Chapter 4

The Zenga curve in the context of reliability analysis

4.1 Introduction

As mentioned in Section 2.3, the focal theme of interest in reliability analysis is the modelling and analysis of lifetime data. To enable this, certain concepts such as failure rate, mean residual life function etc. which are capable of describing the failure pattern are formulated and are used to obtain lifetime models. If \( X \) represents the lifetime of a component or device, a random variable which has received much interest in reliability analysis is the truncated random variable \( X \mid X > x \) as well as \( X \mid X \leq x \). The average values namely, \( \mu^+(x) = E(X \mid X > x) \) and \( \mu^-(x) = E(X \mid X \leq x) \) represents the average lifetime of components which has attained age \( x \) and the average lifetime of components which has failed before attaining age \( x \). The former, namely \( \mu^+(x) \), is the vitality function and is extensively studied by Kupka & Loo (1989) and Nair & Rajesh (2000).

The Zenga curve, defined in (3.1), is given in terms of \( \mu^+(x) \) and \( \mu^-(x) \). Observing that one can write (3.1) as

\[
A(x) = \frac{\mu^+(x) - \mu^-(x)}{\mu^+(x)}.
\]

\( A(x) \) shall be interpreted as the difference in average age of components which has survived beyond age \( x \) from those which has failed before attaining age \( x \), expressed in terms of average
age of components exceeding age $x$. $A(x)$ shall be viewed as a measure of proportional change in average age while switching over from survival before and after attaining age $x$. In this sense, the Zenga curve has a lot of significance in the study of reliability of components. If $m(x) = E(X - x | X > x)$ represents the mean residual life function and $r(x) = E(x - X | X \leq x)$ represents the mean waiting time, Zenga (2008) has represented (3.1) in the form

$$A(x) = \frac{r(x) + m(x)}{m(x) + x}.$$ 

Using the relationships

$$\mu^{-}(x) = \frac{\mu - \bar{F}(x) [x + m(x)]}{F(x)}$$

and

$$\mu^{+}(x) = x + m(x)$$

in the definition of $A(x)$, one can get alternate representation for $A(x)$ as

$$A(x) = \frac{1}{F(x)} \left[ 1 - \frac{\mu}{x + m(x)} \right],$$

where $\mu = E(X)$ represents the average lifetime.

Although several representations for the Zenga curve are feasible, the representation in terms of the quantile function given in (3.3) is more mathematically tractable. Further very little work seems to have been done on the Zenga curve in the quantile framework. Motivated by this, in the present chapter we look into the problem of (i) determining the possible relationships of the curve with other inequality measures as well as reliability concepts ii) characterization of probability distributions using these relationships iii) classification of lifetime distributions using the Zenga curve and iv) examining the behaviour of the curve using an empirical data on survival times. Some of the results of this chapter are included in Nair & Sreelakshmi (2012).
4.2 Zenga curve and other inequality measures

The relationship between the Zenga curve and the Lorenz curve was examined in section 3.2. It was observed there exists an explicit relationship between them, which is given in (3.5). In view of this relationship the knowledge of one of them enables to determine the other and hence the results especially characterization theorems for the Lorenz curve can be reformulated in terms of the Zenga curve. Bonferroni (1930) has proposed another measure of inequality which is referred to as the Bonferroni curve. For a nonnegative random variable $X$ admitting an absolutely continuous distribution, the Bonferroni curve is defined as

$$B(p) = \frac{1}{\mu p} \int_0^p Q(u)du.$$  

Observing that the Bonferroni curve is connected to the Lorenz curve through the relationship $L(p) = pB(p)$, from (3.5), we get

$$I(p) = \frac{1 - B(p)}{1 - pB(p)}$$  

or

$$B(p) = \frac{1 - I(p)}{1 - pI(p)}.$$  

In view of the above relationships it is inherent that $I(p)$ and $B(p)$ determine each other uniquely.

Another inequality measure extensively used in Informetrics is the Leimkuhler curve defined by

$$K(p) = \frac{1}{\mu} \int_p^1 Q(u)du.$$
The Lorenz curve $L(p)$ and $K(p)$ are connected through the relationship

