CHAPTER FIVE

MEASURES OF UNCERTAINTY FOR DOUBLY TRUNCATED RANDOM VARIABLES

5.1 Introduction

In the previous chapter, we have discussed some measures of uncertainty for the right truncated random variables and characterized certain models arising out of them. But in reliability/survival analysis there may be situations in which the data is doubly truncated. As pointed out in chapter 1, a doubly truncated failure time arises if an individual is potentially observed only if its failure time falls within a certain interval, unique to that individual. In such type of truncation, the individual whose event time lies within a certain time interval are only observed. Thus an individual whose event time is not in this interval is not observed and therefore the information on this subject outside the interval is not available to the investigator (see Ruiz and Navarro (1996), Efron and Petrosian (1999), Betensky and Martin (2003), Navarro and Ruiz (1996, 2004), Sankaran and Sunoj (2004) and Bairamov and Gebizlioglu (2005)). Such types of truncation happen in lifetime studies also. Therefore the properties related to these type of datasets are important both in reliability and survival analysis. In addition, the measures of uncertainty and reliability are closely related. All measures of uncertainty have much relevance in characterizing and classifying life distributions according to the behavior of them.

* Some of the results in this Chapter have been communicated to an International Journal.
Accordingly in the present chapter, we focus on characterizing some probability models based on different measures of uncertainty and conditional expectations of doubly truncated random variables. Because of the wide applicability of conditional expectations for interval truncated data in survival studies and reliability life testing, in the present chapter, we study the different uncertainty measures considered in the previous chapter to the doubly truncated case and examine its relationships. We also extend these studies in weighted models. Many of the results that we have obtained in the present chapter are generalizations of some of the existing results.

5.2 Definitions and properties

5.2.1 Geometric vitality function

Kupka and Loo (1989) studied the vitality function extensively in connection with their studies on ageing process. It provides a useful tool in modeling lifetime data. Recently, Nair and Rajesh (2000) defined a conditional geometric vitality function and it has been found a useful tool in modeling and analysis of lifetime data. For a non-negative rv \( X \) representing the lifetime of a component with an absolutely continuous df \( F(t) \) and \( E(\log X) < \infty \), then the geometric vitality function of a left truncated rv is

\[ \log G(t) = E\bigl(\log X | X > t\bigr). \quad (5.1) \]

In reliability theory, (5.1) gives the geometric mean of the lifetimes of components, which has survived \( t \) units of time. Nair and Rajesh studied this measure in detail and characterize some probability distributions based on it. Using (5.1), a straightforward generalization of geometric vitality function for a doubly truncated rv \( \{ X | t_1 \leq X \leq t_2 \} \), where \( (t_1, t_2) \in D = \{(u, v) \in \mathbb{R}^2; F(u) < F(v)\} \) is given by

\[ \log G(t_1, t_2) = E\bigl(\log X | t_1 < X < t_2\bigr) \]

\[ = \frac{1}{(F(t_2) - F(t_1))} \int_{t_1}^{t_2} (\log x) f(x) dx \quad (5.2) \]
log \( G(t_1, t_2) \) gives the geometric mean life of a rv truncated at two points \( t_1 \) and \( t_2 \). It is clear that when \( t_2 \to \infty \) (5.2) reduces to (5.1). The following properties are immediate from the definition (5.2),

\[
\begin{align*}
(1) & \lim_{t_1 \to 0} \lim_{t_2 \to \infty} \log G(t_1, t_2) = E(\log X), \text{ and } \\
(2) & m(t_1, t_2) \geq \log G(t_1, t_2) \text{ for all } t_1 < t_2,
\end{align*}
\]

where \( m(t_1, t_2) = E(X | t_1 < X < t_2) \).

Denoting \( h_i(t_1, t_2) = \frac{f(t_1)}{(R(t_1) - R(t_2))} \) and \( h_2(t_1, t_2) = \frac{f(t_2)}{(R(t_1) - R(t_2))} \) as the GFR functions of Navarro and Ruiz (1996), (2.1) is related to \( h_i(t_1, t_2) \); \( i = 1, 2 \) as

\[
h_i(t_1, t_2) = \frac{\frac{\partial}{\partial t_1} \left( \log G(t_1, t_2) \right)}{\log \left( \frac{G(t_1, t_2)}{t_1} \right)}
\]

and

\[
h_2(t_1, t_2) = \frac{\frac{\partial}{\partial t_2} \left( \log G(t_1, t_2) \right)}{\log \left( \frac{t_2}{G(t_1, t_2)} \right)}.
\]

for all \( (t_1, t_2) \in D \).

