean of the truncated random variable can be advantageously utilized to measure inequality of the truncated distributions. This concept is being extended to the bivariate setup in chapter 5.Apart from this the bivariate exponential, bivariate Pareto and the bivariate finite range models are characterized using the form of the bivariate geometric vitality function.

Even though a lot of work has been carried out on the problem of estimation of the Lorenz curve and Gini-index in the classical frame work, only very little work seems to have been done in this area using Bayesian concepts. In chapter 6 we look into problem of estimation of the Lorenz curve, Gini-index and variance of logarithms for the Pareto distribution using Bayesian techniques. Estimation is carried out in two situations namely when the scale parameter is known and the scale parameter is unknown. Also a comparison of the estimates is done using data generated from the Pareto population. It is established that the estimates provided by the Bayesian procedure are better than the classical estimates from the point of view of reduction in variance. Utilizing a relationship between the Lorenz curve and the TTT transform, discussed in section (1.5), we also provide estimators for the TTTtransform in the Pareto situation.

Chapter II

Characterization of probability distributions based on truncated versions of certain measures of income inequality

2.1 Introduction

As pointed out in the previous chapter, the Lorenz curve, defined by (1.3.1), and the Gini-index, defined by (1.4.1), has been extensively used in literature as reasonable measures of income inequality. Properties of these measures as well as characterization of probability distributions using this concept had been a hot area of research during the middle of the twentieth century. Apart from these measures, the variance of logarithms as well as the entropy indices are also advantageously used to measure income inequality. However an in-depth study on the truncated version of these measures does not seem to have been undertaken so far. Recently a lot of interest seems to have been evoked in using certain concepts in Reliability theory such as the failure rate and mean residual life function, for the study of income distributions. In the present chapter we look into the problem of characterization of probability distributions using the truncated versions of the above-mentioned concepts.

2.2 The Lorenz Curve

The Lorenz curve of a distribution of income is defined as that fraction of the total income owned by the lowest p^{th} fraction of the population as a function of p ,$(0 \leq p \leq 1)$. Assume that $X$is a non-negative random variable with distribution function $F(x)$ such that $E(X) < \infty$. Denote by

$$\frac{-L'(t)}{1-L(t)} = \frac{t}{t+r(t)} \frac{\bar{F}'(t)}{\bar{F}(t)}.$$

In view of (1.2.6) and (1.2.20) the above equation simplifies to

$$\frac{L'(t)}{1-L(t)} = \frac{t\left(1+r'(t)\right)}{r(t)\left(t+r(t)\right)}, \text{ as claimed.}$$

Observing that (2.2.2) is a differential equation of the first order in $L(t)$ or $r(t)$, $L(t)$ can be solved in terms of $r(t)$ or vice versa. Hence the knowledge of the Lorenz curve is sufficient to determine the mean residual life function and that of $r(t)$ is sufficient to determine $L(t)$.

The above relationship can be advantageously used to obtain a characterization result for the Pareto type-1 distribution in terms of a functional relationship between the Lorenz curve and the mean residual life function, which is given as theorem 2.2.

Theorem 2.2

Let X be a non-negative random variable admitting an absolutely continuous distribution with $E(X)<\infty$. If $L(t)$ represents the Lorenz curve and $r(t)$ the mean residual life function, then the relationship

$$\frac{L'(t)}{1-L(t)} = \frac{1}{r(t)} \qquad (2.2.6)$$

holds for all real $t \geq 0$ if and only if X follows the Pareto type-1 distribution with survival function

$$\bar{F}(x) = \left(\frac{\alpha}{x}\right)^{a}, \; x \geq \alpha, \; a>1 \qquad (2.2.7)$$

Proof

When (2.2.6) holds using (2.2.2) we get

$$t\left(1+r'(t)\right) = t + r(t)$$

or

$$t\,r'(t) - r(t) = 0.$$

This gives

$$r(t) = k\,t \text{ , with } k>0.$$

Using the relation (1.2.21) we get (2.2.7) as claimed.

Conversely when the distribution of X is specified by (2.2.7) by direct calculations we get

$$L(t) = 1-\left(\frac{t}{\alpha}\right)^{-(a-1)},\; r(t) = \frac{t}{a-1}$$ and the validity of (2.2.6) is straightforward.

Our next theorem provides a characterization result for a family of distributions using a possible relationship between the Lorenz curve and the mean residual life function.

Theorem 2.3

Let X be a non-negative random variable admitting an absolutely continuous distribution such that $E(X)<\infty$. The relationship

$$\frac{L'(t)}{1-L(t)} = \frac{k\,t}{r(t)\,\left(t+r(t)\right)},\quad k > 0. \tag{2.2.8}$$

holds for all real $t\geq 0$ if and only if X follows any one of the following distributions according as $k=1$, $k>1$ and $k<1$ respectively.

(i) the exponential distribution with survival function

$$\bar{F}(x) = e^{-\lambda x},\, x\geq 0,\; \lambda>0 \tag{2.2.9}$$

(ii) the Pareto distribution with survival function

$$\bar{F}(x) = \left(\frac{\alpha}{x+\alpha}\right)^{a},\, x\geq 0,\; a>1,\; 0<\alpha<\infty \tag{2.2.10}$$

(iii) the finite range distribution with survival function

$$\bar{F}(x) = \left(1-\frac{x}{R}\right)^{c},\, 0<x<R,\; c>1 \tag{2.2.11}$$

Proof

When (2.2.8) holds using (2.2.2) we get

$$t\,r'(t)+t = k\,t.$$

The above equation gives

$$r(t) = (k-1)\,t+c.$$

From Mukherjee and Roy (1986), reviewed in section 1.2, the above relation is characteristic to (2.2.9) for $k=1$, (2.2.10) for $k>1$ and (2.2.11) for $k<1$. Hence X follows any one of the three distributions.

The if part of the theorem follows from the expressions for $L(t)$ and $r(t)$ given below.

Distribution	$L(t)$	$r(t)$
Exponential	$1-(1+t\lambda)\ e^{-t\lambda}$	$\frac{1}{\lambda}$
Pareto	$1-\alpha^{a-1}\,(t+\alpha)^{-a}\,(at+\alpha)$	$\frac{t+\alpha}{a-1}$
Finite range	$1-\frac{1}{R}\,(R+ct)\left(1-\frac{t}{R}\right)^{c}$	$\frac{R-t}{c+1}$

The following theorem provides a characterization result for the Pearson family of distributions by the form of $\frac{L'(t)}{L(t)}$.

Theorem 2.4

For the random variable considered in theorem 2.3, the relationship

$$\frac{L'(t)}{L(t)} = \frac{kt}{a_0+a_1t+a_2t^2},\ k,a_0,a_1,a_2>0,\ a_2>\frac{k}{2} \qquad (2.2.12)$$

holds for all real $t\geq 0$ if and only if X belongs to the Pearson family of distributions specified by

$$\frac{f'(t)}{f(t)} = \frac{-(t+d)}{b_0+b_1t+b_2t^2} \text{ with } d=b_1,\ d,b_0,b_1,b_2>0, 2b_2>1. \qquad (2.2.13)$$

Proof

When (2.2.12) holds we have

$$\frac{L'(t)}{L(t)} = \frac{k\,t}{a_0 + a_1\,t + a_2\,t^2}.$$

Using the definition (2.2.2) we get

$$\frac{t\,f(t)}{\mu}\left(a_0 + a_1\,t + a_2\,t^2\right) = \frac{k\,t}{\mu}\int_0^t x\,f(x)\,dx.$$

Differentiating with respect to t and rearranging the terms we get

$$\frac{f'(t)}{f(t)} = \frac{(k - 2a_2)\,t - a_1}{a_0 + a_1\,t + a_2\,t^2}$$

or

$$\frac{f'(t)}{f(t)} = \frac{-\,(t+d)}{b_0 + b_1\,t + b_2\,t^2}, \text{as claimed.}$$

with $d = \dfrac{a_1}{2a_2 - k}$, $b_0 = \dfrac{a_0}{2a_2 - k}$, $b_1 = \dfrac{a_1}{2a_2 - k}$ and $b_2 = \dfrac{a_2}{2a_2 - k}$.

Conversely when (2.2.13) holds we have

$$f'(t)\left(b_0 + b_1\,t + b_2\,t^2\right) = -\,(t+d)\;f(t)$$

or

$$\frac{d}{dt}\left\{f(t)\left(b_0 + b_1 t + b_2 t^2\right)\right\} - f(t)\,(2b_2 t + b_1) = -(t+d)\;f(t).$$

Integrating with respect to t and simplifying we get

$$f(t)\left(b_0 + b_1 t + b_2 t^2\right) = -(1 - 2b_2)\int_0^t x f(x)dx - (d - b_1)\int_0^t f(x)dx.$$

Using the definition of $L(t)$ and also applying the condition $d = b_1$ we get

$$L'(t)\left(b_0 + b_1 t + b_2 t^2\right) = (2b_2 - 1)\;t\,L(t)$$

or

$$\frac{L'(t)}{L(t)} = \frac{k\,t}{b_0 + b_1\,t + b_2\,t^2}, \text{ where } k = 2b_2 - 1.$$

This is of the form (2.2.12).

2.3 The Gini-index

In this section we look into the problem of characterization of probability distributions using the truncated form of the Gini-index considered in section (1.4). First we establish a relationship between the Gini-index and the vitality function, defined by (1.2.36).

Theorem 2.5

Let X be a non-negative, non-degenerate random variable admitting an absolutely continuous distribution and with a finite mean. If $m(t)$ represents the vitality function, $\overline{F}(t)$ the survival function and $G(t)$ the truncated Gini-index then the following relationship holds

$$(1-G(t))\ m(t) = t + \frac{1}{\overline{F}^2(t)} \int_t^\infty \overline{F}^2(x)dx \qquad (2.3.1)$$

Proof

From (1.4.4) we have

$$2\int_t^\infty \left(\frac{F(x)-F(t)}{\overline{F}(t)}\right) \frac{xf(x)}{\overline{F}(t)}\, dx - \int_t^\infty \frac{xf(x)}{\overline{F}(t)}\, dx = G(t) \int_t^\infty \frac{xf(x)}{\overline{F}(t)}\, dx.$$

The above equation can be written as

$$\frac{2}{\overline{F}(t)} \int_t^\infty x\, f(x)\, F(x)\, dx - \frac{2\left(1-\overline{F}(t)\right)}{\overline{F}(t)} \int_t^\infty x\, f(x)\, dx - \int_t^\infty x\, f(x)\, dx = G(t) \int_t^\infty x\, f(x) dx$$

or

$$(1-G(t)) \int_t^\infty x\, f(x)\, dx = \frac{2}{\overline{F}(t)} \int_t^\infty x\, f(x)\, \overline{F}(x)\, dx. \qquad (2.3.2)$$

Integrating the right hand side of equation (2.3.2) by parts we get

$$(1-G(t)) \int_t^\infty x\, f(x)\, dx = t\,\overline{F}(t) + \frac{1}{\overline{F}(t)} \int_t^\infty \overline{F}^2(x)\, dx.$$

Using equation (1.2.36) we get

$$(1-G(t))\ m(t) = t + \frac{1}{\overline{F}^2(t)} \int_t^\infty \overline{F}^2(x)\, dx \quad \text{as claimed.}$$

The above relationship provides a handy tool to compute the Gini-index for distributions, which have a closed form for the survival function and the vitality function. The following theorem provides a relationship between the Gini-index $G(t)$, the vitality function $m(t)$ and the failure rate $h(t)$.

Theorem 2.6

For the random variable X considered in theorem 2.5 if $h(t)$ represents the failure rate, $G'(t)$ the first derivative of the Gini-index and $m(t)$ the vitality function then the following relationship holds.

$$G'(t)\ m(t) = h(t)\left[\ (G(t)-1)\ m(t) + t\left(G(t)+1\right)\ \right] \quad (2.3.3)$$

Proof

Using (2.3.1) we have

$$\left(1-G(t)\right)\ m(t) = t + \frac{1}{\overline{F}^2(t)} \int_t^\infty \overline{F}^2(x)dx.$$

Differentiating with respect to t and simplifying we get

$$\left(1-G(t)\right)\ m'(t) - m(t)\ G'(t) = \frac{f(t)}{\overline{F}^3(t)} \int_t^\infty \overline{F}^2(x)\ dx.$$

Using the relationship (1.2.6) and (2.3.1) we get

$$\left(1-G(t)\right)\ m'(t) - m(t)\ G'(t) = 2\ h(t)\ \left((1-G(t))m(t)-t\right).$$

Since $m(t)=t+r(t)$, the above equation can be written as

$$\left(1-G(t)\right)\ \left(1+r'(t)\right)\ - m(t)\ G'(t) = 2h(t)r(t)-2t\ h(t)\ G(t)-2h(t)\ G(t)-2h(t)\ r(t)\ G(t).$$

or

$$G'(t)\ m(t) = h(t)\ \left[\ G(t)\ r(t)-r(t)+2\ t\,G(t)\right].$$

Using (1.2.36), (2.3.3) is immediate from the above equation.

In the sequel we look into the situation where the truncated Gini-index is a constant.

Theorem 2.7

Let X be a non-negative random variable admitting an absolutely continuous distribution such that $E(X)<\infty$. If $G(t)$ is the truncated Gini-index defined by (1.4.4), then the relationship

$$G(t) = k \tag{2.3.4}$$

where k is a constant holds for all real $t \geq 0$ if and only if X follows the Pareto type-1 distribution with survival function (2.2.7).

Proof

When (2.3.4) holds using (2.3.1) we have

$$(1-k)\ m(t) = t + \frac{1}{\overline{F}^2(t)} \int_t^{\infty} \overline{F}^2(x)\,dx$$

or

$$(1-k)\ m(t)\,\overline{F}^2(t) = t\,\overline{F}^2(t) + \int_t^{\infty} \overline{F}^2(x)\,dx.$$

Differentiating both sides with respect to t and simplifying we get

$$-2\,(1-k)\ m(t)\,f(t) + (1-k)\ m'(t)\,\overline{F}(t) = -2\,t\,f(t).$$

Using (1.2.6) and (1.2.36) in the above equation we get

$$-2\,(1-k)\ t\,h(t) - (1-k)\,(1+r'(t)) = -2\,t\,h(t).$$

Using (1.2.20), the above equation simplifies to

$$r(t) = \frac{2\,k}{1-k}\,t.$$

The desired form for $\overline{F}(t)$ is immediate upon using (1.2.21).

Conversely when the distribution of X is specified by (2.2.7) by direct calculations we get

$$G(t) = \frac{1}{2a-1}$$ and the sufficiency part follows.

Bhattacharjee (1993) stress the role of anti-aging distributions in Reliability theory as reflecting the features of skewness and heavy tails typical of wealth distributions. Our next result provides a characterization theorem for the exponential distribution using a functional relationship between the Gini-index and the vitality function.

Theorem 2.8

Let X be a non-negative random variable admitting an absolutely continuous distribution such that $E(X) < \infty$. If $G(t)$ be the truncated Gini-index and $m(t)$ be the vitality function, then a relationship of the form

$$G(t)\, m(t) = G(0)\, m(0) \qquad (2.3.5)$$

holds for all real $t \geq 0$ if and only if X follows the exponential distribution with survival function specified in (2.2.9).

Proof

When (2.3.5) holds, from (2.3.1) we have

$$\left(m(t) - G(0)\, m(0)\right) \bar{F}^2(t) = t\, \bar{F}^2(t) + \int_t^\infty \bar{F}^2(x)\, dx.$$

Differentiating the above equation with respect to t, in view of (1.2.6) we can write

$$-2\, t\, h(t) = m'(t) - 2\,(\, m(t) - G(0)\, m(0)\,)\, h(t).$$

Using (1.2.36) and (1.2.20) the above equation can be written as

$$\left(1 + r'(t)\right)\left(2\, G(0)\, m(0) - r(t)\right) = 0. \qquad (2.3.6)$$

(2.3.6) gives

$$r(t) = 2\, G(0)\, m(0) \qquad (2.3.7)$$

or

$$r(t) = -t + c. \qquad (2.3.8)$$

The later solution (2.3.8) leads to the trivial case where the distribution is degenerate. Observing that the constancy of the mean residual life function is characteristic to the exponential model, the only if part follows from (2.3.7).

The sufficiency part follows from the expressions for $G(t)$ and $m(t)$ namely $G(t) = \dfrac{1}{2 + 2\,\lambda\, t}$ and $m(t) = t + \dfrac{1}{\lambda}$.

The following theorem provides a characterization for a family of distributions using a possible relationship between the truncated Gini-index and the vitality function.

Theorem 2.9

For the random variable X considered in theorem 2.8 the relationship

$$(1-G(t))\, m(t) = a + bt \quad , a > 0,\ b > 0. \qquad (2.3.9)$$

holds for all real $t \geq 0$ if and only if X follows (2.2.9), (2.2.10) and (2.2.11) according as $b=1$, $b>1$ and $b<1$ respectively.

Proof

When (2.3.9) holds, we have from (2.3.1)

$$t\,\overline{F}^2(t) + \int_t^\infty \overline{F}^2(x)\, dx = (a + bt)\, \overline{F}^2(t).$$

Differentiating the above equation with respect to t we get

$$-2tf(t) = -2\, a\, f(t) - 2\, b\, t\, f(t) + b\, \overline{F}(t).$$

Using (1.2.6) the above equation can be written as

$$-2th(t) = -2\, a\, h(t) - 2\, b\, t\, h(t) + b.$$

This gives

$$h(t) = \frac{1}{A + Bt} \qquad (2.3.10)$$

where $A = \dfrac{2a}{b}$ and $B = \dfrac{2\,(b-1)}{b}$

The only if part follows from Mukherjee and Roy (1986) reviewed in section 1.2. That is (2.3.10) is characteristic to the exponential distribution for $B=1$, the Pareto distribution for $B>1$ and the finite range distribution for $B<1$. Using the relationship between b and B the only if part follows.

The if part of the theorem follows from the expression for $G(t)$ and $m(t)$ given below.

Distribution	$m(t)$	$G(t)$
Exponential	$t+\frac{1}{\lambda}$	$\frac{1}{2+2\lambda t}$
Pareto	$\frac{\alpha+at}{a-1}$	$\left(\frac{a}{2a-1}\right)\left(\frac{\alpha+t}{\alpha+at}\right)$
Finite range	$\frac{R+ct}{c+1}$	$\left(\frac{c}{c+1}\right)\left(\frac{R-t}{R+ct}\right)$

We now look into the problem of characterizing the family of distributions considered in theorem 2.9 using a possible relationship between the truncated Gini-index and the income gap ratio defined by (1.4.6).

Theorem 2.10

Let X be a non-negative random variable admitting an absolutely continuous distribution such that $E(X) < \infty$. Let $\beta^*(t)$ be defined as in (1.4.6). Then the relationship

$$G(t) = k\beta^*(t) \qquad ,0 < k < 1 \qquad (2.3.11)$$

holds for all real $t \geq 0$ if and only if X follows the distributions specified by (2.2.9), (2.2.10) and (2.2.11) according as $k=\frac{1}{2}$, $k>\frac{1}{2}$ and $k<\frac{1}{2}$ respectively.

Proof

When (2.3.11) holds using (2.3.1) we have

$$\left(1-k\beta^*(t)\right) m(t) = t + \frac{1}{\bar{F}^2(t)} \int_t^\infty \bar{F}^2(x)\, dx.$$

Using (1.4.6), the above equation simplifies to

$$m(t)\,\overline{F}^2(t) - k\,m(t)\,\overline{F}^2(t) + k\,t\,\overline{F}^2(t) = t\,\overline{F}^2(t) + \int_t^\infty \overline{F}^2(x)\,dx.$$

Differentiating the above equation with respect to t and rearranging the terms we get

$$-2\,(1-k)\;m(t)\;h(t)+(1-k)\;m'(t)-2k\,t\,h(t)+k = -2\,t\,h(t).$$

Using (1.2.36) and (1.2.20) the above equation reduces to

$$2\,k\,t\,\frac{(1+r'(t))}{r(t)}-(1-k)\;(1+r'(t))\;-2\,k\,t\,\frac{(1+r'(t))}{r(t)}+k = 0.$$

This gives

$$r'(t) = \frac{2k-1}{1-k}$$

or

$$r(t) = \frac{2k-1}{1-k}\,t + c \tag{2.3.12}$$

From Mukherjee and Roy (1986), reviewed in section 1.2, (2.3.12) is characteristic to the exponential distribution for $k=\frac{1}{2}$, the Pareto distribution for $k>\frac{1}{2}$ and the finite range distribution for $k<\frac{1}{2}$. Hence X follows any one of the three distributions.

The if part of the theorem follows from the expression for $G(t)$ and $\beta^*(t)$ given below.

Distribution	$\beta^*(t)$	$G(t)$
Exponential	$\frac{1}{1+\lambda\,t}$	$\frac{1}{2+2\lambda\,t}$
Pareto	$\frac{\alpha+t}{\alpha+a\,t}$	$\left(\frac{a}{2a-1}\right)\left(\frac{\alpha+t}{\alpha+at}\right)$
Finite range	$\frac{R-t}{R+c\,t}$	$\left(\frac{c}{c+1}\right)\left(\frac{R-t}{R+ct}\right)$

The following theorem provides a characterization for the Pareto type-1 distribution using a possible relationship between $G(t)$ and $\beta^*(t)$.

Theorem 2.11

For the random variable X considered in theorem 2.10 the relationship

$$G(t) = \frac{\beta^*(t)}{2-\beta^*(t)} \qquad (2.3.13)$$

holds for all real $t \geq 0$ if and only if X follows the Pareto type-1 distribution with survival function specified by (2.2.7).

Proof

When (2.3.13) holds, we have from (2.3.1)

$$\left(1 - \frac{\beta^*(t)}{2-\beta^*(t)}\right) m(t) = t + \frac{1}{\overline{F}^2(t)} \int_t^\infty \overline{F}^2(x)\, dx.$$

The above equation simplifies to

$$\frac{2\, t\, m(t)}{t + m(t)} = t + \frac{1}{\overline{F}^2(t)} \int_t^\infty \overline{F}^2(x)\, dx.$$

Differentiating both sides with respect to t we get

$$2\, t^2\, m'(t) + 2\, m^2(t)\, \left(1-t\, h(t)\right) + 2\, t^3\, h(t) = 0.$$

Using the relationship $m'(t) = 1+r'(t)$ and equation (1.2.20) we get

$$r'(t)\, \left(2\, t^2 + 2\, t\, r(t)\right) - 2\, t\, r(t) - 2\, r^2(t) = 0.$$

The solution of the above differential equation is

$$r(t) = (a-1)\, t.$$

Using the relationship (1.2.21) we get the desired form for $\overline{F}(t)$.

Conversely for the Pareto distribution specified by (2.2.7) by direct calculation we get $G(t) = \frac{1}{2a-1}$ and $\beta^*(t) = \frac{1}{a}$ from which (2.3.13) is immediate.

Using the Gini-index for the random variable $X_r(t) = X \mid X < t$, considered by Sen (1986), we give below a characterization for the Power function distribution as theorem 2.12.