$$K(p) = 1 - \frac{1}{\mu} \int_0^{1-p} Q(u)du$$

$$= 1 - L(1 - p).$$

Using (3.5) and the above expression, we get

$$I(p) = \frac{1}{p} \left( 1 - \frac{1 - p}{K(1 - p)} \right)$$

or

$$K(p) = \frac{p}{1 - (1 - p)I(1 - p)}.$$

The above relations enables one to evaluate $K(p)$ through the knowledge of $I(p)$ and vice versa.
### Table 4.1: The three curves for different distributions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Lorenz curve</th>
<th>Bonferroni curve</th>
<th>I(p) curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
<td>$p^{\frac{1}{\beta}+1}$</td>
<td>$\frac{1}{p^{\beta}}$</td>
<td>$\frac{1-p^{\frac{1}{\beta}}}{1-p^{\frac{1}{\beta}+1}}$</td>
</tr>
<tr>
<td>Exponential</td>
<td>$p + (1 - p) \log(1 - p)$</td>
<td>$1 + \frac{(1-p)}{p} \log(1 - p)$</td>
<td>$\frac{1}{p[1-(\log(1-p))^{-1}]}$</td>
</tr>
<tr>
<td>Pareto II</td>
<td>$\frac{c}{p(c-1)} \left(1 - (1 - p)^{1 - \frac{1}{c}}\right)$</td>
<td>$\frac{c(1-1-1)^{\frac{1}{c}}}{p} - (c - 1)$</td>
<td>$\frac{c[1-(1-p)^{\frac{1}{c}}-1]}{p[c(1-(1-p)^{\frac{1}{c}}-1)+1]}$</td>
</tr>
<tr>
<td>Pareto I</td>
<td>$\frac{\alpha}{p} \left[1 - (1 - p)^{1 - \frac{1}{\alpha}}\right]$</td>
<td>$\frac{\alpha}{p} \left[1 - (1 - p)^{1 - \frac{1}{\alpha}}\right]$</td>
<td>$\frac{1-(1-p)^{\frac{1}{\alpha}}}{p}$</td>
</tr>
<tr>
<td>Rescaled beta</td>
<td>$\frac{c}{p(c+1)} \left(1 - (1 - p)^{\frac{1}{c}}\right)$</td>
<td>$\frac{c(1-1-1)^{\frac{1}{c}}}{p} + (c + 1)$</td>
<td>$\frac{c[1-(1-p)^{\frac{1}{c}}]}{p[c(1-(1-p)^{\frac{1}{c}})+1]}$</td>
</tr>
<tr>
<td>Govindarajulu</td>
<td>$p^{\beta+1} \left[\frac{\beta+2-\beta p}{\beta+2}\right]$</td>
<td>$p^\beta \left[\frac{\beta+2-\beta p}{\beta+2}\right]$</td>
<td>$\frac{2-p^\beta(\beta+2-\beta p)}{p^2[2-p^{\beta+1}(\beta+2-\beta p)]}$</td>
</tr>
</tbody>
</table>
The following graphs give $I(p)$ curve for different distributions

Figure 4.1: power distribution; $\beta = 2$

Figure 4.2: Govindarajulu distribution; $\beta = 6$

Figure 4.3: unit exponential distribution

Figure 4.4: Pareto I distribution; $\alpha = 3$

Figure 4.5: Pareto II distribution; $c = 3$
4.3 Relationship between the Zenga curve and certain reliability measures

The utility of reliability concepts in Economic analysis had been the focal theme of investigation in several works. Concepts such as moments of residual life, vitality function etc. has been advantageously used to identify the models to represent income data. In this section, we establish certain relationships between the Zenga curve and certain reliability concepts such as mean residual quantile function and reversed mean residual quantile function. Note that the approach used here is the representation using the quantile functions. These relationships are used subsequently to arrive at characterization results for certain distributions. If $M(p)$ represents the mean residual quantile function reviewed in section 2.3.2, there exists the relationship

$$
\int_0^p Q(u)du = \mu (M(p) + Q(p)(1 - p)) .
$$

(4.1)

Using the definition of $I(p)$ curve given in (3.3) and (4.1), it follows that

$$
I(p) = \frac{M(p) + Q(p) - \mu}{p (M(p) + Q(p))} .
$$

(4.2)

From (2.8) and (4.2), we have,

$$
I(p) = p^{-1} \left[ 1 + \frac{\mu}{\int_0^p \frac{M(u)}{1-u} du} \right]^{-1} .
$$

(4.3)

Rearranging the terms in the above equation, we get

$$
\int_0^p \frac{M(u)}{1-u} du = \frac{\mu p I(p)}{1 - p I(p)} .
$$
Differentiating the above expression with respect to $p$, we get

\[
\frac{M(p)}{1-p} = \mu \left[ pI'(p) + I(p) \right] \left[ 1 - pI(p) \right]^2
\]

or

\[
M(p) = (1-p)\mu \frac{d}{dp} \left[ \frac{1}{1-pI(p)} \right].
\] (4.4)

Similarly from the definition of reversed mean residual quantile function $R(p)$ given in section 2.3.4, we get

\[
\int_0^p Q(u)du = p(Q(p) - R(p)).
\] (4.5)

From (4.5) and (2.9), we get

\[
\int_0^p Q(u)du = p \int_0^p \frac{R(u)}{u}du.
\]

Substituting the above expression in (3.3) we have

\[
I(p) = \frac{\mu - \int_0^p \frac{R(u)}{u}du}{\mu - p \int_0^p \frac{R(u)}{u}du}.
\]

or

\[
\int_0^p \frac{R(u)}{u}du = \mu \frac{1 - I(p)}{1 - pI(p)}.
\]

Differentiating the above expression with respect to $p$, we get

\[
R(p) = p\mu \left[ \frac{(p - 1)I'(p) + I(p)(1 - I(p))}{(1 - pI(p))^2} \right].
\] (4.6)

(4.6) provides the relationship between $I(p)$ curve and reversed mean residual quantile function.
Examples:

1. For the uniform distribution with quantile function

\[ Q(p) = bp, \]

by direct calculations we get,
\[ I(p) = \frac{1}{1+p} \text{ and } M(p) = \frac{b}{2}(1 - p). \]

The relationship (4.4) is immediate since \( M(p) = \mu (1 - p) \) where \( \mu \) is the mean.

2. Rohde (2009) has examined the potential of the truncated Pareto distribution as a suitable model for income data and studied the properties of the corresponding distribution. Subsequently Sarabia et al. (2010b) showed that the model proposed by Rohde is a reparameterization of the model proposed by Aggarwal (1984) and they have discussed some important economic properties of the model.