Table 5.1 provides the relationships between geometric vitality function \( \log G(t_1, t_2) \) and GFR functions \( h_i = h_i(t_1, t_2) \); \( i = 1, 2 \) of certain probability distributions.
### Table 5.1: Relationships between geometric vitality function and GFR functions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$R(t)$</th>
<th>$\log G(t_1, t_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>$\exp(-\lambda t); t &gt; 0, \lambda &gt; 0$</td>
<td>$\frac{1}{\lambda} [h_1 \log t_1 - h_2 \log t_2 + A(t_1, t_2)]$</td>
</tr>
<tr>
<td>Beta</td>
<td>$(1 - Rt)^d; 0 &lt; t &lt; \frac{1}{R}, d &gt; 0, R &gt; 0$</td>
<td>$\frac{1}{Rd} [(1 - Rt_1) h_1 \log t_1 - (1 - Rt_2) h_2 \log t_2 + A(t_1, t_2) - R]$</td>
</tr>
<tr>
<td>Pareto II</td>
<td>$(1 + pt)^{-q}; t &gt; 0, p &gt; 0, q &gt; 0$</td>
<td>$\frac{1}{pq} [(1 + pt_1) h_1 \log t_1 - (1 + pt_2) h_2 \log t_2 + A(t_1, t_2) + p]$</td>
</tr>
<tr>
<td>Power</td>
<td>$1 - (t/b)^c; 0 \leq t \leq b, b &gt; 0, c &gt; 0$</td>
<td>$\frac{1}{c} [t_2 h_1 \log t_2 - t h_1 \log t_1 - 1]$</td>
</tr>
<tr>
<td>Pareto I</td>
<td>$(k/t)^e; t &gt; k, k &gt; 0, c &gt; 0$</td>
<td>$\frac{1}{c} [t_1 h_1 \log t_1 - t h_2 \log t_2 - 1]$</td>
</tr>
</tbody>
</table>

Here $h_i = h_i(t_1, t_2) = \frac{f(t)}{R(t_i) - R(t_j)}; i = 1, 2$ and $A(t_1, t_2) = E \left( \frac{1}{X} | t_1 < X < t_2 \right)$.

**Theorem 5.1:** If $h_i(t_1, t_2); i = 1, 2$ satisfy the properties given in Navarro and Ruiz (1996), then the geometric vitality function (5.2) determine distribution uniquely.

**Proof:** The proof follows from (5.5), (5.6) and Theorem 4.1 of Navarro and Ruiz (1996) (see Chapter 1).

### 5.2.2 Measure of uncertainty

Combining the residual entropy (1.35) defined by Ebrahimi and Pellerey (1995) and the past entropy (4.1) defined by Di Crescenzo and Longobardi (2002), we introduce a new measure of uncertainty which generalize (1.35) and (4.1) to the doubly truncated random variables. Defining a rv $\left( X | t_1 \leq X \leq t_2 \right)$ which represents the life of a unit between $t_1$ and $t_2$, a measure of uncertainty for the doubly truncated rv is given by
\[ H(t_1,t_2) = - \int_{t_1}^{t_2} \frac{f(x)}{(R(t_1) - R(t_2))} \log \left( \frac{f(x)}{(R(t_1) - R(t_2))} \right) dx. \] (5.7)

Clearly, \( \lim_{t_1 \to 0} H(t_1,t_2) = H(t_2) \) and \( \lim_{t_2 \to \infty} H(t_1,t_2) = H(t_1) \).

Equation (5.7) can be written as

\[
H(t_1,t_2) = 1 - \frac{1}{(R(t_1) - R(t_2))} \int_{t_1}^{t_2} f(x) \left( \log h(x) \right) dx
\]

\[
+ \frac{1}{(R(t_1) - R(t_2))} \left[ R(t_2) \log R(t_2) - R(t_1) \log R(t_1) \right] + \log \left( R(t_1) - R(t_2) \right). \] (5.8)

\[
H(t_1,t_2) = 1 - \frac{1}{(F(t_2) - F(t_1))} \int_{t_1}^{t_2} f(x) \left( \log \lambda(x) \right) dx
\]

\[
- \frac{1}{(F(t_2) - F(t_1))} \left[ F(t_2) \log F(t_2) - F(t_1) \log F(t_1) \right] + \log \left( F(t_2) - F(t_1) \right), \] (5.9)

where (5.8) and (5.9) are the expressions of \( H(t_1,t_2) \) in terms of hazard rate \( h(t) \) and reversed hazard rate \( \lambda(t) \), respectively.

Now using (4.1), (1.35) and (5.7), the Shannon entropy (1.34) can be decomposed as

\[
H = F(t_1)H(t_1) + \left( R(t_1) - R(t_2) \right) H(t_1,t_2) + R(t_2)H(t_2) - \left[ F(t_1) \log F(t_1) \right.
\]

\[
+ \left( R(t_1) - R(t_2) \right) \log \left( R(t_1) - R(t_2) \right) + R(t_2) \log R(t_2) \]. \] (5.10)

The identity (5.10) can be interpreted in the following way. The uncertainty about the failure of an item can be decomposed into 4 parts: (i) the uncertainty about the failure time in \( (0,t_1) \) given that the item has failed before \( t_1 \), (ii) the uncertainty about the failure time in the interval \( (t_1,t_2) \) given that the item has failed after \( t_1 \) but before \( t_2 \), (iii) the uncertainty about the failure time in \( (t_2,\infty) \) given that it has failed after \( t_2 \) and (iv) the uncertainty of the item that has failed before \( t_1 \) or in between \( t_1 \) and \( t_2 \) or after \( t_2 \).