Theorem 2.12

Let X be a non-negative random variable admitting an absolutely continuous distribution such that $E(X) < \infty$. Let $G_r(t)$ represent the Gini-index defined in (1.4.5). Then the relationship

$$G_r(t) = k \quad ,0 < k < 1 \tag{2.3.14}$$

where k is a constant, holds for all real $t \geq 0$ if and only if X follow the Power function distribution with survival function

$$\overline{F}(x) = 1-\left(\frac{x}{b}\right)^c \ ,0 \leq x \leq b \ \ c, b \geq 0 \tag{2.3.15}$$

Proof

When (2.3.14) holds, using (1.4.5) we have

$$1 - \frac{2}{E(X \mid X<t)} \int_0^t y\left(1-\frac{F(y)}{F(t)}\right) \frac{f(y)}{F(t)}\, dy = k.$$

The above equation can be written as

$$2\int_0^t y\left(1-\frac{F(y)}{F(t)}\right)\frac{f(y)}{F(t)}\, dy = \frac{1-k}{F(t)}\int_0^t y\, f(y)\, dy.$$

or

$$2\int_0^t y\, f(y)\, dy-(1-k)\int_0^t y\, f(y)\, dy = 2\int_0^t y\, \frac{F(y)}{F(t)}\, f(y)\, dy.$$

This gives

$$(k+1)\ F(t)\int_0^t y\, f(y)\, dy = 2\int_0^t y\, F(y)\, f(y)\, dy. \tag{2.3.16}$$

Differentiating (2.3.16) with respect to t and simplifying we get

$$(k+1)\int_0^t y\, f(y)\, dy = (1-k)\ t\, F(t). \tag{2.3.17}$$

Differentiating (2.3.17) with respect to t we get

$$(k+1)\ t\, f(t) = (1-k)\ t\, f(t)+(1-k)\ F(t)$$

or

$$\frac{f(t)}{F(t)} = \frac{(1-k)}{2k}\frac{1}{t}$$

or

$$\frac{d}{dt}\log F(t) = \frac{(1-k)}{2k}\frac{1}{t}.$$

The above equation gives

$$F(t) = t^{\frac{1-k}{2k}}\, c.$$

From which

$$\bar{F}(t) = 1 - t^{\frac{1-k}{2k}}\, c \quad \text{as desired.}$$

Conversely for the Power function distribution specified by (2.3.15) by direct calculations we get $G_r(t) = \frac{1}{2c+1}$, as claimed.

2.4 The Variance of Logarithms

The Variance of logarithms, denoted by V_L is defined as the variance applied to the distribution of logarithm of incomes. Let X be a non-negative random variable admitting an absolutely continuous distribution function $F(x)$. The variance of logarithms (Gibrat (1931)) is defined as

$$V_L = E\left(\log x - E(\log x)\right)^2. \tag{2.4.1}$$

For the random variable $X_t(t) = X \mid X > t$, V_L takes the form

$$V_L(t) = \frac{1}{\bar{F}(t)}\int_t^{\infty} (\log x)^2 f(x)\, dx - \left(\log G(t)\right)^2 \tag{2.4.2}$$

where $\log G(t)$ is the geometric vitality function defined in Nair and Rajesh (2000), namely

$$\log G(t) = E(\log X \mid X > t)$$

$$= \frac{1}{\bar{F}(t)}\int_t^{\infty} \log x\, f(x)\, dx. \tag{2.4.3}$$

We first look into the situation where $V_L(t)$ is a constant.

Theorem 2.13

Let X be a non-negative random variable admitting an absolutely continuous distribution function with finite mean and let $V_L(t)$ denote the truncated variance of logarithm defined in (2.4.2), then the relationship

$$V_L(t) \quad = k \quad , k > 0 \qquad (2.4.4)$$

where k is a constant holds for all real $t \geq 0$ if and only if X follows the Pareto type-1 distribution with survival function specified by (2.2.7).

Proof

When (2.4.4) holds, using (2.4.2) we have

$$\int_t^\infty (\log x)^2 f(x)\, dx - \bar{F}(t)\, (\log G(t))^2 = k\, \bar{F}(t).$$

Differentiating the above equation with respect to t and rearranging the terms we get

$$-(\log t)^2 f(t) - 2\, \bar{F}(t) \log G(t) \frac{G'(t)}{G(t)} + f(t)\, (\log G(t))^2 = -k\, f(t).$$

Using (1.2.6) we get

$$-(\log t)^2 h(t) - 2 \log G(t) \frac{G'(t)}{G(t)} + h(t)\, (\log G(t))^2 = -k\, h(t) \qquad (2.4.5)$$

From (2.4.3) we have

$$\bar{F}(t) \log G(t) = \int_t^\infty \log x\, f(x)\, dx. \qquad (2.4.6)$$

Differentiating (2.4.6) with respect to t we get

$$\frac{G'(t)}{G(t)} - h(t) \log G(t) = -h(t) \log t$$

or

$$\frac{G'(t)}{G(t)} = h(t) \log G(t) - h(t) \log t. \qquad (2.4.7)$$

Using (2.4.7) in equation (2.4.5) we get

$$-(\log t)^2\, h(t) + 2\, h(t)\, \log t\, \log G(t) - h(t)\, (\log G(t))^2 = -k\, h(t).$$

The above equation simplifies to

$$-(\log t)^2 + 2\, \log t\, \log G(t) - (\log G(t))^2 = -k$$

or

$$(\log G(t) - \log t)^2 = k$$

or

$$\log \frac{G(t)}{t} = P, \text{ where } P = \sqrt{k} \text{ is a constant.}$$

The only if part follows from Nair and Rajesh (2000).

Conversely for the Pareto distribution specified in (2.2.7) by direct calculation, we get $V_L(t) = \frac{1}{a^2}$, as stipulated in the theorem.

Our next result provides a characterization theorem for the Pareto distribution using a functional relationship between the truncated version of the variance of logarithms and the geometric vitality function.

Theorem 2.14

For the random variable X considered in theorem 2.13 let $V_L(t)$ denote the truncated variance of logarithm, and $\log G(t)$ the geometric vitality function. Then a relationship of the form

$$V_L(t) = \left(\log \frac{G(t)}{t}\right)^2 \tag{2.4.8}$$

holds for all real $t \geq 0$ if and only if X follow the Pareto type-1 distribution with survival function specified in (2.2.7).

Proof

When (2.4.8) holds, we have from (2.4.2)

$$\int_t^{\infty} (\log x)^2 f(x)\, dx - \bar{F}(t)\, (\log G(t))^2 = \bar{F}(t) \left(\log \frac{G(t)}{t}\right)^2.$$

Differentiating the above equation with respect to t we get

$$- (\log t)^2 \; f(t) - 2\,\overline{F}(t)\,\log G(t)\,\frac{G'(t)}{G(t)} + f(t)\,(\log G(t))^2$$

$$= 2\,\overline{F}(t)\,(\log G(t) - \log t)\left(\frac{G'(t)}{G(t)} - \frac{1}{t}\right) - f(t)\left(\log\frac{G(t)}{t}\right)^2.$$

Using (1.2.6) the above equation can be written as

$$- (\log t)^2 \; h(t) - 2\,\log G(t)\,\frac{G'(t)}{G(t)} + h(t)\,(\log G(t))^2$$

$$= 2\,(\log G(t) - \log t)\left(\frac{G'(t)}{G(t)} - \frac{1}{t}\right) - h(t)\left(\log\frac{G(t)}{t}\right)^2.$$

or

$$- (\log t)^2 \; h(t) - 2\,\log G(t)\,\frac{G'(t)}{G(t)} + h(t)\,(\log G(t))^2$$

$$= 2\,\log G(t)\,\frac{G'(t)}{G(t)} - 2\log t\,\frac{G'(t)}{G(t)} - h(t)\left((\log G(t))^2 + (\log t)^2 - 2\log G(t)\log t\right).$$

The above equation simplifies to

$$- 4\,\log G(t)\,\frac{G'(t)}{G(t)} + 2\,h(t)\,(\log G(t))^2 = -2\log t\,\frac{G'(t)}{G(t)} - \frac{2}{t}\,\log G(t)$$

.

Using (2.4.7) we have

$$-4\,(\log G(t))^2\,h(t) + 4\,\log G(t)\,h(t)\,\log t + 2\,h(t)\,(\log G(t))^2$$

$$= 2\,(\log t)^2\,h(t) - \frac{2}{t}\,\log G(t) + \frac{2}{t}\,\log t$$

or

$$-2\,h(t)\,(\log G(t))^2 + 4\,\log G(t)\,h(t)\,\log t = (\log t)^2\,h(t) - \frac{1}{t}\,\log G(t) + \frac{1}{t}\,\log t.$$

This gives

$$(\log t)^2 + (\log G(t))^2 - 2\,\log t\,\log G(t) = \frac{1}{t\,h(t)}\,\log G(t) + \frac{1}{t\,h(t)}\log t$$

or

$$\log\frac{G(t)}{t} = \frac{1}{t\,h(t)}. \qquad (2.4.9)$$

Using (1.7.3), equation (2.4.9) reduces to

$$\frac{1}{t} = \frac{d}{dt}\log G(t).$$

The above equation gives

$$\log G(t) = \log t + e^{\log c}$$

or

$$\log\frac{G(t)}{t} = c$$

The only if part follows from Nair and Rajesh (2000).

Conversely for the Pareto distribution specified by (2.2.7) by direct calculation we get $V_L(t) = \frac{1}{a^2}$ and $\log\frac{G(t)}{t} = \frac{1}{a}$ from which (2.4.8) is immediate.

The following theorem provides a characterization result for the Pareto distribution using a functional relationship between the Geometric vitality function defined by (1.7.1) and the vitality function defined by (1.2.36).

Theorem 2.15

Let X be a non-negative random variable admitting an absolutely continuous distribution function such that $E(X) < \infty$. If $m(t)$ represents the vitality function and $\log G(t)$ the geometric vitality function, then the relationship

$$\log G(t) - \log m(t) = k \quad , k > 0 \tag{2.4.10}$$

where k is a constant holds for all real $t \geq 0$ if and only if X follows the Pareto type-1 distribution with survival function specified by (2.2.7).

Proof

When (2.4.10) holds using (2.4.3) we have

$$\int_t^{\infty} \log x\, f(x)\, dx - \bar{F}(t)\log\left[\frac{1}{\bar{F}(t)}\int_t^{\infty} \log x\, f(x)\, dx\right] = k\,\bar{F}(t).$$

Differentiating with respect to t we get

$$-\log t\, f(t) - \bar{F}(t) \frac{1}{\frac{1}{\bar{F}(t)}\int_t^{\infty} x\, f(x)\, dx} \left\{ \frac{-t\, f(t)}{\bar{F}(t)} + \int_t^{\infty} x\, f(x)\, dx\, \frac{t\, f(t)}{\bar{F}^2(t)} \right\} + f(t) \log\left(\frac{1}{\bar{F}(t)} \int_t^{\infty} x\, f(x)\, dx \right)$$

$$= -\, k\, f(t).$$

The above equation simplifies to

$$-\log t\, f(t) + f(t) \log m(t) + \frac{t\, f(t)}{m(t)} - f(t) = -k\, f(t)$$

or

$$\log m(t) + \frac{t}{m(t)} = 1 - k + \log t. \qquad (2.4.11)$$

Differentiating (2.4.11) with respect to t we get

$$\frac{m'(t)}{m(t)} + \frac{m(t) - t\, m(t)}{m^2(t)} = \frac{1}{t}$$

or

$$m'(t) \left[\frac{1}{m(t)} - \frac{t}{m^2(t)} \right] + \frac{1}{m(t)} = \frac{1}{t}.$$

Solving the above differential equation we get

$$m(t) = \frac{t}{d} \text{ where } d > 0$$

or

$$r(t) = \left(\frac{1}{d} - 1 \right) t.$$

Using (1.2.21) we get the required form for $\bar{F}(t)$.

Conversely for the Pareto distribution specified in (2.2.7) we have $m(t) = \dfrac{a\, t}{a-1}$ and $\log G(t) = \dfrac{1}{a} + \log t$ from which (2.4.10) is immediate.

2.5 The Theil's entropy

The utility of the Theil's entropy in the context of measurement of income inequality is highlighted in section (1.6). For the random variable $X_t(t) = X \mid X > t$, the Theils entropy defined by (1.6.3) takes the form

$$R_F(t) = \frac{1}{m(t)\bar{F}(t)} \int_t^\infty x \log x\, f(x)\, dx - \log m(t) \qquad (2.5.1)$$

The following theorem focuses attention on the constancy of (2.5.1).

Theorem 2.16

Let X be a non-negative random variable admitting an absolutely continuous distribution function such that $E(X) < \infty$. If $R_F(t)$ denotes the truncated Theils entropy defined in (2.5.1) then the relationship

$$R_F(t) = k \qquad (2.5.2)$$

where k is a constant, holds if and only if X follows the Pareto distribution specified in (2.2.7).

Proof

When (2.5.2) holds we have from (2.5.1)

$$\int_t^\infty x \log x\, f(x)\, dx = (k + \log m(t))\; m(t)\; \bar{F}(t).$$

Differentiating the above equation with respect to t and simplifying the resulting expression we get

$$-t \log t\, h(t) = m'(t) + (k + \log m(t))\left(m'(t) - m(t)\; h(t)\right).$$

In view of (1.2.20) and (1.2.37) the above equation gives

$$t \log t\,(1 + r'(t)) = \left(t\,(k + \log[t + r(t)]) - r(t)\right)(1 + r'(t))$$

Solving the above differential equation we get

$$r(t) = k\,t$$

where k is a constant. Using the relation (1.2.21) we get (2.2.7) as claimed. Conversely, for the Pareto distribution specified in (2.2.7), by direct calculations we get

$$R_F(t) = \frac{1}{a-1} + \log \frac{a-1}{a}$$

and the sufficiency part follows.

Chapter III

The bivariate Gini-index

3.1 Introduction

The need for including more than one attribute in the analysis of economic inequality is emphasized in the works of Atkinson and Bourguignon (1982,1989), Kolm (1977), Maasoumi (1986), Maasoumi and Nickelsburg (1988), Mosler (1994a), Rietveld (1990) and Slottje (1987). Arnold (1987) has given a definition for the Lorenz curve in the bivariate setup. The problem of extending the Gini-index to higher dimensions was also considered by Koshevoy and Mosler (1996,1997). The utility of the truncated form of the Gini-index in the univariate setup is being mentioned in section (1.4). In the present chapter, we propose a measure of income inequality for the truncated variable in the bivariate setup and look into the problem of characterizing certain bivariate probability distributions using this measure.

3.2 Bivariate Gini-index

As pointed out in section (1.2), in the reliability context, the failure rate for a two dimensional random vector is defined in two ways. [(1.2.13) and (1.2.14)]. Analogous to the vector valued failure rate, we propose a definition for the Gini-index in the bivariate setup and look into the problem of characterization of probability distributions by the form of the bivariate Gini-index.

Let $X=(X_1,X_2)$ represent a bivariate random vector, where X_1 and X_2 represents two attributes of measuring income in a population. The random variable $Y_1=X_1| \ X_2>t_2$ corresponds to the distribution of X_1 subject to the condition that X_2 is greater than an amount equal to

t_2. Using the terminology used by Ord, Patil and Taillie (1983), quoted in section (1.4), one can define the Gini-index for the random variable Y_1 as

$$G_1(t_1,t_2) = 2\int_{t_1}^{\infty} F(x_1,t_1,t_2)\, dF_1(x_1,t_1,t_2) - 1 \qquad (3.2.1)$$

where $F(x_1,t_1,t_2)$ is the distribution function of Y_1 namely

$$F(x_1,t_1,t_2) = \int_{t_1}^{x_1} \frac{f(y_1 \mid X_2 > t_2)}{\bar{F}(t_1 \mid X_2 > t_2)}\, dy_1$$

and $F_1(x_1,t_1,t_2)$ is the first moment distribution of Y_1 given by

$$F_1(x_1,t_1,t_2) = \frac{\int_{t_1}^{x_1} y_1 \frac{f(y_1 \mid X_2 > t_2)}{\bar{F}(t_1 \mid X_2 > t_2)}\, dy_1}{\int_{t_1}^{\infty} y_1 \frac{f(y_1 \mid X_2 > t_2)}{\bar{F}(t_1 \mid X_2 > t_2)}\, dy_1}.$$

Similarly for the random variable $Y_2 = X_2 \mid X_1 > t_1$ the Gini-index turns out to be

$$G_2(t_1,t_2) = 2\int_{t_2}^{\infty} F(x_2,t_1,t_2)\, dF_2(x_2,t_1,t_2) - 1 \qquad (3.2.2)$$

where $F(x_2,t_1,t_2)$ is the distribution function of Y_2 defined by

$$F(x_2,t_1,t_2) = \int_{t_2}^{x_2} \frac{f(y_2 \mid X_1 > t_1)}{\bar{F}(t_2 \mid X_1 > t_1)}\, dy_2$$

and $F_2(x_2,t_1,t_2)$ is the first moment distribution of Y_2 given by

$$F_2(x_2,t_1,t_2) = \frac{\int_{t2}^{x_2} y_2 \frac{f(y_2 \mid X_1 > t_1)}{\bar{F}(t_2 \mid X_1 > t_1)}\, dy_2}{\int_{t_2}^{\infty} y_2 \frac{f(y_2 \mid X_1 > t_1)}{\bar{F}(t_2 X_1 > t_1)}\, dy_2} \quad .$$

Definition 3.1

For a non-negative random vector $X = (X_1, X_2)$ admitting an absolutely continuous distribution function, we define the bivariate Gini index for the truncated distribution as the vector

$$G(t_1,t_2) = \left(G_1(t_1,t_2), G_2(t_1,t_2) \right) \qquad (3.2.3)$$

where $G_1(t_1,t_2)$ and $G_2(t_1,t_2)$ are defined as in (3.2.1) and (3.2.2) respectively.

Let $X=(X_1,X_2)$ represent two attributes of income, say income from the land and income from the employment. Suppose among the population of individuals whose income in one of the components, say X_2, exceeds a certain threshold value say t_2, then $G_1(t_1,t_2)$ measures the disparity of income in source one. Similarly, $G_2(t_1,t_2)$ measures the disparity of income from the second source subject to the condition that the income from the other source exceeded a threshold value say t_1. Hence $G(t_1,t_2)$ can be viewed as a measure of inequality when the two factors are taken into consideration simultaneously.

3.3 Characterization Theorems

In this section, we discuss characterization theorems associated with some bivariate models based on the functional form of the bivariate truncated Gini-index. We first establish a relationship between the Gini-index defined by (3.2.3) and the vitality function defined by (1.2.39), which is useful for the calculation of bivariate Gini-index for particular distributions as well as for establishing characterization theorems, in the sequel.

Theorem 3.1

Let $X=(X_1,X_2)$ be a non-negative, non-degenerate random vector admitting an absolutely continuous distribution function. If $m_i(t_1,t_2), i=1,2$ represents the components of the bivariate vitality function defined by (1.2.40) and $G_i(t_1,t_2)$, $i=1,2$ represents the components of the bivariate Gini index defined by (3.2.3), then the following relationship holds.

$$(1-G_i(t_1,t_2))\ m_i(t_1,t_2) = t_i + \frac{1}{\overline{F}^2(t_1,t_2)} \int_{t_i}^{\infty} \overline{F}^2(x_i,t_2)\, dx_i, \quad i = 1,2 \quad (3.3.1)$$

Proof

From the definition (3.2.1), we get

$$-\frac{G_1(t_1,t_2)}{\overline{F}(t_1,t_2)}\int_{t_1}^{\infty} x_1 \frac{\partial}{\partial x_1}\overline{F}(x_1,t_2)\,dx_1 = 2\int_{t_1}^{\infty} x_1\left(\frac{\overline{F}(t_1,t_2)-\overline{F}(x_1,t_2)}{\overline{F}(t_1,t_2)}\right)\frac{-\frac{\partial}{\partial x_1}\overline{F}(x_1,t_2)}{\overline{F}(t_1,t_2)}\,dx_1$$

$$+\int_{t_1}^{\infty}\frac{x_1\frac{\partial}{\partial x_1}\overline{F}(x_1,t_2)}{\overline{F}(t_1,t_2)}\,dx_1$$

The above equation can be written as

$$(1-G_1(t_1,t_2))\,\frac{1}{\overline{F}(t_1,t_2)}\int_{t_1}^{\infty} x_1\frac{\partial}{\partial x_1}\overline{F}(x_1,t_2)\,dx_1 = \frac{2}{\overline{F}^2(t_1,t_2)}\int_{t_1}^{\infty} x_1\,\overline{F}(x_1,t_2)\frac{\partial}{\partial x_1}\overline{F}(x_1,t_2)\,dx_1 \tag{3.3.2}$$

or

$$(1-G_1(t_1,t_2))\,\frac{1}{\overline{F}(t_1,t_2)}\left(-t_1\overline{F}(t_1,t_2)-\int_{t_1}^{\infty}\overline{F}(x_1,t_2)dx_1\right) =$$

$$\frac{2}{\overline{F}^2(t_1,t_2)}\left(-t_1\overline{F}^2(t_1,t_2)-\int_{t_1}^{\infty}\overline{F}^2(x_1,t_2)dx_1\right).$$

This gives

$$(1-G_1(t_1,t_2))\left(t_1+\frac{1}{\overline{F}(t_1,t_2)}\int_{t_1}^{\infty}\overline{F}(x_1,t_2)\,dx_1\right) = t_1+\frac{1}{\overline{F}^2(t_1,t_2)}\int_{t_1}^{\infty}\overline{F}^2(x_1,t_2)\,dx_1 .$$

The above equation can be written as

$$(\,1-G_1(t_1,t_2)\,)\,(\,t_1+r_1(t_1,t_2)\,) = t_1+\frac{1}{\overline{F}^2(t_1,t_2)}\int_{t_1}^{\infty}\overline{F}^2(x_1,t_2)\,dx_1 .$$

(3.3.1) with $i=1$ is immediate from the above equation. The proof for $i=2$ is similar.

In the following theorem we look into the property of the bivariate Gini-index from the point of view of truncation invariance.