The distribution is specified by

\[ F(x) = \eta - \alpha \frac{x^{-\frac{1}{2}}}{\eta^2}; \quad \frac{\alpha}{\eta^2} \leq x \leq \frac{\alpha}{(\eta - 1)^2} \]

where \( \alpha = \eta(\eta - 1)\mu \). The quantile function associated with \( F(x) \) is

\[ Q(p) = \frac{\alpha}{(\eta - p)^2}. \]

The \( I(p) \) curve and \( R(p) \) simplifies to

\[ R(p) = \frac{\alpha p}{\eta(\eta - p)^2} \quad (4.7) \]

and

\[ I(p) = \frac{1}{\eta} (= c, \text{ a constant}). \]
From (4.5), we get
\[ R(p) = \frac{p\mu c(1-c)}{(1-cp)^2} \]
which is same as the \( R(p) \) given in (4.6) with \( \alpha = \frac{\mu(1-c)}{c^2} \) and \( \eta = \frac{1}{c} \).

In the sequel, we look into the problem of characterizing probability distributions using possible relationships between \( I(p) \) and certain reliability concepts. Our first characterization result pertains to the class of distributions considered in Nair & Sankaran (2009) using a relationship between the \( I(p) \) curve and hazard quantile function \( H(p) \).

**Theorem 4.1.** Let \( X \) be a nonnegative continuous random variable with \( E(X) < \infty \). Then there exists a function \( g(.) \) satisfying
\[
I(p) = \frac{H(p)g \left[ Q(p) \right]}{p(\mu + H(p)g \left[ Q(p) \right])}
\]
if and only if
\[
\frac{f' \left[ Q(p) \right]}{f \left[ Q(p) \right]} = \frac{\mu - Q(p) - g' \left[ Q(p) \right]}{g \left[ Q(p) \right]}. \tag{4.8}
\]

**Proof.** When (4.8) holds, using (4.2) we have
\[
\frac{M(p) + Q(p) - \mu}{p(M(p) + Q(p))} = \frac{H(p)g \left[ Q(p) \right]}{p(\mu + H(p)g \left[ Q(p) \right])}.
\]
The above equation simplifies to,
\[
M(p) + Q(p) = \mu + H(p)g \left[ Q(p) \right]. \tag{4.9}
\]
Nair & Sankaran (2009) showed that (4.9) holds if and only if the density quantile function has the form
\[
\frac{f' \left[ Q(p) \right]}{f \left[ Q(p) \right]} = \frac{\mu - Q(p) - g' \left[ Q(p) \right]}{g \left[ Q(p) \right]},
\]
Remark 4.1. The above general class of distributions include the Pearson, beta, gamma distributions etc. For instance, for the exponential distribution with quantile function,

\[ Q(p) = -\frac{1}{\lambda} \ln(1 - p), \quad \lambda > 0, \]

the following relation exists between \( I(p) \) and \( H(p) \).

\[ I(p) = \frac{1}{p \left[ 1 + (Q(p)H(p))^{-1} \right]}. \]

For the gamma distribution specified by the p.d.f.

\[ f(x) = \frac{m^n}{n} e^{-mx} x^{n-1}, \quad x > 0, \]

the form of \( g[Q(p)] \) can be identified as \( g[Q(p)] = \frac{Q(p)}{m} \) and the \( I(p) \) curve is related to \( H(p) \) through the relationship

\[ I(p) = \frac{1}{p \left[ \frac{\mu m}{H(p)Q(p)} + 1 \right]}. \]

Next two theorems provide characterization results using the mean residual and reversed mean residual quantile functions.

**Theorem 4.2.** For a nonnegative continuous random variable \( X \), the relationship

\[ pI(p) = \frac{A - M(p)}{B - M(p)} \tag{4.10} \]

holds if and only if \( X \) follows the distribution specified by the quantile function

\[ Q(p) = \frac{\mu B}{B - A} + C(1 - p)^{\frac{B - A}{n}} \tag{4.11} \]
provided $C(A - B) > 0$.

**Proof.** When (4.10) holds, using (3.5), we get

$$\frac{p - L(p)}{1 - L(p)} = \frac{A - M(p)}{B - M(p)}$$

This gives,

$$L(p) = \frac{A - pB - (1 - p)M(p)}{A - B}.$$ 

Using the definition of $L(p)$ and $M(p)$, we have

$$\frac{1}{\mu} \int_0^p Q(u)du = \frac{A - pB}{A - B} + \frac{1 - p}{A - B} \left( \frac{1}{1 - p} \int_0^1 Q(u)du - Q(p) \right).$$

Differentiating the above expression with respect to $p$ and rearranging the terms, we get

$$q(p) - \frac{A - B}{(1 - p)\mu}Q(p) - \frac{B}{1 - p} = 0$$

The solution of the above differential equation is,

$$Q(p) = \frac{\mu B}{B - A} + C(1 - p)^{\frac{B - A}{\mu}}.$$ 

For $Q(p)$ is an increasing function, $C(A - B) > 0$. The proof of the converse is straightforward and hence omitted. \(\Box\)

**Remark 4.2** Setting $B = 0$ in (4.10) and (4.11), we get

$$pI(p) = 1 - \frac{A}{M(p)}$$

and

$$Q(p) = C(1 - p)^{\frac{B}{\mu}}.$$
Put $C = k$ and $A = \frac{\mu}{\sigma}$ in the above expression, we get Pareto distribution of first kind with quantile function,

$$Q(p) = k(1 - p)^{-\frac{1}{\alpha}}.$$

Further setting $Q(0) = 0$ in (4.11), we get

$$B\mu = C(A - B).$$

Using the above expression in (4.11) we get,

$$Q(p) = C \left[ (1 - p)^{-\frac{B}{C} - 1} \right].$$

If $C = \frac{b}{a}$ and $B = -\frac{b}{a+1}$, we get the quantile function of generalised Pareto distribution. For the ranges $0 < a < 1$ and $-1 < a < 0$, we get Pareto II and Rescaled Beta distributions respectively as special cases. But exponential distribution is not a special case.

