A STUDY ON THE KERRIDGE'S INACCURACY MEASURE AND RELATED CONCEPTS

Thesis submitted to the

Cochin University of Science and Technology

for the award of the Degree of

Doctor of Philosophy

under the Faculty of Science

By

SMITHA S.

DEPARTMENT OF STATISTICS

COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY

COCHIN 682 022

JANUARY 2010

Certificate

Certified that the thesis entitled 'A study on the Kerridge's inaccuracy measure and related concepts' is a bonafide record of work done by Smt. Smitha S. under my guidance in the Department of Statistics, Cochin University of Science and Technology and that no part of it has been included anywhere previously for the award of any degree or title.

Cochin-22
14th January

Dr. K.R.Muraleedharan Nair
Supervising Guide

Declaration

This thesis contains no material which has been accepted for the award of any other Degree or Diploma in any University and to the best of my knowledge and belief, it contains no material previously published by any other person, except where due references are made in the text of the thesis.

Cochin-22 Smitha S.
14th January 2010

Acknowledgement

I take this opportunity with much pleasure to thank all the people who have helped me through the course of my journey towards producing this thesis.

I am heartily thankful to my guide, Dr K.R. Muraleedharan Nair, Professor, Department of Statistics, Cochin University of Science and Technology whose help, stimulating suggestions and encouragement helped me in all the time of research and writing of this thesis.

I owe my deepest gratitude to Dr. N. Unnikrishnan Nair, Emeritus Professor and Former Vice Chancellor, Cochin University of Science and Technology for the valuable suggestions and advice.

I am grateful to Dr. N Balakrishna, Head, Department of Statistics, CUSAT who supported me in any respect during the completion of the thesis.

I also wish to acknowledge my indebtedness to Dr. P.G. Sankaran, Reader, Department of Statistics, CUSAT and all other faculty members of the department for their assistance and useful comments.

I am deeply indebted to Dr. G. Rajesh, Departments of Mathematics and Statistics, D.B. Pampa College Parumala for his timely support during the process of this work.

I owe a lot to my friends especially to research scholars, who helped, inspired and encouraged me whenever it was need. The co-operation and help rendered by the non-teaching staffs will be remembered always.

The financial support of the Cochin University of Science and Technology is gratefully acknowledged.

This thesis would not have been possible without the confidence, endurance and support of my family. I wish to thank my parents and brother, my in laws, and finally my husband and daughter for their affection, understanding and patience.

Above all, I am thankful to the choicest blessings of almighty at all times.

Smitha S

CONTENTS

Chapter 1

INTRODUCTION

The Shannon's entropy, introduced by Shannon (1948), has been extensively used as a quantitative measure of uncertainty associated with a random phenomena. If $A_1, A_2 ... A_n$ are mutually exclusive and exhaustive events in a sample space with respective probabilities $p_1, p_2 ... p_n$, the Shannon's entropy is defined as

$$H_n(P) = -\sum_{i=1}^{n} p_i \log p_i .$$

$H_n(P)$ is being interpreted as a measure of uncertainty concerning the outcome of the experiment or a measure of information conveyed through the knowledge of the probabilities associated with the events.

Observing that the Shannon's entropy satisfies several properties, the earlier work on the Shannon's entropy was centered around characterizing $H_n(P)$ based on several postulates. The works of Khinchin (1953), Tverberg (1958), Chaundy and Mcleod (1960), Lee (1964), Mathai and Rathie (1975), Ebanks et al. (1998), Yeung (2002), and Csiszar (2008) in this direction. Another aspect of interest that has received much attention among researchers is the identification of probability distributions that maximizes the Shannon's entropy subject to some restrictions on the underlying random variable. The books by Kapur (1989, 1994) provide a more or less exhaustive review of various maximum entropy models.

In the continuous setup if $f(x)$ denotes the probability density function of a random variable X with support $[a,b]$, the continuous analogue of Shannon's entropy takes the form

$$H(F) = -\int_a^b f(x)\log f(x)\,dx.$$

Ebrahimi and Pellerey (1995) have extended the definition of Shannon's entropy to the left truncated situation and they used this measure to introduce a new partial ordering for life distributions. Ebrahimi (1996) has given an upper bound for this measure in terms of the mean residual life function, $m(t)$, namely

$$H(F;t) \le 1 + \log m(t),$$

where $m(t) < \infty$. Nair and Rajesh (1998), Sankaran and Gupta (1999), Asadi and Ebrahimi (2000) and Belzuence et al. (2004) have looked into the problem of characterizing probability distributions using the functional form of the residual entropy function. Rajesh and Nair (1998) have defined the residual entropy function in discrete time domain and have shown that it determines the distribution uniquely. Further, it is established that the constancy of the same is characteristic to the geometric distribution. Di Crescenzo and Longobardi (2002) have shown that in many realistic situations uncertainty is not necessarily related to the future but can also refer to the past. For considering such situations, they proposed the past entropy defined over $(0,t)$. Recently, Nanda and Paul (2006,a) have proposed some ordering properties based on this measure.

Kullback and Leibler (1951) have extensively studied the concept of directed divergence, which aims at discrimination between two populations. Aczel and Daroczy (1975) laid down an axiomatic foundation to this concept. Ebrahimi and Kirmani (1996, b) extended this concept and has given a measure of discrimination between two residual lifetime distributions. Further, they proved that the constancy of this measure is a characteristic property of the proportional hazards model. Along the similar lines of the measure proposed by Ebrahimi and Kirmani (1996, b), Di Crescenzo and Longobardi (2004) have examined the problem of discrimination between the past lives.

Another useful measure for discrimination among distributions is the notion of affinity studied by Matusita (1954). Affinity focuses attention on the likeness of distribution and has properties similar to that of Kullback- Leibler divergence measure. Kirmani (1968) has shown that the affinity between two distributions is related to the idea of distance between distributions. In testing hypothesis, it is desirable to know bounds of errors, because even if the most powerful test is adopted, it is often the case that we cannot obtain the exact value of the power of the test. However, we can easily get them in terms of affinity. The relative Renyi entropy, also known as Chernoff distance, finds application in several branches of learning as a potential measure of distance between two populations. Asadi et. al (2005) have studied the application of this measure in the context of reliability studies.

The notion of inaccuracy was introduced by Kerridge (1961) and can be viewed as a generalization of the Shannon's entropy. Suppose that the experimenter asserts that the probability of the i^{th} eventuality is q_i when the true probability is p_i. Then the inaccuracy of the observer, as proposed by Kerridge (1961), can be measured by

$$I(P,Q) = -\sum_{i=1}^{n} p_i \log q_i ,$$

where $P = (p_1, p_2, ..., p_n)$ and $Q = (q_1, q_2, ..., q_n)$ are two discrete probability distributions such that $p_i \geq 0$, $q_i \geq 0$ and $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1$. In fact, the Kerridge's inaccuracy measure can be expressed as the sum of a measure of uncertainty and a measure of discrimination between two populations. When an experimenter states the probabilities of various events in an experiment, the statement can lack precision in two ways: one is resulting from incorrect information and the other from vagueness in the statement. Kerridge (1961) proposed the "inaccuracy measure" that can take accounts for these two types of errors. Nath (1968) extended Kerridge's inaccuracy to the case of continuous situation and discussed some properties. If $F(x)$ is the actual distribution function corresponding to the observations and $G(x)$ is the distribution function assigned by the experimenter and $f(x)$ and $g(x)$ are the corresponding density functions, the inaccuracy measure is defined as

$$I(F,G) = -\int_0^\infty f(x)\log g(x)dx .$$

He also extended this measure to inaccuracies of order 'r'. Nair and Gupta (2007) extended the definition of measure of inaccuracy to the truncated situation. Recently, Taneja et. al (2009) proposed the uniqueness property of the dynamic inaccuracy measure defined by Nair and Gupta (2007) and some properties of this measure were also studied. In addition, the concept of inaccuracy has its application in statistical inference, estimation and coding theory.

Even though concepts such as failure rate, mean residual life function, vitality function etc are extensively used in reliability studies for modeling lifetime data, recently a lot of interest has evoked in using entropy concepts to describe the stability of components. In life time studies, the data is generally truncated. Hence there is scope for extending information theoretic concepts to the truncated situation. Motivated by this, in the present study, we extend the definition of inaccuracy, affinity and Chernoff distance to the truncated situation. Further we also look into the problem of characterization of probability distributions using the functional form of these measures.

After the present introductory chapter, in Chapter 2 we give a brief review of the existing literature in the area of study. In Chapter 3, we extend the definition of inaccuracy to the truncated situation and provide characterization results for certain probability distributions. The inaccuracy measure is generalized to inaccuracies of order 'r' in Chapter 4. Characterizations of distributions in the context of proportional hazards model and proportional reversed hazards model using the functional form of the generalized inaccuracy measure are also given in this chapter. In Chapter 5, we extend the notion of Chernoff distance to the truncated situation and obtain characterization results using functional form of the truncated Chernoff distance. We also discuss affinity in the truncated situation, which is a special case of the Chernoff distance, in this chapter. Residual inaccuracy measure and affinity in discrete setup are the subject matter of Chapter 6.Towards the end of this chapter we also give a plan for future study in this area.

Chapter 2

REVIEW OF LITERATURE

In the present chapter, we give a brief review of some of the existing work in reliability theory, entropy, inaccuracy and related areas, which are of use in subsequent chapters.

2.1 Some basic concepts in Reliability

The concepts of entropy, inaccuracy and affinity find a lot of application in fields of reliability. The basic concepts in reliability are the survival function, hazard rate, mean residual life function and their generalizations. In the sequel, we give the definitions and discuss some basic properties associated with these concepts.

Survival function

Let X be a non-negative random variable, defined on a probability space (Ω, F, P), with distribution function $F(x) = P(X \le x)$. In the reliability context, X generally represents the length of life of a device measured in some units of time. The function

$$\overline{F}(x) = P(X > x)$$

$$= 1 - F(x), \tag{2.1}$$

is called the survival (reliability) function. $\overline{F}(x)$ gives the probability that the device will operate with out failure for a time x. It is a non-increasing continuous function with $\overline{F}(0) = 1$ and $\lim_{x\to\infty} \overline{F}(x) = 0$. One of the major problems of interest in reliability analysis is that of determination of the functional form of the survival function using data on failure times.

If X is a discrete random variable with support $N^{+}=\{0,1,2,...\}$, the survival function corresponding to X is defined as

$$\begin{aligned}\overline{F}(x) &= P(X > x)\\ &= \sum_{j=x+1}^{\infty} p(j)\,,\ x=0,1,2,... \end{aligned} \tag{2.2}$$

where $p(x)$ is the probability mass function associated with X. Note that $\overline{F}(x)$ is a non-increasing step function with $\overline{F}(x)=1$ for $x<0$.

Hazard rate

Defining the right extremity L of $F(x)$ by $L=\inf\{x: F(x)=1\}$, for $x<L$, the hazard rate $h(x)$ of X is defined as

$$h(x)=\lim_{\Delta x\to 0^{+}} \frac{P\left(x<X\le x+\Delta x \mid X>x\right)}{\Delta x}. \tag{2.3}$$

$h(x).\Delta x$ can be interpreted as the probability of failure in the interval $[x, x+\Delta x)$, given that the component has survived up to time x, as the length of the interval $\Delta x \to 0$. The hazard rate is also referred to as the failure rate, instantaneous hazard rate in reliability, force of mortality in demographic studies and the age specific hazard rate in epidemiology. When X admits an absolutely continuous distribution with probability density function $f(x)$, equation (2.3) reduces to

$$\begin{aligned} h(x) &= \frac{f(x)}{\overline{F}(x)}\\ &= -\frac{d}{dx}\left(\log \overline{F}(x)\right). \end{aligned} \tag{2.4}$$

In the general set up, for a random variable X with support $-\infty<X<\infty$, Kotz and Shanbhag (1980) defines the hazard rate as the Radon-Nikodym derivative with respect to Lebesgue measure on $\{x: F(x)<1\}$, of the hazard measure

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

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

$$\overline{F}(x)=\prod_{\Delta x<x}\left(1-H(\Delta x)\right)\exp\left(-H_C(-\infty,x)\right) , \tag{2.5}$$

where H_C is the continuous part of H. When X is a non-negative random variable admitting an absolutely continuous distribution function $F(x)$, then equation (2.5) reduces to

$$\overline{F}(x)=\exp\left(-\int_0^x h(t)dt\right). \tag{2.6}$$

In the view of equation (2.6), $h(x)$ determines the distribution uniquely and the constancy of the same is characteristic to the exponential model [Galambos and Kotz (1978)]. For further characterizations of probability distributions based on the functional form of hazard rate, we refer to Mukherjee and Roy (1986) and Azlarov and Volodin (1986).

In the discrete setup, Barlow et al. (1963) defines the hazard rate for a random variable X in the support of non-negative integers as

$$\begin{aligned} h(x) &= \frac{P(X=x)}{P(X \geq x)} \\ &= \frac{f(x)}{\overline{F}(x-1)}. \end{aligned} \tag{2.7}$$

Equation (2.7) gives the conditional probability of the failure of a device at time x, given that it has not failed up to time $x-1$. Equation (2.7) can also be written as

$$h(x)=\frac{\overline{F}(x-1)-\overline{F}(x)}{\overline{F}(x-1)}. \tag{2.8}$$

It is established that the hazard rate function $h(x)$ defined in equation (2.7) uniquely determines the distribution. Xekalaki (1983), Gupta and Gupta (1983) and Hitha and Nair (1989) have extensively studied the problem of characterization of probability distributions using the form of hazard rate.

It may be observed that in the continuous situation, the hazard rate can be unbounded, where as in the discrete case it is always finite. Xie, Gaudoin and Bracquemond (2002) have observed that $h(x)$ defined in equation (2.7) cannot grow exponentially, which is common in the case for components during the wear-out lifetime period. Further, in the discrete case the cumulative hazard function $H(x)=\sum_{i=1}^{x}h(i)$ is not equivalent to $-\log\overline{F}(x)$ as in the continuous case.

In view of the above, several authors including Roy and Gupta (1998), and Xie, Gaudoin and Bracquemond (2002) have proposed an alternative definition of hazard rate function in discrete time.