On partially differentiating \( H(t_1,t_2) \) with respect to \( t_1 \) and \( t_2 \), we get
\[ \frac{\partial}{\partial t_1} H(t_1, t_2) = h_i(t_1, t_2) \left[ \log h_i(t_1, t_2) + H(t_1, t_2) - 1 \right] \]  
(5.11)

and

\[ \frac{\partial}{\partial t_2} H(t_1, t_2) = h_i(t_1, t_2) \left[ 1 - \log h_i(t_1, t_2) + H(t_1, t_2) \right]. \]  
(5.12)

When \( H(t_1, t_2) \) is increasing in \( t_1 \) and in \( t_2 \), then, (5.11) and (5.12) together implies

\[ 1 - \log h_i(t_1, t_2) \leq H(t_1, t_2) \leq 1 - \log h_i(t_1, t_2). \]  
(5.13)

Thus when the uncertainty measure is increasing, then it lies between \((1 - \log h_i(t_1, t_2))\) and \((1 - \log h_i(t_1, t_2))\). We can also write the bounds (5.13) as

\[ h_i(t_1, t_2) \leq \exp \left( 1 - H(t_1, t_2) \right) \leq h_i(t_1, t_2). \]

Table 5.2 provides the relationships between the measure of uncertainty \( H(t_1, t_2) \), the conditional expectation \( m(t_1, t_2) = E(X | t_1 < X < t_2) \) and GFR functions \( h_i = h_i(t_1, t_2) \); \( i = 1, 2 \) for various distributions.

**Table 5.2: Relationships between \( H(t_1, t_2) \), the conditional expectation and GFR functions for various distributions**

<table>
<thead>
<tr>
<th>Distribution</th>
<th>( R(t) )</th>
<th>( H(t_1, t_2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>( \exp(-\lambda t) ; t &gt; 0, \lambda &gt; 0 )</td>
<td>( \lambda m(t_1, t_2) - \lambda t_2 - \log h_i(t_1, t_2) ) or ( \lambda m(t_1, t_2) - \lambda t_1 - \log h_i(t_1, t_2) )</td>
</tr>
<tr>
<td>Beta</td>
<td>( (1 - Rt)^d ; 0 &lt; t &lt; \frac{1}{R}, d &gt; 0, R &gt; 0 )</td>
<td>( -(d-1)E \left[ \log(1-RX)</td>
</tr>
<tr>
<td>Pareto II</td>
<td>( (1 + pt)^{-q} ; t &gt; 0, p &gt; 0, q &gt; 0 )</td>
<td>( (q+1)E \left[ \log(1 + pX)</td>
</tr>
<tr>
<td>Power</td>
<td>( 1 - \left( \frac{t}{b} \right)^c ; 0 \leq t \leq b, b &gt; 0, c &gt; 0 )</td>
<td>( 1 + \log G(t_1, t_2) + t_i h_i \log(t_i / b) - t_2 h_i \log(t_2 / b) - \log \left( \frac{c}{(t_2 / b)^{t_i} - (t_1 / b)^{t_i}} \right) )</td>
</tr>
<tr>
<td>Pareto I</td>
<td>( (k/t)^c ; t &gt; k, k &gt; 0, c &gt; 0 )</td>
<td>( 1 + \log G(t_1, t_2) + t_i h_i \log(k / t_i) - t_k h_i \log(k / t_k) - \log \left( \frac{c}{(k / t_i)^{t_k} - (k / t_k)^{t_k}} \right) )</td>
</tr>
</tbody>
</table>
5.2.3 Conditional measure of uncertainty

As an extension of the definition (4.25) given in the previous chapter, we define the conditional measure of uncertainty for the doubly truncated rv as

\[ M(t_1, t_2) = -E \left[ \log f(X) \big| t_1 < X < t_2 \right] \]

\[ = \frac{-1}{(R(t_1) - R(t_2))} \int_{t_1}^{t_2} f(x) \left( \log f(x) \right) dx \]  \hspace{1cm} (5.14)

where \((t_1, t_2) \in D\). Using (5.14), \(M(t_1, t_2)\) can be easily related to \(H(t_1, t_2)\) through the relation

\[ M(t_1, t_2) = H(t_1, t_2) - \log \left( R(t_1) - R(t_2) \right). \]  \hspace{1cm} (5.15)

Differentiating (5.15) with respect to \(t_1\) and \(t_2\) respectively provide the relationships with GFR functions, which are given by

\[ \frac{\partial M(t_1, t_2)}{\partial t_1} = \frac{\partial H(t_1, t_2)}{\partial t_1} + h_1(t_1, t_2) \]

and

\[ \frac{\partial M(t_1, t_2)}{\partial t_2} = \frac{\partial H(t_1, t_2)}{\partial t_2} - h_2(t_1, t_2). \]

The various relationships between the conditional measure of uncertainty \(M(t_1, t_2)\) for doubly truncated random variables and GFR functions \(h_i = h_i(t_1, t_2); \ i = 1, 2\) for some probability models are given in Table 5.3.