**Theorem 4.3.** For a non negative random variable $X$ with reversed mean residual quantile function $R(p)$, the relationship

$$I(p) = \frac{1 - \beta R(p)}{1 - \beta p R(p)}; \beta > 0 \quad (4.12)$$

holds if and only if $X$ follows power distribution specified by the quantile function

$$Q(p) = \sigma p^{\frac{1}{\phi}}; \sigma, \phi > 0. \quad (4.13)$$

**Proof.** For the quantile function given in (4.13), direct calculations give

$$R(p) = \frac{\sigma}{\phi + \frac{1}{p^{\frac{1}{\phi}}}}.$$
and

\[ I(p) = \frac{1 - \beta R(p)}{1 - \beta p R(p)} \]

where \( \beta = \frac{\phi + 1}{\sigma} \).

Conversely, suppose that (4.12) holds, using (3.5) and (4.12), we have

\[
\frac{p - L(p)}{p (1 - L(p))} = \frac{1 - \beta R(p)}{1 - \beta p R(p)}.
\]

The above equation gives,

\[ L(p) = \beta p R(p). \]

That is,

\[
\frac{1}{\mu} \int_0^p Q(u) du = \beta p \left( Q(p) - \frac{1}{p} \int_0^p Q(u) du \right).
\]

Differentiating the above equation with respect to \( p \), we get

\[
\frac{q(p)}{Q(p)} = \frac{1}{p \beta \mu}.
\]

The solution to the above differential equation is

\[ Q(p) = C p^{\frac{1}{\mu}}. \quad (4.14) \]

Put \( C = \sigma \) and \( \beta = \frac{\phi}{\mu} \), we get the quantile function given in (4.13) and the theorem follows.

### 4.4 Classification of Lifetime distributions

In the reliability context, concepts of ageing describe how a component or system improves or deteriorates with age. As an extension to the work of Chandra & Singpurwalla (1981), Klefsjö (1984) and Kochar & Xu (2009) have discussed the ageing properties such as IFR, IFRA, NBUE
and HNBUE based on the Lorenz curve and their related partial orderings. In this section we obtain certain limits for \( I(p) \) using certain criteria based on ageing. Our first result focuses attention on a necessary and sufficient condition for a distribution to be IFR.

**Theorem 4.4.** Let \( X \) be a continuous random variable with distribution function \( F(x) \), finite mean \( \mu \) and \( I(p) \) curve denoted by \( I_F(p) \). Assume that \( G(.) \) is the distribution function of a random variable following exponential distribution with same mean \( \mu \) and \( I(p) \) curve specified by

\[
I_G(p) = \frac{1}{p \left[ 1 - (\ln(1 - p))^{-1} \right]}.
\]

\( F \) is IFR if and only if

\[
I_F(p) \leq I_G(p).
\]

**Proof.** Barlow & Proschan (1975) has shown that \( F \) is IFR if and only if \( F \leq_c G \). Hence

\[
\begin{align*}
F \text{ is IFR } & \iff F \leq_c G \\
& \iff \int_0^1 Q_F(u)du \leq \int_0^1 Q_G(u)du \\
& \iff \int_0^p Q_F(u)du \geq \int_0^p Q_G(u)du \\
& \iff I_F(p) \leq I_G(p).
\end{align*}
\]

We have discussed some ageing concepts using the quantile based representation in chapter 2. Based on these definitions, lifetime distributions can be classified using the Zenga curve. The next two theorems provide sufficient conditions for distributions belonging to different ageing classes in terms of the \( I(p) \) curve.

**Theorem 4.5.** A nonnegative continuous random variable \( X \) is IMRL (DMRL) if

\[
I(p) \geq (\leq) \frac{1}{p} \left[ \frac{Q(p)}{Q(p) + \mu} \right].
\]
Proof. Using (2.16) and (4.4), we get

\[ X \text{ is IMRL(DMRL)} \Rightarrow (1 - u)\mu \frac{d}{du} \frac{1}{1 - uI(u)} \geq (\leq) \frac{1}{H(u)} \]

\[ \Rightarrow \frac{d}{du} \frac{1}{1 - uI(u)} \geq (\leq) \frac{1}{(1 - u)\mu H(u)} = \frac{q(u)}{\mu}. \]  \hspace{1cm} (4.15)

On integration from 0 to \( p \), (4.15) becomes

\[ \frac{1}{1 - pI(p)} \geq (\leq) \frac{Q(p)}{\mu} + 1. \]

This implies

\[ I(p) \geq (\leq) \frac{1}{p} \left[ \frac{Q(p)}{Q(p) + \mu} \right] \]

as claimed.

\[ \square \]

**Theorem 4.6.** Let \( X \) be a lifetime random variable with finite positive mean \( \mu \). A sufficient condition for \( X \) to be UBAE (UWAE) is that

\[ I(p) \geq (\leq) \frac{1}{p} \left\{ 1 - \frac{\mu}{M(1) \log(1 - p)} \right\}^{-1}. \]

Proof. From (2.17) and (4.4), we have

\[ X \text{ is UBAE(UWAE)} \Rightarrow \frac{d}{du} \frac{1}{1 - uI(u)} \geq (\leq) \frac{M(1)}{\mu(1 - u)}. \]  \hspace{1cm} (4.16)

Integrating (4.16) from 0 to \( p \), we get

\[ \frac{1}{1 - pI(p)} \geq (\leq) 1 - \frac{M(1) \log(1 - p)}{\mu} \Rightarrow I(p) \geq (\leq) \frac{1}{p} \left\{ 1 - \frac{\mu}{M(1) \log(1 - p)} \right\}^{-1}. \]

The proof is complete.  \[ \square \]
Analogous to the above theorems, one can find the sufficient conditions for ageing classes like increasing hazard rate average, IHRA (decreasing hazard rate average, DHRA), harmonic new better than used in expectation, HNBUE (harmonic new worse than used in expectation, HNWUE) in terms of $I(p)$ curve using the basic definitions given in Nair & Vineshkumar (2011) and the relationships discussed in section 4.3. The results are mentioned below. The proof of the results are analogous to that of theorem 4.5 and theorem 4.6 and hence not included.