For a discrete distribution with survival function $\overline{F}(x)$, the alternative hazard rate function $h^{*}(x)$ is defined as

$$h^{*}(x)=\log\left(\frac{\overline{F}(x-1)}{\overline{F}(x)}\right);\ x=1,2,... \tag{2.9}$$

with this definition one can have a simple relationship connecting equations (2.7) and (2.9) as

$$h(x)=1-e^{-h^{*}(x)}. \tag{2.10}$$

Reversed hazard rate

The concept of reversed hazard rate has been introduced by Keilson and Sumita (1982) and extensively studied by Shaked and Shanthikumar (1994). For a non-negative random variable X, the reversed hazard rate is defined as

$$\lambda(x) = \lim_{\Delta x \to 0} \frac{P(x - \Delta x < X \le x \mid X \le x)}{\Delta x}. \tag{2.11}$$

$\lambda(x).\Delta x$ can be interpreted as the probability of failure in the interval $(x - \Delta x, x]$, given that the failure had occurred in $[0, x]$ as $\Delta x \to 0$. When the probability density function of X , $f(x)$, exists equation (2.11) can be written as

$$\lambda(x) = \frac{f(x)}{F(x)}$$
$$= \frac{d \log F(x)}{dx}.$$

The reversed hazard rate uniquely determines $F(x)$ through the relation

$$F(x) = \exp\left(-\int_x^\infty \lambda(t)\, dt\right). \tag{2.12}$$

The problem of ordering of life distributions using the reversed hazard rate has been addressed by Shaked and Shanthikumar (1994). Block et al. (1998) established that there is no non-negative random variable having an increasing reversed hazard rate distribution. They also describe some useful properties for k out of n systems in terms of the reversed hazard rate. Observing that the hazard rate $h(x)$ and reversed hazard rate $\lambda(x)$ are functionally related through the relationship

$$\lambda(x) = \frac{h(x).\overline{F}(x)}{F(x)}.$$

Finkelstein (2002) has shown that

$$\lambda(x) = \frac{h(x)}{\exp\left(\int_0^x h(t)\, dt\right) - 1}.$$

Nair et al. (2005) characterized certain probability models using a possible relationship between reversed hazard rate and conditional expectation.

Mean residual life function

The mean residual life function (MRLF) or life expectancy at age x represents the average life time 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

$$m(x) = E(X - x \mid X \geq x), \tag{2.13}$$

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

$$m(x) = \frac{1}{\overline{F}(x)} \int_x^\infty \overline{F}(t) dt . \tag{2.14}$$

Further, the following relationship holds between $m(x)$ and $h(x)$ when both exist.

$$m'(x) = h(x).m(x) - 1, \tag{2.15}$$

where $m'(x)$ denote the derivative of $m(x)$. Also knowledge of the mean residual life function completely determines the survival function through the relationship

$$\overline{F}(x) = \frac{m(0)}{m(x)}.\exp\left(-\int_0^x \frac{dt}{m(t)} \right), \tag{2.16}$$

for every x in $(0, L)$. Thus $\overline{F}(x)$, $h(x)$ and $m(x)$ are all equivalent in the sense that given one of them, the other two can be determined. It is easy to see that the constancy of $h(x)$ or $m(x)$ characterizes the exponential distribution. For further characterizations using the functional form of mean residual life function, we refer to Mukharjee and Roy (1986) and Sullo and Rutherford (1977). Gupta and Kirmani

(2004) have observed that the ratio of the hazard rate and the mean residual life function determines the distribution uniquely. The mean residual life function finds application in actuarial science for setting rates and benefits for life insurance. In the biomedical setting, researchers analyze survivorship studies by the mean residual life function.

Vitality function

The concept of vitality function was introduced by Kupka and Loo (1989) as a Borel measurable function defined on the real line. It is closely related to mean residual life function and it is defined as

$$v(x) = E(X \mid X \geq x)$$

$$v(x) = \frac{1}{\overline{F}(x)} \int_x^\infty t\, dF(t). \tag{2.17}$$

Obviously

$$v(x) = x + m(x) \tag{2.18}$$

and

$$v'(x) = m(x).h(x),$$

where $v'(x)$ denotes the derivative of $v(x)$. For characterizations of probability distributions using vitality function, we refer to Nair and Sankaran (1991) and Ruiz and Navarro (1994). Nair and Rajesh (2000) defines the geometric vitality function $G(t)$ for $t > 0$ as

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

$$= \frac{1}{\overline{F}(t)} \int_t^\infty \log x\, f(x) dx\ . \tag{2.19}$$

They have discussed the properties of geometric vitality function and further characterized some probability distributions using the functional form of $\log G(t)$. The doubly truncated situation was considered in Sunoj et al. (2009)

and exponential distribution, Pareto distribution, beta distribution and power distribution have been characterized using the functional form of the geometric vitality function.

2.2 Proportional Hazards model (PH model)

Cox (1972) has introduced and extensively studied a dependence structure among two distributions, which is referred as the proportional hazards model (PH model). Let $F(x)$ and $G(x)$ be two distribution functions. Denote the hazard rates associated with $F(x)$ and $G(x)$ by $h_1(x)$ and $h_2(x)$ respectively. $F(x)$ is said to be proportional hazards model of $G(x)$ if the relationship

$$h_2(x) = \theta h_1(x), \tag{2.20}$$

where θ is some real constant, holds for all real $x > 0$. From the definitions of hazard rates, equation (2.20) is equivalent to

$$\overline{G}(x) = \left[\overline{F}(x)\right]^{\theta}, \text{ for all } \theta > 0 .$$

It may be noted that if $F(x)$ is the proportional hazards model of $G(x)$, then $G(x)$ will be the proportional hazards model of $F(x)$. The importance of proportional hazards models from the point of view of modeling statistical data has been studied by Gupta et al. (2001). The model finds application in variety of fields of study such as reliability, survival analysis, medicine, economics etc. Ebrahimi and Kirmani (1996) and Nair and Gupta (2007) looked into the problem of characterization of probability distributions under the proportional hazards model assumption.

2.3 Proportional Reversed Hazards model (PRH model)

Like the proportional hazards model, Gupta et al. (1998) proposed a dual model called proportional reversed hazards model (PRH model), defined by the relationship

$$\lambda_2(x) = \phi\lambda_1(x), \tag{2.21}$$

where $\lambda_1(x)$ and $\lambda_2(x)$ are the reversed hazard rates associated with distribution functions $F(x)$ and $G(x)$ and $\phi > 0$. Equation (2.21) can be expressed in terms of distribution functions as

$$G(x) = \left[F(x)\right]^{\phi}. \tag{2.22}$$

Proportional reversed hazards model is useful in the analysis of left censored or right truncated data. Sengupta et.al (1999) illustrated that proportional reversed hazards model leads to a better fit for some data sets than proportional hazards model. Di Crescenzo (2000) has obtained some results on the proportional reversed hazards model concerning ageing characteristics and stochastic orders.

2.4 Weighted distributions

Rao (1965) introduced the concept of weighted distributions in connection with modeling statistical data in situations where the usual practice of using standard distributions for the purpose was not found appropriate. Jain et al. (1989), Gupta and Kirmani (1990) and Nanda and Jain (1999) used the weighted distribution in many practical problems to model unequal sampling probabilities. Mathematically weighted distribution is defined as follows. Let (Ω, F, P) be a probability space and $X : \Omega \to R$ be a random variable, where $R = (a,b)$ is the subset of the real line with $a > 0$ and $b > a$ can be finite or infinite. Assume that the distribution function $F(x)$ is absolutely continuous with probability density function $f(x)$ and let $w(x)$ is a non-negative function of X such that $\mu = E\left(w(X)\right) < \infty$. The random variable X_w with probability density function

$$f_w(x) = \frac{w(x) f(x)}{\mu}, x > 0 \tag{2.23}$$

is said to have the weighted distribution associated with the distribution of X. It arises when the observations generated from a stochastic process are recorded according to some weight function.

The basic problem when one uses a weighted distribution, as a tool for modeling is the identification of the appropriated weight function that fits the data. When the weight function depends on the length of the unit of interest, the resulting distribution is called length-biased distribution. When $w(x) = x$, the probability density function of the length-biased random variable X_L turns out to be

$$f_L(x) = \frac{x\, f(x)}{\mu}, \; x > 0, \;\; \mu = E\left(w(X)\right) < \infty . \tag{2.24}$$

Length-biased sampling situations may occur in clinical trials, reliability, queuing models, survival analysis and population studies where a proper sampling frame is absent. In such situations, items are sampled at rate proportional to their length so that larger values of the quantity being measured are sampled with higher probabilities. Cox (1962) provided the statistical interpretation of the length- biased distribution in the context of renewal theory. Numerous works on various aspects of length-biased sampling are available in Patil and Rao (1977), Rao (1965), Sen and Khattree (1996), Oluyede (1999, 2000, 2002), Van Es et.al (2000), El Barmi and Nelson (2002) and Sankaran and Nair (1993). Gupta and Keating (1986) proposed some standard relationships between original and length-biased random variable using reliability concepts. They are

$$\overline{F_L}(t) = \left(\frac{t + m(t)}{\mu}\right)\overline{F}(t), \tag{2.25}$$

$$h_L(t) = \left(\frac{t}{t + m(t)}\right)h(t), \tag{2.26}$$

where $\overline{F_L}(t)$ and $h_L(t)$ denotes the survival function and hazard rate corresponding to the length-biased models and $\overline{F}(t)$ and $h(t)$ denote the survival function and hazard rate of the original distribution.

Another important particular case of weighted distribution is the equilibrium distribution, which arises when $w(x) = \frac{1}{h(x)}$, where $h(x)$ is the hazard rate . Let X be a continuous random variable with probability density function $f(x)$ and distribution function $F(x)$. Associated with X, a random variable X_E can be defined with probability density function

$$f_E(x) = \frac{\overline{F}(x)}{\mu}, \; X_E > 0, \; \mu = E\left(w(X)\right) < \infty \tag{2.27}$$

$f_E(x)$ is called an equilibrium distribution. The reliability implication of this model, the relationship between various characteristics of $F_E(x)$ with those of $F(x)$ and some characteristics are available in Gupta (1979) and Gupta and Kirmani (1990). The major relationships are

$$h_E(x) = \frac{1}{m(x)} \tag{2.28}$$

$$m(x) = \frac{m_E(x)}{1 + m_E^{'}(x)} \tag{2.29}$$

and

$$\overline{F_E}(x) = \mu^{-1}\,\overline{F}(x)\,m(x), \tag{2.30}$$

where $h_E(x)$ and $m_E(x)$ are the hazard rate and mean residual life function corresponding to the equilibrium model.

2.5 Shannon's entropy

The notion of entropy was originally developed by Shannon (1948), an electrical engineer in Bell Telephone Laboratory. At the same time, Wiener (1948) also considered the communication situation and came up independently with results similar to those of Shannon.

Consider a random experiment with n possible outcomes having probabilities $p_1, p_2, ...p_n$. The Shannon's entropy is defined as

$$H_n\left(P\right) = -\sum_{i=1}^{n} p_i \log p_i . \qquad (2.31)$$

Equation (2.31) measures the extent of uncertainty concerning the outcome of the experiment. When $p_i = 1$ for some i, $H_n\left(P\right) = 0$, which implies that there is no uncertainty about the predictability of the random variable. As a convention $0 \log 0$ is taken as zero. On the other hand, if we consider equation (2.31) after the experiment has been carried out, then equation (2.31) can be viewed as a measure of the amount of information conveyed by the realization of the experiment.

If X is a continuous random variable having a cumulative distribution function $F(x)$ and $f(x) = F^{'}(x)$ denote its density function, then the continuous analogue of Shannon's entropy takes the form

$$H(F) = -\int_0^\infty f(x) \log f(x) dx \qquad (2.32)$$

$$= -E\left(\log f(x)\right).$$

Equation (2.32) is commonly referred to in literature as the Shannon information measure. In life testing experiments, one has information about the current age of the component under consideration. In such cases a more realistic approach for measuring the uncertainty about remaining lifetime of the unit was developed by Ebrahimi and Pellerey (1995). If $X_t = X - t \,|\, X > t$ represents the residual life of a component, then the probability density function of X_t is $\frac{f\left(x+t\right)}{\overline{F}\left(t\right)}$. The Shannon's entropy associated with X_t takes the form

$$H(F;t) = -\int_t^\infty \frac{f(x)}{\overline{F}(t)} \log\left(\frac{f(x)}{\overline{F}(t)}\right) dx \,, \overline{F}(t) > 0 \,. \qquad (2.33)$$

Equation (2.33) can be written as

$$H(F;t) = \log \overline{F}(t) - \frac{1}{\overline{F}(t)} \int_t^{\infty} f(x) \log f(x) dx .$$

The residual entropy function can also be expressed in terms of the hazard rate through the relationship

$$H(F;t) = 1 - \frac{1}{\overline{F}(t)} \int_t^{\infty} f(x) \log h(x) dx . \tag{2.34}$$

It may be noticed that $-\infty \le H(F;t) \le \infty$ and that $H(F;0)$ reduces to Shannon's entropy given by equation (2.32) defined over $(0,\infty)$. Belzunce et al. (2004) has shown that, under certain conditions, the residual entropy function determines the distribution uniquely.

Units having less uncertainty in life times are more reliable and hence the residual entropy function has much relevance in the study of stability of units/ components. Ebrahimi (1996), Nair and Rajesh (1998), Sankaran and Gupta (1999), Asadi and Ebrahimi (2000) and Belzunce et al. (2004) have characterized several lifetime distributions using the functional form of residual entropy function. Ordering and classification of life distributions using this concept are discussed in Ebrahimi and Pellerey (1995) and Ebrahimi (1996). Rajesh and Nair (1998) have defined the properties of the residual entropy function in the discrete time domain. Further, characterization results associated with the geometric distribution using functional form of residual entropy function are also obtained.