**Table 5.3:** Relationships between \(M(t_1, t_2)\) and GFR functions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>(R(t))</th>
<th>(M(t_1, t_2))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>(\exp(-\lambda t); t &gt; 0, \lambda &gt; 0)</td>
<td>(\lambda m(t_1, t_2) - \log \lambda)</td>
</tr>
<tr>
<td>Beta</td>
<td>((1 - Rt)^d; 0 &lt; t &lt; \frac{1}{R}, d &gt; 0, R &gt; 0)</td>
<td>(-\log Rd - (d + 1)E \left[ \log(1 - RX) \big</td>
</tr>
<tr>
<td>Distribution</td>
<td>Density Function</td>
<td>Condition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Pareto II</td>
<td>$(1 + pt)^{-q}; \ t &gt; 0, p &gt; 0, q &gt; 0$</td>
<td>$t &gt; 0$</td>
</tr>
<tr>
<td>Power</td>
<td>$1 - \left(\frac{t}{b}\right)^c; 0 \leq t \leq b, b &gt; 0, c &gt; 0$</td>
<td>$0 \leq t \leq b$</td>
</tr>
<tr>
<td>Pareto I</td>
<td>$\left(\frac{k}{t}\right)^c; t &gt; k, k &gt; 0, c &gt; 0$</td>
<td>$t &gt; k$</td>
</tr>
<tr>
<td>Weibull</td>
<td>$\exp\left(-t^c\right); t &gt; 0, p &gt; 0$</td>
<td>$t &gt; 0$</td>
</tr>
<tr>
<td>Rayleigh</td>
<td>$\exp\left(-t^2\right); t &gt; 0$</td>
<td>$t &gt; 0$</td>
</tr>
</tbody>
</table>

### 5.3 Characterizations

In this section we prove certain characterization theorems for some life distributions and certain family of distributions using GFR functions, geometric vitality function (5.2) and conditional Shannon’s measure of uncertainties (5.7) and (5.14).

The following theorem gives a characterization to a family of distributions such as exponential, Pareto II and beta using a possible relation connecting the geometric vitality function and the GFR functions.

**Theorem 5.2:** Let $X$ be a rv with support $(0, \infty)$ admitting an absolutely continuous distribution function $F(t)$ with respect to Lebesgue measure. Then a relation of the form

$$\log G(t_1, t_2) = \frac{1}{k}\left[(1 + Ct_1)h_1(t_1, t_2)\log t_1 - (1 + Ct_2)h_2(t_1, t_2)\log t_2 + A(t_1, t_2) + C\right]$$  

(5.16)

where $k > 0$ and $C$ are constants holds for all $(t_1, t_2) \in D$ if and only if $X$ follows exponential distribution for $C = 0$, Pareto II distribution for $C > 0$ and Beta distribution for $C < 0$ with distribution functions (2.18), (2.17) and (2.19) respectively.
Proof: Assume that the relation (5.16) holds. From the definitions of $A(t_1,t_2), h_i(t_1,t_2)$; $i = 1,2$ and $\log G(t_1,t_2)$, (5.16) becomes

$$
\int_{t_1}^{t_2} f(x) \log x \, dx = \frac{1}{k} \left[ (1 + Ct_1) f(t_1) \log t_1 - (1 + Ct_2) f(t_2) \log t_2 \right. \\
\left. + \int_{t_1}^{t_2} \frac{1}{x} f(x) \, dx + C \left( R(t_1) - R(t_2) \right) \right]. 
$$

(5.17)

Differentiating (5.17) with respect to $t_i; i = 1,2$ and on simplification, we obtain

$$
\frac{f'(t_i)}{f(t_i)} = \frac{(k + C)}{(1 + Ct_i)}, \text{ for } (t_1,t_2) \in D
$$

or

$$
\frac{d}{dt} \log f(t) = -\frac{(k + C)}{(1 + Ct)}. 
$$

(5.18)

From (5.18.), it follows that $X$ follows exponential, Pareto II and Beta distributions according as $C = 0, C > 0$ and $C < 0$.

The converse part is obtained from Table 5.1.

Theorem 5.3 gives a characterization to the exponential distribution using the functional relation connecting the conditional measure of uncertainty and the conditional moment function $m(t_1,t_2)$.

**Theorem 5.3:** For a non-negative rv $X$, a relation of the form

$$
M(t_1,t_2) - \mu^{\mu_m(t_1,t_2)} = k, 
$$

(5.19)

where $k > 0$ is a constant, holds for all $(t_1,t_2) \in D$ if and only if $X$ follows exponential distribution with distribution function (2.18).

Proof: Assume (5.19) holds. From (5.14), we can write

$$
- \int_{t_1}^{t_2} f(x) \left( \log f(x) \right) \, dx - \mu^{-1} \int_{t_1}^{t_2} x f(x) \, dx = k \left( R(t_1) - R(t_2) \right). 
$$

(5.20)
Differentiating (5.20) with respect to $t_i; i = 1, 2$ gives

$$\log f(t_i) = -k - \mu^{-1}t_i$$

(5.21)

or $f(t) = K \exp(-\mu^{-1}t)$, which provides the result.

The converse part is obtained from Table 5.3.

The following two theorems identify Pareto I and Power distributions using the functional relation connecting conditional measure of uncertainty and geometric vitality function.

**Theorem 5.4:** For a non-negative random variable $X$ in the support $(k, \infty)$, $k > 0$, admitting an absolutely continuous df, then a relation of the form

$$M(t_1, t_2) - (c + 1) \log G(t_1, t_2) = K,$$

(5.22)

where $K > 0$ and $c$ are constants, $c > 0$, holds for $k < t_1 < t_2$ with $F(t_1) < F(t_2)$ if and only if $X$ follows a Pareto I with df (4.20).