Theorem 3.2

Let $X=(X_1,X_2)$ be a non-negative random vector admitting an absolutely continuous distribution with respect to Lebesgue measure in the support of $(a_1,\infty)\times(a_2,\infty)$. The relation

$$G_i(t_1,t_2) = \frac{1}{a_i + b_i \log t_j} \qquad i,j = 1,2,\ i \neq j \qquad (3.3.3)$$

where a_i, b_i are constants, holds for all real $t_1, t_2 \geq 0$ if and only if X is distributed as the bivariate Pareto type- I distribution with survival function specified by

$$\bar{F}(t_1,t_2) = \left(\frac{t_1}{a_1}\right)^{-\alpha_1}\left(\frac{t_2}{a_2}\right)^{-\alpha_2}\left(\frac{t_1}{a_1}\right)^{-\theta\log\left(\frac{t_2}{a_2}\right)} ,t_i > a_i,\ 0 < a_i < \infty\ ,i=1,2 \quad (3.3.4)$$

Proof

When (3.3.3) holds, with $i=1$, using equation (3.3.2) we get

$$\left(1-\frac{1}{a_1+b_1\log t_2}\right)\bar{F}(t_1,t_2)\int_{t_1}^{\infty} x_1 \frac{\partial}{\partial x_1}\bar{F}(x_1,t_2)\,dx_1 = 2\int_{t_1}^{\infty} x_1 \bar{F}(x_1,t_2)\frac{\partial}{\partial x_1}\bar{F}(x_1,t_2)\,dx_1.$$

Differentiating with respect to t_1 and rearranging the terms we get

$$\left(1 + \frac{1}{a_1 + b_1 \log t_2}\right) t_1 \bar{F}(t_1,t_2) = -\int_{t_1}^{\infty} x_1 \frac{\partial}{\partial x_1}\bar{F}(x_1,t_2)\,dx_1. \qquad (3.3.5)$$

Differentiating (3.3.5) with respect to t_1 we get

$$\left(1 + \frac{1}{a_1 + b_1 \log t_2}\right)\left(\bar{F}(t_1,t_2) + t_1\frac{\partial}{\partial t_1}\bar{F}(t_1,t_2)\right) = t_1\frac{\partial}{\partial t_1}\bar{F}(t_1,t_2).$$

This gives

$$\frac{t_1}{\bar{F}(t_1,t_2)}\frac{\partial\bar{F}(t_1,t_2)}{\partial t_1} + (1 + a_1 + b_1 \log t_2) = 0.$$

The solution of the above partial differential equation is

$$\bar{F}(t_1,t_2) = t_1^{-1-a_1-b_1\log t_2}\, c_1(t_2) \qquad (3.3.6)$$

where $c_1(t_2)$ is independent t_1. Proceeding on similar lines with $i=2$ in (3.3.3) one can also obtain

$$\bar{F}(t_1,t_2) = t_2^{-1-a_2-b_2\log t_1}\, c_2(t_1) \qquad (3.3.7)$$

where $c_2(t_1)$ is independent of t_2.

When $t_1 = 1$, in (3.3.6) we get

$$c_1(t_2) = \overline{F}_2(t_2)$$

Thus, (3.3.6) becomes

$$\overline{F}(t_1, t_2) = \overline{F}_2(t_2)\ t_1^{-1-a_1-b_1 \log t_2} \qquad (3.3.8)$$

In a similar manner, when $t_2 = 1$, (3.3.7) reads as

$$c_2(t_1) = \overline{F}_1(t_1)$$

so that

$$\overline{F}(t_1, t_2) = \overline{F}_1(t_1)\ t_2^{-1-a_2-b_2 \log t_1} \qquad (3.3.9)$$

When $t_2 = 1$, in (3.3.8) we get

$$\overline{F}_1(t_1) = t_1^{-1-a_1}$$

and from (3.3.9) we have

$$\overline{F}(t_1, t_2) = t_1^{-1-a_1}\ t_2^{-1-a_2-b_2 \log t_1} \qquad (3.3.10)$$

Similarly when $t_1 = 1$ in (3.3.10) we have

$$\overline{F}_2(t_2) = t_2^{-1-a_2}$$

and from (3.3.8) we get

$$\overline{F}(t_1, t_2) = t_2^{-1-a_2}\ t_1^{-1-a_1-b_1 \log t_2} \qquad (3.3.11)$$

From (3.3.10) and (3.3.11) we get

$$t_1^{-1-a_1}\ t_2^{-1-a_2-b_2 \log t_1} = t_2^{-1-a_2}\ t_1^{-1-a_1-b_1 \log t_2}.$$

Taking logarithm on both sides and rearranging the terms in the above equation, we get

$$b_1 = b_2 = b \text{ (say)}$$

From (3.3.11) we get the desired form for $\overline{F}(t_1, t_2)$.

Conversely, when the distribution of X is specified by (3.3.4), by direct calculations we get

$$G_i(t_1, t_2) = \frac{1}{2\alpha_i - 1 + 2\theta \log\left(\frac{t_j}{a_j}\right)}, \; i, j = 1,2, \; i \neq j.$$

so that the conditions of the theorem are satisfied.

In the following theorem, we look into the situation where $(1-G_i(t_1,t_2))\ m_i(t_1,t_2)$ is linear in t_i, $i=1,2$.

Theorem 3.3

Let $X=(X_1,X_2)$ be a non-negative, non-degenerate random vector admitting an absolutely continuous distribution function. The relationship

$$(1-G_i(t_1,t_2))\ m_i(t_1,t_2) = At_i + B_i(t_j)\ ,i,j=1,2\ ,i\neq j \qquad (3.3.12)$$

where $B_i(t_j)$ are non-negative functions of t_j holds for all $t_1,t_2\geq 0$ if and only if X follows

(i) the Gumbels bivariate exponential distribution with survival function

$$\overline{F}(t_1,t_2) = e^{-\lambda_1 t_1-\lambda_2 t_2-\theta t_1 t_2}\ ,\ \lambda_1,\lambda_2>0\ ,t_1,t_2>0\ ,0\leq\theta\leq\lambda_1\lambda_2 \qquad (3.3.13)$$

if $A = 1$

(ii) the bivariate Pareto type-II distribution with survival function

$$\overline{F}(t_1,t_2) = (1+a_1t_1+a_2t_2+bt_1t_2)^{-c}\ ,t_1,t_2>0\ ,a_1,a_2,c>0$$

$$,0\leq b\leq(c+1)\,a_1a_2 \qquad (3.3.14)$$

if $A > 1$ and

(iii) the bivariate finite-range distribution with survival function

$$\overline{F}(t_1,t_2) = (1-p_1t_1-p_2t_2+qt_1t_2)^{d}\ ,\ 0<t_1<\frac{1}{p_1},0<t_2<\frac{1-p_1t_1}{p_2-qt_1}$$

$$,p_1,\ p_2>0\ ,\ 1-d\leq\frac{q}{p_1p_2}\leq 1\ ,d>0 \qquad (3.3.15)$$

if $A < 1$.

Proof

When (3.3.12) holds with $i=1$, using (3.3.1) we have

$$t_1 + \frac{1}{\overline{F}^2(t_1,t_2)}\int_{t_1}^{\infty}\overline{F}^2(x_1,t_2)\,dx_1 = At_1 + B_1(t_2)$$

or

$$\int_{t_1}^{\infty} \bar{F}^2(x_1,t_2)\,dx_1 = \left((A-1)t_1 + B_1(t_2) \right) \bar{F}^2(t_1,t_2).$$

Differentiating with respect to t_1, we get

$$-\bar{F}^2(t_1,t_2) = (A-1)\,\bar{F}^2(t_1,t_2) + \left((A-1)\,t_1 + B_1(t_2) \right) 2\bar{F}(t_1,t_2)\,\frac{\partial}{\partial t_1}\bar{F}(t_1,t_2)$$

The above equation simplifies to

$$-\frac{1}{\bar{F}(t_1,t_2)}\,\frac{\partial \bar{F}(t_1,t_2)}{\partial t_1} = \frac{A}{2(A-1)\,t_1 + B_1(t_2)}.$$

Denoting by $h = \left(h_1(t_1,t_2)\,, h_2(t_1,t_2) \right)$, the vector valued failure rate discussed in Johnson and Kotz (1975), using (1.2.15) the above equation gives

$$h_1(t_1,t_2) = \frac{1}{\dfrac{2\,(A-1)}{A}\,t_1 + \dfrac{2\,B_1(t_2)}{A}}.$$

Proceeding on similar lines with $i=2$, we also get

$$h_2(t_1,t_2) = \frac{1}{\dfrac{2\,(A-1)}{A}\,t_2 + \dfrac{2\,B_2(t_1)}{A}}.$$

The above expressions for $h_j(t_1,t_2)\,, j = 1, 2$ are reciprocal linear in t_j. The rest of the proof follows from Roy (1989) reviewed in section 1.2.

The if part of the theorem follows from the expressions for $(1-G_1(t_1,t_2))\;m_1(t_1,t_2)$ given by $t_1 + \dfrac{1}{2(\lambda_1 + t_2\theta)}$, $\left(\dfrac{2c}{2c-1}\right)t_1 + \dfrac{(1+a_2t_2)}{(2c-1)(a_1+bt_2)}$ and $\left(\dfrac{2d}{2d+1}\right)t_1 + \dfrac{(1-p_2t_2)}{(2d+1)(p_1-qt_2)}$ respectively for distributions specified by (3.3.13), (3.3.14) and (3.3.15) with similar expression for $(1-G_2(t_1,t_2))\;m_2(t_1,t_2)$. Hence the condition of the theorem holds.

Corollary 3.1

When $\theta = 0$ in (3.3.13), we have

$$\bar{F}(t_1,t_2) = \exp(\,-\lambda_1 t_1 - \lambda_2 t_2\,) \qquad (3.3.16)$$

so that X_1 and X_2 are independent and exponentially distributed. In this set up the relation takes the form

$$(1-G_i(t_1,t_2))\ m_i(t_1,t_2) = t_i + \frac{1}{2\lambda_i}, \quad i=1,2$$

which is characteristic to (3.3.16).

Corollary 3.2

When $b=0$ in (3.3.14), we have

$$\overline{F}(t_1,t_2) = (1 + a_1 t_1 + a_2 t_2)^{-c}, t_1,t_2>0, c>0 \qquad (3.3.17)$$

which is the model obtained by Lindley and Singpurwalla (1986) under a different set of conditions. In this case, the property

$$(1-G_i(t_1,t_2))\ m_i(t_1,t_2) = \left(\frac{2c}{2c-1}\right) t_i + \frac{(1+a_j t_j)}{a_i(2c-1)}, i,j=1,2, i\neq j.$$

is characteristic to (3.3.17). It may be noted that the right-hand side of the above equation is a linear function of t_1 and t_2.

The following theorem provides a characterization for the three distributions considered in the above theorem based on the relationship between the bivariate Gini-index defined in (3.2.3) and the bivariate mean residual life function defined in (1.2.25).

Theorem 3.4

For the random vector X considered in theorem 3.3 the relationship

$$(1-G_i(t_1,t_2))\ m_i(t_1,t_2) = t_i + k\, r_i(t_1,t_2), i=1,2, i\neq j, k>0. \qquad (3.3.18)$$

holds for all $t_1, t_2 \geq 0$ if and only if X follows any one of the three distributions specified by (3.3.13), (3.3.14) and (3.3.15) respectively according as $k=\frac{1}{2}$, $k<\frac{1}{2}$ and $k>\frac{1}{2}$.

Proof

When (3.3.18) holds with $i=1$, using (3.3.1), we have

$$t_1 + \frac{1}{\overline{F}^2(t_1,t_2)} \int_{t_1}^{\infty} \overline{F}^2(x_1,t_2)\, dx_1 = t_1 + k\, r_1(t_1,t_2)$$

or

$$\int_{t_1}^{\infty} \overline{F}^2(x_1,t_2)\, dx_1 = k\, r_1(t_1,t_2)\, \overline{F}^2(t_1,t_2).$$

Differentiating with respect to t_1 we get

$$2\, k\, r_1(t_1,t_2) \frac{1}{\overline{F}(t_1,t_2)} \frac{\partial \overline{F}(t_1,t_2)}{\partial t_1} + k \frac{\partial r_1(t_1,t_2)}{\partial t_1} = -1.$$

Using the relationships (1.2.15) and (1.2.30), the above equation can be written as

$$h_1(t_1,t_2)\; r_1(t_1,t_2) = \frac{1-k}{k}.$$

Proceeding on similar lines with $i=2$, one can also get

$$h_2(t_1,t_2)\; r_2(t_1,t_2) = \frac{1-k}{k}.$$

The rest of the proof follows from Roy (1989), reviewed in section 1.2.

The if part of the theorem follows from the expressions for $(\,1-G_1(t_1,t_2)\,)\; m_1(t_1,t_2)$ and $r_1(t_1,t_2)$ given below with similar expressions for $(\,1-G_2(t_1,t_2)\,)\; m_2(t_1,t_2)$ and $r_2(t_1,t_2)$.

Distribution	$(\,1-G_1(t_1,t_2)\,)\; m_1(t_1,t_2)$	$r_1(t_1,t_2)$
(i) exponential	$t_1+\frac{1}{2}\left(\frac{1}{\lambda_1+t_2\theta}\right)$	$\frac{1}{\lambda_1+\theta t_2}$
(ii) Pareto	$t_1+\frac{c-1}{2c-1}\left(\frac{t_1}{c-1}+\frac{(1+a_2t_2)}{(a_1+bt_2)(c-1)}\right)$	$\frac{t_1}{c-1}+\frac{1+a_2t_2}{(a_1+bt_2)(c-1)}$
(iii) finite-range	$t_1+\frac{d+1}{2d+1}\left(\frac{-t_1}{d+1}+\frac{(1-p_2t_2)}{(p_1-qt_2)(d+1)}\right)$	$\frac{-t_1}{d+1}+\frac{1-p_2t_2}{(p_1-qt_2)(d+1)}$

Instead of the mean residual life function if we consider the bivariate failure rate, we get a characterization for the three distributions considered in theorem 3.4, which we state as theorem 3.5 below.

Theorem 3.5

For the random vector X considered in theorem 3.4, the relationship

$$(1-G_i(t_1,t_2))\; m_i(t_1,t_2) = t_i + k\,\frac{1}{h_i(t_1,t_2)},\quad i=1,2 \tag{3.3.19}$$

holds for all $t_1,t_2 \geq 0$ if and only if X follows any one of the three distributions specified by (3.3.13),(3.3.14) and (3.3.15) respectively according as $k = \frac{1}{2}$, $k > \frac{1}{2}$ and $k < \frac{1}{2}$.

Proof

When (3.3.19) holds using (3.3.1) we have

$$t_1 + \frac{1}{\bar{F}^2(t_1,t_2)} \int_{t_1}^{\infty} \bar{F}^2(x_1,t_2)\, dx_1 = t_1 + \frac{k}{h_1(t_1,t_2)}$$

or

$$\int_{t_1}^{\infty} \bar{F}^2(x_1,t_2)\, dx_1 = k\,\bar{F}^2(t_1,t_2)\, z(t_1,t_2)$$

where $z(t_1,t_2) = \dfrac{1}{h_1(t_1,t_2)}$.

Differentiating with respect to t_1, we get

$$-\bar{F}^2(t_1,t_2) = k\,\bar{F}^2(t_1,t_2)\,\frac{\partial z(t_1,t_2)}{\partial t_1} + 2\,k\,\bar{F}(t_1,t_2)\,\frac{\partial \bar{F}(t_1,t_2)}{\partial t_1}\, z(t_1,t_2)$$

or

$$-1 = k\,\frac{\partial z(t_1,t_2)}{\partial t_1} + 2\,k\,z(t_1,t_2)\,\frac{1}{\bar{F}(t_1,t_2)}\,\frac{\partial \bar{F}(t_1,t_2)}{\partial t_1}.$$

Using (1.2.15), the above equation gives

$$-1 = k\,\frac{\partial z(t_1,t_2)}{\partial t_1} - 2\,k\,z(t_1,t_2)\,h_1(t_1,t_2)$$

or

$$\frac{\partial z(t_1,t_2)}{\partial t_1} = \frac{2k-1}{k}.$$

Solving the above partial differential equation, we get

$$z(t_1,t_2) = \left(\frac{2k-1}{k}\right) t_1 + c_1(t_2)$$

where $c_1(t_2)$ is independent t_1. This gives

$$h_1(t_1,t_2) = \frac{1}{\left(\frac{2k-1}{k}\right) t_1 + c_1(t_2)}.$$

Proceeding on similar lines with $i=2$, we get

$$h_2(t_1,t_2) = \frac{1}{\left(\frac{2k-1}{k}\right) t_2 + c_2(t_1)}.$$

The rest of the proof follows from Roy (1989), mentioned in section 1.2.

The if part of the theorem follows from the expressions for $(1-G_1(t_1,t_2))\,m_1(t_1,t_2)$ and $h_1(t_1,t_2)$ given below with similar expressions for $(1-G_2(t_1,t_2))\,m_2(t_1,t_2)$ and $h_2(t_1,t_2)$.

Distribution	$(1-G_1(t_1,t_2))\,m_1(t_1,t_2)$	$h_1(t_1,t_2)$
(i) exponential	$t_1 + \frac{1}{2}\left(\frac{1}{\lambda_1 + t_2\theta}\right)$	$\frac{1}{\lambda_1 + \theta t_2}$
(ii) Pareto	$t_1 + \frac{(1+a_1t_1+a_2t_2+bt_1t_2)}{(2c-1)\,(a_1+bt_2)}$	$\frac{c\,(a_1+bt_2)}{(1+a_1t_1+a_2t_2+bt_1t_2)}$
(iii) finite-range	$t_1 + \frac{1-p_1t_1-p_2t_2+qt_1t_2}{(1+2d)\,(p_1-qt_2)}$	$\frac{d\,(p_1-qt_2)}{1-p_1t_1-p_2t_2+qt_1t_2}$

The following theorem provides a characterization result for the bivariate Pareto type-1 distribution using a possible relationship between the Gini-index and the vitality function in the bivariate setup.

Theorem 3.6

For the random vector X considered in theorem 3.5, the relationship

$$G_i(t_1,t_2) = \frac{m_i(t_1,t_2) - t_i}{m_i(t_1,t_2) + t_i}, i = 1,2 \tag{3.3.20}$$

holds for all $t_1,t_2 \geq 0$ if and only if X follow the bivariate Pareto distribution with survival function (3.3.4).

Proof

When (3.3.20) holds with $i=1$, using (3.3.1), we get

$$\frac{2\, t_1\, m_1(t_1,t_2)}{t_1 + m_1(t_1,t_2)} = t_1 + \frac{1}{\overline{F}^2(t_1,t_2)} \int_{t_1}^{\infty} \overline{F}^2(x_1,t_2)\, dx_1 \tag{3.3.21}$$

or

$$2\, t_1\, m_1(t_1,t_2)\, \overline{F}^2(t_1,t_2) = t_1\, (t_1 + m_1(t_1,t_2))\overline{F}^2(t_1,t_2) + (t_1 + m_1(t_1,t_2)) \int_{t_1}^{\infty} \overline{F}^2(x_1,t_2)\, dx_1 .$$

Differentiating with respect to t_1 and rearranging the terms we get

$$2\, t_1\, m_1(t_1,t_2) \frac{1}{\overline{F}(t_1,t_2)} \frac{\partial \overline{F}(t_1,t_2)}{\partial t_1} + t_1 \frac{\partial m_1(t_1,t_2)}{\partial t_1} + m_1(t_1,t_2) =$$

$$2t_1^2 \frac{1}{\overline{F}(t_1,t_2)} \frac{\partial \overline{F}(t_1,t_2)}{\partial t_1} + 2t_1 - (m_1(t_1,t_2) + t_1) + \left(1 + \frac{\partial m_1(t_1,t_2)}{\partial t_1}\right) \frac{1}{\overline{F}^2(t_1,t_2)} \int_{t_1}^{\infty} \overline{F}^2(x_1,t_2) dx_1 .$$

Using the relationships (1.2.15) and (3.3.21), the above equation can be written as

$$-2t_1 m_1(t_1,t_2)\, h_1(t_1,t_2)(t_1 + m_1(t_1,t_2)) + t_1 (t_1 + m_1(t_1,t_2)) \frac{\partial m_1(t_1,t_2)}{\partial t_1} + m_1(t_1,t_2)(t_1 + m_1(t_1,t_2))$$

$$= -2\, t_1^2\, h_1(t_1,t_2)\, (t_1 + m_1(t_1,t_2)) + t_1\, (t_1 + m_1(t_1,t_2)) - m_1(t_1,t_2)\, (t_1 + m_1(t_1,t_2)) +$$

$$\left(1 + \frac{\partial m_1(t_1,t_2)}{\partial t_1}\right) + \left(t_1 m_1(t_1,t_2) - t_1^2\right).$$

or

$$-2\, t_1\, m_1^2(t_1,t_2)\, h_1(t_1,t_2) + 2\, t_1^2 \frac{\partial}{\partial t_1} m_1(t_1,t_2) + 2\, m_1^2(t_1,t_2) = -2\, t_1^3\, h_1(t_1,t_2).$$

Using (1.2.44) and simplifying, we get

$$-2t_1 m_1^2(t_1,t_2)\frac{\partial}{\partial t_1} m_1(t_1,t_2) + \left(2t_1^2 \frac{\partial}{\partial t_1} m_1(t_1,t_2) + 2m_1^2(t_1,t_2)\right)\left(m_1(t_1,t_2) - t_1\right) = -2t_1^3 \frac{\partial}{\partial t_1} m_1(t_1,t_2)$$

or

$$-2\, t_1 \frac{\partial}{\partial t_1} m_1(t_1,t_2) + 2\, t_1^2 \frac{\partial}{\partial t_1} \log m_1(t_1,t_2) + 2\, m_1(t_1,t_2) - 2t_1 = 0.$$

On solving the above partial differential equation, we get

$$m_1(t_1,t_2) = c_1(t_2)\, t_1$$

where $c_1(t_2)$ is a function of t_2 alone. This gives

$$r_1(t_1,t_2) = (\, c_1(t_2) - 1\,)\, t_1\ ,\ c_1(t_2) \text{ is a function of } t_2.$$

Proceeding on similar lines with $i=2$, in (3.3.20), we get

$$r_2(t_1,t_2) = (\, c_2(t_1) - 1\,)\, t_2\ ,\ c_2(t_1) \text{ is a function of } t_1.$$

Using the pair of identities for the survival function of $X=(X_1,X_2)$ in terms of the components of the bivariate MRLF, specified in (1.2.28), (1.2.29) and inserting the values of $r_1(t_1,t_2)$ and $r_2(t_1,t_2)$, we get

$$\bar{F}(t_1,t_2) = \frac{a_1 a_2}{t_1 t_2} \exp\left\{- \frac{1}{(c_1(a_2)-1)}\log\left(\frac{t_1}{a_1}\right) - \frac{1}{(c_2(t_1)-1)}\log\left(\frac{t_2}{a_2}\right)\right\} \quad (3.3.22)$$

and

$$\bar{F}(t_1,t_2) = \frac{a_1 a_2}{t_1 t_2} \exp\left\{- \frac{1}{(c_1(t_2)-1)}\log\left(\frac{t_1}{a_1}\right) - \frac{1}{(c_2(a_1)-1)}\log\left(\frac{t_2}{a_2}\right)\right\} \quad (3.3.23)$$

Equating (3.3.22) and (3.3.23), we get

$$\exp\left\{ - \frac{1}{(c_1(a_2)-1)} \log\left(\frac{t_1}{a_1}\right) - \frac{1}{(c_2(t_1)-1)} \log\left(\frac{t_2}{a_2}\right)\right\}$$

$$= \exp\left\{ - \frac{1}{(c_1(t_2)-1)} \log\left(\frac{t_1}{a_1}\right) - \frac{1}{(c_2(a_1)-1)} \log\left(\frac{t_2}{a_2}\right)\right\} \quad (3.3.24)$$

Setting $\alpha_i = \dfrac{1}{c_i(a_j)\ -1}$, $i=1,2$, $i \neq j$, (3.3.24) can be written as

$$\exp\left\{-\alpha_1\log\left(\frac{t_1}{a_1}\right)-\frac{1}{(c_2(t_1)-1)}\log\left(\frac{t_2}{a_2}\right)\right\} = \exp\left\{-\frac{1}{(c_1(t_2)-1)}\log\left(\frac{t_1}{a_1}\right)-\alpha_2\log\left(\frac{t_2}{a_2}\right)\right\} \qquad (3.3.25)$$

or

$$\left(\frac{1}{(c_1(t_2)-1)}-\alpha_1\right)\log\left(\frac{t_1}{a_1}\right) = \left(\frac{1}{(c_2(t_1)-1)}-\alpha_2\right)\log\left(\frac{t_2}{a_2}\right).$$

Dividing both sides by $\log\left(\frac{t_1}{a_1}\right)\log\left(\frac{t_2}{a_2}\right)$ we get

$$\left(\frac{1}{(c_1(t_2)-1)}-\alpha_1\right)\left(\log\left(\frac{t_2}{a_2}\right)\right)^{-1} = \left(\frac{1}{(c_2(t_1)-1)}-\alpha_2\right)\left(\log\left(\frac{t_1}{a_1}\right)\right)^{-1}.$$

This means that each quantity should be a constant, say θ, independent of t_1 and t_2. Thus,

$$\frac{1}{(c_1(t_2)-1)} = \alpha_1 + \theta\log\left(\frac{t_2}{a_2}\right)$$

and

$$\frac{1}{(c_2(t_1)-1)} = \alpha_2 + \theta\log\left(\frac{t_1}{a_1}\right).$$

Inserting values of $\frac{1}{(c_1(t_2)-1)}$ and $\frac{1}{(c_2(t_1)-1)}$ in (3.3.24), the required distribution is obtained.