1) If $X$ is IHRA (DHRA), then

$$I(p) \geq (\leq) \frac{z(p) - p}{p(z(p) - 1)},$$

where $z(p) = -\frac{1}{\mu} \int_0^p \frac{\log(1-u)}{H(u)} du$.

2) If $X$ is HNBUE (HNWUE), then

$$I(p) \geq (\leq) \frac{1}{p} \left\{ 1 - \left[ \frac{1}{1-p} - \int_0^p \frac{1}{(1-u)^2} e^{-\frac{Q(u)}{\mu}} du \right]^{-1} \right\}.$$

### 4.5 Illustration

In the context of reliability theory, the form of the hazard function enables to find the appropriate model for the lifetime data. In the sequel we illustrate the behaviour of the Zenga curve empirically using a survival data considering the quantile model having linear hazard quantile form with quantile function specified by

$$Q(p) = \frac{1}{a + b} \log \left[ \frac{a + bp}{a(1-p)} \right]$$

We fit the model specified by (4.17) to the survival data given in Bryson & Siddiqui (1969) and examine the behaviour of $I(p)$ curve. The data contains survival time of 43 patients suffering from chronic granulocytic leukemia. We fit the model (4.17) for the data by the method of L
moments.
The first two L moments are given by

\[
l_1 = \frac{1}{a+b} \log \left( \frac{a+b}{a} \right)
\]

\[
l_2 = \frac{a \log \left( \frac{a+b}{a} \right) + b}{b^2}
\]

Equating the population L moments to the sample L moments, the estimates of parameters are evaluated as

\[
a = 0.000666573, \quad b = 0.000972694.
\]

Here the data is divided into 5 groups. Corresponding \( x \) values with respect to the values of \( p \) obtained by \( p_i = \frac{i}{5}; i = 1, 2, ..., 5 \) are used to get the observed and expected frequencies. Thus the chisquare value obtained here is 3.2 which is admissible so that the model (4.17) fits well to the data. This fact is also evidenced by the Q-Q plot given as figure 4.6.
Chapter 4. The Zenga curve in the context of reliability analysis

The plot of $\hat{I}(p)$ for different values of $p$ is given below.

![Plot of $\hat{I}(p)$](image)

Figure 4.7:

From figure 4.7, one can observe that the curve is bathtub shaped. Further one can compute the average survival time of the least fortunate $p$100% of the patients is $I(p)$ 100% lower than that of remaining $(1 - p)$ 100% of the patients suffering chronic granulocytic leukemia and for example $I(.8) = .70$ can be interpreted as the average survival time of the least fortunate 80% of patients is 70% lower than that of remaining 20% of the patients.

4.6 Quantile based income models

In this section, we consider three distributions expressed in terms of quantile functions, which are potential models to represent income data. It may be noticed that only very little work has been done in modelling income data using quantile functions and hence the properties of these models are examined in detail. Inequality measures such as Lorenz curve, Bonferroni curve etc. are calculated for these models. Further characterization results associated with these models are also discussed. Even though the results discussed below have no specific reference to the Zenga curve, the results are useful in view of the relationships between the inequality measures
4.6.1 Govindarajulu distribution

In the context of reliability analysis, Govindarajulu (1977) proposed a lifetime model with quantile based representation

\[ Q(p) = \theta + \sigma \left\{ (\beta + 1) p^\beta - \beta p^{\beta+1} \right\}; \quad \theta, \sigma, \beta > 0. \]  

(4.18)

When \( \theta = 0 \), (4.18) reduces to

\[ Q(p) = \sigma \left\{ (\beta + 1) p^\beta - \beta p^{\beta+1} \right\}; \quad \sigma, \beta > 0. \]

(4.19)

We now look into some popular measures of income inequality for the Govindarajulu model. Using (4.18), \( L(p) \) simplifies to

\[ L(p) = \frac{\beta + 2}{\beta + 2} \theta + \sigma \left\{ \theta p + \sigma p^{\beta+1} \left( \frac{\beta + 2 - \beta p}{\beta + 2} \right) \right\}. \]

In the special case when \( \theta = 0 \), we get

\[ L(p) = p^{\beta+1} \left( \frac{\beta + 2 - \beta p}{\beta + 2} \right). \]

(4.20)

For \( \beta = 1 \), the above expression becomes the Lorenz curve of the rescaled beta distribution. The Bonferroni curve is given by

\[ B(p) = p^3 \left( \frac{\beta + 2 - \beta p}{\beta + 2} \right). \]

(4.21)

It may be observed that as \( \beta \to 0 \), \( B(p) \to 1 \). Also as \( \beta \) gets large, \( B(p) \to 0 \). This means that as the value of \( \beta \) decreases, there is a tendency to reach maximum equality. The Bonferroni
index defined by
\[ B_1 = \int_0^1 B(p) dp, \]
and for the model (4.19), \( B_1 \) simplifies to
\[ B_1 = \frac{1}{(\beta + 1)(\beta + 2)}. \]

The expression for \( R(p) \) for the model (4.19) is given by
\[ R(p) = p^\beta \left( \beta + 1 - \beta p - \frac{\beta + 2 - \beta p}{\beta + 2} \right). \quad (4.22) \]

From (4.21) and (4.22), we can see that the ratio of Bonferroni curve to the reversed mean residual quantile function is in the bilinear form. The following theorem provides a characterization result for the Govindarajulu distribution based on a relationship between \( B(p) \) and \( R(p) \).