**Proof:** Assuming (5.22), then by using (5.14), and (5.2), (5.22) we have

$$\frac{-1}{(R(t_1) - R(t_2))} \int f(x) \left( \log f(x) \right) dx - \frac{(c + 1)}{(R(t_1) - R(t_2))} \int f(x) \log x dx = K$$

(5.23)

or

$$\int f(x) \left( \log f(x) \right) dx - (c + 1) \int f(x) \log x dx = K \left( R(t_1) - R(t_2) \right).$$

(5.24)

Differentiating (5.24) with respect to $t_i; i = 1, 2$ and simplifying we get $f(t) = K t^{-(c+1)}$, which corresponds to Pareto type I with $K = ck^c$. The converse part can be easily verified by direct calculation and it is obtained from Tables 5.1 and 5.3

**Theorem 5.5:** For a non-negative random variable $X$ having an absolutely continuous df $F(t)$ then a relation

$$M(t_1, t_2) + (c - 1) \log G(t_1, t_2) = k, \text{ a constant}$$

(5.25)
is satisfied for \( 0 < t_1 < t_2 < b \) with \( F(t_1) < F(t_2) \) and \( c > 1 \) if and only if \( X \) follows Power distribution with distribution function (4.5).

**Proof:** The proof is similar to that of Theorem 5.4.

The following theorem characterizes log exponential family using the functional relation connecting the conditional measure of uncertainty and geometric vitality function.

**Theorem 5.6:** The distribution of \( X \) belongs to one-parameter log exponential family (1.29) if and only if

\[
M(t_1, t_2) = \log A(\theta) - \theta \log G(t_1, t_2) - \log m_c(t_1, t_2),
\]

where \( \log m_c(t_1, t_2) = E[\log C(X)|t_1 < X < t_2] \) with \( E(\log C(X)) < \infty \), for all \( (t_1, t_2) \in D \).

**Proof:** Assume (5.26) holds. From the definition (5.14), we get

\[
-1 \frac{1}{(R(t_1) - R(t_2))} \int_{t_1}^{t_2} f(x)(\log f(x)) dx = \log A(\theta) - \theta \int_{t_1}^{t_2} f(x) \log x dx
\]

\[
+ \frac{1}{(R(t_1) - R(t_2))} \int_{t_1}^{t_2} f(x)(\log C(x)) dx.
\]

Multiplying both sides of (5.27) by \((R(t_1) - R(t_2))\), we obtain

\[
- \int_{t_1}^{t_2} f(x)(\log f(x)) dx = \log A(\theta)(R(t_1) - R(t_2)) - \theta \int_{t_1}^{t_2} f(x) \log x dx + \int_{t_1}^{t_2} f(x)(\log C(x)) dx.
\]

Differentiating (5.28) with respect to \( t_i; i = 1, 2 \) and simplifying, we get (1.29).

Conversely assume (1.29), by direct calculation and using the definition of \( \log m_c(t_1, t_2) \), we obtain (5.26).
The next theorem characterizes exponential family using the possible relation connecting \( M(t_1, t_2) \) and \( m(t_1, t_2) \).

**Theorem 5.7:** The df of a non-negative rv \( X \) belongs to one parameter exponential family (1.28) if and only if the relation

\[
M(t_1, t_2) + \theta m(t_1, t_2) + m_c(t_1, t_2) + D(\theta) = 0
\]  
(5.29)

where \( m_c(t_1, t_2) = E[C(X) | t_1 < X < t_2] \) with \( E(C(X)) < \infty \) is satisfied for all \((t_1, t_2) \in D\).

**Proof:** The proof is similar to that of the Theorem 5.6.

The next result characterizes generalized Pearson family of distributions using the relation connecting the \( r \)th order conditional moment functions and the GFR functions. For a doubly truncated rv \( (X|t_1 < X < t_2) \), the conditional moment function of order \( r \) is given by

\[
m'(t_1, t_2) = E \left( X^{r} | t_1 < X < t_2 \right) = \frac{1}{(R(t_1) - R(t_2))} \int_{t_1}^{t_2} x^r f(x) dx, \quad (t_1, t_2) \in D. \]  
(5.30)

**Theorem 5.8:** The df of a rv \( X \) belongs to generalized Pearson family of distributions (1.32) if and only if its \( r \)th order conditional moments satisfies a recurrence relation of the form

\[
m'(t_1, t_2) = (b_{or} + b_{o}t_2 + b_{2}t_2^2)h_{r-2}(t_1, t_2) - (b_{or} + b_{o}t_1 + b_{2}t_2^2)h_{r-2}(t_1, t_2)
\]

\[-a_{or}m^{r-2}(t_1, t_2) - a_{r}m^{r-1}(t_1, t_2) - b_{or}(r-2)m^{r-3}(t_1, t_2)
\]

where \( b_{or} = \frac{b_i}{a_2}; \quad i = 0, 1, 2, \quad a_{or} = \frac{(a_o + (r-1)b_1)}{a_2}, \quad a_{r} = \frac{(a_o + rb_2)}{a_2} \) provided \( a_2 \neq 0 \)

and

\[
m'(t_1, t_2) = (b_{or} + b_{o}t_2 + b_{2}t_2^2)h_{r-1}(t_1, t_2) - (b_{or} + b_{o}t_1 + b_{2}t_2^2)h_{r-1}(t_1, t_2)
\]

\[-a_{or}m^{r-1}(t_1, t_2) - b_{or}(r-1)m^{r-2}(t_1, t_2)
\]  
(5.31)
where \( b_i = \frac{b_i}{((r+1)b_2 + a_i)} \); \( i = 0,1,2 \), \( a_0 = \frac{(b_0 + rb_i)}{((r+1)b_2 + a_i)} \), provided \( ((r+1)b_2 + a_i) \neq 0 \) and \( a_2 = 0 \).