The if part of the theorem follows from the expressions for $G_1(t_1,t_2)$ and $m_1(t_1,t_2)$ given by $\frac{1}{2\alpha_1-1+2\theta\log\left(\frac{t_2}{a_2}\right)}$ and $t_1\left(\frac{\alpha_1+\theta\log\left(\frac{t_2}{a_2}\right)}{\alpha_1+\theta\log\left(\frac{t_2}{a_2}\right)-1}\right)$ respectively,with similar expressions for $G_2(t_1,t_2)$ and $m_2(t_1,t_2)$.

The following theorem provides a characterization result for the three distributions considered in theorem 3.6 based on a functional relationship between the bivaraite Gini-index defined in (3.2.3) and the bivariate vitality function defined in (1.2.39).

Theorem 3.7

For the random vector X considered in theorem 3.6, the relation

$$G_i(t_1,t_2) = k\left(1-\frac{t_i}{m_i(t_1,t_2)}\right) \quad ,i=1,2,\ k>0 \qquad (3.3.26)$$

holds for all $t_1,t_2 \geq 0$ if and only if X follows any one of the three distributions specified by (3.3.13), (3.3.14) and (3.3.15) respectively according as $k=\frac{1}{2}$,$k>\frac{1}{2}$ and $k<\frac{1}{2}$.

Proof

When (3.3.26) holds with $i=1$ using (3.3.1) we have

$$\left(1 - k\left(1-\frac{t_1}{m_1(t_1,t_2)}\right)\right) m_1(t_1,t_2) = t_1 + \frac{1}{\bar{F}^2(t_1,t_2)} \int_{t_1}^{\infty} \bar{F}^2(x_1,t_2)\, dx_1 .$$

Using the relation (1.2.43), we get

$$t_1 + r_1(t_1,t_2) - k\, r_1(t_1,t_2) = t_1 + \frac{1}{\bar{F}^2(t_1,t_2)} \int_{t_1}^{\infty} \bar{F}^2(x_1,t_2)\, dx_1$$

or

$$(1-k)\ r_1(t_1,t_2)\, \bar{F}^2(t_1,t_2) = \int_{t_1}^{\infty} \bar{F}^2(x_1,t_2)\, dx_1 .$$

Differentiating with respect to t_1, we get

$$2(1-k)\ r_1(t_1,t_2)\, \bar{F}(t_1,t_2)\, \frac{\partial \bar{F}(t_1,t_2)}{\partial t_1} + (1-k)\, \frac{\partial r_1(t_1,t_2)}{\partial t_1}\, \bar{F}^2(t_1,t_2) = -\bar{F}^2(t_1,t_2).$$

Using the equations (1.2.30) and (1.2.15), the above equation takes the form

$$-2\ (1-k)\ r_1(t_1,t_2)\ h_1(t_1,t_2) + (1-k)\ \left(r_1(t_1,t_2)\ h_1(t_1,t_2) - 1\right) = -1.$$

This gives,

$$h_1(t_1,t_2)\ r_1(t_1,t_2) = \frac{k}{1-k}.$$

Similar expression can be obtained for $h_2(t_1,t_2)\ r_2(t_1,t_2)$. The rest of the proof is analogous to that of theorem 3.4.

The if part of the theorem follows from the expressions for $G_1(t_1,t_2)$ and $\left(1 - \frac{t_1}{m_1(t_1,t_2)}\right)$ given below.

Distribution	$G_1(t_1,t_2)$	$\left(1 - \frac{t_1}{m_1(t_1,t_2)}\right)$
(i) exponential	$\frac{1}{2(1+t_1\lambda_1+t_1t_2\theta)}$	$\frac{1}{\lambda_1+\theta t_2}$
(ii) Pareto	$\frac{c\,(1+a_1t_1+a_2t_2+bt_1t_2)}{(2c-1)\,(1+a_2t_2+ca_1t_1+bct_1t_2)}$	$\frac{(1+a_1t_1+a_2t_2+bt_1t_2)}{(1+a_2t_2+ca_1t_1+bct_1t_2)}$
(iii) finite-range	$\frac{d\,(1-p_1t_1-p_2t_2+qt_1t_2)}{(1+2d)\,(1+dp_1t_1-p_2t_2-dqt_1t_2)}$	$\frac{(1-p_1t_1-p_2t_2+qt_1t_2)}{(1+dp_1t_1-p_2t_2-dqt_1t_2)}$

Chapter IV

Bivariate residual entropy function

4.1 Introduction

The residual entropy function has been extensively used in Reliability theory as a measure of the stability of a component or a device. Characterization of probability distributions based on certain relationships between the residual entropy function and other Reliability concepts are discussed in Nair and Rajesh (1998). In addition to the common measures of income inequality such as variance, coefficient of variation, Lorenz curve, Gini-index etc, the Shannon's entropy has been advantageously used as a handy tool to measure income inequality. The utility of this measure is highlighted in the works of Theil (1967) and Hart (1971). Ord, Patil and Taillie (1983) has used the truncated form of the entropy measure as a measure for examining the inequality of income of persons whose income exceeds a specified limit. In this chapter we extend this concept to higher dimensions and look into the problem of characterizing some bivariate models based on the functional form of the residual entropy function.

4.2.Bivariate residual entropy function

In the univariate setup, Ebrahimi (1996) defines the residual entropy function as the Shannon's entropy associated with the residual life distribution, namely

$$H(f,t) = -\int_t^\infty \frac{f(x)}{\bar{F}(t)} \log\frac{f(x)}{\bar{F}(t)}\, dx. \qquad (4.2.1)$$

(4.2.1) can also be written as

$$H(f,t) = 1 - \frac{1}{\bar{F}(t)} \int_t^\infty f(x) \log h(x)\, dx. \qquad (4.2.2)$$

One of the main problems encountered while extending a univariate concept to higher dimensions is that it cannot be done in a unique way. Accordingly several extensions are possible for (4.2.1) in the bivariate setup. A natural extension for (4.2.1) to the bivariate setup can be obtained by replacing $f(x)$ by $f(x_1,x_2)$ and $\bar{F}(x)$ by $\bar{F}(x_1,x_2)$. This is given below as definition 4.1

Definition 4.1

Let $X=(X_1,X_2)$ be a non-negative bivariate random vector admitting an absolutely continuous distribution function with probability density function $f(x_1,x_2)$and survival function $\bar{F}(x_1,x_2)$ .For $\underline{t} = (t_1,t_2)$in R_2^+, we define the bivariate residual entropy function through the relation

$$H(f,t_1,t_2) = -\int_{t_1}^{\infty}\int_{t_2}^{\infty} \frac{f(x_1,x_2)}{\bar{F}(t_1,t_2)} \log \frac{f(x_1,x_2)}{\bar{F}(t_1,t_2)}\, dx_1\, dx_2 \quad (4.2.3)$$

If $Y=(Y_1,Y_2)$, where $Y_j = X_j - t_j \mid X_1 > t_1, X_2 > t_2$, $j=1,2$, the survival function and density function of Y are respectively

$$\bar{G}(y_1,y_2) = \frac{\bar{F}(y_1+t_1, y_2+t_2)}{\bar{F}(t_1,t_2)}$$

and

$$g(y_1,y_2) = \frac{f(y_1+t_1, y_2+t_2)}{\bar{F}(t_1,t_2)}$$

The Shannon's entropy associated with Y, namely

$$H(g) = -\int_0^{\infty}\int_0^{\infty} g(y_1,y_2) \log g(y_1,y_2)\, dy_1\, dy_2. \quad (4.2.4)$$

(4.2.4) simplify to (4.2.3) under the transmission $x_j = y_j + t_j$, $j=1,2$. So (4.2.3) can be viewed as the Shannon's entropy associated with the residual life distribution .If $X=(X_1,X_2)$ represents the life times of the components in a two-component system, (4.2.3) can be viewed as a measure of the stability of the system when the components have survived up to time $\underline{t} = (t_1, t_2)$. (4.2.3) can also be written as

$$H(f, t_1, t_2) = \log \bar{F}(t_1, t_2) - \frac{1}{\bar{F}(t_1, t_2)} \int_{t_1}^{\infty}\int_{t_1}^{\infty} f(x_1, x_2) \log f(x_1, x_2)\, dx_1\, dx_2 \quad (4.2.5)$$

In the sequel, we obtain a representation for the residual entropy function in terms of the bivariate failure rate. Denote by

$$b(t_1, t_2) = \frac{f(t_1, t_2)}{\bar{F}(t_1, t_2)} \quad (4.2.6)$$

the (scalar) failure rate [Basu(1971)]and

$$h(t_1, t_2) = (h_1(t_1, t_2), h_2(t_1, t_2)) \quad (4.2.7)$$

with

$$h_j(t_1, t_2) = -\frac{\partial}{\partial t_j} \log \bar{F}(t_1, t_2), \; j = 1, 2$$

the vector valued failure rate (Johnson and Kotz (1975)).Observing that (4.2.3) can be written as

$$H(f, t_1, t_2) = \log \bar{F}(t_1, t_2) - \frac{1}{\bar{F}(t_1, t_2)} \int_{t_1}^{\infty}\int_{t_1}^{\infty} f(x_1, x_2) \log b(x_1, x_2)\, dx_1\, dx_2$$

$$- \frac{1}{\bar{F}(t_1, t_2)} \int_{t_1}^{\infty}\int_{t_1}^{\infty} f(x_1, x_2) \log \bar{F}(x_1, x_2)\, dx_1\, dx_2 \quad (4.2.8)$$

we get

$$H(f, t_1, t_2) = 1 - \frac{1}{\bar{F}(t_1, t_2)} \int_{t_1}^{\infty}\int_{t_1}^{\infty} [h_1(x_1, x_2) h_2(x_1, x_2) - b(x_1, x_2) \log b(x_1, x_2)]\, \bar{F}(x_1, x_2)\, dx_1\, dx_2$$

(4.2.9)

Since $b(t_1, t_2)$ does not determine the distribution uniquely, in view of (4.2.9), the bivariate residual entropy function does not determine the distribution uniquely.

If $X = (X_1, X_2)$ represents the lifetime of the components in a two component system, $Y_1 = X_1 \mid X_2 > t_2$ corresponds to the life length of the first component subject to the condition that the second component has survived up to time t_2. The residual entropy of Y_1 namely

$$H_1(f,t_1,t_2) = -\int_{t_1}^{\infty} \frac{f(x_1 \mid X_2 > t_2)}{\bar{F}(t_1 \mid X_2 > t_2)} \log \frac{f(x_1 \mid X_2 > t_2)}{\bar{F}(t_1 \mid X_2 > t_2)}\, dx_1,$$

simplifies to

$$H_1(f,t_1,t_2) = \log \bar{F}(t_1 \mid X_2 > t_2) - \frac{1}{\bar{F}(t_1 \mid X_2 > t_2)} \int_{t_1}^{\infty} f(x_1 \mid X_2 > t_2) \log f(x_1 \mid X_2 > t_2)\, dx_1 \tag{4.2.10}$$

Observing that $\bar{F}(t_1,t_2) = \bar{F}(t_1 \mid X_2 > t_2)\, \bar{F}_2(t_2)$, where $\bar{F}_i(t_i) = P(X_i > t_i)$, $i = 1,2$ (4.2.10) simplifies to

$$H_1(f,t_1,t_2) = 1 + \frac{1}{\bar{F}(t_1,t_2)} \int_{t_1}^{\infty} \frac{\partial \bar{F}(x_1,t_2)}{\partial x_1} \log h_1(x_1,t_2)\, dx_1 \,. \tag{4.2.11}$$

Similarly, the residual entropy function of $Y_2 = X_2 \mid X_1 > t_1$ simplifies to

$$H_2(f,t_1,t_2) = 1 + \frac{1}{\bar{F}(t_1,t_2)} \int_{t_2}^{\infty} \frac{\partial \bar{F}(t_1,x_2)}{\partial x_1} \log h_1(t_1,x_2)\, dx_2 \,. \tag{4.2.12}$$

Analogous to the definition of vector valued failure, rate we give below an alternative definition for the residual entropy function.

Definition 4.2

For a non-negative random vector $X = (X_1, X_2)$ admitting an absolutely continuous distribution function, the bivariate residual entropy function is defined as the vector

$$H(f,t_1,t_2) = \left(H_1(f,t_1,t_2)\, , H_2(f,t_1,t_2) \right) \tag{4.2.13}$$

where $H_1(f,t_1,t_2)$ and $H_2(f,t_1,t_2)$ are given by (4.2.11) and (4.2.12).

If $X = (X_1, X_2)$ is a bivariate random vector representing the wealth of two populations, with the random variable X_1 representing the wealth of the first population and the random variable X_2 representing the wealth of the second population, then $H_1(f,t_1,t_2)$ measures the expected uncertainty contained in the distribution of the amount of wealth of the first population subject to the condition that the wealth of the second population exceeds a threshold t_2. Similarly $H_2(f,t_1,t_2)$ measures the expected uncertainty contained in the distribution of the amount of wealth of the second population subject to the condition that the wealth of the first population exceeds a threshold t_1. Hence $H(f,t_1,t_2)$can be viewed as a measure of the expected uncertainty contained in the distribution of the amount of wealth when the two factors are taken into consideration simultaneously.

4.3 Characterization theorems

In this section, we look into the problem of characterizing certain bivariate models based on the functional form of $H(f,t_1,t_2)$. We first examine the situation where $H(f,t_1,t_2)$ is constant in t.

Theorem 4.1

Let $X=(X_1,X_2)$be a non-negative, non-degenerate random vector admitting an absolutely continuous distribution function with respect to Lebsegue measure such that $H(f,t_1,t_2) < \infty$. A relation of the form

$$H(f,t_1,t_2) = (\ c_1\ ,c_2\ ) \tag{4.3.1}$$

holds if and only if X is distributed as the bivariate exponential distribution with independent (exponential) marginals.

Proof

When (4.3.1) holds with $i=1$, using (4.2.11) we have

$$(c_1-1)\ \overline{F}(t_1,t_2) = \int_{t_1}^{\infty} \frac{\partial\overline{F}(x_1,t_2)}{\partial x_1} \log h_1(x_1,t_2)\ dx_1\,.$$

Differentiating the above equation with respect to t_1, we get

$$(c_1-1)\ \frac{\partial\overline{F}(t_1,t_2)}{\partial t_1} = -\ \frac{\partial\overline{F}(t_1,t_2)}{\partial t_1}\ \log h_1(t_1,t_2)\,.$$

Since $\frac{\partial \overline{F}(t_1,t_2)}{\partial t_1} > 0$, the above equation gives

$$\log h_1(t_1,t_2) = \lambda_1 > 0$$

where $\lambda_1 = e^{(1-c_1)}$. Proceeding along similar lines with (4.2.12), we get

$$h_2(t_1,t_2) = e^{(1-c_2)} = \lambda_2 > 0.$$

From Galambos and Kotz (1978), we have

$$\overline{F}(t_1,t_2) = \exp(-\lambda_1 t_1 - \lambda_2 t_2), t_1,t_2 \geq 0. \qquad (4.3.2)$$

as claimed.

Conversely, when the distribution of X is specified by (4.3.2), by direct calculations we get

$$H_i(f,t_1,t_2) = 1 - \log\lambda_i, i=1,2$$

so that the condition of the theorem holds.

The following theorem looks into the situation where $H(f,t_1,t_2)$ is log linear in t_i.

Theorem 4.2

For the random vector X considered in the theorem 4.1, the relation

$$H_i(f,t_1,t_2) = \log(at_i + b_i(t_j)), i,j = 1,2, i \neq j \qquad (4.3.3)$$

where $b_i(t_j)$ are non-negative, non-increasing functions of $t_j > 0$ holds if and only if X is distributed as

(i) the Gumbels bivariate exponential distribution with survival function

$$\overline{F}(t_1,t_2) = e^{-\lambda_1 t_1 - \lambda_2 t_2 - \theta t_1 t_2}, \lambda_1, \lambda_2 > 0, t_1, t_2 > 0, 0 \leq \theta \leq \lambda_1\lambda_2 \qquad (4.3.4)$$

if $a = 0$

(ii) the bivariate Pareto type-II distribution with survival function

$$\overline{F}(t_1,t_2) = (1 + a_1 t_1 + a_2 t_2 + b t_1 t_2)^{-c}, t_1, t_2 > 0, a_1, a_2, c > 0$$
$$, 0 \leq b \leq (c+1) a_1 a_2 \qquad (4.3.5)$$

if $a > 0$ and

(iii) the bivariate finite-range distribution with survival function

$$\bar{F}(t_1,t_2) = (1 - p_1t_1 - p_2t_2 + q\,t_1t_2)^{d},\ 0 < t_1 < \frac{1}{p_1}, 0 < t_2 < \frac{1-p_1t_1}{p_2 - qt_1}$$

$$,\ p_1,\ p_2 > 0,\ 1-d \le \frac{q}{p_1p_2} \le 1, d > 0 \quad (4.3.6)$$

if $a < 0$.

Proof

When (4.3.3) holds with $i=1$, we have

$$\bar{F}(t_1,t_2)\ \log(at_1 + b_1(t_2)) = \bar{F}(t_1,t_2) + \int_{t_1}^{\infty} \frac{\partial \bar{F}(x_1,t_2)}{\partial x_1} \log h_1(x_1,t_2)\, dx_1 \ . \quad (4.3.7)$$

Differentiating (4.3.7) with respect to t_1 and rearranging the terms, we get

$$h_1(t_1,t_2)\left[\ \log(at_1 + b_1(t_2)) + \log h_1(t_1,t_2)\ - 1\ \right] = \frac{a}{at_1 + b_1(t_2)}. \quad (4.3.8)$$

Denoting by

$$c_1(t_1,t_2) = h_1(t_1,t_2)\ [at_1 + b_1(t_2)], \quad (4.3.9)$$

(4.3.8) takes the form

$$c_1(t_1,t_2)\ [\log c_1(t_1,t_2) - 1] = a \quad (4.3.9)$$

Differentiating with respect to t_1, (4.3.9) reads as

$$\log c_1(t_1,t_2)\ \frac{\partial c_1(t_1,t_2)}{\partial t_1} = 0 \quad (4.3.10)$$

The solution to (4.3.10) is

$$c_1(t_1,t_2) = z_1(t_2) \quad (4.3.11)$$

where $z_1(t_2)$ is independent t_1.

Differentiating (4.3.9) with respect to t_2 and proceeding along the same line, we also get

$$c_1(t_1,t_2) = z_2(t_1) \quad (4.3.12)$$

where $z_2(t_1)$ is a function of t_1 alone.

For (4.3.12) and (4.3.11) to hold simultaneously, we should have

$$c_1(t_1,t_2) = k_1 \quad (4.3.13)$$

where k_1 is a constant. Similarly, when (4.3.3) holds with $i=2$, we also get

$$c_1(t_1,t_2) = k_2 \tag{4.3.14}$$

where k_2 is a constant.

We now show that the values of k_1 and k_2 in equations (4.3.13) and (4.3.14) are the same. With $c_1(t_1,t_2)=k_1$, (4.3.9) read as

$$k_1\ (\log k_1 - 1) = A.$$

Similarly, we can also have

$$k_2\ (\log k_2 - 1) = A.$$

This gives

$$\frac{k_1}{k_2} = \frac{\log k_2 - 1}{\log k_1 - 1}$$

If $k_1 > k_2$, $\frac{k_1}{k_2} > 1$ so that

$$\frac{\log k_2 - 1}{\log k_1 - 1} > 1$$

or

$$\log k_2 - 1 > \log k_1 - 1$$

since $\log k_2 - 1$ and $\log k_1 - 1$ must be of the same sign for $\frac{k_1}{k_2}>1$. This gives $k_2 > k_1$ which is a contradiction. Similarly $k_1 < k_2$ also leads to a contradiction. Assume

$$k_1 = k_2 = k\text{(say)}.$$

This gives

$$c_1(t_1,t_2) = k$$

so that

$$h_1(t_1,t_2) = \frac{k}{at_1 + b_1(t_2)}.$$

Similarly, we can also have

$$h_2(t_1,t_2) = \frac{k}{at_2 + b_2(t_1)}.$$

The rest of the proof follows from Roy (1989), reviewed in section 1.2. The if part of the theorem follows from the expressions for $H_1(f,t_1,t_2)$ given by $1-\log(\alpha_1+\theta t_2), 1-\log\left(\frac{e^{1+\frac{1}{c}}}{c}t_1 + \frac{e^{1+\frac{1}{c}}}{c}\frac{1+a_2t_2}{c(a_1+bt_2)}\right)$ and $\log\left(\frac{-e^{1-\frac{1}{d}}}{d}t_1 + \frac{e^{1-\frac{1}{d}}}{d}\frac{1-p_2t_2}{(p_2-qt_2)}\right)$ respectively for distributions specified by (4.3.4), (4.3.5) and (4.3.6) with the similar expression for $H_2(f,t_1,t_2)$. Hence the condition of the theorem holds.

The following theorem gives a characterization of $H_i(f,t_1,t_2)$ with the bivariate failure rate $h_i(t_1,t_2), i=1,2$.

Theorem 4.3

For the random vector X considered in theorem 4.2, a relationship of the form

$$H_i(f,t_1,t_2) = k - \log h_i(t_1,t_2) \quad ,i=1,2 \qquad (4.3.15)$$

where $h_i(t_1,t_2)$ are the components of the bivariate failure rate, holds for all real $t_1,t_2 \geq 0$ if and only if X follow any one of the three distributions specified by (4.3.4), (4.3.5) and (4.3.6) respectively according as $k=1, k>1$ and $k<1$.