**Theorem 4.7.** Let \( X \) be a nonnegative random variable with finite positive mean \( \mu \). The relationship
\[ B(p) = \left[ \frac{A - Bp}{C - Dp} \right] R(p); \quad A, B, C, D > 0 \quad (4.23) \]
holds if and only if \( X \) follow the Govindarajulu distribution with quantile function specified by
\[ Q(p) = \sigma \left\{ (\beta + 1) p^\beta - \beta p^{\beta+1} \right\}. \]

**Proof.** By direct calculations using (4.21) and (4.22), we get \( B(p) \) is of the form (4.23) with
\[ A = (\beta + 2)^2, B = C = \beta (\beta + 2), D = \beta (\beta + 1). \]

Conversely, suppose that (4.23) holds.

Using the following relation between \( L(p) \) and \( R(P) \)
\[ R(p) = \mu \left[ L'(p) - \frac{1}{p} L(p) \right], \]
in view of (4.23) and the fact that \( B(p) = p^{-1} L(p) \), we get

\[
\frac{L'(p)}{L(p)} = \frac{C - Dp + \mu(A - Bp)}{\mu p(A - Bp)}.
\]

Integrating the above equation from 0 to \( p \), we get

\[
L(p) = Z p^{\frac{C}{\mu}} + 1 (A - Bp)^{\frac{\mu}{\mu} - \frac{D}{2\mu}}, \tag{4.24}
\]

where \( Z \) is the constant of integration. Since the Lorenz curve determines the distribution uniquely, it is clear from (4.24) that the model is Govindarajulu distribution when \( A = (\beta + 2), B = \beta, C = \beta\mu(\beta + 2) \) and \( D = \beta\mu(\beta + 1) \). \( \Box \)

For the model (4.19), \( G \) simplifies to

\[
G = \frac{\beta^2 + 4\beta + 1}{\beta^2 + 4\beta + 3}.
\]

To describe the size distribution of income, Esteban (1986) introduced the concept of income share elasticity which provides the rate of change of total income at each income level. The income share elasticity is defined as

\[
\pi(x) = 1 + \frac{x f'(x)}{f(x)}.
\]

Setting \( x = Q(p) \) and using (2.3), we get

\[
E(p) = \pi[Q(p)] = 1 - \frac{Q(p)q'(p)}{q(p)}.
\]

For Govindarajulu model,

\[
E(p) = 1 - \frac{\sigma p^\delta(\beta + 1 - \beta p)[\beta(1 - p) - 1]}{1 - p}.
\]
It may be observed from the above expression that the income share elasticity of Govindarajulu distribution is non monotonic in behaviour.

Another measure which has attracted a lot of interest is the poverty index described in Sen (1976). Nair & Vineshkumar (2010) has given the following definition for the above notions in terms of quantile function. The expression for the truncated Gini index and Sen index are respectively given as

$$\eta(p) = 1 - \frac{2p}{p} \int_0^p \frac{Q(u)(p-u)du}{p^2}$$

and

$$S(p) = p[g(p) + (1 - g(p)) \eta(p)]$$

where \( g(p) \) is the income gap ratio given by

$$g(p) = 1 - \frac{1}{pQ(p)} \int_0^p Q(u)du.$$

For the model (4.19), the expression for \( \eta(p) \) and \( S(p) \) are given as

$$\eta(p) = 1 - \frac{\beta + 2 - \beta p}{(\beta + 2)(\beta + 1 - \beta p)}$$

and

$$S(p) = p \left[ 1 - \frac{p^2\beta^2 - p(2\beta^2 + 4\beta + 1) + (\beta^2 + 4\beta + 5)}{(\beta + 2)(\beta + 1 - \beta p)^2} \right].$$

**Remark 4.3** When the division of population is based on the mean income \( \mu \) instead of \( x \), we can represent the Zenga measure as a function of Frigyes’ measure defined by

$$F = (F_1, F_2, F_3) = \left( \frac{\mu}{m_1}, \frac{m_2}{m_1}, \frac{m_2}{\mu} \right)$$
where \( m_1 = E(X|X < \mu) \) and \( m_2 = E(X|X > \mu) \). Using (3.1), the Zenga measure is

\[
A(\mu) = 1 - \frac{1}{F_2}.
\]

The \( I(p) \) curve corresponding to (4.19) simplifies to

\[
I(p) = \frac{2 - p^\beta (\beta + 2 - \beta p)}{2 - p^{\beta+1}(\beta + 2 - \beta p)}.
\]

The following graphs show the behavior of the \( I(p) \) curve for different values of \( \beta \).