**Proof:** Case I: When \( a_2 \neq 0 \), using (5.30), (1.24) and (1.25), (5.31) becomes

\[
\int_{t_1}^{t_2} x f(x)dx = (b_{0r} + b_1r, t_2 + b_2r, t_2^2) i_1^{-2} f(t_1) - (b_{0r} + b_1r, t_1 + b_2r, t_1^2) i_1^{-2} f(t_1) \\
- a_{0r} \int_{t_1}^{t_2} x^{-2} f(x)dx - a_{0r} \int_{t_1}^{t_2} x^{-1} f(x)dx - b_{0r} (r-1) \int_{t_1}^{t_2} x^{-3} f(x)dx = 0.
\]

Differentiating (5.32) with respect to \( t_i \); \( i = 1,2 \) and simplifying, we get

\[
\frac{f(t)}{f(t)} = \frac{\left( A_0 + A_1 i + A_2 i^2 \right)}{\left( B_0 + B_1 i + B_2 i^2 \right)}.
\]

From (5.33) it follows that the distribution of \( X \) belongs to generalized Pearson family with \( A_0 = (a_{0r} - (r-1)b_{0r}) \), \( A_1 = (a_{0r} - rb_{0r}) \), \( A_2 = 1 \) and \( B_i = b_{0r} \); \( i = 0,1,2 \).

Similarly we can prove the case \( a_2 = 0 \) as that of the case \( a_2 \neq 0 \).

Conversely assume (1.32). Multiplying both sides of it by \( x^{-2} \) and on integrating over the limits \( t_1 \) to \( t_2 \) we get

\[
(b_0 + b_1r, t_2 + b_2r, t_2^2) i_1^{-2} f(t_1) - (b_0 + b_1r, t_1 + b_2r, t_1^2) i_1^{-2} f(t_1) - \left( (r-1)b_2 + a_0 \right) \int_{t_1}^{t_2} x^{-2} f(x)dx \\
- (rb_2 + a_1) \int_{t_1}^{t_2} x^{-1} f(x)dx - b_0 (r-2) \int_{t_1}^{t_2} x^{-3} f(x)dx - a_2 \int_{t_1}^{t_2} x x f(x)dx = 0.
\]

Multiplying both sides of (5.34) by \( (R(t_1) - R(t_2))^{-1} \) and using (5.30), we obtain the required form. Substitute \( a_2 = 0 \) in (5.34) and following the similar steps we get (5.31) for \( a_2 = 0 \).
Remarks 5.1: 1) When $t_1 = 0$, Theorem 5.8 reduces to the Theorem 2.8 given in Chapter 2.

2) When $a_2 = 0$, this theorem reduces to that for the Pearson family of distributions (1.30).

5.4 Weighted models

Now we study the application of these uncertainty measures in the context of weighted distributions. We examine the functional relationships of the GFR functions and the uncertainty measures in the context of weighted distributions and prove some useful characterizations arising out of it. For the weighted rv $X_w$, the functional relationship connecting the GFR functions are

$$F^w(t_2) - F^w(t_1) = P(t_1 < X_w < t_2) = \frac{E[w(X)|t_1 < X < t_2]}{\mu_w}(F(t_2) - F(t_1)) \quad (5.35)$$

and

$$h^{\prime\prime}_i(t_1, t_2) = \frac{w(t_i)h_i(t_1, t_2)}{E[w(X)|t_1 < X < t_2]} ; \quad i = 1, 2 \quad (5.36)$$

where $h^{\prime\prime}_i(t_1, t_2) = \frac{f^{\prime\prime}(t_i)}{F^w(t_2) - F^w(t_1)} ; \quad i = 1, 2$ and $(t_1, t_2) \in D$.

Remark 5.2: When $w(t) = t$, (5.35) and (5.36) reduces to the forms given in Sankaran and Sunoj (2004).

Next few theorems prove the relationship connecting the ratio of the distribution functions of weighted and original models and the GFR functions.

Theorem 5.9: If \[ \alpha(t_1, t_2) = \left( \frac{F^w(t_2) - F^w(t_1)}{F(t_2) - F(t_1)} \right) \]

for $(t_1, t_2) \in D$, then $\alpha(t_1, t_2)$ determines $F(t)$. 

94
Proof: Differentiating $\alpha(t_1, t_2)$ with respect to $t_1$ and $t_2$ respectively, we obtain

$$\frac{\partial}{\partial t_1} \alpha(t_1, t_2) = h_1(t_1, t_2) \left( \alpha(t_1, t_2) - \frac{w(t_1)}{\mu_w} \right)$$

(5.37)

and

$$\frac{\partial}{\partial t_2} \alpha(t_1, t_2) = h_2(t_1, t_2) \left( \frac{w(t_2)}{\mu_w} - \alpha(t_1, t_2) \right).$$

(5.38)

(see Navarro et al. (2001) and Sunoj and Maya (2006). From (5.37) and (5.38), we get $h_i(t_1, t_2); i = 1, 2$ as

$$h_i(t_1, t_2) = \frac{\frac{\partial}{\partial t_i} \alpha(t_1, t_2)}{\left( \alpha(t_1, t_2) - \frac{w(t_i)}{\mu_w} \right)}.$$ 

(5.39)

$$h_2(t_1, t_2) = \frac{\frac{\partial}{\partial t_2} \alpha(t_1, t_2)}{\left( \frac{w(t_2)}{\mu_w} - \alpha(t_1, t_2) \right)}.$$ 

(5.40)

Now (5.39), (5.40) and the Theorem 4.1 in Navarro and Ruiz (1996) (see Chapter 1), implies the required result.