Proof

When (4.3.15) holds, using (4.2.11), we can write

$$[k-\log h_1(t_1,t_2)]\,\bar{F}(t_1,t_2) = \bar{F}(t_1,t_2) + \int_{t_1}^{\infty} \frac{\partial \bar{F}(x_1,t_2)}{\partial x_1}\log h_1(x_1,t_2)\,dx_1.$$

Differentiating with respect to t_1 and rearranging the terms, we get

$$-\frac{1}{h_1^2(t_1,t_2)}\frac{\partial h_1(t_1,t_2)}{\partial t_1} = k-1.$$

If $u(t_1,t_2) = \dfrac{1}{h_1(t_1,t_2)}$, the above equation turn out to be

$$\frac{\partial u(t_1,t_2)}{\partial t_1} = k-1$$

whose solution is

$$u(t_1,t_2) = (k-1)\, t_1 + c_1$$

where c_1 is a constant. This gives,

$$h_1(t_1,\, t_2) = [(k-1)\, t_1 + c_1]^{-1}.$$

Proceeding along similar lines, one can also get

$$h_2(t_1,\, t_2) = [(k-1)\, t_2 + c_2]^{-1}.$$

where c_2 is a constant. This shows that the components of the vector valued failure rate are reciprocal linear. The rest of the proof is analogous to that of theorem 4.2.

The if part follows from the expression for $H_1(f_1,\, t_1,\, t_2)$ given by

$1\text{-}\log(\alpha_1+\theta t_2)$, $1+\dfrac{1}{c}+\log\left(\dfrac{1+a_1t_1+a_2t_2+b\,t_1\,t_2}{c\,(a_1+bt_2)}\right)$ and

$1\text{-}\dfrac{1}{d}+\log\left(\dfrac{1-p_1t_1-p_2t_2+qt_1t_2}{d(p_1-qt_2)}\right)$ and that of $h_1(t_1,t_2)$ given by $\alpha_1+\theta t_2$,

$\dfrac{c\,(a_1+bt_2)}{1+a_1\,t_1+a_2\,t_2+b\,t_1t_2}$ and $\dfrac{d\,(p_1-q\,t_2)}{1-p_1\,t_1-p_2\,t_2+q\,t_1\,t_2}$ when the distribution is specified by (4.3.4), (4.3.5) and (4.3.6) respectively ,with similar expression for $H_2(f,\, t_1,\, t_2)$ and $h_2(t_1,\, t_2)$.

Chapter V

Bivariate Geometric Vitality function

5.1 Introduction

The vitality function, extensively studied by Kupka and Loo (1989) in connection with their studies on ageing process, provides a useful tool in modelling lifetime data. Kotz and shanbag (1980) has used this concept, without specifying the name, to obtain characterization results for some lifetime distributions. Where as the hazard rate reflects the risk of sudden death within a life span, the vitality function provides a more realistic measure of the failure pattern in the sense that it is expressed in terms of increased average life span. The vitality function, defined by

$$m(t) = E(X | X > t) \qquad (5.1.1)$$

can be interpreted as the average lifespan of components whose age exceeds t. Bhattacharjee (1993) points out the relevance of this concept in income studies. (5.1.1) can be viewed as the average income of persons whose income exceeds the level t. Based on the geometric mean of the residual lifetime of the components, Nair and Rajesh (2000) has introduced a new measure, namely geometric vitality function. In connection with income studies, the geometric vitality function defined in Nair and Rajesh (2000) shall be interpreted as the geometric mean of the income of people whose income is greater than a threshold t. In other words, it represents the geometric mean of the income of wealthy people.

In the present chapter, we extend this concept to the bivariate setup. Further, we look into the problem of characterizing certain bivaraite models using the functional form of the geometric vitality function.

5.2 Bivariate geometric vitality function

For a random variable X admitting an absolutely continuous distribution function with respect to Lebsegue measure on $(-\infty, L)$, where

$$L = \inf\{x : F(x) = 1\},$$

Nair and Rajesh (2000) defines the geometric vitality function through the relationship

$$\begin{aligned}\log G(t) &= E(\log X \mid X \geq t) \\ &= \frac{1}{\bar{F}(t)} \int_t^{\infty} \log x\, f(x)\, dx. \qquad (5.2.1)\end{aligned}$$

Let $X = (X_1, X_2)$ represents a bivariate random vector measuring two attributes of income in a population. The random variable $Y_1 = X_1 \mid X_2 > t_2$ corresponds to the distribution of X_1 subject to the condition that X_2 is greater than an amount equal to t_2. One can define the geometric vitality function $G_1(t_1, t_2)$ for the random variable Y_1 through the relationship

$$\log G_1(t_1, t_2) = \frac{1}{\bar{F}(t_1 \mid X_2 > t_2)} \int_{t_1}^{\infty} \log x_1\, f(x_1 \mid X_2 > t_2)\, dx_1 \qquad (5.2.2)$$

(5.2.2) can also be written as

$$\log\left(\frac{G_1(t_1, t_2)}{t_1}\right) = \frac{1}{\bar{F}(t_1, t_2)} \int_{t_1}^{\infty} \frac{\bar{F}(x_1, t_2)}{x_1}\, dx_1. \qquad (5.2.3)$$

Similarly for the random variable $Y_2 = X_2 \mid X_1 > t_1$, the geometric vitality function $G_2(t_1, t_2)$ turns out to be

$$\log G_2(t_1, t_2) = \frac{1}{\bar{F}(t_2 \mid X_1 > t_1)} \int_{t_2}^{\infty} \log x_2\, f(x_2 \mid X_1 > t_1)\, dx_2 \qquad (5.2.4)$$

(5.2.4) simplifies to

$$\log\left(\frac{G_2(t_1, t_2)}{t_2}\right) = \frac{1}{\bar{F}(t_1, t_2)} \int_{t_2}^{\infty} \frac{\bar{F}(t_1, x_2)}{x_2}\, dx_2. \qquad (5.2.5)$$

Definition 5.1

For a non-negative random vector $X=(X_1,X_2)$ admitting an absolutely continuous distribution function, the bivariate geometric vitality function is defined as the vector

$$\log G(t_1,t_2) = (\ \log G_1(t_1,t_2)\ ,\log G_2(t_1,t_2)\ ) \qquad (5.2.6)$$

where $\log G_1(t_1,t_2)$ and $\log G_2(t_1,t_2)$ are given by (5.2.2) and (5.2.4) respectively.

Let $X=(X_1\ ,\ X_2)$ represents a bivariate random vector representing income of two populations, with the random variable X_1 representing the income of the first population and the random variable X_2 representing the income of the second population. Then $G_1(t_1,t_2)$ defines the geometric mean of the first population subject to the condition that the income of the second population exceeds a threshold t_2. Similarly, $G_2(t_1,t_2)$ provides the geometric mean of the second population when the income of the first population exceeds a threshold t_1. Hence $G(t_1,t_2)$can be viewed as an index, which represents the geometric mean of X, under the conditions specified.

5.3 Characterization Theorems

In this section, we look into the problem of characterizing some well-known bivariate models by the form of the bivariate geometric vitality function. In the following theorem, we look into the situation where $\log\left(\frac{G_i(t_1,t_2)}{t_i}\right)$,i = 1,2 is locally constant.

Theorem 5.1

Let $X=(X_1,X_2)$ be a non-negative random vector admitting an absolutely continuous distribution function with respect to Lebesgue measure. The relation

$$\log\left(\frac{G_i(t_1,t_2)}{t_i}\right) = \frac{1}{a + b\log(t_j)}\ ,\ i,j=1,2\ ,i\neq j\ ,a,b>0 \quad (5.3.1)$$

holds for all $t_1,t_2 \geq 0$ if and only if X follows the Pareto type-1 distribution specified by

$$\overline{F}(t_1,t_2) = \left(\frac{t_1}{a_1}\right)^{-\alpha_1}\left(\frac{t_2}{a_2}\right)^{-\alpha_2}\left(\frac{t_1}{a_1}\right)^{-\theta\log\left(\frac{t_2}{a_2}\right)}\ ,t_i \geq a_i\ ,i=1,2 \quad (5.3.2)$$

Proof

When (5.3.1) holds with $i=1$ using (5.2.3) we have

$$\frac{1}{\overline{F}(t_1,t_2)}\int_{t_1}^{\infty}\frac{\overline{F}(x_1,t_2)}{x_1}\,dx_1 = \frac{1}{a+ b\log(t_2)}.$$

Differentiating with respect to t_1 we get

$$-\frac{\overline{F}(t_1,t_2)}{t_1} = \left(\frac{1}{a + b\log(t_2)}\right)\frac{\partial\overline{F}(t_1,t_2)}{\partial t_1}$$

or

$$-\frac{1}{\overline{F}(t_1,t_2)}\frac{\partial\overline{F}(t_1,t_2)}{\partial t_1} = \left(\frac{a+b\log(t_2)}{t_1}\right).$$

Using (1.2.15) we get

$$h_1(t_1,t_2) = \frac{a + b\log(t_2)}{t_1}.$$

Proceeding on similar lines with $i=2$, we also get

$$h_2(t_1,t_2) = \frac{a+ b\log(t_1)}{t_2}.$$

Using the pair of identities for the survival function given by (1.2.16) and (1.2.17), we get

$$\overline{F}(t_1,t_2) = \exp\left(-(a+b\log(a_2))\log\left(\frac{t_1}{a_1}\right) - (a+b\log(t_1))\log\left(\frac{t_2}{a_2}\right)\right). \quad (5.3.3)$$

and

$$\overline{F}(t_1,t_2) = \exp\left(-(a+b\log(a_1))\log\left(\frac{t_2}{a_2}\right) - (a+b\log(t_2))\log\left(\frac{t_1}{a_1}\right)\right). \quad (5.3.4)$$

Equating (5.3.3) and (5.3.4) and setting $\alpha_i = a + b \log a_j, i, j = 1,2,\ i \neq j$ we get

$$\exp\left(-\alpha_1 \log\left(\frac{t_1}{a_1}\right) - (a + b\log(t_1))\log\left(\frac{t_2}{a_2}\right)\right) = \exp\left(-\alpha_2 \log\left(\frac{t_2}{a_2}\right) - (a + b\log(t_2))\log\left(\frac{t_1}{a_1}\right)\right) \tag{5.3.5}$$

(5.3.5) can also be written as

$$(a + b\log(t_2) - \alpha_1)\left(\log\left(\frac{t_2}{a_2}\right)\right)^{-1} = (a + b\log(t_1) - \alpha_2)\left(\log\left(\frac{t_1}{a_1}\right)\right)^{-1} \tag{5.3.6}$$

For (5.3.6) to hold we should have both sides equal to a constant, say θ. This gives

$$a + b\log(t_2) = \alpha_1 + \theta \log\left(\frac{t_2}{a_2}\right) \tag{5.3.7}$$

and

$$a + b\log(t_1) = \alpha_2 + \theta \log\left(\frac{t_1}{a_1}\right). \tag{5.3.8}$$

Using (5.3.7) and (5.3.8) in equation (5.3.5) we get the required distribution (5.3.2)

The if part of the theorem follows from the expression of $\log\left(\frac{G_1(t_1,t_2)}{t_1}\right)$ given by $\frac{1}{\alpha_1 + \theta \log(\frac{t_2}{a_2})}$ with similar expression for $\log\left(\frac{G_2(t_1,t_2)}{t_2}\right)$.

The following theorem provides a characterization result for a family of distributions using a possible relationship between the bivariate geometric vitality function and the reciprocal moment of X_1 given $X_2 > t_2$.

Theorem 5.2

Let $X=(X_1,X_2)$ be a non-negative random vector admitting an absolutely continuous distribution function with respect to Lebesgue measure. The relationship

$$\log\left(\frac{G_i(t_1,t_2)}{t_i}\right) = A + B_i(t_j)\ R(t_1,t_2) \ ,i,j=1,2\ ,i\neq j. \qquad (5.3.9)$$

where $B_i(t_j)$ are non-negative functions of t_j and

$$R(t_1,t_2) = \frac{1}{\overline{F}(t_1 \mid X_2>t_2)} \int_{t_1}^{\infty} \frac{1}{x_1} f(x_1 \mid X_2>t_2)\ dx_1 \qquad (5.3.10)$$

is the reciprocal moment of X_1 given $X_2>t_2$ holds for all $t_1,t_2\geq 0$ if and only if X is distributed as

(i) the Gumbels bivariate exponential distribution with survival function

$$\overline{F}(t_1,t_2) = e^{-\lambda_1 t_1-\lambda_2 t_2-\theta t_1 t_2} \ ,\ \lambda_1,\lambda_2>0\ ,t_1,t_2>0\ ,0\leq\theta\leq\lambda_1\lambda_2 \quad (5.3.11)$$

if $A=0$

(ii) the bivariate Pareto type-II distribution with survival function

$$\overline{F}(t_1,t_2) = (1+a_1t_1+a_2t_2+bt_1t_2)^{-c} \ ,t_1,t_2>0\ ,a_1,a_2,c>0 \ ,0\leq b\leq(c+1)\,a_1a_2 \qquad (5.3.12)$$

if $A>0$ and

(iii) the bivariate finite-range distribution with survival function

$$\overline{F}(t_1,t_2) = (1-p_1t_1-p_2t_2+q\,t_1t_2)^{d}\ ,\ 0<t_1<\frac{1}{p_1},0<t_2<\frac{1-p_1t_1}{p_2-qt_1} \ ,p_1,\ p_2>0\ ,\ 1-d\leq\frac{q}{p_1p_2}\leq 1, d>0 \quad (5.3.13)$$

if $A<0$.

Proof

When (5.3.10) holds with $i=1$, using (5.2.3) we have

$$\frac{1}{\overline{F}(t_1,t_2)}\int_{t_1}^{\infty}\frac{\overline{F}(x_1,t_2)}{x_1}\,dx_1 = (A+B_1(t_2))\ R(t_1,t_2)$$

or

$$\int_{t_1}^{\infty} \frac{\overline{F}(x_1,t_2)}{x_1}\, dx_1 = A\,\overline{F}(t_1,t_2) - B_1(t_2) \int_{t_1}^{\infty} \frac{\frac{\partial \overline{F}(x_1,t_2)}{\partial x_1}}{x_1}\, dx_1$$

Differentiating with respect to t_1 we get

$$-\frac{\overline{F}(t_1,t_2)}{t_1} = A\,\frac{\partial \overline{F}(t_1,t_2)}{\partial t_1} + B_1(t_2)\,\frac{\frac{\partial \overline{F}(t_1,t_2)}{\partial t_1}}{t_1}$$

Using (1.2.15), the above equation gives

$$h_1(t_1,t_2) = \frac{1}{A\,t_1 + B_1(t_2)}.$$

Proceeding on similar line with $i=2$, we get

$$h_2(t_1,t_2) = \frac{1}{A\,t_2 + B_2(t_1)}.$$

The rest of the proof follows from Roy (1989).

Conversely when the distribution of $X=(X_1,X_2)$ is specified above by (5.3.11), (5.3.12) and (5.3.13) calculations yield the expressions for $\log\left(\frac{G_1(t_1,t_2)}{t_1}\right)$ as $\frac{1}{\lambda_1+\theta t_2}, \frac{1}{c} + \frac{(1+a_2t_2)}{c(a_1+bt_2)}\,R(t_1,t_2)$ and $\frac{-1}{d} + \frac{(1-p_2t_2)}{d(p_1-qt_2)}\,R(t_1,t_2)$ respectively, so that the condition of the theorem holds.

The following theorem provides a characterization result for the bivariate Pareto type-1 distribution specified in (5.3.2) using a possible relationship between the bivariate geometric vitality function and the bivariate residual entropy function defined in (4.2.13).

Theorem 5.3

Let $X=(X_1,X_2)$ be a non-negative random vector admitting an absolutely continuous distribution function with components of the geometric vitality function $\log G_i(t_1,t_2)$ and the residual entropy function $H_i(f,t_1,t_2), i=1,2$. Then the relationship

$$H_i(f,t_1,t_2) - \log G_i(t_1,t_2) = 1 - \log\left(a + b\log t_j\right), i,j=1,2, i\neq j \quad (5.3.14)$$

holds for all real $t_1,t_2 \geq 0$ if and only if X follows the Pareto type-1 distribution with survival function specified in (5.3.2).

Proof

When (5.3.14) holds with $i=1$ using (4.2.11) and (5.2.2) we have

$$1+\frac{1}{\bar{F}(t_1,t_2)}\int_{t_1}^{\infty}\frac{\partial \bar{F}(x_1,t_2)}{\partial x_1}\log h_1(x_1,t_2)dx_1 + \frac{1}{\bar{F}(t_1,t_2)}\int_{t_1}^{\infty}\log x_1\frac{\partial \bar{F}(x_1,t_2)}{\partial x_1}dx_1 = 1-\log(a+b\log t_2)$$

or

$$\int_{t_1}^{\infty}\frac{\partial \bar{F}(x_1,t_2)}{\partial x_1}\log h_1(x_1,t_2)\,dx_1 + \int_{t_1}^{\infty}\log x_1\frac{\partial \bar{F}(x_1,t_2)}{\partial x_1}\,dx_1 = -\bar{F}(t_1,t_2)\log(a+b\log t_2).$$

Differentiating with respect to t_1 and rearranging the terms we get

$$\frac{\partial \bar{F}(t_1,t_2)}{\partial t_1}\log h_1(t_1,t_2) - \log t_1\frac{\partial \bar{F}(t_1,t_2)}{\partial t_1} = -\log(a+b\log t_2)\frac{\partial \bar{F}(t_1,t_2)}{\partial t_1}.$$

The above equation can be written as

$$h_1(t_1,t_2) = \frac{a + b\log t_2}{t_1}.$$

Proceeding on similar lines with $i=2$, we get

$$h_2(t_1,t_2) = \frac{a + b\log t_1}{t_2}.$$

Rest of the proof is analogous to that of theorem 5.1 and hence omitted.

If part of the theorem follows from the expression for $\log\left(\frac{G_i(t_1,t_2)}{t_i}\right)$ and $H_i(f,t_1,t_2)$ for the Pareto distribution specified by (5.3.2)

$$\frac{1}{\alpha_1+\theta\log(\frac{t_2}{a_2})} \text{ and } 1 + \frac{1}{\alpha_1+\theta\log(\frac{t_2}{a_2})} - \log\left(\frac{\alpha_1+\theta\log(\frac{t_2}{a_2})}{t_1}\right), i,j=1,2, i\neq j$$

respectively.

The following theorem provides a characterization of the bivariate weibull distribution using a possible relationship between the geometric vitality function and the bivariate residual entropy function.

Theorem 5.4

For the random vector $X=(X_1,X_2)$ considered in the above theorem the relationship

$$H_i(f,t_1,t_2) + (\beta-1)\log G_i(f,t_1,t_2) = 1-\log\beta\left(\alpha_i+bt_j^{\beta}\right),\ i,j=1,2,\ i\neq j \quad (5.3.15)$$

holds for all real $t_1,t_2\geq 0$ if and only if X follows the bivariate weibull distribution with survival function specified by

$$\bar{F}(t_1,t_2) = \exp\left(-\alpha_1 t_1^{\beta}-\alpha_2 t_2^{\beta}-\theta t_1^{\beta} t_2^{\beta}\right),\ t_1,t_2\geq 0,\ \alpha_1,\alpha_2,\ \beta>0,\ \theta>0 \quad (5.3.16)$$

Proof

When (5.3.15) holds with $i=1$ using (5.2.2) we get

$$H_1(f,t_1,t_2)\ \bar{F}(t_1,t_2) -(\beta-1)\int_{t_1}^{\infty}\log x_1 \frac{\partial \bar{F}(x_1,t_2)}{\partial x_1}dx_1 = \bar{F}(t_1,t_2)\left(1-\log\beta\left(\alpha_1+bt_2^{\beta}\right)\right)$$

Differentiating with respect to t_1 and rearranging the terms we get

$$\frac{\partial H_1(f,t_1,t_2)}{\partial t_1}-H_1(f,t_1,t_2)h_1(t_1,t_2)-(\beta-1)\log(t_1h_1(t_1,t_2))=-\left(1-\log\beta\left(\alpha_1+bt_2^{\beta}\right)\right)h_1(t_1,t_2)$$

Using the relation

$$\frac{\partial H_1(f,t_1,t_2)}{\partial t_1} = h_1(t_1,t_2)\left(H_1(f,t_1,t_2) - \log h_1(t_1,t_2) -1\right)$$

we get

$$h_1(t_1,t_2) = \frac{c_1\left(\alpha_1+bt_2^{\beta}\right)}{t_1^{\beta-1}},$$ where $c_1>0$ is a constant.

Proceeding on similar lines with $i=2$ one can also get

$$h_2(t_1,t_2) = \frac{c_2\left(\alpha_2+bt_1^{\beta}\right)}{t_2^{\beta-1}},$$ where $c_2>0$ is a constant.

Using the pair of identities for the survival function of $\bar{F}(x_1,x_2)$ specified in (1.2.16) and (1.2.17) we get

$$\bar{F}(t_1,t_2) = \exp\left(-\left(\frac{c\alpha_1}{2-\beta}\right) t_1^{2-\beta} - \frac{c\left(\alpha_2 + bt_1^{\beta}\right)}{2-\beta} t_2^{2-\beta}\right) \quad (5.3.17)$$

and

$$\bar{F}(t_1,t_2) = \exp\left(-\left(\frac{c\alpha_2}{2-\beta}\right) t_2^{2-\beta} - \frac{c\left(\alpha_1 + bt_2^{\beta}\right)}{2-\beta} t_1^{2-\beta}\right) \quad (5.3.18)$$

Equating (5.3.17) and (5.3.18) and rearranging the terms

$$\exp\left(- \alpha_1 t_1^{2-\beta} - \frac{c\left(\alpha_2 + bt_1^{\beta}\right)}{2-\beta} t_2^{2-\beta}\right) = \exp\left(- \alpha_2 t_2^{2-\beta} - \frac{c\left(\alpha_1 + bt_2^{\beta}\right)}{2-\beta} t_1^{2-\beta}\right) (5.3.19)$$

or

$$\left(\frac{c\left(\alpha_1 + bt_2^{\beta}\right)}{2-\beta} - \alpha_1 \right) t_1^{2-\beta} = \left(\frac{c\left(\alpha_2 + bt_1^{\beta}\right)}{2-\beta} - \alpha_2 \right) t_2^{2-\beta}$$

Dividing both sides by $t_1^{2-\beta}\, t_2^{2-\beta}$ we get

$$\left(\frac{c\left(\alpha_1 + bt_2^{\beta}\right)}{2-\beta} - \alpha_1 \right) t_2^{2-\beta} = \left(\frac{c\left(\alpha_2 + bt_1^{\beta}\right)}{2-\beta} - \alpha_2 \right) t_1^{2-\beta} \text{ for all } t_1, t_2 \quad (5.3.20)$$

For equation (5.3.20) to hold each term should be a constant. This gives

$$\frac{c\left(\alpha_1 + bt_2^{\beta}\right)}{2-\beta} = \alpha_1 + \theta\, t_2^{2-\beta} \quad (5.3.21)$$

and

$$\frac{c\left(\alpha_2 + bt_1^{\beta}\right)}{2-\beta} = \alpha_2 + \theta\, t_1^{2-\beta}. \quad (5.3.22)$$

Inserting (5.3.21) and (5.3.22) in (5.3.19) we get the required distribution.