In many realistic situations uncertainty is not necessarily related to future but may also refer to past. In such a situation, Di Crescenzo and Longobardi (2002) considered the truncated distribution in $(0,t)$ and defined the measure

$$\overline{H}(F;t) = -\int_0^t \frac{f(x)}{F(t)} \log\left(\frac{f(x)}{F(t)}\right) dx , \tag{2.35}$$

where $F(x)$ is the distribution function. This measure is generally referred to as the past entropy. In terms of the reversed hazard rate $\lambda(x)$, the past entropy can be written as

$$\overline{H}(F;t) = 1 - \frac{1}{F(t)} \int_0^t f(x) \log \lambda(x)\, dx. \tag{2.36}$$

Renyi (1961) defines entropies of order α as

$$H_\alpha(F) = \frac{1}{1-\alpha} \log\left(\sum_{x=0}^{\infty} f^\alpha(x) \right); \alpha > 0, \alpha \neq 1. \tag{2.37}$$

For a continuous non-negative random variable X admitting an absolutely continuous distribution, equation (2.37) takes the form

$$H_\alpha(F) = \frac{1}{1-\alpha} \log\left(\int_0^\infty f^\alpha(x) dx \right); \alpha > 0, \alpha \neq 1. \tag{2.38}$$

When $\alpha \to 1$, equation (2.38) reduces to the Shannon's entropy given in equation (2.32). For the random variable $(X - t)$ truncated at $t > 0$, Renyi's entropy measure takes the form

$$H_\alpha(F;t) = \frac{1}{1-\alpha} \log\left(\int_t^\infty \left(\frac{f(x)}{\overline{F}(t)} \right)^\alpha dx \right). \tag{2.39}$$

Further, when $\alpha \to 1$, $H_\alpha(F;t)$ simplifies to the residual entropy function (2.33). For properties and characterization of (2.39) we refer to Rajesh (2001) and Abraham and Sankaran (2005). Khinchin (1957) generalized Shannon's entropy given by equation (2.32), by choosing a convex function ϕ with $\phi(1) = 0$ and defined the measure

$$H_\phi(X) = \int_0^\infty f(x) \phi(f(x)) dx. \tag{2.40}$$

Nanda and Paul (2006) generalized Shannon’s entropy given in equation (2.32) for two particular choices of ϕ and has defined entropies of order β as

$$H_1^{\beta}(X)=\frac{1}{\beta-1}\left(1-\int_0^{\infty} f^{\beta}(x)\,dx\right),\quad \beta\neq 1 \text{ and } \beta>0 \tag{2.41}$$

and

$$H_2^{\beta}(X)=\frac{1}{1-\beta}\ln\left(\int_0^{\infty} f^{\beta}(x)\,dx\right),\quad \beta\neq 1 \text{ and } \beta>0 \tag{2.42}$$

As $\beta\to 1$, equations (2.41) and (2.42) will reduce to Shannon’s entropy given in equation (2.32). For a unit which has survived up to an age t, equations (2.41) and (2.42) takes the form

$$H_1^{\beta}(X;t)=\frac{1}{\beta-1}\left(1-\int_t^{\infty}\left(\frac{f(x)}{\overline{F}(t)}\right)^{\beta} dx\right) \tag{2.43}$$

and

$$H_2^{\beta}(X;t)=\frac{1}{1-\beta}\ln\int_t^{\infty}\left(\frac{f(x)}{\overline{F}(t)}\right)^{\beta} dx\,. \tag{2.44}$$

It may be noted that when $\beta\to 1$, equations (2.43) and (2.44) become residual entropy function defined by equation (2.33). Nanda and Paul (2006) has also discussed some ordering and ageing properties in terms of the generalized entropy function. Further, they have obtained some characterization results for distributions based on the form of the generalized residual entropy function.

2.6 Kullback- Leibler divergence measure

Kullback and Leibler (1951) have studied a measure of information involving two probability distributions associated with the same experiment, which finds application in the field of information theory as well as in several other branches of learning. Aczel and Daroczy (1975) laid down an axiomatic foundation to this concept. This measure is also referred to as cross entropy, relative information etc.

Consider two discrete probability distributions $P=(p_1,p_2,...p_n)$ and $Q=(q_1,q_2,...q_n)$ with $p_{i,}\ q_i \geq 0$ and $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1$. The Kullback-Leibler divergence measure between P and Q is defined as

$$D_n\left(P,Q\right)=\sum_{i=1}^{n} p_i \log\left(\frac{p_i}{q_i}\right). \tag{2.45}$$

This measure is always non-negative and is zero if $p_i = q_i$. Further this measure cannot be viewed as a true distance measure, since it is neither symmetric nor it satisfies the triangle inequality. Kannappan and Rathie (1973) followed by Mathai and Rathie (1975) have obtained some characterization results based on certain postulates which naturally leads to equation (2.45). The concept of generalized directed divergence is discussed in Kapur (1968) and Rathie (1971). The continuous analogue of the measure given in equation (2.45) turns out to be

$$D(P,Q)=\int_{-\infty}^{\infty} f(x)\log\left(\frac{f(x)}{g(x)}\right)dx, \tag{2.46}$$

where $f(x)$ and $g(x)$ are the probability density functions corresponding to the probability measures P and Q.

If X and Y be two non-negative random variables admitting absolutely continuous distribution functions $F(x)$ and $G(x)$ respectively, then equation (2.46) takes the form

$$D(X,Y)=D(F,G)=\int_{0}^{\infty} f(x)\log\left(\frac{f(x)}{g(x)}\right)dx. \tag{2.47}$$

Ebrahimi and Kirmani (1996 a) have modified the definition of Kullback - Leibler measure in order to accommodate the current age of the system. If X and

Y represents the lifetime of components in a two component system and t is a specified unit of time, equation (2.47) has been modified as

$$D(X,Y;t) = D(F,G;t) = \int_t^\infty \frac{f(x)}{\overline{F}(t)} \log\left(\frac{f(x)/\overline{F}(t)}{g(x)/\overline{G}(t)}\right) dx \,. \tag{2.48}$$

The above equation can also be written as

$$D(F,G;t) = \int_t^\infty \frac{f(x)}{\overline{F}(t)} \log\left(\frac{f(x)}{\overline{F}(t)}\right) dx - \int_t^\infty \frac{f(x)}{\overline{F}(t)} \log\left(\frac{g(x)}{\overline{G}(t)}\right) dx \,.$$

$$= I(F,G;t) - H(F;t),$$

where $H(F;t)$ is the residual entropy function considered by Ebrahimi and Pellerey (1995) and $I(F,G;t)$ is the truncated inaccuracy measure explained in Section 2.8. Ebrahimi and Kirmani (1996 a) claims that for each fixed $t > 0$, equation (2.48) will have all the same properties of the measure defined in equation (2.47). In particular, $D(F,G;t) \geq 0$ with equality if and only if the probability density functions of residual life functions are equal almost everywhere. They have also studied properties of $D(F,G;t)$. Ebrahimi and Kirmani (1996 b) have further proved that the constancy of (2.48) with respect to t is a characteristic property for distributions coming under the proportional hazards model, described in Section 2.2.

2.7 Affinity between distributions

The concept of affinity between distributions was introduced and extensively studied in a series of work by Matusita (1954, 1955, 1957, 1961, 1967). Some results on affinity are also given in Kirmani (1968). This measure has been widely used in literature as a useful tool for discrimination among distributions. Affinity is symmetric in distributions and has direct relationships with error probability when classification or discrimination is concerned.

Consider two discrete distributions $P=(p_1, p_2,..., p_n)$ and $Q=(q_1,q_2,...,q_n)$; $p_i \geq 0$, $q_i \geq 0$, $i=1,2,...n$. Then affinity [Mathai and Rathie (1975)] between P and Q is defined as

$$\rho(P,Q)=\sum_{i=1}^{n}[p_i q_i]^{\frac{1}{2}}. \tag{2.49}$$

If X and Y are two non-negative random variables and if $f(x)$ and $g(x)$ are the corresponding probability density functions, then the affinity between F and G is defined as

$$\rho(F,G)=\int_0^{\infty}\sqrt{f(x)g(x)}dx. \tag{2.50}$$

This measure is also called Bhattacharyya coefficient [Bhattacharyya (1946)]. $\rho(F,G)$ lies between zero and one. It may be noted that, the smaller the affinity the larger the discrepancy among distributions. Ikeda (1963) has established that

$$D(F,G)\geq 1-\rho^2(F,G),$$

where $D(F,G)$ is the Kullback-Leibler divergence measure considered in equation (2.47). Recently Majernik (2004) has shown that

$$H_E(F,G)=2\left(1-\rho(F,G)\right),$$

where $H_E(F,G)$ is the Hellinger's distance defined by

$$H_E(F,G)=\int_0^{\infty}\left(\sqrt{f(x)}-\sqrt{g(x)}\right)^2 dx.$$

Matusita (1967) has extended the notion of affinity concerning two distributions to the case of several distributions, as

$$\rho_n\left(F_1,F_2,...,F_n\right)=\int\left(f_1(x), f_2(x),..., f_n(x)\right)^{\frac{1}{n}}dx,$$

where $F_1, F_2, ..., F_n$ are distributions defined in the same sample space and $f_1(x), f_2(x), ..., f_n(x)$ are their respective density functions. Also ρ_n satisfies the following properties

(i) $\rho_n(F_1, F_2, ..., F_n) \geq 0$.

(ii) $\rho_n(F_1, F_2, ..., F_n) = 1$, when and only when $F_1 = F_2 = ... = F_n$.

Affinity is used to study difference between populations or to classify populations. George and Mathai (1974) used this concept in population studies. Ramkumar (1975) used this to study the distribution of populations by religious affiliations. Matusita (1967) discussed the application of affinity in cluster analysis.

2.8 The concept of inaccuracy

The concept of inaccuracy was introduced by Kerridge (1961). This can be viewed as a generalization of the idea of entropy. It has been extensively used as a useful tool for measurement of error in experimental results. In expressing statement about probabilities of various events in an experiment, two kinds of errors are possible, namely, one resulting from the lack of enough information or vagueness in experimental results (eg: missing observation or insufficient data) and the other from incorrect information (eg: mis-specifying the model). All estimation and inference problems are concerned with making statements, which may be inaccurate in either or both of these ways. The error due to vagueness can be explained by using Shannon's measure of uncertainty. Kerridge (1961) discusses the theory and application of the concept of inaccuracy to statistical inference.

Suppose that the experimenter asserts that the probability of the i^{th} eventuality is q_i when the true probability is p_i. Then the inaccuracy of the observer, as proposed by Kerridge (1961), can be measured by

$$I(P,Q) = -\sum_{i=1}^{n} p_i \log q_i , \qquad (2.51)$$

where $P=(p_1,p_2,...,p_n)$ and $Q=(q_1,q_2,...,q_n)$ are two discrete probability distributions such that $p_i \geq 0$, $q_i \geq 0$ and $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1$.

Properties of inaccuracy measure

In the discrete set up, Kerridge (1961) has discussed the properties of $I\left(P,Q\right)$ which are listed below.

(i) $I(P,Q)=0$ if and only if $p_i = q_i = 1$ for one value, and $p_i = q_i = 0$, for all other i. This means that zero inaccuracy implies a correct statement made with complete certainty.

(ii) There is an infinite value of $I(P,Q)$ if $q_i = 0$, $p_i \neq 0$ for any i.

(iii) The value of $I(P,Q)$ is minimum for fixed p_i, when $p_i = q_i$ for all i.

(iv) If variations both of p_i's and q_i's are considered, the point $p_i = q_i = n^{-1}$ for all i, is a minimax point.

(v) If two sets of alternatives are asserted to have probabilities, which are independent, the inaccuracy of the joint assertion is the sum of the separate inaccuracies.

Nath (1968) extended Kerridge's inaccuracy to the case of continuous situation and discussed some properties. Let $F(x)$ be the actual distribution function corresponding to the observations and $G(x)$ be the distribution function assigned by the experimenter and $f(x)$ and $g(x)$ be the corresponding density functions. The inaccuracy measure is defined as

$$I(F,G) = -\int_{0}^{\infty} f(x)\log g(x)dx. \qquad (2.52)$$

Nath (1968) has also discussed the following properties for the inaccuracy measure expressed in equation (2.52).

(i) The inaccuracy measure can be written as

$$I(F,G) = -\int_0^\infty f(x)\log f(x)dx + \int_0^\infty f(x)\log\left(\frac{f(x)}{g(x)}\right)dx. \qquad (2.53)$$

The first term on the right side of equation (2.53) represents the error due to uncertainty, which is the Shannon's entropy [Shannon (1948)], while the second term is the Kullback-Leibler measure considered in equation (2.47) representing the error due to wrongly specifying the distribution as $G(x)$. Note that this becomes zero when $F(x) \equiv G(x)$. In this situation $I(F,G)$ can be thought of as the generalization of the Shannon's entropy. $I(F,G)$ achieves its minimum value only when $f(x) = g(x)$. Further, $I(F,G)$ is not necessarily invariant under the transformation of co-ordinates.

(ii) A necessary and sufficient condition for $I(F,G)$ to be finite is that F and G should be absolutely continuous with respect to each other. The condition is however not sufficient. Further, the finiteness of $I(F,G)$ and $H(F)$ may depend upon a parameter whereas the value of $D(F,G)$ so derived may be independent of that parameter.

(iii) $I(F,G) < \infty$, does not imply $I(G,F) < \infty$.

(iv) $I(F,G) < \infty$, $I(G,K) < \infty$ does not imply $I(F,K) < \infty$.

(v) Since we are considering finite measures, $I(F,G)$ always exists though its value may be $+\infty$ or even $-\infty$.

(vi) $I(F,K) - I(F,G) = D(F,K) - D(F,G)$.