The behavior of the curve can be easily ascertained from the sign of \( I'(p) \).

\[
I'(p) \leq 0 \Rightarrow 2p(1 - p)\beta p + (\beta + 2 - \beta p)(2p - 2(1 - p)\beta) \leq p^{\beta+1}(\beta + 2 - \beta p)^2 \quad (4.25)
\]

Since

\[
p^{\beta+1} < \frac{1}{1 + \frac{1 - p}{\beta - 1}},
\]

(4.25) becomes

\[
2p(1 - p)\beta p + (\beta + 2 - \beta p)(2p - 2(1 - p)\beta) \leq \left[1 + \frac{1 - p}{\beta + 1}\right]^{-1}(\beta + 2 - \beta p)^2 \leq 0.
\]
The above condition is true only when $\beta > 0$. Thus we can say that for all $\beta$, $I(p)$ curve is decreasing.

**Remark 4.4**

From theorem 4.7 and the relation between $B(p)$ and $I(p)$, it can be noted that the expression connecting $I(p)$ curve and $R(p)$ is not in simple form.

That is, we have

\[ I(p) = \frac{1 - B(p)}{1 - pB(p)} \]

and for model (4.19)

\[ I(p) = \frac{1 - k(p)R(p)}{1 - pk(p)R(p)} \]

where $k(p) = \frac{A-Bp}{C-Dp}$. But when $k(p) = 1$, the above expression connecting $I(p)$ and $R(p)$ provides a characterization for the power distribution.

**Estimation**

Gilchrist (2000) provides a detailed discussion on various estimation procedures of parameters in the quantile function. However L moments can be used for estimating the parameters in an easier manner. We fit the three parameter GovindaRajulu model given in (4.18) to a real data set using the method of L moments. For the GovindaRajulu model (4.18), the first three L moments are given as

\begin{align*}
    l_1 &= \theta + \frac{2\sigma}{\beta + 2} \\
    l_2 &= \frac{2\beta\sigma}{(\beta + 2)(\beta + 3)} \\
    l_3 &= \frac{2\beta\sigma(\beta - 2)}{\beta^3 + 9\beta^2 + 26\beta + 24}.
\end{align*}

In this method, we equate the sample L-moments to population L-moments to obtain the esti-
mates of $\theta, \beta,$ and $\sigma$. To illustrate the application of the model in practical situation, we consider a data collected from the site of Beuro of Economic Analysis. The data includes 255 values denoting quarterly state personal incomes of Michigan state from the year 1948 up to 2011, third quarter.

The estimates for the parameters are obtained as

$$\hat{\theta} = 14257.3; \hat{\sigma} = 3.77; \hat{\beta} = 336762.$$  

Here the data is divided into 10 groups. Corresponding $x$ values with respect to the values of $u_i = \frac{i}{10}; i = 1, 2, \ldots, 10$ are used to get the observed and expected frequencies. Thus the chisquare value obtained here is 8.4 which is admissible so that the model follows Govindarajulu distribution. The Q-Q plot for the model is given as figure 4.3 below. The graph also reveals the appropriateness of the model.

### 4.6.2 Quantile model with linear hazard quantile form

Development of new models assigning different functional forms for various concepts in reliability theory is a potential area of research. Many well known distributions that exist in the literature have been arisen in this way. Recently, Nair & Vineshkumar (2011) have proposed a new quantile function using linear form of the hazard quantile function. In this section, we in-
vestigate its application in modelling income data. The distribution is specified by the quantile function

\[ Q(p) = \frac{1}{a + b} \log \left( \frac{a + bp}{a(1 - p)} \right) \]  

When \( a = \frac{1}{\lambda} \) and \( b = 0 \), (4.26) becomes,

\[ Q(p) = -\frac{1}{\lambda} \log(1 - p). \]

The above quantile function corresponds to the exponential distribution. When \( a = b = \frac{1}{2\sigma} \), the quantile function takes the form

\[ Q(p) = \sigma \log \left( \frac{1 + p}{1 - p} \right) \]

and the distribution is the half logistic distribution. Also for \( a = \frac{\lambda}{1-u} \) and \( b = -\frac{a\lambda}{1-u} \), the distribution is exponential-geometric with quantile function

\[ Q(p) = \frac{1}{\lambda} \log \left( \frac{1 - pu}{1 - p} \right). \]

Moving onto the inequality measures as defined earlier, Lorenz curve for the class of models given in (4.26) takes the form

\[ L(p) = \frac{(a + bp) \log(a + bp) - \log a(1 + a) + (b - 1) \log(1 - p)}{(a + b) \log \left( \frac{a + b}{a} \right)}. \]

Also we get the income share elasticity as

\[ E(p) = 1 - \frac{a + b(2p - 1)}{(a + b)^2(1 - p)} \log \left( \frac{a + bp}{a(1 - p)} \right). \]

It may be observed that the income share elasticity of the distributions in the class (4.26) is decreasing. Unlike Govindarajulu distribution, the Gini index of this model does have a simple
expression. The Gini index for the model (4.26) simplifies to,

\[ G = \frac{1}{4b} \left[ 2(a + b)^2 \log(a + b) - b \left( 2a + 5b + 4 \log a(1 - a) - 4 \right) - 2a^2 \log a \right]. \]

The Zenga curve is obtained as

\[ I(p) = \frac{p(a + b) \log \left( \frac{a + b}{a} \right) - (a + bp) \log(a + bp) + \log a(1 + a) - (b - 1) \log(1 - p)}{p [1 - (a + bp) \log(a + bp) + \log a(1 + a) - (b - 1) \log(1 - p)]}. \]

The \( I(p) \) curve of the above class is always decreasing for any values of the parameters. Figure 4.9 plots the \( I(p) \) curve for different values of the parameters.