Conversely when the distribution of X is specified by (5.3.16) by direct calculations we get

$$H_1(f,t_1,t_2) = 1 - \log\beta\left(\alpha_1 + \theta t_2^{\beta}\right) + \frac{\beta-1}{\bar{F}(t_1,t_2)} \int_{t_1}^{\infty} \log x_1 \frac{\partial \bar{F}(x_1,t_2)}{\partial x_1} dx_1.$$

The validity of (5.3.15) is straightforward.

The following theorem provides a characterization of the bivariate Pareto type-1 distribution using a possible relationship between the bivariate geometric vitality function and the bivariate mean residual life function.

Theorem 5.5

For the random vector $X=(X_1,X_2)$ considered in theorem 5.4 the relationship

$$\log\left(\frac{G_i(t_1,t_2)}{t_i}\right) = \frac{r_i(t_1,t_2)}{t_i + r_i(t_1,t_2)}, i=1,2. \tag{5.3.23}$$

holds for all real $t_1,t_2 \geq 0$ if and only if X follows the bivariate Pareto type-1 distribution specified by (5.3.1).

Proof

When (5.3.23) holds with $i=1$, using (5.2.3) we get

$$\frac{1}{\overline{F}(t_1,t_2)} \int_{t_1}^{\infty} \frac{\overline{F}(x_1,t_2)}{x_1} dx_1 = \frac{r_1(t_1,t_2)}{t_1+r_1(t_1,t_2)}$$

or

$$-(t_1+r_1(t_1,t_2)) \int_{t_1}^{\infty} \frac{\overline{F}(x_1,t_2)}{x_1} dx_1 = \int_{t_1}^{\infty} \overline{F}(x_1,t_2)\, dx_1 .$$

Differentiating with respect to t_1 we have

$$-(t_1+r_1(t_1,t_2)) \frac{\overline{F}(t_1,t_2)}{t_1} + \left(1+\frac{\partial r_1(t_1,t_2)}{\partial t_1}\right) \int_{t_1}^{\infty} \frac{\overline{F}(x_1,t_2)}{x_1} dx_1 = -\overline{F}(t_1,t_2)$$

or

$$\left(\frac{t_1+r_1(t_1,t_2)}{t_1}\right) - \left(1 + \frac{\partial r_1(t_1,t_2)}{\partial t_1}\right)\left(\frac{t_1+r_1(t_1,t_2)}{t_1}\right) = 1$$

The solution of the above partial differential equation is

$$r_1(t_1,t_2) = c_1(t_2)\, t_1$$

where $c_1(t_2)$ is independent t_1.

Proceeding on similar line with $i=2$, we also get

$$r_2(t_1,t_2) = c_2(t_1)\, t_2$$

The rest of the proof is analogous to that of theorem 3.6 and hence omitted.

The if part of the theorem follows from the expression for $\log\left(\frac{G_1(t_1,t_2)}{t_1}\right)$ and $\frac{r_1(t_1,t_2)}{t_1+r_1(t_1,t_2)}$ given by $\frac{1}{\alpha_1 + \theta \log(\frac{t_2}{a_2})}$, with similar expressions for $\log\left(\frac{G_2(t_1,t_2)}{t_2}\right)$ and $\frac{r_2(t_1,t_2)}{t_2+r_2(t_1,t_2)}$.

Chapter VI

Estimation of certain measures of income inequality using Bayesian techniques

6.1 Introduction

The problem of estimation of Lorenz curve and Gini-index in the classical framework has been investigated by Gastwirth (1972), Moothathu (1985a, 1990). However only very little work seems to have been done using the Bayesian techniques while estimating measures of income inequality. Motivated by this in the present chapter we look into the problem of estimation of the Lorenz curve, Gini-index and the variance of logarithms using ideas from Bayesian inference. Further a comparison of the estimates obtained from the classical and Bayesian methods is under taken using their variances.

6.2 The model

Let $X_1, X_2, \ldots X_n$ is a random sample from the Pareto distribution, with density function

$$f(x, a, \theta) = a\,\theta^{a}\,x^{-(a+1)},\ x \geq \theta > 0,\ a > 1. \quad (6.2.1)$$

For the model (6.2.1), by straight forward calculations the Lorenz curve, Gini-index and the variance of logarithms simplifies receptively to

$$L(p) = 1 - (1-p)^{1-\frac{1}{a}},\ 0 < p < 1 \quad (6.2.2)$$

$$G = (2a-1)^{-1} \quad (6.2.3)$$

and

$$V = \frac{1}{a^2} \quad (6.2.4)$$

6.3 Estimation of Lorenz curve

In this section we obtain the Bayes estimate of the Lorenz curve under the two alternatives namely (i) when the scale parameter θ is known and (ii) when θ is unknown.

Estimation with known θ

For the model (6.2.1), when θ is known, the likelihood function can be written as

$$l(\underline{x} \mid a, \theta) = C_1 a^n e^{-at} \tag{6.3.1}$$

where

$$t = \sum_{i=1}^{n} \ln\left(\frac{x_i}{\theta}\right).$$

Estimation of the parameters a and θ in the Bayesian framework has been studied by Malik (1970), Zellener (1971), Sinha and Howlder (1980), Arnold and Press (1983) and Jeevanand and Nair (1992). We presently look into the problem of estimating $L(p)$ as such. Since θ is known the form of the likelihood function provides a conjugate prior with density function

$$g(a) = C_2 a^{r-1} e^{-t'a}, \; r, \; t', a > 0 \tag{6.3.2}$$

The symbol C with various suffixes stands for the normalizing constants. The posterior density from (6.3.1) and (6.3.2) turns out to be

$$f(a \mid \underline{x}) = C_3 a^{m-1} e^{-aT}, \; a \geq 0 \tag{6.3.3}$$

where $T = t + t'$, $m = n + r$.

From (6.2.2) we have the representation of a as

$$a = \frac{\log(1-p)}{\log(1-p) - \log(1-L)} \tag{6.3.4}$$

and

$$\frac{da}{dL} = -\frac{\log(1-p)}{(1-L)\left(\log(1-p) - \log(1-L)\right)^2} \tag{6.3.5}$$

Using (6.3.4) and (6.3.5) the posterior distribution of L turns to be

$$f(L|\underline{x}) = [C_4(p,0)]^{-1} \frac{(1-p)^{\frac{-T}{\log(1-p)-\log(1-L)}}\left(\frac{\log(1-p)}{\log(1-p)-\log(1-L)}\right)^m}{(1-L)(\log(1-p)-\log(1-L))}, 0<L<p \quad (6.3.6)$$

where

$$C_4(p,d) = \int_0^p \frac{L^d}{(1-L)} \frac{(1-p)^{\frac{-T}{\log(1-p)-\log(1-L)}}\left(\frac{\log(1-p)}{\log(1-p)-\log(1-L)}\right)^m}{(\log(1-p)-\log(1-L))} dL. \quad (6.3.7)$$

(6.3.7) can be solved by numerical integration. One can have estimators for L by specifying appropriate loss functions and (6.3.6). Under the quadratic loss function the Bayes estimator turns out to be

$$\widehat{L}_1 = \frac{C_4(p,1)}{C_4(p,0)} \quad (6.3.8)$$

with Bayes risk

$$R(L_1,\widehat{L}_1) = \frac{C_4(p,2)}{C_4(p,0)} - \left(\widehat{L}_1\right)^2 \quad (6.3.9)$$

Estimation with unknown θ

The most general and perhaps a more realistic situation is when both the shape and scale parameters are unknown. In this case the likelihood can be written as

$$l(\underline{x}|a,\theta) = C_5\ a^n\ \theta^{na}\ e^{-az} \quad (6.3.10)$$

where

$$z = \sum_{i=1}^{n} \log x_i .$$

The kernel of the likelihood suggests the following the joint prior density of a and θ (conjugate prior)

$$g(a,\theta) = C_6\ a^r\ \theta^{n'a}\ e^{-az'}, a>1,\ 0\le\theta\le\theta_0,\ \theta_0,r,z' > 0 \quad (6.3.11)$$

where r, n' and z' are prior parameters.

Using (6.3.10) and (6.3.11) the posterior distribution turn out to be

$$f(a,\theta \mid \underline{x}) = C_7\ a^m\ \theta^{Na-1}\ e^{-aZ},\ a>1,\ 0\le\theta\le\min(\theta_0, X_{(1)}) \quad (6.3.12)$$

where

$$Z = z + z' \ , \ m = n + r , \ N = n + n' \ , \ X_{(1)} = \min(X_1, \ldots, X_n)$$

Further from (6.2.2) we get the representation of a as

$$a = \frac{\log(1-p)}{\log(1-p) - \log(1-L)} \qquad (6.3.13)$$

Under the transformation $u = \theta$ and (6.3.13) we get the Jacobian of transformation as

$$|J| = \begin{vmatrix} \frac{da}{dL} & \frac{da}{du} \\ \frac{d\theta}{dL} & \frac{d\theta}{du} \end{vmatrix}$$

$$= \frac{\log(1-p)}{(1-L)\ (\log(1-p) - \log(1-L))^2}$$

Now using the transformation (a, θ) to (L, u) we get the joint posterior density of L and u, as

$$f(L, u|\underline{x}) = C_8 \frac{(1-p)^{\frac{-z}{\log(1-p)-\log(1-L)}}\ u^{N\left(\frac{\log(1-p)}{\log(1-p)-\log(1-L)}\right)-1} \left(\frac{\log(1-p)}{\log(1-p) - \log(1-L)}\right)^m \log(1-p)}{(1-L)\left(\log(1-p) - \log(1-L)\right)^2}, 0 < L < p$$

(6.3.14)

Integrating out u from (6.3.14) we get the posterior density of L as

$$f(L|\underline{x}) = \int_0^{\theta_0} f(L, u|\underline{x})\ du .$$

This gives

$$f(L|\underline{x}) = C_8 \left(\frac{\log(1-p)}{\log(1-p) - \log(1-L)}\right)^{m-1} \frac{\log(1-p)}{(1-L)\ (\log(1-p) - \log(1-L))^2}$$

$$\times \exp\left(- \left(\frac{\log(1-p)}{\log(1-p) - \log(1-L)}\right) z \right) \theta_0^{\left(\frac{\log(1-p)}{\log(1-p)-\log(1-L)}\right)N} .$$

The above equation can be written as

$$f(L|\underline{x}) = C_8 \left(\frac{\log(1-p)}{\log(1-p)-\log(1-L)}\right)^{m-1} \frac{\log(1-p)}{(1-L)\ (\log(1-p)-\log(1-L))^2}$$

$$\times \exp\left(- \left(\frac{\log(1-p)}{\log(1-p)-\log(1-L)}\right) A\right) \qquad (6.3.15)$$

where

$$A = Z - N\log\theta_0$$

Rearranging the terms in (6.3.15) we get

$$f(L|\underline{x}) = \left[C_9(p,0)\right]^{-1} \frac{(1-p)^{\frac{-A}{\log(1-p)-\log(1-L)}}\left(\frac{\log(1-p)}{\log(1-p)-\log(1-L)}\right)^m}{(1-L)\ (\log(1-p)-\log(1-L))}, 0 < L < p \qquad (6.3.16)$$

where

$$C_9(p,d) = \int_0^p \frac{L^d}{(1-L)} \frac{(1-p)^{\frac{-A}{\log(1-p)-\log(1-L)}}\left(\frac{\log(1-p)}{\log(1-p)-\log(1-L)}\right)^m}{(\log(1-p)-\log(1-L))}\, dL \qquad (6.3.17)$$

Under the quadratic loss, the Bayes estimate and the corresponding risk of the Lorenz curve is given by

$$\widehat{L_2} = \frac{C_9(p,1)}{C_9(p,0)} \qquad (6.3.18)$$

and

$$R(L_2,\widehat{L_2}) = \frac{C_9(p,2)}{C_9(p,0)} - \left(\widehat{L_2}\right)^2 \qquad (6.3.19)$$

6.4 Estimation of Gini-index

In this section we obtain the Bayes estimate of the Gini-index given by (6.2.3), under the two alternatives namely (i) the scale parameter θ is known and (ii) θ is unknown.

Estimation with known θ

From (6.2.3) we have

$$a = \frac{G+1}{2G} \tag{6.4.1}$$

and

$$\frac{da}{dG} = -\frac{1}{2\,G^2} \tag{6.4.2}$$

From (6.4.1) and (6.4.2), in view of (6.3.3) we get the posterior distribution of G as

$$f(\,G|\underline{x}\,) = \frac{[C_{10}(0)]^{-1}}{2G^2}\left(\frac{G+1}{2G}\right)^{m-1} e^{-\left(\frac{G+1}{2G}\right)T}, \; 0 < G < 1 \tag{6.4.3}$$

where

$$[C_{10}(d)] = \int_0^1 \frac{G^d}{2\,G^2}\left(\frac{G+1}{2\,G}\right)^{m-1} e^{-\left(\frac{G+1}{2G}\right)T} dG \tag{6.4.5}$$

Bayes estimate of G under quadratic loss function is

$$\hat{G}_1 = \frac{C_{10}(1)}{C_{10}(0)} \tag{6.4.5}$$

and the expected loss when (6.4.5) is used as estimate is

$$R(G_1, \hat{G}_1) = \frac{C_{10}(2)}{C_{10}(0)} - \left(\hat{G}_1\right)^2 \tag{6.4.6}$$

To evaluate (6.4.5) and (6.4.6) we seek numerical integration.

Estimation when θ is unknown.

From (6.2.3) we have

$$a = \frac{G+1}{2G} \tag{6.4.7}$$

Taking $u=\theta$ and from (6.4.7), the Jacobian of transformation is

$$|J| = \frac{1}{2G^2}$$

Using the transformation (a, θ) to (G, u), In view of (6.3.12) we get the joint posterior density of (G, u) as

$$f(G,u|\underline{x}) = \frac{C_{11}}{2G^2}\left(\frac{G+1}{2G}\right)^m u^{\left(\frac{G+1}{2G}\right)N-1} e^{-\left(\frac{G+1}{2G}\right)Z}, \; 0 < G < 1, \; 0 < u < \theta_0 \qquad (6.4.8)$$

Integrating out u from (6.4.8) we get the posterior density of G as

$$f(G|\underline{x}) = \int_0^{\theta_0} f(G,u|\underline{x})\, du \qquad (6.4.9)$$

or

$$f(G|\underline{x}) = \frac{C_{11}}{2G^2}\left(\frac{G+1}{2G}\right)^{m-1} \theta_0^{\left(\frac{G+1}{2G}\right)N} e^{-\left(\frac{G+1}{2G}\right)Z}$$

The above equation can be written as

$$f(G|\underline{x}) = \frac{[C_{12}(0)]^{-1}}{2G^2}\left(\frac{G+1}{2G}\right)^{m-1} e^{-\left(\frac{G+1}{2G}\right)A}, \; 0 < G < 1, A = Z - N\log\theta_0 \qquad (6.4.10)$$

where

$$C_{12}(d) = \int_0^1 \frac{G^d}{2G^2}\left(\frac{G+1}{2G}\right)^{m-1} e^{-\left(\frac{G+1}{2G}\right)A}\, dG \qquad (6.4.11)$$

Bayes estimate of G under the quadratic loss function is

$$\hat{G}_2 = \frac{C_{12}(1)}{C_{12}(0)} \qquad (6.4.12)$$

and the expected loss when (6.4.12) is used as estimate is

$$R(G_2,\hat{G}_2) = \frac{C_{12}(2)}{C_{12}(0)} - \left(\hat{G}_2\right)^2 \qquad (6.4.13)$$

The estimates can now be obtained using numerical integration.

6.5 Estimation of the variance of logarithms

In this section we provide the Bayes estimate of the variance of logarithms (6.2.4) under the two alternative namely (i) the scale parameter θ is known and (ii) θ is unknown.

Estimation with known θ

From (6.2.4), we can find

$$a = \frac{1}{\sqrt{V}} \text{ Since } V > 0 \tag{6.5.1}$$

and

$$\frac{da}{dV} = -\frac{1}{2v^{\frac{3}{2}}} \tag{6.5.2}$$

Using (6.5.1), (6.5.2), and the posterior density (6.3.3) we get the posterior distribution of Vas

$$f(V|\underline{x}) = \frac{[C_{13}(0)]^{-1}}{2V^{\frac{3}{2}}} \left(\frac{1}{\sqrt{V}}\right)^{(m-1)} \exp\left(-\frac{1}{\sqrt{V}}\, T\right),\ 0 < V < 1 \tag{6.5.3}$$

where

$$C_{13}(d) = \int_0^1 \frac{V^d}{2V^{\frac{3}{2}}} \left(\frac{1}{\sqrt{V}}\right)^{(m-1)} \exp\left(-\frac{1}{\sqrt{V}}T\right) dv \tag{6.5.4}$$

The above integral can be written as

$$C_{13}(d) = \left(\frac{1}{T}\right)^{m-2d} \overline{|m-2d,T} \tag{6.5.6}$$

where $\overline{|a,z}$ is the incomplete Gamma function, defined by

$$\overline{|a,z} = \int_z^{\infty} t^{z-1} e^{-t}\, dt \tag{6.5.7}$$

Bayes estimate of Vunder the quadratic loss function is

$$\widehat{V}_1 = \frac{C_{13}(1)}{C_{13}(0)} \tag{6.5.8}$$

and the Bayes risk is

$$R(V_1, \widehat{V}_1) = \frac{C_{13}(2)}{C_{13}(0)} - \left(\widehat{V}_1\right)^2 \tag{6.5.9}$$

To evaluate (6.5.8) and (6.5.9) we seek numerical integration.

Estimation with unknown θ

From (6.2.4) we get

$$a = \frac{1}{\sqrt{V}}$$

Taking $u=\theta$, the above equation provides the Jacobian of transformation J as

$$J = \frac{1}{2v^{\frac{3}{2}}}$$

Using the transformation (a, θ) to (V,u) and using the posterior density (6.3.12) we get the joint posterior distribution of (V,u) as

$$f(\, V,u|\underline{x}\,) = \frac{C_{14}}{2V^{\frac{3}{2}}}\left(\frac{1}{\sqrt{V}}\right)^{m} u^{\frac{1}{\sqrt{V}}N-1}\, e^{-\frac{1}{\sqrt{V}}Z}\,,\ 0<V<1,\ 0<u<\theta_0 \qquad (6.5.10)$$

Integrating out u from (6.5.10) we get the posterior density of V as

$$f(\, V|\underline{x}\,) = \int_0^{\theta_0} f(V,u|\underline{x})\, du \qquad (6.5.11)$$

or

$$f(V|\underline{x}) = \frac{[C_{15}(0)]^{-1}}{2V^{\frac{3}{2}}}\left(\frac{1}{\sqrt{V}}\right)^{m-1} e^{-\frac{1}{\sqrt{V}}A}\,,\ A = Z - N\log\theta_0,\ 0 < V < 1 \qquad (6.5.12)$$

where

$$C_{15}(d) = \int_0^1 \frac{V^d}{2V^{\frac{3}{2}}}\left(\frac{1}{\sqrt{V}}\right)^{(m-1)} \exp\left(-\frac{1}{\sqrt{V}}\, A\right) dv \qquad (6.5.13)$$

(6.5.13) can be written as

$$C_{15}(d) = \left(\frac{1}{A}\right)^{m-2d} \overline{|m-2d, A} \qquad (6.5.14)$$

where $\overline{|a,z}$ is defined in (6.5.7).

Bayes estimate of V under the quadratic loss function is

$$\widehat{V}_2 = \frac{C_{15}(1)}{C_{15}(0)} \qquad (6.5.15)$$

and the Bayes risk is

$$R(V_2,\widehat{V}_2) = \frac{C_{15}(2)}{C_{15}(0)} - \left(\widehat{V}_2\right)^2 \qquad (6.5.16)$$

To evaluate (6.5.15) and (6.5.16) we seek numerical integration.

6.6. Discussion

In this section we compare the Bayes estimate of the Lorenz curve and the Gini-index with corresponding classical estimates of the same in terms of their variance (empirical value is taken). For this purpose the criteria used is the risks improvement (RI) factor, defined as

$$\text{RI }(\%) = \frac{\text{Variance of the m.l.e} - \text{Variance of the new estimate}}{\text{Variance of the m.l.e}} \times 100$$

In the absence of the real data we compare the estimates empirically by generating observations from the Pareto distribution. The samples of sizes n=15, 20, 25, 35 were generated for different values of the parameter. First we compare the Bayes estimate of Lorenz curve defined in (6.3.8) with the maximum likelihood estimate of Lorenz curve (Moothathu (1990)) defined by

$$\hat{\beta}_r = \max\left(0,\ 1-(1-p)^{1-\frac{1}{a_r}}\right) \qquad (6.6.1)$$

and

$$a_r = \left(\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{X_i}{\theta}\right)\right)^{-1}$$

The comparison of the Bayes estimate of Lorenz curve with the classical estimate is presented in Table 6.1.Also we find the average, variance and bias of the estimate.

Now we compare the Bayes estimate of Gini-index defined in (6.4.5) with the UMVUE estimate (Moothathu (1990)) defined by

$$\hat{\lambda}_n = {}_1F_1\left(1;\ n;\ \frac{T_n}{2}\right) - 1 \qquad (6.6.2)$$

where ${}_1F_1(a;b;x)$ is the Kummer's function defined by

$${}_1F_1(a;\ b;\ x) = \sum_{r=0}^{\infty}\frac{(a)_r}{(b)_r}\frac{x^r}{r!}$$

with $T_n = \sum_{i=1}^{n}\log\left(\frac{X_i}{\theta}\right)$.

The results are presented in Table 6.2.

Finally to study the robustness of the hyper-parameters on the posterior density and Bayes estimates, we adopted the procedure given by Sinha (1980). He suggested that a Bayes estimate is robust with respect to its hyper-parameters if it leads to a high $\left(\frac{\min}{\max}\right)$ index of the estimate for the varying value of those hyper-parameters and is robust with respect to the posterior density if the graphs of the posterior densities for different values of the hyper-parameters coincide. For this, we calculate the $\left(\frac{\min}{\max}\right)$ index for each of the hyper-parameters by keeping the others fixed for different values of the population parameters for a sample of size n=15, 20, 25, 35. Also we made several plots of the posterior density. Table 6.3 gives the $\left(\frac{\min}{\max}\right)$ index of the estimate of Lorenz curve when θ unknown for various values of the hyper-parameters when $n'=0,1,2$ and $r',z'=0,1,2,3,4$. Table 6.4 gives the $\left(\frac{\min}{\max}\right)$ index of the estimate of Gini-index when θ unknown for various values of the hyper-parameters specified above. Table 6.5 and Table 6.6 gives the $\left(\frac{\min}{\max}\right)$ index of the estimates of Lorenz curve and Gini-index when θ known for various values of the hyper-parameters as $t' = r = 0,1,2,3,4$. Also we give the posterior plots for the Lorenz curve and Gini-index under the two alternatives (i) when θ known and (ii) when θ unknown corresponding to the hyper-parameter taking the values $r=0,1,2,3$ keeping the other hyper-parameters fixed as n', r' = $0,1,2$.