Recently, Nair and Gupta (2007) extended the definition of measure of inaccuracy to the truncated situation. The inaccuracy measure for random variables truncated below for some $t > 0$ is defined as

$$I(F,G;t) = -\int_t^\infty \frac{f(x)}{\overline{F}(t)} \log\left(\frac{g(x)}{\overline{G}(t)}\right) dx, \tag{2.54}$$

and the measure truncated above some $t > 0$ is defined as

$$I^*(F,G;t) = -\int_0^t \frac{f(x)}{F(t)} \log\left(\frac{g(x)}{G(t)}\right) dx. \tag{2.55}$$

A physical meaning of equation (2.54) is that $I(F,G;t)$ represents the expected inaccuracy in the conditional distribution of $X - t \mid X > t$ when its actual density is specified as $\frac{g(x)}{\overline{G}(t)}$. For the convenience in the sequel we denote $I(F,G;t)$ and $I^*(F,G;t)$ by $I(t)$ and $I^*(t)$ respectively. We can easily obtain the functional relationship between hazard rates and $I(t)$ as

$$h_1(t) = \frac{I'(t) + h_2(t)}{I(t) + \log h_2(t)}, \tag{2.56}$$

where $I'(t) = \frac{d}{dt} I(t)$, $h_1(t)$ and $h_2(t)$ are the hazard rates of F and G respectively. A similar result exists between reversed hazard rates and $I^*(t)$, namely

$$\lambda_1(t) = \frac{\lambda_2(t) - I^{*'}(t)}{I^*(t) + \log \lambda_2(t)}. \tag{2.57}$$

Nair and Gupta (2007) also established that if F and G are absolutely continuous distribution functions such that G is the proportional hazards model of F, then $I(t)$ has the log linear form

$$I(t) = \log\left(\frac{at+b}{\theta(a+1)}\right) + \frac{a + \theta(a+1)}{a+1}, \tag{2.58}$$

for all $t > 0, a > -1$ and $b > 0$ if and only if F has generalized Pareto distribution specified by

$$\overline{F}(x) = \left(1 + \frac{ax}{b}\right)^{-\left(1+\frac{1}{a}\right)} . \tag{2.59}$$

Recently Taneja et al. (2009) proposed the uniqueness property of the dynamic inaccuracy measure defined in equation (2.54) and some properties of this measure are also studied.

Chapter 3

MEASURE OF INACCURACY

3.1 Introduction

The identification of an appropriate probability distribution for lifetimes is one of the basic problems encountered in reliability theory. Although several methods such as probability plots, goodness of fit procedures etc are available in literature to find an appropriate model followed by the observations, they fail to provide an exact model. A method to attain this goal is to utilize a suitable characteristic property of the model. A property P is said to be characteristic to a distribution if P holds under the distributional assumption and the only distribution for which P holds is the underlying distribution. Thus a characterization theorem enables one to uniquely determine the distribution. Most of the work on characterization of distributions in the reliability context centers around the hazard rate or the mean residual life function. In a variant approach, Ebrahimi (1996) proposed the residual entropy function as a useful tool to analyze the stability of a component/system. Following this several papers appeared employing information measures like time dependent Kullback-Leibler directed distance and their generalizations in characterizing life distributions. As pointed out in the chapter two, the inaccuracy measure can be thought of as a generalization of Shannon's entropy. Therefore, there is a scope for extending the results based on Shannon's entropy and its modifications as applicable for the inaccuracy measure to suit the context of reliability. Motivated by this, in the present chapter we look into the problem of characterization of probability distributions using truncated versions of the inaccuracy measure.

Recently, Nair and Gupta (2007) have extended the inaccuracy measure defined by equation (2.52) to the truncated situation as given in equation (2.54). For the random variable, $X-t \mid X>t$ the truncated inaccuracy measure has the form

$$I(F,G;t) = -\int_t^\infty \frac{f(x)}{\overline{F}(t)} \log\left(\frac{g(x)}{\overline{G}(t)}\right) dx .$$

It may be observed that $I(F,G;t)$ considered above can be decomposed as

$$I(F,G;t) = H(F;t) + D(F,G;t), \tag{3.1}$$

where $H(F;t)$ is the residual entropy function defined in equation (2.33) and $D(F,G;t)$ is the modified Kullback- Leibler divergence measure considered in equation (2.48). If $F(x)$ is the actual distribution corresponding to the observations and $G(x)$ is the distribution function assigned by the experimenter, equation (3.1) asserts that the residual inaccuracy measure $I(F,G;t)$ is the sum of the residual entropy function of $F(x)$, which measures uncertainty, and a measure of discrimination between $F(x)$ and $G(x)$. For convenience in the sequel, we denote $I(F,G;t)$ by $I(t)$.

3.2 Characterization of probability distributions using the functional form of inaccuracy measure

In this section we consider the problem of characterizing distribution functions $F(x)$ and $G(x)$ based on given functional forms for the inaccuracy measure $I(t)$.

Theorem: 3.1

Let X and Y be two non-negative continuous random variables with distribution functions $F(x)$ and $G(x)$ respectively and $I(t)$ be as defined in equation (2.54). Further assume that $I(t)$ is independent of t for all $t > 0$. Then $F(x)$ is exponential if and only if $G(x)$ is exponential.

Proof

Let $I(t) = c$, where c is a positive constant.

Further assume that $F(x)$ is exponential with survival function

$$\overline{F}(x) = e^{-\alpha x}; \alpha > 0, x \geq 0. \tag{3.2}$$

Using the relation (2.56) namely

$$h_1(t) = \frac{I^{'}(t) + h_2(t)}{I(t) + \log h_2(t)} \tag{3.3}$$

we get

$$\alpha\left(c + \log h_2(t)\right) = h_2(t).$$

Differentiating the above equation with respect to t, we get

$$h_2^{'}(t)\left(\alpha\left(h_2(t)\right)^{-1} - 1\right) = 0.$$

This gives either $h_2^{'}(t) = 0$ or $h_2(t) = \alpha$. In either case $h_2(t) =$ a constant. Since the constancy of hazard rate is characteristic to the exponential distribution, one can conclude that $G(x)$ is also exponential.

Conversely assume $\overline{G}(x) = e^{-\beta x}, \beta > 0.$

From equation (3.3)

$$h_1(x) = \frac{\beta}{c + \log \beta}.$$

Using equation (2.6) we get

$$\overline{F}(x) = \exp\left(\frac{-\beta x}{c + \log \beta}\right).$$

This shows that $F(x)$ is exponential.

Remark: 3.1

If $\overline{F}(x)$ and $\overline{G}(x)$ are the survival functions of two random variables following the exponential distribution, it is immediate that $\overline{G}(x)=\left[\overline{F}(x)\right]^{\theta}$ for some θ. In other words $\overline{G}(x)$ is the proportional hazards model of $\overline{F}(x)$ and in this situation $I(t)$ is constant. However, it is not necessary that $\overline{G}(x)$ is the proportional hazards model of $\overline{F}(x)$ or $F(x)$ is exponential for $I(t)$ to be a constant, as the next result shows.

Theorem: 3.2

For the random variables X and Y considered in Theorem 3.1, assume that $I(t)$, defined in equation (2.54), is independent of t for all $t>0$. Then F has the finite range distribution specified by the survival function

$$\overline{F}(x)=\left(\frac{c+\log\alpha-\log x}{c+\log\alpha-\log k}\right)^{\alpha},\ \frac{\alpha}{k}e^{c}<X<\alpha e^{c},$$

if and only if G has the Pareto distribution with survival function

$$\overline{G}(x)=\left(\frac{k}{x}\right)^{\alpha},\ x>k>0,\ \alpha>0.$$

The proof of the theorem is analogous to that of Theorem 3.1 and hence omitted.

Remark: 3.2

The difference between the models used in Theorem 3.1 and Theorem 3.2 , both giving constant inaccuracy, is that where as in Theorem: 3.1 the decomposition (3.1) yields constant values for $H(F;t)$ and $D(F,G;t)$ while these measures are functions of x in Theorem:3.2.

A random variable X has the Generalized Pareto Distribution (GPD) if its survival function has the form [Lai and Xie (2006)]

$$\overline{F}(x) = \left(1 + \frac{ax}{b}\right)^{-(1+\frac{1}{a})}, \; x > 0, a > -1, b > 0. \tag{3.4}$$

The importance of this distribution in reliability modeling lies in the fact that it has a linear mean residual life in the form $m(x) = b + ax$. Further the family is rich in the sense that it contains the Lomax distribution $(a > 0)$, rescaled beta $(-1 < a < 0)$, the exponential $(a \to 0)$ and the uniform distribution. Hall and Wellner (1981) have established that the Generalized Pareto Distribution is the only family of distributions that has linear mean residual life function.

The next theorem provides a characterization result for the family of distributions specified in equation (3.4) based on a functional form for the inaccuracy measure.

Theorem: 3.3

Let X and Y be two non-negative continuous random variables with distribution functions $F(x)$ and $G(x)$ respectively and $I(t)$ denote the truncated inaccuracy measure. Assume that $I(t)$ is a linear function of t. Then X follow the generalized Pareto distribution if and only if Y follow the exponential distribution.

Proof

Assume that $I(t) = a + bt$, $b \neq 0$ and that Y follow the exponential distribution with parameter α.

Using equation (3.3), we get

$$h_1(t) = \frac{b + \alpha}{a + bt + \log \alpha}.$$

Using equation (2.6) namely

$$\overline{F}(x) = \exp\left(-\int_0^x h_1(t)dt\right),$$

we get

$$\overline{F}\left(x\right) = \left(1 + \frac{bx}{a + \log\alpha}\right)^{-(1+\frac{\alpha}{b})}$$

$$= \left(1 + \frac{bx}{A}\right)^{-(1+\frac{\alpha}{b})}, \text{ where } A = a + \log\alpha\ . \tag{3.5}$$

Hence X follows the generalized Pareto distribution.

Conversely let X follow the generalized Pareto distribution with survival function (3.4). Then by direct computations we get

$$h_1(x) = \frac{b+\alpha}{A+bx}.$$

Equation (3.3) now becomes

$$\frac{b+\alpha}{A+bt} = \frac{b + h_2(t)}{a + bt + \log h_2(t)}$$

or

$$\frac{a+bt+\log\alpha}{b+\alpha} = \frac{a+bt+\log h_2(t)}{b+h_2(t)}. \tag{3.6}$$

Since the last equation holds for all $t > 0$ and the left side is linear in t, the right side also must be linear. Thich is possible only when $h_2\left(t\right) =$ a constant. Thus G is exponential.

Note:

When G follows the exponential distribution with parameter α and F follows the exponential distribution with parameter $\dfrac{\alpha-1}{\log\alpha}$ then $I(t) = m(t) = \dfrac{\log\alpha}{\alpha-1}$, where $m(t)$ is the mean residual life function of F.

Regarding the inaccuracy measures for past life, defined in (2.55), one can obtain similar results. In the next theorem we look into the form of $F(x)$ and $G(x)$ when the truncated inaccuracy measure for past life, $I^*(t)$ is independent of t.

Theorem: 3.4

For the random variables X and Y considered in Theorem 3.3, assume that $I^*(t)$, defined in equation (2.55) is independent of t for all $t>0$. Then Y follows the power distribution specified by

$$G(x)=x^c,\ 0<x<1,\ c>0, \tag{3.7}$$

if and only if X has a finite range distribution given by

$$F(x)=\left(1-\frac{\log x}{\log c+k}\right)^{-c},\ 0<x<1. \tag{3.8}$$

Proof

Let $I^*(t)=k$, where k is a constant and that Y be distributed as in equation (3.7). From equation (2.57), we get

$$\lambda_1(x)=c\left[x\left(\log c-\log x+k\right)\right]^{-1}.$$

This gives

$$F(x)=\exp\left(-\int_x^1 \lambda_1(t)dt\right)$$

$$=\exp\left(-\int_x^1 \frac{c}{t\left(\log\left(\frac{c}{t}\right)+k\right)}dt\right). \tag{3.9}$$

Take $\log\left(\frac{c}{t}\right)=u$, then equation (3.9) becomes

$$F(x) = \exp\left(-\int_{\log c}^{\log\left(\frac{c}{x}\right)} \frac{c}{u+k} du \right) . \tag{3.10}$$

Solving equation (3.10), we get equation (3.8). The proof of the converse is analogous to that of Theorem 3.3 and hence omitted.

3.3 Characterization of probability distributions under the proportional hazards model assumption

In this section we look into the problem of characterization of probability distributions by the form of the inaccuracy measure under the assumption that $\overline{F}(x)$ and $\overline{G}(x)$ satisfy the condition for being a proportional hazards model. Assume that

$$\overline{G}(x) = \left(\overline{F}(x)\right)^{\theta},\ \theta > 0 .$$

Then equation (2.56) becomes

$$\left(I(t) + \log(\theta h_1(t))\right) h_1(t) = I^{'}(t) + \theta h_1(t)$$

or

$$I^{'}(t) = h_1(t)\left(I(t) + \log(\theta h_1(t)) - \theta\right). \tag{3.11}$$

Multiplying both sides by $\overline{F}(t)$, the above equation can be written as

$$\frac{d}{dt}\left(\overline{F}(t) I(t)\right) = f(t)\left(\log(\theta h_1(t)) - \theta\right) . \tag{3.12}$$

Integrating equation (3.12) over the range (t, ∞), we get

$$I(t) = -\int_{t}^{\infty} \frac{f(x)}{\overline{F}(t)} \left(\log\left(\theta h_1(x)\right) - \theta\right) dx .$$

Hence, if $(Y, \overline{G})$ is the proportional hazards model of $(X, \overline{F})$, we have the relationship

$$I(t) = E\left(\left(\theta - \log\left(\theta h_1(x)\right)\right) \mid X > t\right). \tag{3.13}$$

Remark: 3.3

By direct calculation, the modified Kullback-Leibler divergence measure $D(F,G;t)$, discussed in Chapter 2, is a constant under the proportional hazards model assumption. Hence, from the decomposition of $I(t)$ in equation (3.1), the inaccuracy in the proportional hazards model at different time points varies with

$$H(t) = 1 - E\left(\log h(x) \mid X > t\right), \tag{3.14}$$

only. Consequent to this and the characterization of proportional hazard models by the constancy of $D(F,G;t)$ given in Ebrahimi and Kirmani (1996,b), we can write

$$I(t) = H(t) + K(\theta), \tag{3.15}$$

where $K(\theta)$ is a constant, independent of t. From equation (3.15), we have

$$I'(t) = H'(t).$$

Hence $H(t)$ is an increasing function of t if and only if $I(t)$ is increasing in t. Belzunce et al. (2004) proved that an increasing $H(t)$ determines $F(t)$ uniquely. Thus, we have the following result.

Theorem: 3.5

If X has an absolutely continuous distribution function $F(x)$ and an increasing inaccuracy measure $I(t)$, then $F(x)$ is uniquely determined by $I(t)$.

Remark: 3.4

Functional form of $I(t)$ characterizing various continuous distributions is given in Table 3.1 given below.