Theorem 5.10: Under length-biased sampling, for a non-negative rv $X$ with pdf $f(t)$ and df $F(t)$, the ratio

$$\frac{F^L(t_2) - F^L(t_1)}{F(t_2) - F(t_1)} = 1 + t_1(1 + Ct_1)h_1(t_1, t_2) - t_2(1 + Ct_2)h_2(t_1, t_2),$$

(5.41)

where $\left( F^L(t_2) - F^L(t_1) \right)$ is the df corresponding to the length-biased model, holds for all $(t_1, t_2) \in D$, if and only if $X$ follows Pareto II (2.17), exponential (2.18) and beta (2.19) according as $C > 0$, $C = 0$ and $C < 0$.

Proof: From (5.35), under the weight function $w(t) = t$, we have
Comparing (5.41) and (5.42), we have

$$\frac{F^L(t_2) - F^L(t_1)}{F(t_2) - F(t_1)} = \frac{m(t_1, t_2)}{\mu} = \frac{1}{\mu \left( R(t_1) - R(t_2) \right)} \int_{t_1}^{t_2} x f(x) dx. \quad (5.42)$$

Comparing (5.41) and (5.42), we have

$$\frac{1}{\mu} \int_{t_1}^{t_2} x f(x) dx = \left( R(t_1) - R(t_2) \right) + t_c (1 + C t_c) f(t_c) - t_c (1 + C t_c) f(t_c). \quad (5.43)$$

On differentiating (5.43) with respect to $t_i; i = 1, 2$, we get

$$f'(t_i) = \frac{(1 + 2\mu C)}{f(t_i)} \left( 1 + Ct_i \right). \quad (5.44)$$

Integrating (5.44), we obtain the densities of Pareto II, exponential and beta distributions according as $C > 0, C = 0$ and $C < 0$.

Conversely, substituting (2.17) in the definition of conditional moment function $m(t_1, t_2)$ and division by $\mu$, yields the required form for the Pareto II distribution. The case is similar to that of exponential and beta distributions.

**Theorem 5.11:** The ratio of the relation

$$\frac{F^L(t_2) - F^L(t_1)}{F(t_2) - F(t_1)} = 1 - g^*(t_1) h_2(t_1, t_2) + g^*(t_1) h_1(t_1, t_2), \quad (5.45)$$

where $g^*(t_i) = \frac{g(t_i)}{\mu}; i = 1, 2$, holds for all $(t_1, t_2) \in D$ for the family (1.31).

**Proof:** Integrating (2.24) over the limits $t_1$ to $t_2$ and by dividing $(F(t_2) - F(t_1))$, we get

$$g(t_2) h_2(t_1, t_2) - g(t_1) h_1(t_1, t_2) = \mu - m(t_1, t_2). \quad (5.46)$$

Equations (5.42) and (5.46) together imply the required result. The converse part is obtained by direct calculation.

**Corollary 5.1:** When $g(t) = b_0 + b_1 t + b_2 t^2$ in Theorem 5.11, reduces to that of Pearson family of distributions (1.30).
Theorem 5.12: Let $X$ be a weighted rv associated to $X$ and $w(t) = t$, then the ratio of the relationship

$$\frac{F^-(t_1) - F^-(t_2)}{F(t_1) - F(t_2)} = K \left[ 1 + t_1 h_1(t_1, t_2) - t_2 h_2(t_1, t_2) + n'(t_1, t_2) \right],$$

(5.47)

where $K = \mu c(\theta)$, a constant and $n'(t_1, t_2) = E \left[ X \frac{a'(X)}{a(X)} | t_1 < X < t_2 \right]$ holds for all $(t_1, t_2) \in D$ for the class of distributions (Sankaran and Gupta (2005)) defined by

$$f(t, \theta) = \begin{cases} a(t)(c(\theta)) \exp(-c(\theta)t); & a < t < b \smallskip \text{or otherwise} \\ 0; & \text{otherwise} \end{cases}$$

(5.48)

Proof: Assuming (5.47) and using (5.42), we have

$$\int_{t_1}^{t_2} xf(x) dx = K \left[ (R(t_1) - R(t_2)) + t_1 f(t_1) - t_2 f(t_2) - \int_{t_1}^{t_2} x \frac{a'(x)}{a(x)} f(x) dx \right]$$

(5.49)

differentiating (5.49) with respect to $t_i$, $i = 1, 2$ and on simplification, we get (5.48).

Substitution of (5.49) in (5.42) and on simplification, yields the converse part of the theorem.

In view of the form-invariance property for the generalized Pearson family of distributions (1.32), the analogous statement for Theorem (5.8) in the context of size-biased model is immediate. This is stated in the following theorem.