The conclusions from the empirical study are

1.The bias and the expected loss become smaller as the sample size increases.

2.The Bayes estimate in both cases performs better than that of the classical estimate.

3.The Bayes estimate and the posterior density are robust to all values of the hyper-parameters.

In the same way in Table 6.5 we find the average, variance and bias of the Bayes estimate of variance of logarithm. Also in table 6.6 and table 6.7 respectively gives the $\left(\frac{\min}{\max}\right)$ index of the estimate of variance of logarithm when θ known and unknown for various values of the hyper-parameters. Figure 6.5 and 6.6 gives the posterior plot under the two situations. The plot shows that Bayes estimate and the posterior density are robust to all values of the hyper parameters.

6.7 Estimation of Total time on test transform

The importance and role of total time on test transform in income studies as well as various relationships with the Lorenz curve and Gini-index are discussed in section 1.5.In this section we find the Bayes estimate of the total time on test transform of the Pareto type-1 model under the two alternative namely when the scale parameter is known and unknown.

Estimation with known θ

Let $X_1, X_2, \ldots X_n$ is a random sample from the Pareto distribution, with density function (6.2.1). Then by straightforward calculations the total time on test transform simplifies to

$$\phi(t) = 1 - \frac{(1-t)^{\frac{a-1}{a}}}{a}, \quad a \geq 1, 0 \leq t \leq 1 \qquad (6.7.1)$$

From the above equation we get

$$a = -\frac{\log(1-t)}{\text{Pr}\,oduct\log\left[\frac{(-1+\phi)\,\log(1-t)}{(1-t)}\right]} \qquad (6.7.2)$$

where Pr*oduct*log[*z*] gives the principal solution for w in $z = w e^{w}$,which can be viewed as a generalization of logarithm.

$$\frac{da}{d\phi} = \frac{\exp\left(\operatorname{Pr\mathit{oductlog}}\left[\frac{(-1+\phi)\ \log(1-t)}{1-t}\right]\right)(1-t)}{(-1+\phi)^2\left(1 + \operatorname{Pr\mathit{oductlog}}\left[\frac{(-1+\phi)\ \log(1-t)}{1-t}\right]\right)} \qquad (6.7.3)$$

From (6.7.1) and (6.7.2), in view of (6.3.3) we get the posterior distribution of ϕ as

$$f(\phi\,|\,\underline{x}) = \frac{[C_{16}(t,0)]^{-1}\ (1-t)^{\frac{T}{\operatorname{Pr\mathit{oductlog}}\left[\frac{(-1+\phi)\log(1-t)}{(1-t)}\right]}}\left(\frac{-\log(1-t)}{\operatorname{Pr\mathit{oductlog}}\left[\frac{(-1+\phi)\log(1-t)}{(1-t)}\right]}\right)^{m}}{(-1+\phi)\left(1+\operatorname{Pr\mathit{oductlog}}\left[\frac{(-1+\phi)\log(1-t)}{(1-t)}\right]\right)}$$

$$0 < \phi < 1 \qquad (6.7.4)$$

where

$$C_{16}(t,d) = \int_0^1 \frac{\phi^d}{(1-\phi)}\,\frac{(1-t)^{\frac{T}{\operatorname{Pr\mathit{oductlog}}\left[\frac{(-1+\phi)\log(1-t)}{(1-t)}\right]}}\left(\frac{-\log(1-t)}{\operatorname{Pr\mathit{oductlog}}\left[\frac{(-1+\phi)\log(1-t)}{(1-t)}\right]}\right)^{m}}{(-1+\phi)\left(1+\operatorname{Pr\mathit{oductlog}}\left[\frac{(-1+\phi)\log(1-t)}{(1-t)}\right]\right)} \qquad (6.7.5)$$

(6.7.5) can be solved by numerical integration. One can have estimators for ϕ by specifying appropriate loss function and using (6.7.4). Under the quadratic loss function the Bayes estimator turns out to be

$$\hat{\phi}_1 = \frac{C_{16}(t,1)}{C_{16}(t,0)} \qquad (6.7.6)$$

with Bayes risk

$$R\,(\phi_1,\hat{\phi}_1) = \frac{C_{16}(t,2)}{C_{16}(t,0)} - \left(\hat{\phi}_1\right)^2 \qquad (6.7.7)$$

Estimation with unknown θ

From (6.7.1) we get

$$a = -\frac{\log(1-t)}{\Pr oduct\log\left[\frac{(-1+\phi)\log(1-t)}{(1-t)}\right]} \qquad (6.7.8)$$

Taking $u=\theta$, the above equation provides the Jacobian of transformation J as

$$J = \frac{\exp\left(\Pr oduct\log\left[\frac{(-1+\phi)\log(1-t)}{1-t}\right]\right)(1-t)}{(-1+\phi)^2\left(1+\Pr oduct\log\left[\frac{(-1+\phi)\log(1-t)}{1-t}\right]\right)} \qquad (6.7.9)$$

Using the transformation (a,θ) to (ϕ,u) we get the joint posterior density of (ϕ,u) as

$$f(\phi,u\,|\,\underline{x}) = C_{17}\left(\frac{-\log(1-t)}{\Pr oduct\log\left[\frac{(-1+\phi)\log(1-t)}{(1-t)}\right]}\right)^m u^{N\left(\frac{-\log(1-t)}{\Pr oduct\log\left[\frac{(-1+\phi)\log(1-t)}{(1-t)}\right]}\right)-1}$$

$$\exp\left(-z\left(\frac{-\log(1-t)}{\Pr oduct\log\left[\frac{(-1+\phi)\log(1-t)}{(1-t)}\right]}\right)\right)\frac{\exp\left(\Pr oduct\log\left[\frac{(-1+\phi)\log(1-t)}{1-t}\right]\right)(1-t)}{(-1+\phi)^2\left(1+\Pr oduct\log\left[\frac{(-1+\phi)\log(1-t)}{1-t}\right]\right)}$$

$$0 < \phi < 1,\ 0 \le \theta \le \theta_0 \qquad (6.7.10)$$

Integrating out u from (6.7.10) we get the posterior density of ϕ as

$$f(\phi\,|\,\underline{x}) = \int_0^{\theta_0} f(\phi,u\,|\,\underline{x})\,du$$

This gives

$$f(\phi\,|\,\underline{x}) = \frac{C_{18}[t,0]^{-1}}{(1-\phi)} \frac{(1-t)^{\frac{A}{\mathrm{Productlog}\left[\frac{(-1+\phi)\log(1-t)}{1-t}\right]}}\left(\frac{-\log(1-t)}{\mathrm{Productlog}\left[\frac{(-1+\phi)\log(1-t)}{(1-t)}\right]}\right)^{m}}{1+\mathrm{Productlog}\left[\frac{(-1+\phi)\log(1-t)}{1-t}\right]} ,0 < \phi < 1 \tag{6.7.11}$$

where

$$C_{18}[t,d] = \int_0^1 \frac{\phi^d}{(1-\phi)} \frac{(1-t)^{\frac{A}{\mathrm{Productlog}\left[\frac{(-1+\phi)\log(1-t)}{1-t}\right]}}\left(\frac{-\log(1-t)}{\mathrm{Productlog}\left[\frac{(-1+\phi)\log(1-t)}{(1-t)}\right]}\right)^{m}}{1+\mathrm{Productlog}\left[\frac{(-1+\phi)\log(1-t)}{1-t}\right]}\, d\phi \tag{6.7.12}$$

Bayes estimate of ϕ under quadratic loss function is

$$\hat{\phi}_2 = \frac{C_{18}[t,1]}{C_{18}[t,0]} \tag{6.7.13}$$

and the expected loss when (6.7.13) is used as estimate is

$$R(\phi_2,\hat{\phi}_2) = \frac{C_{18}[t,2]}{C_{18}[t,1]} - \left(\hat{\phi}_2\right)^2 \tag{6.7.14}$$

To evaluate (6.7.13) and (6.7.14) we seek numerical integration.

In table 6.7 we calculate the mean, variance and bias of the estimate of total time on test transform. Also in table 6.6 and table 6.7 respectively give the $\left(\frac{\min}{\max}\right)$ index of the estimate of total time on test transform when θ known and unknown for various values of the hyper-parameters. Figure 6.5 and 6.6 gives the posterior plot under the two situations. The plot shows that Bayes estimate and the posterior density are robust to all values of the hyper parameters.

Future Study

Several problems have opened out during the present investigation. Most of the works on residual entropy function, Gini-index, etc are centered around the continuous case and only very little work seems to have been done in the discrete domain. It seems that analogous to the continuous case, definition and properties of the above-mentioned concepts in the discrete time domain can be formulated.

Properties of the entropy indices for weighted distributions, as well as mixtures of distributions, are another problem to be investigated. Apart from this the behaviour of these indices in the context of additive and multiplicative damage models is yet to be studied.

Estimation of measures of income inequality, of some other distributions, such as exponential, Pareto type-II and finite range in the Bayesian framework is yet another problem to be examined.

A related concept is the distance between two populations, studied by Kullback and Leibler (1951) .A study of the properties and characterizations based on the truncated form of the Kullback-Leibler directed divergence and other distance functions seems to be in order. Hopefully some answers to these problems will be presented in a future work.

Table 6.1

a=2, θ =1, n=20, p = 0.2 when θ is known .Actual Value of L is .10553					
Bayes Estimate $\hat{L}$					Classical Estimate $\hat{\beta}$
t'=r=0		t'=r=1	t'=r=2	t'=1, r=2	
0.10183		0.10612	0.10349	0.10607	0.11045
0.10152		0.10514	0.10434	0.10282	0.11601
0.10141		0.10692	0.10201	0.10829	0.11092
0.10126		0.10527	0.10281	0.10789	0.11040
0.10194		0.10684	0.10326	0.10336	0.11753
Average	0.10159	0.10606	0.10318	0.10569	0.11306
Variance	$8.157E^{-8}$	$7.0552E^{-7}$	$7.3847E^{-7}$	$6.35253E^{-6}$	$11.7947E^{-6}$
Bias	-0.00394	0.00053	-0.00235	0.00016	0.00753
RI	99.3084	99.4018	99.3739	99.4614	

Table 6.2

a=2, θ=1, n=20 when θ is known. Actual Value of G is 0.33333					
Bayes Estimate $\hat{G}$					Classical Estimate $\hat{\lambda}$
t'=r=0		t'=r=1	t'=r=2	t'=1, r=2	
0.32745		0.33475	0.33587	0.33894	0.28029
0.32301		0.33792	0.32764	0.33049	0.27463
0.33108		0.33996	0.33589	0.34559	0.24559
0.32971		0.32830	0.32604	0.33159	0.23881
0.33512		0.33797	0.33939	0.32508	0.26477
Average	0.32921	0.33578	0.33297	0.33434	0.26082
Variance	0.0000200	0.0000209	0.0000337	0.0000640	0.000325
Bias	-0.00412	0.00245	-0.00036	0.00101	-0.07251
RI	93.8462	93.5692	89.6308	80.3077	

Table 6.3

Estimate of Lorenz curve when θ unknown.

(n=20,a=2,θ=1,p=. 2)

n'	r / z'	0	1	2	3	4	$\frac{Min}{Max}$
0	0	.101695	.106014	.10035	.108859	.108241	.921835
	1	.104647	.106446	.101539	.104534	.108743	.933752
	2	.100634	.100553	.102999	.100895	.105017	.957493
	3	.100202	.102584	.102552	.101697	.108098	.926955
	4	.105605	.105226	.105563	.104355	.101017	.956555
	$\frac{Min}{Max}$	.948838	.944639	.950617	.926841	.93326	
1	0	.107518	.108515	.109367	.100819	.107325	.921841
	1	.109447	.109182	.102144	.104657	.105725	.933274
	2	.103078	.108894	.106754	.106647	.102472	.941025
	3	.103759	.104986	.106965	.105032	.101627	.950096
	4	.105396	.108608	.10428	.101193	.105696	.931727
	$\frac{Min}{Max}$	.941807	.961569	.933956	.945352	.946909	
2	0	.106483	.107941	.104393	.10231	.104433	.947833
	1	.103724	.104268	.106464	.100827	.105416	.947053
	2	.10622	.102937	.109393	.104251	.104398	.969092
	3	.100152	.104624	.102921	.104415	.107506	.931595
	4	.102798	.10082	.104811	.104701	.105383	.956701
	$\frac{Min}{Max}$	.940545	.934029	.940837	.962999	.97109	

Table 6.4

Estimate of Gini-index when θ unknown.

(n=20,a=2, θ=1)

n'	r / z'	0	1	2	3	4	$\frac{Min}{Max}$
0	0	.33641	.32783	0.32518	0.31575	0.32727	0.93857
	1	.31217	.33874	0.31826	0.33839	0.33688	0.92254
	2	.33968	0.34901	0.34885	0.32433	0.32209	0.92286
	3	.33610	0.3338	0.34057	0.34037	0.32965	0.96792
	4	.33889	0.33735	0.33601	0.33899	0.34856	0.96403
	$\frac{Min}{Max}$	.91903	.93931	0.91231	0.90632	0.92406	
1	0	0.32394	0.32828	0.33967	0.34342	0.32997	0.94327
	1	0.33891	0.33387	0.34112	0.32258	0.32563	0.94564
	2	0.32484	0.32429	0.33439	0.34925	0.32943	0.92856
	3	0.34264	0.34598	0.32610	0.34367	0.34302	0.94254
	4	0.33484	0.34223	0.34281	0.33775	0.32111	0.93669
	$\frac{Min}{Max}$	0.94542	0.93732	0.95126	0.93863	0.93611	
2	0	0.33871	0.32288	0.33882	0.33619	0.32025	0.94519
	1	0.33476	0.32339	0.33385	0.32867	0.34679	0.93256
	2	0.32765	0.32010	0.32729	0.32538	0.32255	0.97696
	3	0.32509	0.33754	0.32089	0.33929	0.33027	0.95811
	4	0.33113	0.32597	0.32746	0.33525	0.32678	0.97234
	$\frac{Min}{Max}$	0.95979	0.94835	0.94709	0.95898	0.92347	

Table 6.5

Estimate of Lorenz curve when θ known.

(n=20,a=2, θ =1,p=. 2)

r / t'	0	1	2	3	4	$\frac{Min}{Max}$
0	.10543	0.10219	0.10477	0.10272	0.10683	0.95657
1	0.10503	0.10289	0.10699	0.10783	0.10787	0.95383
2	0.10548	0.10469	0.10504	0.10249	0.10329	0.97165
3	0.10207	0.10578	0.10574	0.10657	0.10278	0.95777
4	0.10428	0.10378	0.10532	0.10525	0.10705	0.97412
$\frac{Min}{Max}$	0.96767	0.96606	0.97925	0.95013	0.95281	

Table 6.6

Estimate of Gini-index when θ known.

(n=20,a=2, θ =1)

r / t'	0	1	2	3	4	$\frac{Min}{Max}$
0	0.33327	0.32978	0.3330	0.34146	0.34621	0.95253
1	0.32199	0.33906	0.32691	0.33553	0.32622	0.94965
2	0.33801	0.33913	0.33646	0.34214	0.32683	0.95527
3	0.34729	0.33215	0.32878	0.3250	0.32916	0.93584
4	0.35301	0.32767	0.34210	0.34348	0.32743	0.92755
$\frac{Min}{Max}$	0.91213	0.96438	0.95558	0.94619	0.94226	

Table 6.7

$a=2, \theta=1, n=20$ When θ is known, Actual Value of V is 0.25					
Bayes Estimate $\hat{V}$					
$t'=r=0$		$t'=r=1$	$t'=r=2$	$t'=1, r=2$	$t'=2, r=1$
0.24765		0.24743	0.25984	0.25714	0.25141
0.25497		0.25506	0.24482	0.24113	0.24892
0.24532		0.24134	0.25953	0.24623	0.24951
0.25938		0.25940	0.24278	0.25362	0.25432
0.25030		0.24852	0.24056	0.25058	0.24539
Average	0.25152	0.25035	0.24951	0.24974	0.24991
Variance	0.00003214	0.00004928	0.00008863	0.00003924	0.00001082
Bias	0.001524	0.00035	-0.000494	-0.00026	-0.00009

Table 6.8

Estimate of the variance of logarithms when θ known.

(n=20,a=2,θ=1)

r / t'	0	1	2	3	4	$\frac{Min}{Max}$
0	0.25641	0.25053	0.24643	0.25312	0.24628	0.96049
1	0.24579	0.25163	0.24838	0.25275	0.25267	0.97246
2	0.25213	0.24796	0.24776	0.25346	0.25232	0.97751
3	0.25361	0.24491	0.24623	0.24582	0.25162	0.96569
4	0.24468	0.24605	0.24749	0.25177	0.24541	0.97184
5	0.24727	0.24795	0.24857	0.25342	0.25744	0.96050
6	0.25789	0.25371	0.25153	0.25079	0.25960	0.96606
7	0.25001	0.25471	0.24703	0.25327	0.25126	0.96985
$\frac{Min}{Max}$	0.94878	0.96153	0.97893	0.96986	0.94534	

Table 6.9

Estimate of the variance of logarithm when θ unknown

($n=20, a=2, \theta=1$)

n'	r / z'	0	1	2	3	4	$\frac{Min}{Max}$
0	0	0.24503	0.25217	0.25096	0.25149	0.25299	0.96854
	1	0.25169	0.24822	0.24321	0.25288	0.25192	0.96309
	2	0.25741	0.25990	0.25799	0.24390	0.24982	0.93844
	3	0.25167	0.25427	0.24797	0.25166	0.24479	0.96272
	4	0.25538	0.25362	0.24842	0.25948	0.24302	0.93657
	$\frac{Min}{Max}$	0.95191	0.95506	0.94271	0.93996	0.96059	
1	0	0.24677	0.25400	0.25092	0.24839	0.24984	0.97154
	1	0.25453	0.24268	0.24572	0.24789	0.25247	0.95262
	2	0.25074	0.25132	0.25228	0.24748	0.24683	0.97839
	3	0.24749	0.24311	0.24734	0.25443	0.24395	0.95851
	4	0.24398	0.24648	0.24915	0.24594	0.25011	0.97549
	$\frac{Min}{Max}$	0.95855	0.95543	0.97399	0.96663	0.96625	
2	0	0.25324	0.24876	0.25240	0.24499	0.24506	0.96742
	1	0.24545	0.24889	0.25979	0.24793	0.24848	0.94480
	2	0.24771	0.25413	0.25119	0.25839	0.24721	0.97900
	3	0.24887	0.25028	0.24872	0.24548	0.24265	0.96951
	4	0.25511	0.24370	0.24924	0.25217	0.24976	0.95527
	$\frac{Min}{Max}$	0.96213	0.95896	0.95739	0.94814	0.97153	

Table 6.10

$a=2, \theta=1, n=20, t=0.2$ When θ is known, Actual Value of ϕ is .55279					
Bayes Estimate $\hat{\phi}$					
$t'=r=0$		$t'=r=1$	$t'=r=2$	$t'=1, r=2$	$t'=2, r=1$
0.55784		0.55684	0.55404	0.55299	0.55199
0.55942		0.55982	0.55275	0.56353	0.55826
0.55829		0.55697	0.55688	0.55788	0.55688
0.55079		0.55309	0.55911	0.55092	0.54868
0.55915		0.56442	0.56405	0.55349	0.5587
Average	0.55710	0.55823	0.55737	0.55576	0.55490
Variance	0.00001284	0.00001769	0.00002008	0.00002527	0.00001920
Bias	0.00431	0.00544	0.00458	0.00297	0.00211

Table 6.11

Estimate of the TTT-Transform when θ known.

(n=20,a=2,θ=1,p=. 2)

r / t'	0	1	2	3	4	$\frac{Min}{Max}$
0	0.54493	0.54498	0.55706	0.55536	0.54861	0.98782
1	0.54473	0.55128	0.55564	0.55052	0.54698	0.98037
2	0.54676	0.54066	0.55733	0.54057	0.54497	0.96993
3	0.55051	0.54154	0.54981	0.54173	0.55046	0.98371
4	0.55839	0.54978	0.55979	0.54696	0.54346	0.97083
5	0.54660	0.55289	0.54305	0.54649	0.54452	0.98220
6	0.55649	0.55021	0.54685	0.55915	0.54302	0.97115
7	0.54366	0.55181	0.55698	0.55406	0.55150	0.97609
$\frac{Min}{Max}$	0.97695	0.97788	0.97438	0.96677	0.98762	

Table 6.12

Estimate of the TTTtransform when θ unknown.

(n=20,a=2,θ =1,t=. 2)

n'	r / z'	0	1	2	3	4	$\frac{Min}{Max}$
0	0	0.54727	0.54736	0.55600	0.54009	0.54749	0.97139
	1	0.54252	0.55213	0.54236	0.54283	0.55859	0.97095
	2	0.55935	0.55544	0.54603	0.54665	0.54554	0.97531
	3	0.54864	0.54655	0.55354	0.54355	0.54945	0.98195
	4	0.54107	0.54309	0.55522	0.55058	0.54885	0.97452
	$\frac{Min}{Max}$	0.96732	0.97777	0.97547	0.98095	0.97664	
1	0	0.55042	0.54959	0.54854	0.55609	0.55867	0.98187
	1	0.55729	0.55835	0.55135	0.55547	0.54245	0.97152
	2	0.54409	0.54806	0.54155	0.54593	0.55433	0.97695
	3	0.55463	0.54104	0.54138	0.55558	0.55128	0.97383
	4	0.54085	0.54837	0.55024	0.55809	0.54441	0.96911
	$\frac{Min}{Max}$	0.9705	0.96899	0.98192	0.97821	0.97097	
2	0	0.54727	0.54755	0.55068	0.54607	0.54918	0.99163
	1	0.55363	0.55765	0.54375	0.55378	0.55973	0.97145
	2	0.54227	0.54092	0.54901	0.55186	0.54519	0.98018
	3	0.54998	0.54599	0.55367	0.55603	0.55469	0.98194
	4	0.55275	0.54983	0.54746	0.55795	0.55433	0.98119
	$\frac{Min}{Max}$	0.97948	0.96999	0.98208	0.97871	0.97402	

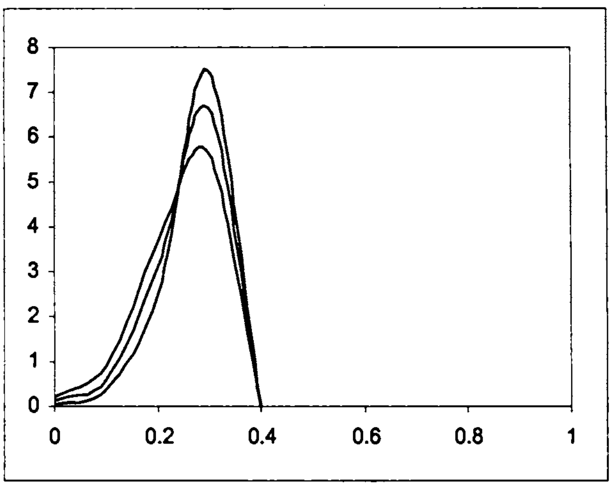

Figure 6.1
Posterior Plot for Lorenz curve
when θ known.