Table 3.1

Distribution	$\overline{F}(x)$	$I(t)$
Generalized Pareto (exponential, Lomax, re-scaled beta)	$\left(1+\frac{\text{ax}}{\text{b}}\right)^{-\left(1+\frac{1}{a}\right)};\ x>0$	$\log(\text{b+at})+(\text{a+1})^{-1}-\log(a+1)+A(\theta)$
Power (Uniform c=1)	$1-x^c,\ 0<x<1.$	$\frac{c-1}{c}+\log\left(\frac{1-t^c}{c}\right)+\frac{(c-1)t^c}{1-t^c}\log t+A(\theta)$
Burr XII	$(1+x^c)^{-k},\ x>0.$	$k^{-1}\left[\log(kc)-c^{-1}\log(1+t^c)+\frac{c-1}{c}\sum_{r=1}^{k-1}\frac{(-1)^{r-1}}{(1+t^c)^k}+A(\theta)\right]$
Exponential geometric(Adamilis&Loukas (1998))	$(1-p)e^{-\lambda x}\left(1-pe^{-\lambda x}\right)^{-1}$	$2-\log\lambda+p^{-1}e^{\lambda t}\log\left[(1-p)e^{-\lambda t}\right].$

In Table 3.1, $A(\theta)=\theta-1-\log\theta$.

Di Cresenzo and Longobardi (2004) has characterized proportional reversed hazards model (See Section 2.4) using Kullback-Leibler measure of discrimination between past lifetime distributions. In this context, the Kullback-Leibler divergence for past life simplifies to

$$D^*(F,G;t)=\phi-1-\log\phi,$$

which is independent of t. Thus the inaccuracy measure of past life is

$$I^*(t)=\phi-1-\log\phi+\overline{H}(F;t)\ , \tag{3.16}$$

where $\overline{H}(F;t)$ is the past entropy as defined in (2.35). Equation (3.16) can also be written as

$$I^*(t)=E\left(\left(\phi-\log\left(\phi\lambda_1(x)\right)\right)\mid X\le t\right). \tag{3.17}$$

Equation (3.17) can now be written as

$$I^*(t)=\int_0^t \frac{f(x)}{F(t)}\left(\phi-\log\left(\phi\lambda_1(x)\right)\right)dx.$$

Differentiating the above expression and rearranging the terms, we have the relationship

$$\lambda_1(t)\left(I^*(t)+\log\left(\phi\lambda_1(t)\right)-\phi\right)=-I^{*'}(t). \tag{3.18}$$

Arguing as in the Theorem 3.5 and using equation (3.18), we have the following theorem.

Theorem: 3.6

If X has an absolutely continuous distribution function $F(x)$ and an increasing inaccuracy measure $I^*(t)$, then $I^*(t)$ uniquely determines $F(x)$.

Remark: 3.5

For the power distribution specified by

$$F(x)=x^c, 0<x<1, c>0,$$

by direct computation using equation (3.16), we get

$$I^*(t)=\log t-c^{-1}-\log(c\phi)+\phi. \tag{3.19}$$

Observing that the above equation is increasing in t, the functional form of $I^*(t)$ given in equation (3.19) holds if and only if X follow the power distribution.

We now characterize some common failure time distributions using a possible relationship between $I(t)$ and the mean residual life function $m(t)$, reviewed in Chapter 2.

Theorem: 3.7

If X is a non-negative random variable admitting an absolutely continuous distribution function F and $(Y,\overline{G})$ is the proportional hazards model of $(X,\overline{F})$, then the relationship

$$I(t) = A(\theta) + \log m(t), \tag{3.20}$$

where $A(\theta)$ is a real valued function of θ, independent of t, and $m(t)$ is the mean residual life function of X holds for every $t > 0$ if and only if X has the generalized Pareto distribution.

Proof

Under the conditions of the theorem, when X has generalized Pareto distribution, using equation (3.11), we get

$$I(t) = (a+1)^{-1} - \log(a+1) + \theta - 1 - \log\theta + \log(b+at),$$

which is of the form (3.20).

Conversely let $I(t)$ be as in equation (3.20).

Differentiating equation (3.20) with respect to t, we get

$$I'(t) = \frac{m'(t)}{m(t)}. \tag{3.21}$$

Further using equations (3.11), (3.20) and (3.21), we have

$$\frac{m'(t)}{m(t)} = h_1(t)\left(C(\theta) + \log(h_1(t)m(t))\right), \tag{3.22}$$

where $C(\theta) = A(\theta) + \log\theta - \theta$.

Using the relation, $h(t)m(t) = 1 + m'(t)$, equation (3.22) becomes

$$m'(t) = \left(1 + m'(t)\right)\left(C(\theta) + \log\left(1 + m'(t)\right)\right).$$

Differentiating with respect to t and rearranging the terms we get

$$m^{''}(t)\left(C(\theta)+\log\left(1+m^{'}(t)\right)\right)=0, \tag{3.23}$$

where $m^{''}(t)=\dfrac{dm^{'}(t)}{dt}$. Equation (3.23) implies if either $m^{''}(t)=0$ or $m^{'}(t)=e^{-C(\theta)}-1$. In both the cases $m^{'}(t)$ is a constant. This shows that $m(t)$ is linear in t. The rest of the proof follows from Hall and Wellner (1981).

Theorem: 3.8

Under the conditions of the Theorem 3.7, the relationship

$$I(t)=K(\theta)-\log h_1(t), \tag{3.24}$$

where $K(\theta)$ is a real valued function of θ and $h_1(t)$ is the hazard rate of F holds if and only if F is generalized Pareto distribution with survival function (3.4).

Proof

The 'if' part of the theorem follows from the expression for $I(t)$ given in equation (3.11). To prove the only if part, assume that equation (3.24) holds. Differentiating equation (3.24) with respect to t, we get

$$I^{'}(t)=-\frac{h_1^{'}(t)}{h_1(t)}. \tag{3.25}$$

Using equations (3.24) and (3.25), in equation (3.11), we get

$$-\frac{h_1^{'}(t)}{\left(h_1(t)\right)^2}=A(\theta), \tag{3.26}$$

where $A(\theta)$ is a function of θ, independent of t. Equation (3.26) can be written as

$$\frac{d}{dt}\left(\frac{1}{h_1(t)}\right)=A(\theta).$$

The solution of the above differential equation is

$$h_1(t) = \frac{1}{Bt + C},$$

where $B = A(\theta)$ and $C^{-1} = h_1(0)$. This is the hazard rate of the generalized Pareto distribution. Since the distribution function is uniquely determined by the hazard rate, the theorem follows.

A parallel result exists for $I^*(t)$ which we state as follows.

Theorem: 3.9

Let X be a non-negative random variable admitting an absolutely continuous distribution function F and let (Y, G) is the proportional reversed hazards model of (X, F). Then the relationship

$$I^*(t) = K(\phi) - \log \lambda_1(t), \tag{3.27}$$

where $K(\phi)$ is a real function of ϕ and $\lambda_1(t)$ is the reversed hazard rate of X holds for all real $t \geq 0$ if and only if X has power distribution with distribution function specified by equation (3.7).

Proof

For the power distribution

$$I^*(t) = \log t - c^{-1} - \log(c\phi) + \phi,$$

which is of the form equation (3.27) with $K(\phi) = \phi - c^{-1} - \log \phi$ and $\lambda_1(t) = ct^{-1}$. This proves the 'if' part.

Conversely assume that the relation (3.27) holds. Differentiating the equation (3.27) with respect to t, we get

$$I^{*'}(t) = -\frac{\lambda_1'(t)}{\lambda_1(t)}. \tag{3.28}$$

From equations (3.28) and (3.18), we get

$$\frac{\lambda_1^{'}(t)}{\left(\lambda_1(t)\right)^2} = K(\phi) - \phi + \log\phi . \tag{3.29}$$

Equation (3.29) can be written as

$$\frac{-d}{dt}\left(\frac{1}{\lambda_1(t)}\right) = K(\phi) - \phi + \log\phi . \tag{3.30}$$

Solving the differential equation (3.30), we obtain

$$\frac{1}{\lambda_1(t)} = At + B ,$$

where $A = \phi - \log\phi - K(\phi)$. This gives $\lambda_1(t) = \frac{1}{At+B}$, using equation (2.12), we conclude that X follows power distribution.

The next theorem focus attention on a characterization result for the Gompertz distribution by the form of $I(t)$ in terms of the vitality function, reviewed in Section 2.1.

Theorem: 3.10

Let X be a non-negative random variable admitting an absolutely continuous distribution function F and with hazard rate $h_1(t)$ and let G be the proportional hazard model of F. Then X has the Gompertz distribution with survival function

$$\overline{F}(x) = \exp\left(\frac{-B}{\log C}\left(C^x - 1\right)\right); x > 0, B > 0, C > 0, \tag{3.31}$$

if and only if

$$I(t) = K(\theta) + \beta v(t), \tag{3.32}$$

for some real function $K(\theta)$ and a real constant $\beta < 0$.

Proof

By direct calculation using equation (3.31) we get $h_1(t) = BC^t$ and

$$H(t) = 1 - \left(\overline{F}(t)\right)^{-1} \int_t^{\infty} \log h_1(x) f(x)\, dx$$
$$= 1 - \left(\overline{F}(t)\right)^{-1} \int_t^{\infty} \left(\log B + x \log C\right) f(x)\, dx .$$

Using the equation (3.15) and the last equation, we get the form of $I(t)$ as stated in the theorem.

Conversely, we assume that equation (3.32) holds. Differentiating equation (3.32) with respect to t, we get

$$I^{'}(t) = h_1(t)\left(I(t) - K(\theta) - \beta t\right). \tag{3.33}$$

Using equation (3.33) in equation (3.11), we have

$$\log h_1(t) = \alpha - \beta t , \tag{3.34}$$

where, $\alpha = \theta - \log\theta - K(\theta)$. Equation (3.34) can be written as $h_1(t) = e^{\alpha-\beta t} = BC^t$, where $B = e^{\alpha}$ and $C = e^{-\beta}$.This is the hazard rate of the Gompertz distribution. Since the hazard rate uniquely determines the distribution, X follows Gompertz distribution.

Roy and Mukharjee (1989) examined the concept of averaging of hazard rate and looked into the problem of the characterization of life distributions using this concept . When one is interested in the pattern of failure of a device in a finite interval, instead of examining the nature of failure at each point in the interval this concept become a handy tool to describe the failure pattern. Rajesh (2001) has obtained characterization results for some lifetime distributions using the residual entropy function and the averages of hazard rate. The arithmetic, geometric and harmonic mean of hazard rates for a non-negative random variable Y with hazard rate $h_2(t)$ are defined as

$$A^*(x) = \frac{1}{x}\int_0^x h_2(t)\,dt\,,$$

$$G^*(x) = \exp\left(\frac{1}{x}\int_0^x \log h_2(t)\,dt\right), \tag{3.35}$$

and

$$H^*(x) = \frac{1}{x}\int_0^x \frac{1}{h_2(t)}\,dt\,.$$

We now look into the problem of characterization of distributions based on the functional form of $A^*(x)$, $G^*(x)$ and $H^*(x)$ in terms of the residual inaccuracy measure.

Theorem: 3.11

Let X and Y be two non-negative random variables admitting absolutely continuous distribution functions such that $(Y,\overline{G})$ is the proportional hazards model of $(X,\overline{F})$. Denote by $A^*(x)$, $G^*(x)$ and $H^*(x)$ are arithmetic, geometric and harmonic mean of hazard rates of Y and let $I(t)$ the residual inaccuracy function. The relationship

$$A^*(t) = G^*(t) = H^*(t) = \exp(\theta - I(t)), \tag{3.36}$$

holds for all real $t > 0$ if and only if X follows the exponential distribution.

Proof

Assume equation (3.36) holds. This gives

$$I(t) + \log G^*(t) = \theta\,. \tag{3.37}$$

Using equation (3.35), equation (3.37) can be written as

$$t\,I(t) + \int_0^t \log h_2(x)\,dx = \theta\,t\,. \tag{3.38}$$

Differentiating with respect to t and using equations (3.11), (3.38) simplifies to

$$I(t)+\log\left(\theta h_1(t)\right)=\theta\,. \tag{3.39}$$

The equation (3.39) can be written as

$$I^{'}(t)+\frac{h^{'}_1(t)}{h_1(t)}=0\,. \tag{3.40}$$

Using equation (3.11), equation (3.40) becomes

$$h^{'}_1(t)=0\,.$$

This gives

$h_1(t)=\lambda$, a constant.

Since the constancy of hazard rate is characteristic to the exponential model, the distribution of X is exponential. From Roy and Mukharjee (1989), the properties $A^*\left(t\right)=G^*\left(t\right)=H^*\left(t\right)$ is characteristic to the exponential model. Hence the sufficiency part holds. Conversely when X follows exponential distribution with parameter λ, by direct calculations we get

$$A^*\left(t\right)=G^*\left(t\right)=H^*\left(t\right)=\lambda\theta$$

and

$$I(t)=\theta-\log\lambda\theta\,.$$

The validity of equation (3.36) can be verified from the above expressions.

Theorem: 3.12

Assume that the conditions of the Theorem 3.11 holds. The relationship

$$I(t)-\left(\frac{a}{\left(a+1\right)\theta}\right)t\,A^*\left(t\right)=k\ , \tag{3.41}$$

where k is a constant, holds for all $t>0$, if and only if F has generalized Pareto distribution with survival function (3.4).

Proof

By direct calculation using equation (3.4), we get,

$$I(t)=\log\left(\frac{b+at}{b}\right)+\left(\frac{a}{(a+1)}+\theta-\log\left(\frac{(a+1)\theta}{b}\right)\right).$$

That is,

$$I(t)=\log\left(\frac{b+at}{b}\right)+k\,, \tag{3.42}$$

where $k=\frac{a}{a+1}+\theta-\log\left(\frac{(a+1)\theta}{b}\right)$, is independent of t. From the definitions of arithmetic mean of hazard rates and the proportional hazards model assumption, we have

$$t\,A^{*}(t)=\int_{0}^{t}\theta h_{1}(x)\,dx\,. \tag{3.43}$$

Since X follows generalized Pareto distribution, we get $h_1(t)$ as

$$h_{1}(t)=\frac{a+1}{b+at}\;. \tag{3.44}$$

Using equation (3.44) in equation (3.43), we have

$$t\,A^{*}(t)=\left(\frac{a+1}{a}\right)\theta\log\left(\frac{b+at}{b}\right)\;. \tag{3.45}$$

Using equations (3.42) in (3.45) we get (3.41).