Theorem 5.13: The df of a non-negative rv $X$ belongs to generalized Pearson family (1.32) under size-biased sampling, if and only if its $r$th order conditional moments satisfies a recurrence relation of the form

$$m'(t_1, t_2) = (q_{r2} + q_{r1}^2) t_2^{-1} h_1(t_1, t_2) - (q_{r1} + q_{r1}^2) t_1^{-1} h_1(t_1, t_2)$$

$$- p_{r0} m_{r-2}(t_1, t_2) - p_{r1} m_{r-1}(t_1, t_2),$$

where $q_i = \frac{q_i}{p_2}$; $i = 1, 2$, $p_{r0} = \frac{p_0 + (r - (\alpha + 1)) q_1}{p_2}$ and $p_{r1} = \frac{p_1 + (r - \alpha) q_2}{p_2}$, provided $p_2 \neq 0$.
and

\[ m'(t_1,t_2) = (q_{1r} + q_{2r},t_2)h_{12}(t_1,t_2) - (q_{1r} + q_{2r},t_1)h_{12}(t_1,t_2) - p_{0r}m_{r-1}(t_1,t_2), \quad (5.50) \]

where \( q_{ir} = \frac{q_{ir}}{(r+1-\alpha)q_i + p_i}; \quad i = 1, 2 \) and \( p_{0r} = \frac{(p_0 + (r-\alpha)q_i)}{(r+1-\alpha)q_i + p_i} \), provided \((r+1-\alpha)q_i + p_i \neq 0\) and \( p_2 = 0 \).

**Proof:** The proof is similar to that of the Theorem 5.8.

**Remarks 5.3:** 1) When \( t = 0 \), Theorem 5.13 reduces to the Theorem 2.9 given in the second Chapter.

2) When \( p_2 = 0 \), this Theorem 5.13 reduces to that of the Pearson family of distributions (2.48).

Now we consider the geometric vitality function for the weighted models. The geometric vitality function corresponding to weighted model is denoted as \( \log G''(t_1,t_2) \) and it is given by

\[ \log G''(t_1,t_2) = E\left( \log X_w \mid t_1 < X_w < t_2 \right), \quad (t_1,t_2) \in D \]

\[ = \frac{1}{\left( F^w(t_1) - F^w(t_2) \right)} \int_{t_1}^{t_2} (\log x)^{w''}(x)dx. \quad (5.51) \]

(5.51) can be written as

\[ \log G''(t_1,t_2) = \frac{1}{m_w(t_1,t_2)(R(t_1) - R(t_2))} \int_{t_1}^{t_2} w(x)f(x)\log x dx, \quad (5.52) \]

where \( m_w(t_1,t_2) = E\left[ w(X) \mid t_1 < X < t_2 \right] \).

**Corollary 5.2:** When \( w(t) = t \), (5.52) reduces to the geometric vitality function of a length-biased model and it is denoted as \( \log G^L(t_1,t_2) \). Substituting this and applying integration by parts, we obtain
Theorem 5.14: For a non-negative rv $X$, the relation

$$\lambda m(t_1, t_2) \log G^s(t_1, t_2) - \log G(t_1, t_2) = 1 + t_1 h_1(t_1, t_2) \log t_1 - t_2 h_2(t_1, t_2) \log t_2$$  \hspace{1cm} (5.54)

holds for all $(t_1, t_2) \in D$ if and only if $X$ follows an exponential distribution (2.18).

Proof: Suppose that the relation (5.54) holds, then by definition,

$$\lambda \int_{t_1}^{t_2} \frac{f(x)}{R(t_1) - R(t_2)} dx - \frac{1}{R(t_1) - R(t_2)} \int_{t_1}^{t_2} f(x) \log x dx$$

$$= 1 + t_1 \frac{f(t_1)}{R(t_1) - R(t_2)} \log t_1 - t_2 \frac{f(t_2)}{R(t_1) - R(t_2)} \log t_2.$$  \hspace{1cm} (5.55)

Multiply both sides of (5.55) by $(R(t_1) - R(t_2))$ and on differentiation with respect to $t_i$; $i = 1, 2$, yields the required result. Converse part can be proved by direct calculation. The measure of uncertainty for the weighted model is denoted as $H^w(t_1, t_2)$ and it is defined by

$$H^w(t_1, t_2) = -\int_{t_1}^{t_2} \frac{f^w(x)}{R^w(t_1) - R^w(t_2)} \log \left( \frac{f^w(x)}{R^w(t_1) - R^w(t_2)} \right) dx$$

$$= \frac{-1}{m_w(t_1, t_2)(R(t_1) - R(t_2))} \int_{t_1}^{t_2} w(x)f(x) \log \left( \frac{w(x)f(x)}{m_w(t_1, t_2)(R(t_1) - R(t_2))} \right) dx$$  \hspace{1cm} (5.56)

and the corresponding conditional measure of uncertainty is

$$M^w(t_1, t_2) = -E\left\{ \log f^w(X) \mid t_1 < X < t_2 \right\}$$

$$= \frac{-1}{m_w(t_1, t_2)(R(t_1) - R(t_2))} \int_{t_1}^{t_2} w(x)f(x) \log \left( \frac{w(x)f(x)}{E(w(X))} \right) dx.$$  \hspace{1cm} (5.57)