![Figure 4.9: \( I(p) \) curve at different values of parameters](image)

**Estimation**

The first two L-moments of this model is obtained as

\[ l_1 = \frac{1}{a + b} \log \left( \frac{a + b}{a} \right) \]

and

\[ l_2 = \frac{a \log \left( \frac{a + b}{a} \right) + b}{b^2}. \]

To illustrate the procedure, we consider the 42 revised annual personal income estimates of United States from the year 1969 to 2010. Revised estimates for 2007-2010 were released June 22, 2011. (data source: U.S. Department of Commerce, Bureau of Economic Analysis,
The parameter estimates of \( a \) and \( b \) are obtained as

\[ a = 0.000092, b = 0.000236. \]

A reasonable model for the distribution of the personal income shall be taken as (4.26) with values of \( a \) and \( b \) be given above. The Q-Q plot for the model is given as figure 4.10 below.

![Q-Q Plot](image)

**Figure 4.10: Q-Q Plot**

### 4.6.3 Power x Pareto distribution

Unlike distribution function, quantile functions have an interesting property that they can be added or multiplied to generate new ones. Gilchrist (2000) considered the multiplied form of Power and Pareto distributions under the name Power x Pareto distribution. The distribution is specified by the quantile function \( Q(p; c, \lambda_1, \lambda_2) \) and specified by

\[
Q(p; c, \lambda_1, \lambda_2) = cp^{\lambda_1}(1 - p)^{-\lambda_2}; c, \lambda_1, \lambda_2 \geq 0. \quad (4.27)
\]
It may be noticed that (4.27) includes Power distribution as $\lambda_2 \to 0$ with distribution function

$$F(x) = \left(\frac{x}{c}\right)^{\lambda_1}; 0 < x < c$$

and the Pareto model specified by

$$F(x) = 1 - \left(\frac{c}{x}\right)^{\frac{1}{\lambda_1}}; x > c > 0$$

when $\lambda_2 \to 0$. Further Log-logistic distribution becomes a special case when $\lambda_1 = \lambda_2 = \lambda$ with the distribution function

$$F(x) = \frac{1}{1 + (c/x)^{\lambda}}; x > 0$$

Since uniform distribution is a special case of power distribution, Uniform x Pareto distribution can be considered as a special case of the quantile model (4.27). The uniform x Pareto distribution is given as

$$Q(p; c, 1, \lambda_2) = cp(1 - p)^{-\lambda_2}. \quad (4.28)$$

Hankin & Lee (2006) has studied the properties of the model (4.27). We look into the economic measures of the model (4.27). The income share elasticity can be obtained as

$$E(p) = \frac{c(1 - p)^{\lambda_2 - 1}}{\lambda_1 + p(\lambda_2 - \lambda_1)} \left[ (\lambda_1 - \lambda_2 - 1) \left( p^2(\lambda_1 - \lambda_2) - 2\lambda_1p \right) + \lambda_1(1 - \lambda_1) \right].$$

The Lorenz curve takes the form

$$L(p) = \frac{\beta_p(\lambda_1 + 1, 1 - \lambda_2)}{\beta(\lambda_1 + 1, 1 - \lambda_2)}; \quad \lambda_2 < 1$$

where

$$\beta_p(\lambda_1 + 1, 1 - \lambda_2) = \int_0^p u^{\lambda_1}(1 - u)^{-\lambda_2}du.$$
Also the Zenga curve of Power x Pareto distribution can be obtained from the relation between
$I(p)$ and $L(p)$ namely
\[
I(p) = \frac{p - L(p)}{p (1 - L(p))}.
\]
The $I(p)$ curve for different values of parameters are given in figure 4.6 below. It may be noted
that the $I(p)$ curve and $E(p)$ take different shapes for different values of $c$, $\lambda_1$ and $\lambda_2$.

![Figure 4.11: $I(p)$ curve at different values of parameters](image)

Next theorem provides a characterization result for the Pareto distribution which is a member
of the family defined by (4.27).

**Theorem 4.8.** A non negative continuous random variable $X$ follows the distribution with
quantile function $Q(p; c, 0, \lambda_2)$ if and only if the relationship holds
\[
1 - E(p) = \frac{A}{1 - p} Q(p); \quad A > 1
\] (4.29) holds for all $p \in (0, 1)$.

**Proof.** Suppose that (4.29) holds. From the definition of income share elasticity, we get
\[
\frac{q'(p)}{q(p)} = \frac{A}{1 - p}.
\]
The solution of the above differential equation is

\[ Q(p) = \frac{\delta}{A - 1} (1 - p)^{-(A-1)} \]

where \( \delta \) is the constant of integration. Setting left end point to \( c \), we get

\[ Q(p) = c(1 - p)^{-(A-1)}, \quad A > 1 \]

as claimed. Proof of the converse is straight forward and hence not pursued here.

\[ \square \]

**Estimation**

Power x Pareto distribution is fitted to a personal income data set by using the method of L-moments. The data includes 179 observations describing the dollar estimates of 179 areas (includes all local areas and states) of United States of the year 2009.

For Power x Pareto model, the first three L-moments are given by

\[ l_1 = C \beta_p(\lambda_1 + 1, 1 - \lambda_2) \]

\[ l_2 = C \beta_p(\lambda_1 + 2, 2 - \lambda_2) \]

and

\[ l_3 = C [3\beta_p(\lambda_1 + 3, 1 - \lambda_2) - \beta_p(\lambda_1 + 2, 1 - \lambda_2) - 2\beta_p(\lambda_1 + 4, 1 - \lambda_2)] . \]

The parameter estimates are obtained as

\[ c = 33.79; \lambda_1 = .8138; \lambda_2 = .6232. \]
The Q-Q plot is given as figure 4.12 below and the graph ensures that the model is a good fit.

Figure 4.12: Q-Q Plot