Conversely assume equation (3.41) holds. Differentiating this equation with respect to t, we get

$$I^{'}(t)=\left(\frac{a}{(a+1)\theta}\right)\left(tA^{*'}(t)+A^{*}(t)\right). \tag{3.46}$$

But from the definition of $A^{*}(x)$ given in equation (3.35) and using the assumption of the theorem, we have the relationship

$$t\,A^{*'}(t)+A^{*}(t)=\theta\,h_1(t)\,. \tag{3.47}$$

Using the equation (3.47), equation (3.46) becomes

$$I^{'}(t)=\left(\frac{a}{a+1}\right)h_1(t). \tag{3.48}$$

From equations (3.11) and (3.48) ,we have

$$I(t)+\log\left(\theta h_1(t)\right)-\theta=\frac{a}{(a+1)}. \tag{3.49}$$

Using equation (3.41) in equation (3.49), we get

$$\left(\frac{at}{(a+1)\theta}\right)A^{*}(t)+\log h_1(t)=k_1\,,$$

which is equivalent to

$$\frac{a}{a+1}\int_0^t h_1(x)dx+\log h_1(t)=k_1\ , \tag{3.50}$$

where $k_1=\dfrac{a}{a+1}+\theta-\log\theta-k$.

Differentiating equation (3.50) with respect to t we get

$$\frac{h_1^{'}(t)}{\left(h_1(t)\right)^2}=-\frac{a}{a+1}.$$

The solution of the above equation is $h_1(t)=\left(c+\dfrac{at}{a+1}\right)^{-1}$, where $c>0$ is the constant of integration. This gives $\overline{F}(t)=\left(1+\dfrac{at}{c(a+1)}\right)^{-\left(1+\frac{1}{a}\right)}$, which is the survival function of generalized Pareto distribution. This completes the proof of the sufficiency part.

3.4 Inaccuracy measure for weighted distributions

As mentioned in Section 2.4, the weighted distributions, defined by Rao (1965), finds a lot of applications in theoretical statistics. The simplest form of the weighted distribution is the length-biased distribution defined in equation (2.24) namely

$$f_L(x) = \frac{x\, f(x)}{\mu}, \; x > 0, \;\; \mu = E\left(w(X)\right) < \infty. \tag{3.51}$$

The inaccuracy for the length-biased random variable X_L associated to a non- negative random variable X is,

$$I_L = -\int_0^\infty f(x) \log f_L(x)\, dx.$$

Using equation (3.51), the above equation can be written as

$$I_L = -\int_0^\infty f(x) \log\left(\frac{x\, f(x)}{\mu}\right) dx\,;\, x > 0, \mu = E(X) < \infty$$

$$= -\int_0^\infty f(x) \log(f(x)) dx - \int_0^\infty f(x) \log x\, dx + \log \mu$$

$$I_L = H(F) - E\left(\log X\right) + \log\left(E\left(X\right)\right). \tag{3.52}$$

The equation (3.52) shows that the inaccuracy of the length-biased variable can be expressed in the form

$I_L =$ Entropy - Geometric mean of X - logarithm of arithmetic mean of X .

The above equation expresses the inaccuracy for the length-biased random variable, X_L as the difference between entropy and the sum of the geometric mean and the logarithm of the arithmetic mean of X .

The measure of residual inaccuracy, when we assume length-biased model instead of actual density can be expressed in terms of residual inaccuracy, geometric vitality function and mean residual life function of original distribution as follows. By definition

$$I_L(t) = -\int_t^{\infty} \frac{f(x)}{\overline{F}(t)} \log\left(\frac{f_L(x)}{\overline{F}_L(t)}\right) dx$$

$$= -\int_t^{\infty} \frac{f(x)}{\overline{F}(t)} \log\left(\frac{xf(x)}{\mu \overline{F}_L(t)}\right) dx$$

$$= M(t) - \log\left(G(t)\right) + \log\left(\mu \overline{F}_L(t)\right), \tag{3.53}$$

where $\log\left(G(t)\right) = \frac{1}{\overline{F}(t)} \int_t^{\infty} f(x) \log x \, dx$ is the geometric vitality function (See equation (2.19)) and $M(t) = -\frac{1}{\overline{F}(t)} \int_t^{\infty} f(x) \log f(x) \, dx$, is the conditional measure of uncertainty proposed by Sankaran and Gupta (1999),which measures the uncertainty contained in $f(t)$ about the predictability of the total lifetime of a unit which has survived to age t. $M(t)$ can be represented as the sum of residual entropy and total hazard rate as

$$M(t) = H(t) - \log \overline{F}(t) .$$

Using the above representation for $M(t)$, equation (3.53) becomes,

$$I_L(t) = H(t) - \log[G(t)] + \log\left(\mu\left(\frac{\overline{F}_L(t)}{\overline{F}(t)}\right)\right) . \tag{3.54}$$

Using equation (2.25), the above equation can be written as,

$$I_L(t) = H(t) - \log\left(G(t)\right) + \log\left(t + m(t)\right). \tag{3.55}$$

But using the relationship between the mean residual life function and the vitality function, given in equation (2.18), the above expression can be rewritten as

$$I_L(t) = H(t) - \log G(t) + \log v(t),$$

where $v(t)$ is the vitality function given in equation (2.17).

Another model of interest in lifetime analysis is the equilibrium distribution discussed in Section 2.4. For the equilibrium distribution, the inaccuracy measure takes the form

$$I_E(t) = -\frac{1}{\overline{F}(t)}\int_t^\infty f(x)\left(\log \overline{F}(x) - \log\left(\overline{F}(t)m(t)\right)\right)dx.$$

$$= 1 + \log\left(m(t)\overline{F}(t)\right) - \log \overline{F}(t)$$

$$= 1 + \log m(t),$$

where $m(t)$ is the mean residual life function. Since $m(t)$ determines the distribution uniquely, $I_E(t)$ determines the distribution F.

In fact

$$\overline{F}(x) = -\exp\left(-\int_0^x \exp(1 - I_E(t))dt\right)\exp\left(1 - I_E(x)\right) I_E^{'}(x). \tag{3.56}$$

Equation (3.56) enables one to characterize distributions by the functional form of $I_E(t)$. The form of $I_E(t)$ which characterizes some distributions is given in the following table.

Table 3.2

Distribution	$\overline{F}(x)$	$I_E(t)$
Generalized Pareto	$\left(1+\frac{ax}{b}\right)^{-\left(1+\frac{1}{a}\right)}, x>0.$	$1+\log(b+at)$
Power	$1-x^c,\ 0<x<1.$	$1+\log\left(1-t^c\right)^{-1}\left[1-t-(c+1)^{-1}\left(1-t^{c+1}\right)\right]$
Gamma	$\frac{\lambda^\alpha e^{-\lambda t}t^{\alpha-1}}{\Gamma\alpha},\ x>0.$	$1+\log\left[\frac{\alpha}{\lambda}-t-\frac{\lambda^{\alpha-1}e^{-\lambda t}t^\alpha}{\overline{F}(t)\Gamma\alpha}\right]$
Exponential geometric	$(1-p)e^{-\lambda x}\left(1-pe^{-\lambda x}\right)^{-1}, x>0$	$1+\log\left[(-\lambda p)^{-1}e^{\lambda t}\left(1-pe^{-\lambda t}\right)\log\left(1-pe^{-\lambda t}\right)\right]$

The characterizations considered in this chapter provide tools for identifying life distributions in terms of different forms of measure of inaccuracy, besides forging relationships between inaccuracy function and basic reliability characteristics. A major difference between the characterizations in this chapter and those in terms of reliability functions currently in use is that, in former, we get a feel of the extent to which the assumed model is inaccurate both in terms of lack of information and mis-specification.

Chapter 4

THE GENERALIZED INACCURACY MEASURE

4.1 Introduction

In this chapter, we discuss a generalization for the concept of inaccuracy, discussed in Chapter three, analogous to the generalization of Shannon's entropy given in Rao (1965), Belzuance et al. (2004) and Nanda and Paul (2006).

Recently Nair and Gupta (2007) has extended the inaccuracy measure defined in equation (2.52) to the truncated situation in the form

$$I(F,G;t) = -\int_t^\infty \frac{f(x)}{\overline{F}(t)} \log\left(\frac{g(x)}{\overline{G}(t)}\right) dx,$$

and has provided characterization results for some well known life time distributions. Motivated by this, we also look into the problem of characterization of probability distributions, using the functional form of the truncated version of the generalized inaccuracy measure.

4.2 Definition and properties

For a non negative random variable X, admitting an absolutely continuous distribution function, Khinchin (1957) has generalized the Shannon's entropy defined by equation (2.32) in the form,

$$H_\phi(F) = \int_0^\infty f(x)\phi(f(x))\, dx, \tag{4.1}$$

where $\phi(.)$ is a convex function satisfying the condition $\phi(1) = 0$. Analogously inaccuracy measure defined in equation (2.52) can be modified as

$$I\left(F,G\right)=\int_{0}^{\infty} f\left(x\right)\left(\phi\left(g\left(x\right)\right)\right)dx. \tag{4.2}$$

Taking $\phi\left(g\left(x\right)\right)=\dfrac{1-g^{r}\left(x\right)}{r}$, equation (4.2) reduces to

$$I_{r}(F,G)=\frac{1}{r}\int_{0}^{\infty} f(x)\left(1-g^{r}\left(x\right)\right)dx\,;\ r>-1,\ r\neq 0\,. \tag{4.3}$$

Our generalization of inaccuracy measure defined in equation (4.3), conforms to the spirit of the extension given in equation (4.1), provided that r is so chosen such that $\dfrac{1-g^{r}\left(x\right)}{r}$ is a convex function. In equation (2.52), all the experimental events contributing to the evaluation of inaccuracy are assigned equal probabilities. This means that events with high and low probabilities have equal weight. It would be more reasonable to have higher probability for events, which impart more sensitivity to $I\left(F,G\right)$ than those with lesser probabilities. This is achieved through our generalization given in equation (4.3). Note that as $r\to 0$, equation (4.3) reduces to equation (2.52).

In Section 4.2, we present some properties of $I_{r}(F,G)$. Characterization results for probability distributions in the context of Cox's proportional hazards model and proportional reversed hazards model are discussed in Section 4.3. In Section 4.4., we consider the generalized inaccuracy measure proposed by Nath (1968) and discuss characterization results for probability distributions using the functional form of this inaccuracy measure.

Properties

(i) The expression for $I_{r}(F,G)$ given in equation (4.3) can be decomposed as the sum of two terms in the form

$$I_{r}(F,G)=H_{r}\left(F\right)+\frac{1}{r}\int_{0}^{\infty} f(x)\left(f^{r}\left(x\right)-g^{r}\left(x\right)\right)dx\,. \tag{4.4}$$

In the above equation, the first term is the generalization of $H(F)$, the Shannon's entropy, given in equation (2.32) and the second term reduces to the Kullback-Leibler divergence measure defined by equation (2.47), as $r \to 0$. Hence equation (4.4) enables one to express the inaccuracy measure as the sum of a measure of uncertainty about F and a measure of discrimination between the distributions.

(ii) $I_r(F,G)$ is minimum, when $f(x)=g(x)$. This is immediate since when $f(x)=g(x)$, $I_r(F,G)$ simplifies to $H_r(F)$, which is the minimum value. Further the error term in equation (4.4) is zero and $I_r(F,G)$ reduces to the generalization of Shannon's entropy measure given in Belzuance et al. (2004), namely

$$\begin{aligned} I_r(F) &= \frac{1}{r}\int_0^\infty f(x)\left(1-f^r(x)\right)dx \\ &= H_r(F). \end{aligned}$$

(iii) $I_r(F,G)=0$ implies $E_f\left(g^r(x)\right)=1$.

This result follows directly from the definition (4.3).

In many practical situations, complete data may not be observable to the experimenter due to various reasons. For instance in lifetime studies, the interest may center around the life time of a unit after a specified period of time, say t. Observing that the probability density function of the random variable $X-t \mid X>t$ and $Y-t \mid Y>t$ respectively $\frac{f(t+x)}{\bar{F}(t)}$ and $\frac{g(t+x)}{\bar{G}(t)}$, the generalized inaccuracy measure in the truncated situation takes the form

$$I_r(F,G;t)=\frac{1}{r}\int_t^\infty \frac{f(x)}{\bar{F}(t)}\left(1-\left(\frac{g(x)}{\bar{G}(t)}\right)^r\right)dx. \tag{4.5}$$

For convenience in notation, we denote $I_r(F,G;t)$ by $I_r(t)$ in the sequel. Differentiating equation (4.5) with respect to t, and using the definition of hazard rate given in chapter two, one can have the representation

$$1-rI_r(t)=\frac{h_1(t)h_2^r(t)-rI_r^{'}(t)}{rh_2(t)+h_1(t)}, \tag{4.6}$$

where $h_1(t)$ and $h_2(t)$ are the hazard rates associated with $F(x)$ and $G(x)$ respectively. Equation (4.6) reveals that the functional form of $I_r(t)$ can mutually characterize $F(x)$ and $G(x)$.

In reliability studies it may happen that the life time may not be observable beyond a specified time point t. Hence the distributions of X and Y may be truncated to the right and one can look at the past life by taking X and Y in the interval $(0,t]$. Here the random variables under consideration are $X_t^*=t-X\,|\,X<t$ and $Y_t^*=t-Y\,|\,Y<t$. In this situation the generalized inaccuracy measure simplifies to

$$I_r^*(t)=\frac{1}{r}\int_0^t\frac{f(x)}{F(t)}\left(1-\left(\frac{g(x)}{G(t)}\right)^r\right)dx\,;\;\; 0<x<t. \tag{4.7}$$

Further, equation (4.7) can also be written as

$$1-rI_r^*(t)=\frac{\lambda_1(t)\lambda_2^r(t)+rI_r^{*'}(t)}{r\lambda_2(t)+\lambda_1(t)}, \tag{4.8}$$

where $\lambda_1(t)$ and $\lambda_2(t)$ are the reversed hazard rates of $F(x)$ and $G(x)$ respectively, reviewed in Section 2.1.