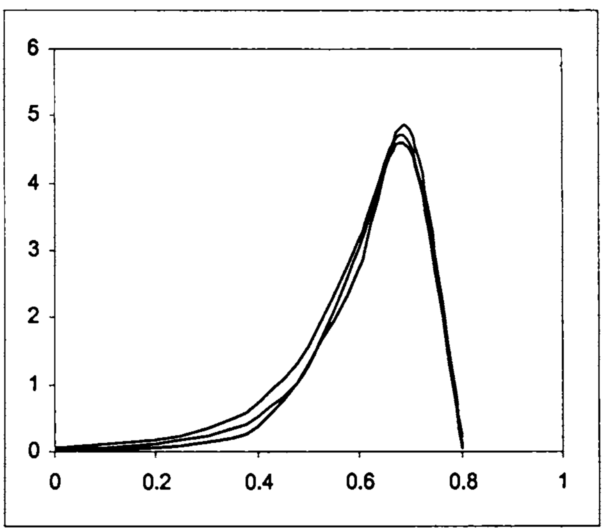

Figure 6.2
Posterior Plot for Lorenz curve
when θ unknown.

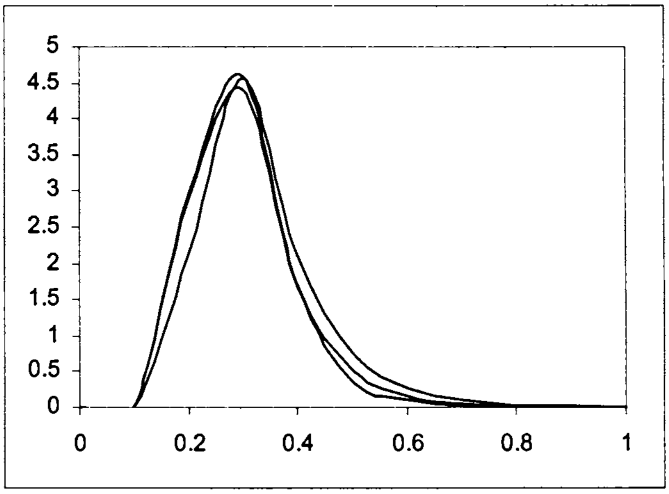

Figure 6.3
Posterior Plot for Gini-index
when θ known.

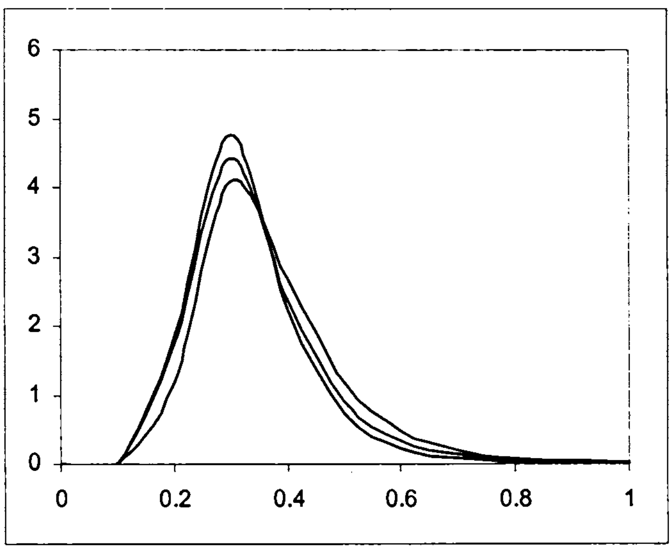

Figure 6.4
Posterior Plot for Gini-index
when θ unknown.

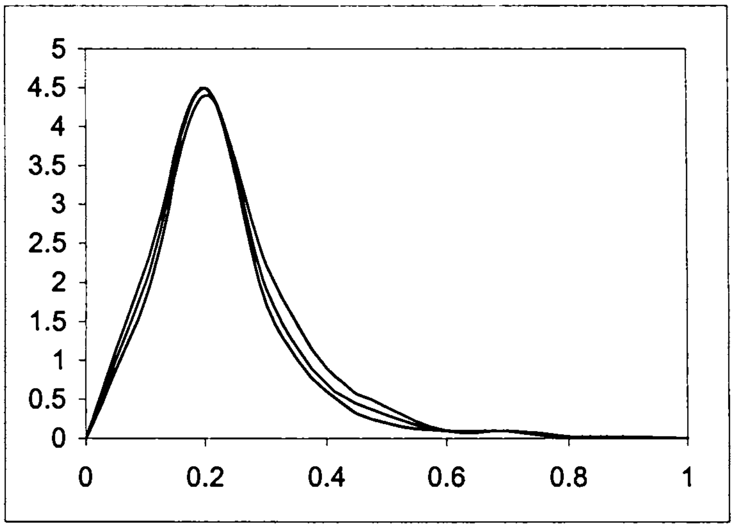

Figure 6.5
Posterior Plot for variance of logarithms when θ known.

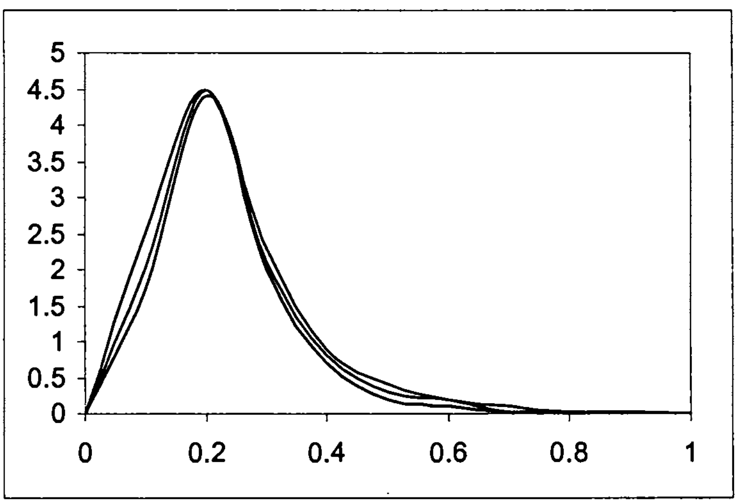

Figure 6.6
Posterior Plot for variance of logarithms when θ unknown

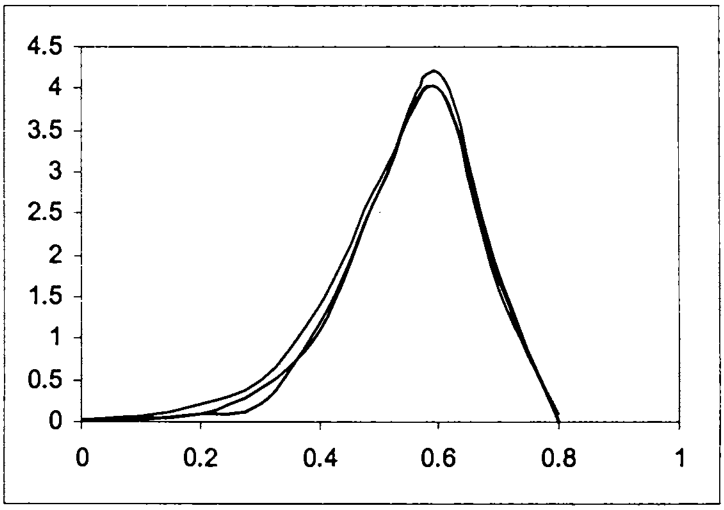

Figure 6.7
Posterior Plot for total time on test transform
when θ known

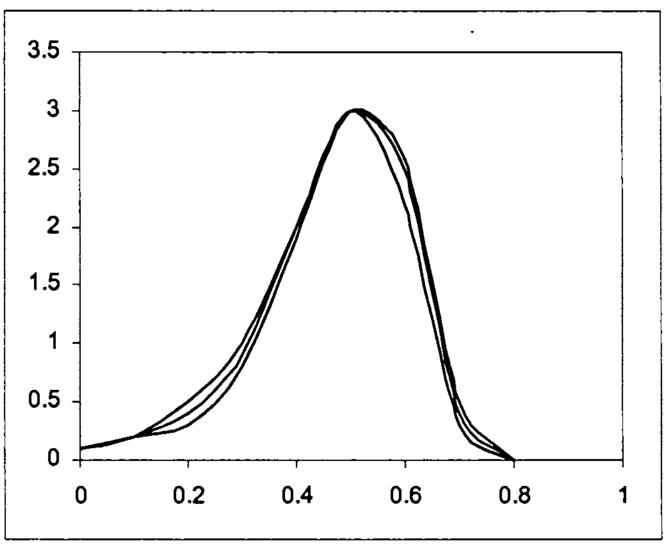

Figure 6.8
Posterior Plot for total time on test transform
when θ unknown

References

1. Ahmed, A.H.N. (1988), Preservation properties for the mean residual life ordering. *Statistical papers* 29,143-150.
2. Anand, S, (1983), *Inequality and Poverty in Malaysia*: Measurement and Decomposition. A World Bank Research publication. Oxford university press, Cambridge.
3. Arnold, B.C. (1987), *Majorization and Lorenz Order*: A Brief Introduction, Springer-Verlag (Lecturer notes in Statistics, No.43). New York.
4. Arnold, B.C and Zahedi, H. (1988), On multivariate mean remaining life functions, *J.Multiv.Anal.*25, 1-9.
5. Arnold, B.C. and Press, S.J. (1983), Bayesian inference for Pareto population, *Journal of Econometrics*, 21,287-306.
6. Atkinson, A.B., and Bourguignon, F. (1982), "The Comparison of Multidimensional Dimensioned Distributions of Economic Status", *Review of Economic Studies*, 49,183-201.
7. Atkinson, A.B., and Bourguignon, F. (1989),"The Design of Direct Taxation and Family Benefits," *Journal of Public Economics*, 41,3-29.
8. Barlow, R.E. (1979), Geometry of the total time on test transform, *Naval Res.Logist.Quart*, 26,393-402.
9. Barlow, R.E. and Campo, R. (1975), Total time on test processes and applications to failure data analysis. *Reliability and fault tree analysis.* SIAM, Philadelphia, 451-481.
10. Barlow, R.E., Bartholomew, D., Bremner, J.M., Brunk, H.D. (1972), *Statistical Inference under order restrictions*, John Wiley and Sons, Inc., New York.
11. Basu, A.P. (1971), Bivariate failure rate, *J. Amer. Statist. Assoc.*, 66, 103-104.
12. Belzunce, F., Candel, J., Ruiz, J.M. (1995), Ordering of truncated distributions through concentration curves. *Sankya, Series A* 57, 375-383.

13. Belzunce, F., Candel, J., Ruiz, J.M. (1998), Ordering and asymptotic properties of residual income distributions. *Sankya Series B* 60 331-348.

14. Bhattacharjee, M.C. (1993), How rich are the rich? Modeling Affluence and inequality via reliability theory. *Sankya Series B* 55,1-26.

15. Bhattacharya, N, and Mahalanobis, P.C. (1967), Regional disparities in house hold consumption in India.*J.Amer.Statist.Assoc.*62, 143-161.

16. Bhattacharya, N. (1963), A property of the Pareto distribution. *Sankya, Series B* 25 195-196.

17. Buchanan, H.B. and Singpurwalla, N.D. (1977), *Some stochastic characterization of multivariate survival.* In the theory and applications of reliability Vol.1, ed.C.P.Tsokos and I.N.Shimi, Academic Press, New York, 329-348.

18. Chakrabarthy, G. (1982), Estimation of Lorenz curve and Gini coefficient from grouped observations. *Artha Vijnana*.24, 120-133.

19. Chandra, M.and Singpurwalla, N.D. (1981), Relationships between some notions which are common to reliability theory and economics. *Mathematics of Operations Research,* 6,113-121.

20. Chatterjee, G.S. and Bhattacharya, N. (1974), On disparities in per capita household consumption in India.*Sankhya C*.36, 183-214.

21. Cox, D.R. (1962), *Renewal theory,* Methuen, London.

22. Dacncelli, L. (1990), *On the behaviour of the* Z_p *concentration curve, Income and wealth Distribution, Inequality and Poverty.*Proceedings, 1989,C.Dagum et al. (eds), Studies in Contemporary Economics, Springer-Verlag, 111-127.

23. Deshpande, J.V. and Suresh, R.P. (1990), On nonmonotonic ageing, *Scand.J.Statist.*17, 257-262.

24. Ebrahimi, N. (1996), How to measure uncertainty in the residual lifetime distribution, *Sankhya*, Series A, 58, 48 - 56.

25. Ebrahimi, N.and Pellerey F. (1995), New partial ordering of survival functions based on the notion of uncertainity, *J.App.Prob*, 32,202-211.

26. Enhorn, H.A. (1962), Changes in concentration of domestic manufacturing establishment output: 1939-1958,*J.Amer.Statist.Assoc.*57, 797-803.

27. Foster, J.E. (1984), On economic poverty: a survey of aggregate measures. *In Advances in Economics*, Vol.3, eds, Basmann R. and Rhodes, Greenwich, CT: JAI Press.

28. Frosini, B.V. (1988), Introduzione alla Statistica, Roma: La Nuova Italia Scientifica.

29. Galambos, J. and Kotz, S. (1978), *Characterizations of Probability Distributions*, Springer-Verlag, New York.

30. Gastwirth, J.L. (1971), A general definition of Lorenz curve.*Econometrica*.39, 1037-1039.

31. Gastwirth, J.L. (1972), The estimation of the Lorenz curve and Gini-index, *Rev.Econ.Statist.*54, 306-216.

32. Gibrat, R. (1931), *Les Inegalities Economiques*.Sirey, Paris.

33. Gini, C. (1912), *Variabilita e Mutabilita*.Bologna.

34. Goldie, C.M. (1977), Convergence theorems for empirical Lorenz curves and their inverses, *Advances in Appl.Prob.9, 765-*791.

35. Hart, P.E. (1971), Entropy and other measures of concentration.*J.Roy.Statist.Soc.*A.134, 73-85.

36. Hart, P.E. (1975), Moment distributions in economics: an exposition, *J.Roy.Statist.Soc.*A.138, 423-434.

37. Hart, P.E. and Paris, S.J. (1956), The analysis of business concentration: a statistical approach, *J.Roy.Statist.Soc.*A.199, 150-181.

38. Hitha, N. (1991), *Some characterizations of Pareto and related populations*, PhD Thesis, Cochin University of Science and Technology, India.

39. Jeevanand, E.S. and Nair, N.U. (1992), Bayes estimate of Pareto parameters in the presence of outliers, *Proceedings of symposium on statistical Inference*, 1992.

40. Johnson, N.L. and Kotz, S. (1975), A vector valued multivariate hazard rate, *J. Multi. Anal.* 5, 53-66.

41. Kakwani, N.C. (1980), *Income Inequality and Poverty: Methods of estimation and policy applications*, World Bank, Oxford University Press, Cambridge

42. Kakwani, N.C. and Podder, N. (1976), Efficient estimation of the Lorenz curve and associated inequality measures.*Econometrica*.44, 137-148.

43. Kendall, M.G. and Stuart, A. (1977), *Advances theory of Statistics*, Vol.1.4th Edn.Charles Griffin and Company, London.

44. Klefsjo, B. (1982), On ageing properties and the total time on test transform, *Scan.J.Statisti*.9, 37-41.

45. Klefsjo, B. (1984), Reliability interpretations of some concepts from economics. *Naval Res.Logist.Qrtrly*, 31,301-308.

46. Kolm, S.C. (1977), "Multidimensional Egalitarianisms," *Quarterly journal of Economics*, 91,1-13.

47. Koshevoy, G., and Mosler, K (1997) "Multivariate Gini indices", *Journal of Multivariate Analysis*, 60,252-276.

48. Koshevoy, G., and Mosler, K. (1996), "The Lorenz Zonoid of a Multivariate Distribution", *Journal of the American Statistical Association*, 91,873-882

49. Kotz, S. and Shanbag, D.N. (1980), Some new approaches to probability distributions, *Adv.Appl.Prob*.12, 903-921.

50. Kullback, S. and Leibler, R.A. (1951) On information and sufficiency, *Ann.Math.Statist*, 22,79-86.

51. Kupka, J. and Loo, S. (1989), The hazard and vitality measures of ageing, *J.Appl.Prob*.12, 903-921.

52. Lindley, D.V. and Singpurwalla, N.D. (1986), Multivariate distribution for the life lengths of components of a system sharing a common environment, *J.Appl.Prob.*23, 418-431.

53. Lorenz, M.O. (1905), Methods of measuring concentration of wealth. *Amer.Statist.Assoc.New series.* 70,209-219.

54. Mailhot (1990), Some properties of truncated distributions connected with log-concavity of distributions functions. *Applicationes Mathematicae* 20,531-542.

55. Malik, H.J. (1970), Bayesian estimation of the Paretian index, *Scandinavian Actuarial Journal*, 53,6-9.

56. Massoumi, E. (1986),"The Measurement and Decomposition of Multi Dimensional Inequality,"*Econometrica*, 54,991-997

57. Massoumi, E., and Nickelsburg, G. (1988),"Multivariate Measures of well-being and an Analysis of Inequality in the Michigan Data," *Journal of Business and Economic Statistics*, 6,327-334.

58. Moothathu, T.S.K. (1985a), Distributions of maximum likelihood estimators of Lorenz curve and Gini index of exponential distribution.*Ann.Inst.Statist.Math.*A, 37,473-479.

59. Moothathu, T.S.K. (1985b), Sampling distributions of Lorenz curve and Gini index of Pareto distribution, *Sankhya B*.47, 247-258.

60. Moothathu, T.S.K. (1985c), *A study of Lorenz curve and Gini-index of some Statistical Models*, Ph.D.thesis, University of Kerala, Trivandrum.

61. Moothathu, T.S.K. (1986), A characterization of power function distribution through a property of the Lorenz curve. *Sankhya* B.48, 262-265.

62. Moothathu, T.S.K. (1989a),On maximum likelihood estimators of certain inequality measures of lognormal distribution, paper presented at *National seminar on Statistical Inference*, June 21-24,1989,University of Poona.

63. Moothathu, T.S.K, (1989b), On unbiased estimation of Gini-index and Yntema-Pietra index of lognormal distribution and their variances, commun. *Statist-Theory Meth.*18, 661-672.

64. Moothathu, T.S.K. (1990), The best estimator and a strongly consistent, asymptotically normal, unbiased estimator of Lorenz curve, Gini-index and theil entropy index of Pareto distributions. *Sankhya* B.

65. Mosler, K. (1994),"*Multidimensional Welfarisms,*" in Models and Measurement of Inequality and Welfare, ed.W.Eichhorn, Berlin: Springer, pp.808-820.

66. Mukherjee, S.P.and Roy, D. (1986), Some characterizations of the exponential and related distributions.*Cal.Statist.Assoc.Bul,* 35, 189-197.

67. Nair, K.R.M. and Nair, N.U. (1988), Bivariate mean residual life, *IEEE Trans.Rel.*38, 362-364.

68. Nair, K.R.M. and Rajesh. (2000), Geometric vitality function and its applications to reliability. *IAPQR Tran.,* 25 (1), 1-8.

69. Nair, K.R.M.andRajesh, G. (1998), Characterization of probability distributions using the residual entropy function, *J. Ind. Statist. Assoc.*, 36, 157-166.

70. Ord, J.K., Patil, G.P, Taillie, C. (1983), Truncated distributions and measures of income inequality. *Sankhya: The Indian Journal of Statistics,* Ser.B.45, 413-430.

71. Pham, T.G., Turkkan, N., (1994), The Lorenz and Scaled Total-Time-on-Test Transform Curves: A unified Approach.*IEEE Trans.Reliability*, vol 43,76-84.

72. Raiffa, H.and Schlaifer, R. (1961), *Applied statistical decision theory,* Harvard University graduate school of business administration, Boston.

73. Rajesh and K.R.Muraleedharan Nair. (1998), Residual entropy function in discrete time.*Far.East.J.Theo.Stat 2 (1),* 1-10.

74. Rietveld, P. (1990),"Multidimensional Inequality Comparisons," *Economics Letters*, 32,187-192

75. Roy, D. (1989)," A characterization of Gumbels bivariate exponential and Lindley and Singpurwalla bivariate Lomax distributions,"*J.Appl.prob*, 27,886-891.

76. Sankaran, P.G. (1992), *Characterization by properties of probability distributions by reliability concepts*, Ph.D.thesis, Cochin University of Science and Technology, India.

77. Sankaran, P.G. and Nair, N.U. (1991), On bivariate vitality functions, *Proceedings of National Symposium on Distribution theory*.

78. Sen, A. (1976), The measurement of poverty: an axiomatic approach.*Econometrica*.44, 219-322.

79. Sen, P.K (1986), The Gini coefficient and poverty indices; some reconcilations.*J.amer.Statist.Assoc*.81, 1050-1057.

80. Shanbhag, D.N. and Kotz, S. (1987), Some new approaches to multivariate probability distributions, *J.Multiv.Anal*.22, 189-211.

81. Shannon, C.E. (1948), A mathematical theory of communication, *Bell system Tech.J*, 279-423,623-656.

82. Shrycock, S. and Siegel, J.S. (1973), *The methods and Materials of demography*, Vol.I.U.S.Government Printing Office, Washington D.C.

83. Sinha, S.K (1980) *Bayesian estimation*, New-age Publications, New Delhi.

84. Sinha, S.K. and Howlader, H.A. (1980), *On the sampling distributions of Bayesian estimators of the Pareto Parameter with proper and improper priors and associated goodness of fit*, Technical Report #103,Department of Statistics, University of Manitoba, Canada.

85. Slottje, D.J. (1987)," Relative Price Changes and Inequality in the Size Distribution of Various Components,"*Jounal of business and Economic Statistics*, 5,19-26.

86. Sullo, P. and Rutherford, D.K. (1977), Characterisation of the power distribution by conditional exceedence. *The American Association Proceeding of the Business and Economics*, 787-792.

87. Suresh, R.P. (1987). *Total time on test transform and nonmonotonic ageing*, Unpublished M.Phil dissertation, University of Poona.

88. Swartz, G.B. (1973), The mean residual lifetime function. *IEEE Trans.Rel*, 22,108-109.

89. Taguchi, T. (1972a), " On the Two-Dimensional Concentration Surface and Extensions of Concentration Coefficient and Pareto Distribution to the Two-Dimensional Case-I," *Annals of the Institute of Statistical Mathematics*, 24,355-382.

90. Taguchi; T (1972b), " On the Two-Dimensional Concentration Surface and Extensions of Concentration Coefficient and Pareto Distribution to the Two-Dimensional Case-II," *Annals of the Institute of Statistical Mathematics*, 24,599-619.

91. Takayama, N. (1979), Poverty, income inequality and their measures: Professor Sen's axiomatic approach reconsidered, *Econometrica*.47, 747-760.

92. Theil, H. (1967), *Economics and Information Theory*. North-Holland, Amsterdam.

93. Thompson, W.A. (1976), Fisherman's luck. *Biometrics* 32, 265-271.

94. Zellener, A. (1971), *An introduction to Bayesian inference in Econometrics*, Wiely, New York.