When $(Y,\overline{G})$ is the proportional hazards model of $(X,\overline{F})$, equation (4.6) takes the form

$$1-rI_r\left(t\right)=\frac{\theta^r h_1^{r+1}\left(t\right)-rI_r^{'}\left(t\right)}{\left(1+r\theta\right)h_1\left(t\right)}. \tag{4.9}$$

In the context of proportional reversed hazards model considered in Section 2.3, equation (4.8) becomes

$$1-rI_r^*\left(t\right)=\frac{\theta^r \lambda_1^{r+1}\left(t\right)+rI_r^{*'}\left(t\right)}{\left(1+r\theta\right)\lambda_1\left(t\right)}. \tag{4.10}$$

$I_r\left(t\right)$ defined in equation (4.5) provides a measure of discrimination between $F\left(x\right)$ and $G\left(x\right)$, where the observation made beyond time t. In general, $F\left(x\right)$ and $G\left(x\right)$ need not be depend on each other. However when there is some dependence structure between two distributions one can arrive at certain characterization results for probability distributions. An extensively studied dependence structure is the proportional hazards model, reviewed in Section 2.2. In this situation we have

$$\overline{G}(x)=\left(\overline{F}(x)\right)^{\theta},\ \theta>0. \tag{4.11}$$

When $\left(Y,\overline{G}\right)$ is the proportional hazards model of $\left(X,\overline{F}\right)$, the expression for $I_r\left(F,G;t\right)$, given in equation (4.5), takes the form

$$1-rI_r\left(F;t\right)=\frac{\theta^r}{\left(\overline{F}\left(t\right)\right)^{1+r\theta}}\int_t^{\infty}\left(f\left(x\right)\right)^{r+1}\left(\overline{F}\left(x\right)\right)^{r(\theta-1)}dx. \tag{4.12}$$

We first examine whether $I_r\left(F;t\right)$ uniquely determine the distribution. The answer to the question is in the affirmative, which we given as Theorem 4.1.

Theorem: 4.1

Let $F(x)$ and $G(x)$ be absolutely continuous distribution functions such that $(Y,\overline{G})$ is the proportional hazards model of $(X,\overline{F})$. Assume that $I_r(F;t)$ is increasing in t. Then $I_r(F;t)$ uniquely determines $F(t)$.

Proof

Using equation (4.11) in equation (4.5), we get

$$1-rI_r\left(F;t\right)=\frac{\theta^r}{\left(\overline{F}\left(t\right)\right)^{1+r\theta}}\int_t^{\infty}\left(\overline{F}\left(x\right)\right)^{r(\theta-1)}f^{r+1}\left(x\right)dx. \tag{4.13}$$

Suppose that $F\left(x\right)$ and $G\left(x\right)$ are distribution functions such that

$$I_r\left(F;t\right)=I_r\left(G;t\right) \text{ , for all } t\geq 0. \tag{4.14}$$

From equations (4.13) and (4.14)

$$\left(\overline{F}\left(t\right)\right)^{-\left(1+r\theta\right)}\int_t^{\infty}\left(\overline{F}\left(x\right)\right)^{r(\theta-1)}f^{r+1}\left(x\right)dx=\left(\overline{G}\left(t\right)\right)^{-\left(1+r\theta\right)}\int_t^{\infty}\left(\overline{G}\left(x\right)\right)^{r(\theta-1)}g^{r+1}\left(x\right)dx,$$

where $f\left(x\right)$ and $g\left(x\right)$ are the probability density functions corresponding to $F\left(x\right)$ and $G\left(x\right)$. Differentiating the above equation with respect to t and using the definition of hazard rate, we get

$$\theta^r\left(h_1\left(t\right)\right)^{r+1}-\left(1-rI_r\left(f;t\right)\right)\left(1+r\theta\right)h_1\left(t\right)=\theta^r\left(h_2\left(t\right)\right)^{r+1}-\left(1-rI_r\left(g;t\right)\right)\left(1+r\theta\right)h_2\left(t\right). \tag{4.15}$$

where $h_1\left(t\right)$ and $h_2\left(t\right)$ are the hazard rates corresponding to $f\left(x\right)$ and $g\left(x\right)$ respectively. To prove $\overline{F}\left(t\right)=\overline{G}\left(t\right)$, it is enough to show that $h_1\left(t\right)=h_2\left(t\right)$, for all $t\left(\geq 0\right)$.

Suppose

$$h_1\left(t\right)>h_2\left(t\right) \text{ with } h_i\left(t\right)\neq 0;\ i=1,2.$$

From equation (4.15) we have

$$\frac{h_1\left(t\right)}{h_2\left(t\right)}=\frac{\theta^r\left(h_2\left(t\right)\right)^r-\left(1-rI_r\left(G;t\right)\right)\left(1+r\theta\right)}{\theta^r\left(h_1\left(t\right)\right)^r-\left(1-rI_r\left(F;t\right)\right)\left(1+r\theta\right)}>1,$$

or

$$\theta^r\left(h_2(t)\right)^r-\left(1-rI_r(G;t)\right)\left(1+r\theta\right)>\theta^r\left(h_1(t)\right)^r-\left(1-rI_r(F;t)\right)\left(1+r\theta\right).$$

Using equation (4.14), we get

$$h_1(t)<h_2(t),$$

which is a contradiction. Similarly, we can show that the inequality $h_1(t)<h_2(t)$ also leads to a contradiction. This gives

$$h_1(t)=h_2(t).$$

4.3 Characterization results

In this section, we look into the problem of characterization of probability distributions using the functional form of $I_r(t)$. First we examine the situation where $I_r(t)$ is independent of t.

Theorem: 4.2

Let $F(x)$ and $G(x)$ be absolutely continuous distribution functions and $I_r(t)$ be as defined in equation (4.5). If $I_r(t)$ is a positive constant $\left(<\frac{1}{r}\right)$, then $F(x)$ is exponential if and only if $G(x)$ is exponential.

Proof

Let $I_r(t)=c$, where c is a positive constant with $c<\frac{1}{r}$ and that F is the exponential distribution with survival function

$$\overline{F}(x)=e^{-\lambda x};\ \lambda>0,\ \mathrm{x}>0.$$

From equation (4.6) we get

$$1 - rc = \frac{\lambda\, h_2^r(t)}{r\, h_2(t) + \lambda}$$

or

$$\left(1 - rc\right)\left(r\, h_2(t) + \lambda\right) = \lambda\, h_2^r(t).$$

Differentiating the above equation with respect to t we get,

$$h_2^{'}\left(t\right)\left(\lambda\, r\, h_2^{r-1}\left(t\right) - r\left(1 - rc\right)\right) = 0.$$

The solution to the above equation is $h_2\left(t\right) = \beta$, where β is a constant. Hence $G(x)$ is exponential.

Conversely assume

$$\overline{G}\left(x\right) = \exp\left(-\beta x\right), \text{ with } \beta > \left(1 - rc\right)^{\frac{1}{r}} > 0.$$

From equation (4.6) we get

$$h_1\left(t\right) = \frac{r\beta\left(1 - rc\right)}{rc + \beta^r - 1}.$$

Using the relationship

$$\overline{F}\left(x\right) = \exp\left(-\int_0^x h_1\left(t\right) dt\right),$$

we get

$$\overline{F}\left(x\right) = \exp\left(-\frac{r\beta\left(1 - rc\right)}{rc + \beta^r - 1} x\right).$$

From the above expression, $F(x)$ is exponential.

The following theorem provides a characterization result for the generalized Pareto model in the context of proportional hazards model.

Theorem: 4.3

Let $F(x)$ and $G(x)$ be two absolutely continuous distribution functions, $f(x)$ and $g(x)$ be the corresponding probability density functions and $h_1(t)$ be the hazard rate of X. Assume $(Y,\overline{G})$ is the proportional hazards model of $(X,\overline{F})$ then the relationship

$$1-rI_r(t)=\beta(h_1(t))^r, \tag{4.16}$$

where β is a constant holds if and only if $F(x)$ is the generalized Pareto distribution with survival function specified by equation (3.4).

Proof

Under the assumptions of the theorem, when X has generalized Pareto distribution, straight forward computations using equation (4.9), gives

$$1-rI_r(t)=\frac{\theta^r(a+1)\left(\frac{b+at}{a+1}\right)^{-r}}{((a+1)(1+r\theta)+ar)}$$

$$=\beta.\left(\frac{b+at}{a+1}\right)^{-r}, \text{ where } \beta=\frac{\theta^r(a+1)}{((a+1)(1+r\theta)+ar)} \text{ is a constant.}$$

Observing that for the generalized Pareto model, the hazard rate is $h_1(t)=\frac{a+1}{b+at}$, the if part follows.

To prove the converse, differentiating equation (4.16) with respect to t we get

$$-rI_r'(t)=r.\beta.(h_1(t))^{r-1}h_1'(t). \tag{4.17}$$

Using equations (4.9) and (4.16), equation (4.17) can be reads as

$$h_1^{r+1}(t)((1+r\theta)\beta-\theta^r)=r\beta(h_1(t))^{r-1}h_1'(t)$$

or

$$\frac{h_1'(t)}{(h_1(t))^2}=\frac{(1+r\theta)\beta-\theta^r}{r\beta}.$$

That is,

$$-\frac{d}{dt}\left(\frac{1}{h_1(t)}\right)=\frac{(1+r\theta)\beta-\theta^r}{r\beta}.$$

The above equation gives

$$\frac{1}{h_1(t)}=\left(\frac{\theta^r-(1+r\theta)\beta}{r\beta}\right)t+c_2$$

or

$$h_1(t)=\frac{1}{c_1 t+c_2}, \tag{4.18}$$

where, $c_1=\dfrac{\theta^r-(1+r\theta)\beta}{r\beta}$ and $c_2^{-1}=h_1(0)$.

Hall and Wellner (1981) have shown that (4.18) is a characteristic property of the generalized Pareto distribution, specified by equation (3.4). The necessary part follows from this result.

Remarks

(i) As $r \to 0$, the result reduces to the Theorem 3.1 in Nair and Gupta (2007), reviewed in Section 2.8.

(ii) The theorem provides a characterization result for the re-scaled beta distribution specified by

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

when $c=-\left(1+\dfrac{1}{a}\right)$ and $R=-\dfrac{b}{a}$ in equation (3.4).

Further it may be noted that the uniform distribution arises as a special case when $c=1$.

(iii) A characterization result for the Lomax distribution specified by

$$\overline{F}(x)=\beta^c(x+\beta)^{-c};\ x>0, \beta>0\ , c>0,$$

can be obtained when $\beta = \dfrac{a+1}{a}$ and $c = \dfrac{b}{a}$ in equation (3.4).

A parallel result exists for $I_r^*(t)$ which we state as follows.

Theorem: 4.4

Let X and Y be non-negative non-degenerate random variables admitting absolutely continuous distribution functions $F(x)$ and $G(x)$ respectively. Further assume that (Y,G) is the proportional reversed hazards model of (X,F). Then the relationship

$$1 - rI_r^*(t) = k\left(\lambda_1(t)\right)^r, \tag{4.19}$$

where $\lambda_1(t)$ is the reversed hazard rate holds for all $t \geq 0$ if and only if X follow the power distribution with distribution function,

$$F(x) = \left(\frac{x}{b}\right)^c;\ b \neq 0. \tag{4.20}$$

Proof

By direct calculation using equation (4.10), we get

$$1 - rI_r^*(t) = \frac{c\theta^r (c/t)^r}{\left(c + (c\theta - 1)r\right)},$$ which is of the form given in equation (4.19)

with $\lambda_1(t) = \dfrac{c}{t}$ and $k = \dfrac{c\theta^r}{\left(c + (c\theta - 1)r\right)}$, and the if part follows.

To prove the converse, differentiating equation (4.19) with respect to t, we get

$$-rI_r^{*'}(t) = k.r.\left(\lambda_1(t)\right)^{r-1} \lambda_1^{'}(t)\ . \tag{4.21}$$

Using equations (4.19) and (4.21) in equation (4.10), we have

$$\left(\lambda_1(t)\right)^{r-1}\left(\left\{\theta^r - (1+\theta r)k\right\}\lambda_1^2(t) - kr\lambda_1^{'}(t)\right) = 0\,.$$

This gives either $\lambda_1(t)=0$

or

$$\frac{\lambda_1'(t)}{(\lambda_1(t))^2}=\frac{\theta^r-(1+r\theta)\beta}{kr}. \tag{4.22}$$

The former solution is inadmissible, in this situation the distribution function F becomes degenerate.

From equation (4.22) we get

$$\lambda_1(t)=\frac{1}{pt+q}, \tag{4.23}$$

where $p=\frac{(1+\theta r)k-\theta^r}{kr}$ and q is the constant of integration.

As $t\to 0$, $q=0$. Using the result (2.12), we obtain the random variable X follows power function distribution.

In the context of equilibrium distribution, reviewed in Section 2.4, $I_r(F,G)$ becomes

$$I_r^E=\frac{1}{r}\left(1-\frac{\mu^{-r}}{r+1}\right), \tag{4.24}$$

where $\mu=E(X)$.The following relationship exists between the generalized inaccuracy measure defined in equation (4.3) and the mean residual life function $m(t)=E(X-t\,|\,X>t)$.

That is,

$$I_r^E(t)=\frac{1}{r}\left(1-\frac{(m(t))^{-r}}{r+1}\right). \tag{4.25}$$

It may be observed that in the context of equilibrium distributions, the knowledge of the mean residual life function of the original distribution determines the inaccuracy measure uniquely. Hence, the characterization results using the form of the mean residual life function can be suitably translated to the inaccuracy measure.

4.4 An alternative measure of inaccuracy

Nath (1968) defines inaccuracy of order r as

$$H_r\left(P,Q\right)=\frac{1}{1-r}\log\left(\sum_{x=0}^{\infty}p(x)q^{r-1}(x)\right);\ r\neq 1, r>0\ . \tag{4.26}$$

For the applications of the measure defined in equation (4.26), we refer to Nath (1968).

If X and Y are two non-negative random variables admitting absolutely continuous distribution functions $F(x)$ and $G(x)$ respectively, then the continuous analogue of equation (4.26) can be taken as

$$H_r\left(F,G\right)=\frac{1}{1-r}\log\left(\int_0^{\infty}f(x)g^{r-1}(x)dx\right);\ r\neq 1, r>0\ . \tag{4.27}$$

Note that as $r\to 1$, equation (4.27) reduces to the inaccuracy measure defined by equation (2.52). In this sense $H_r\left(F,G\right)$ provides a generalization for the inaccuracy measure given in Nath (1968). In the left truncated context, equation (4.27) becomes

$$H_r\left(F,G;t\right)=\frac{1}{1-r}\log\left(\int_t^{\infty}\frac{f(x)}{\overline{F}(t)}\left(\frac{g(x)}{\overline{G}(t)}\right)^{r-1}dx\right);r\neq 1, r>0, t>0\ . \tag{4.28}$$

Further in the right truncated situation, equation (4.27) can be written as

$$\overline{H}_r\left(F,G;t\right)=\frac{1}{1-r}\log\left(\int_0^{t}\frac{f(x)}{F(t)}\left(\frac{g(x)}{G(t)}\right)^{r-1}dx\right);\ r\neq 1, r>0\ ,\ t>0\ . \tag{4.29}$$

For the convenience in the sequel, we denote $H_r\left(F,G;t\right)$ and $\overline{H}_r\left(F,G;t\right)$ by $H_r\left(t\right)$ and $\overline{H}_r\left(t\right)$ respectively. Differentiating equation (4.28) with respect to t, and rearranging the terms we get a relationship between $H_r(t)$ and hazard rates namely

$$(1-r)H_r^{'}(t)=h_1(t)+(r-1)h_2(t)-h_1(t)\left(h_2(t)\right)^{r-1}\exp\left(-(1-r)H_r(t)\right), \quad (4.30)$$

where $h_1(t)$ and $h_2(t)$ are the hazard rates associated with the distribution functions $F(x)$ and $G(x)$ respectively and $H_r^{'}(t)$ denotes the derivative of $H_r(t)$. Similarly, from equation (4.29) one can have the representation

$$(1-r)\overline{H}_r^{'}(t)=\lambda_1(t)\left(\lambda_2(t)\right)^{r-1}\exp\left(-(1-r)\overline{H}_r(t)\right)-\lambda_1(t)+(1-r)\lambda_2(t), \quad (4.31)$$

where $\lambda_1(t)$ and $\lambda_2(t)$ are the reversed hazard rates.

When $\left(Y,\overline{G}\right)$ is the proportional hazards model of $\left(X,\overline{F}\right)$, equation (4.30) takes the form

$$(1-r)H_r^{'}(t)=\left(1+(r-1)\theta\right)h_1(t)-\theta^{r-1}\left(h_1(t)\right)^r\exp\left(-(1-r)H_r(t)\right). \quad (4.32)$$

Proceeding as similar lines, when G is the proportional reversed hazards model of F, equation (4.31) can be written as

$$(1-r)\overline{H}_r^{'}(t)=\phi^{r-1}\left(\lambda_1(t)\right)^r\exp\left(-(1-r)\overline{H}_r(t)\right)-\left(1-(1-r)\phi\right)\lambda_1(t). \quad (4.33)$$

Theorem: 4.5

For the random variables X and Y considered in Theorem 4.4, assume that $\left(Y,\overline{G}\right)$ is the proportional hazards model of $\left(X,\overline{F}\right)$ and $H_r(t)$ is increasing in t. Then $H_r(t)$ uniquely determines $F(t)$.

Proof

The proof of the theorem is similar to that of Theorem 4.1 and hence omitted.

Characterization results

The following theorem focuses attention on the situation when $H_r(t)$ is independent of t.

Theorem: 4.6

Let X and Y be non-negative non-degenerate and absolutely continuous random variables with survival functions $\overline{F}(x)$ and $\overline{G}(x)$ respectively. Assume $(Y,\overline{G})$ is the proportional hazards model of $(X,\overline{F})$. Then the distribution of X is exponential if and only if $H_r(t)=k$, where k is a constant, for every $t>0$.

Proof

Under the assumptions of the theorem, when X follows exponential distribution with survival function

$$\overline{F}(x)=\exp(-\lambda x);\; x>0,\lambda>0,$$

by direct calculation, we get

$$H_r(t)=\frac{1}{r-1}\log\left(1+(r-1)\theta\right)-\log(\lambda\theta)$$

$$=k, \text{ independent of } t.$$

Conversely assume that

$$H_r(t)=k$$

From equation (4.32), we have

$$h_1(t)\left\{1+(r-1)\theta-\theta^{r-1}\exp\left(-(1-r)K\right)h_1^{r-1}(t)\right\}=0.$$

This gives either $h_1(t)=0$ or $h_1(t)=c$, where c is a constant. But the former solution is inadmissible since in this situation X becomes degenerate. From the later solution we conclude that X follows exponential distribution.

In the following theorem, we give a characterization result for a family of distributions using a relationship between $H_r(t)$ and the hazard rate.

Theorem: 4.7

Let X be a non-negative random variable with absolutely continuous distribution function $F(x)$ and let $(Y,\overline{G})$ is the proportional hazards model of $(X,\overline{F})$. The relationship

$$H_r(t)=k(\theta)-\log h_1(t), \tag{4.34}$$

where $k(\theta)$ is a real valued function independent of t and $h_1(t)$ is the hazard rate of X holds for every $t>0$ if and only if X follows any one of the following three distributions.

(i) the exponential distribution with survival function

$$\overline{F}(x)=\exp(-\lambda x);\ x\geq 0,\lambda>0 \tag{4.35}$$

(ii) the Pareto distribution with survival function

$$\overline{F}(x)=\left(\frac{a}{a+x}\right)^b;\quad x\geq 0,\ b>1, 0<a<\infty \tag{4.36}$$

(iii) the Beta distribution with survival function

$$\overline{F}(x)=\left(1-\frac{x}{R}\right)^c;\ 0<x<R, c>1. \tag{4.37}$$

Proof

Assume that equation (4.34) holds and differentiating with respect to t, we get

$$H_r^{'}(t)=-\frac{h_1^{'}(t)}{h_1(t)}\ . \tag{4.38}$$

Substituting equation (4.38) in (4.32) we get

$$-(1-r)\frac{h_1^{'}(t)}{\left(h_1(t)\right)^2}=\left(1+(r-1)\theta\right)-\theta^{r-1}\exp\left(-(1-r)k\right)$$

or

$$\frac{d}{dt}\left(\frac{1}{h_1(t)}\right)=\frac{1}{1-r}\left(1+(r-1)\theta-\theta^{r-1}\exp\left(-(1-r)k\right)\right).$$

This gives

$$\frac{1}{h_1(t)}=\left(\frac{1+(r-1)\theta-\theta^{r-1}\exp\left(-(1-r)k\right)}{1-r}\right)t+c, \tag{4.39}$$

where c is the constant of integration. Equation (4.39) takes the form

$$h_1(t)=(pt+c)^{-1}, \tag{4.40}$$

where $p=\dfrac{1+(r-1)\theta-\theta^{r-1}\exp\left(-(1-r)k\right)}{1-r}$.

From Mukharjee and Roy (1986), equation (4.40) characterizes the exponential distribution for $p=0$, the Pareto distribution for $p>0$ and the beta distribution for $p<0$.

The if part of the theorem follows by direct calculations using the expression for $H_r(t)$ and $h_1(t)$ given in the table given below.

Distribution	$H_r(t)$	$h_1(t)$
Exponential	$\frac{1}{r-1}\log\left(1+(r-1)\theta\right)-\log(\lambda\theta)$	λ
Pareto	$\frac{1}{1-r}\log\left(\frac{\theta^{r-1}b^r(t+a)^{1-r}}{r+b(r\theta-\theta+1)-1}\right)$	$\frac{b}{t+\alpha}$
Beta	$\frac{1}{1-r}\log\left(\frac{\theta^{r-1}c^r(R-t)^{1-r}}{c(r\theta-\theta+1)-r+1}\right)$	$\frac{c}{R-t}$

In the following theorem, we give a characterization for generalized Pareto distribution based on relationship between $H_r(t)$ and mean residual life function.

Theorem: 4.8

For the random variables considered in Theorem 4.7, let $m_1(t)$ be the mean residual life function of X. Then the relation

$$H_r(t) - \log m_1(t) = k, \tag{4.41}$$

where k is a constant holds for all $t \geq 0$ if and only if X follows generalized Pareto distribution.

Proof

Assume that equation (4.41) holds. Differentiating equation (4.41) with respect to t, we get

$$H_r'(t) = \frac{m_1'(t)}{m_1(t)}. \tag{4.42}$$

Using equation (4.42), equation (4.32) can be written as

$$(1-r)m_1'(t) = (1+(r-1)\theta)m_1(t)h_1(t) - \theta^{r-1}\exp(-(1-r)k)(m_1(t)h_1(t))^r. \tag{4.43}$$

Take $\exp(-(1-r)k) = c$ and using the relation $m_1(t)h_1(t) = 1 + m_1'(t)$, we can see that $m(t)$ is a linear function. That is, F is generalized Pareto distribution.

The if part of the theorem follows by direct calculation using $H_r(t)$ and $m_1(t)$.

Chapter 5

CHERNOFF DISTANCE AND AFFINITY FOR TRUNCATED DISTRIBUTIONS*

5.1 Introduction

In the case of distributions that satisfy the regularity conditions, the Cramer-Rao inequality holds and the maximum likelihood estimator converges to normal distribution, whose variance is the inverse of the Fisher information. Therefore, the Fisher information consolidates the amount of accessible information for a regular family of distributions. However, in a non-regular location shift family that is generated by a distribution in R, whose support is not R, the Fisher information diverges and some times cannot be defined. Therefore, in order to characterize the bound of asymptotic performance in estimation, we need an information quantity generalizing Fisher information. Akahira (1996) proposed the limit of the Chernoff distance (relative Renyi entropy) as a substitute information quantity for a non-regular location shift family. Hayashi (2007) has examined the relationship of this measure with Kullback-Leibler divergence measure. In the present chapter, we extend the definition of Chernoff distance, for truncated distributions and examine its properties.

Let X and Y be two non-negative random variables with absolutely continuous distribution functions $F(x)$ and $G(x)$ and with same support. Denote by $f(x)$ and $g(x)$ the corresponding probability density functions. Then the Chernoff distance between $F(x)$ and $G(x)$ is defined as

$$C(F,G) = -\log\left(\int_0^\infty f^{\alpha}(x) g^{1-\alpha}(x)dx\right); \quad 0<\alpha<1. \tag{5.1}$$

* Some of the results in this chapter have been published in **(a)** Smitha S, Nair KRM and Sankaran PG (2007) **(b)** Nair KRM, Sankaran PG and Smitha S.

This measure is an example of the Ali-Silvey class of information theoretic distance measures. The Chernoff distance is always non-negative with zero distance occurring either when $\alpha = 0, 1$ or when the probability distributions are same. In equation (5.1), the parameter α can be interpreted as the weight assigned to the distributions while computing the distance between them. Statistical utility of Chernoff distance is that if one uses a Baye's procedure for testing $f(x)$ against $g(x)$, then $C(F,G)$ is asymptotically $\frac{1}{n}$ times the negative logarithm of the Bayes risk for distinguishing the two. Asadi et al. (2005) have extensively studied the application of this measure in the context of reliability studies.

5.2 Definition and properties

Let X and Y be two non-negative random variables with absolutely continuous distribution functions $F(x)$ and $G(x)$ respectively and with density functions $f(x)$ and $g(x)$. Consider the random variables

$$X_t = X - t \mid X > t \quad \text{and} \quad Y_t = Y - t \mid Y > t \,, \boldsymbol{t} > 0 \,. \tag{5.2}$$

Then the Chernoff distance between X_t and Y_t takes the form

$$C(F,G;t) = C(X,Y;t) = C_t$$

$$= -\log\left(\int_t^{\infty} \left(\frac{f(x)}{\overline{F}(t)} \right)^{\alpha} \left(\frac{g(x)}{\overline{G}(t)} \right)^{1-\alpha} dx \right), \tag{5.3}$$

where $\overline{F}(t)$ and $\overline{G}(t)$ are the survival functions of X and Y. In the right truncated situation, the random variables under consideration are

$$X_t^* = t - X \mid X < t$$

and

$$Y_t^* = t - Y \mid Y < t \,, t > 0 \,,$$

and equation (5.1) becomes

$$C^*(F,G;t) = C^*(X,Y;t) = C_t^* = -\log\left(\int_0^t \left(\frac{f(x)}{F(t)}\right)^{\alpha} \left(\frac{g(x)}{G(t)}\right)^{1-\alpha} dx\right). \tag{5.4}$$

Notice that when $t \to 0$, $C_t \to C(F,G)$ and when $t \to \infty$ $C_t^* \to C(F,G)$, where $C(F,G)$ is defined by equation (5.1).

In terms of the hazard rates, equations (5.3) and (5.4) can be written as

$$C_t^{'} = (\alpha-1)h_2(t) - \alpha h_1(t) + h_1^{\alpha}(t)h_2^{1-\alpha}(t)e^{C_t} \tag{5.5}$$

and

$$C_t^{*'} = \alpha\lambda_1(t) + (1-\alpha)\lambda_2(t) - \lambda_1^{\alpha}(t)\lambda_2^{1-\alpha}(t)e^{C_t^*}, \tag{5.6}$$

where $h_1(t)$ and $h_2(t)$ are the hazard rates , $\lambda_1(t)$ and $\lambda_2(t)$ are the reversed hazard rates of $F(t)$ and $G(t)$ respectively and $C_t^{'}$ and $C_t^{*'}$ represents the derivatives of C_t and C_t^* with respect to t.

We now discuss some properties of the truncated versions of the Chernoff distance defined in equations (5.3) and (5.4).

Theorem: 5.1

If $\phi(.)$ is an increasing function in the argument and $t > 0$, then

$$C\left(X,Y;\phi^{-1}(t)\right) = C\left(\phi(X),\phi(Y);t\right)$$

and

$$C^*\left(X,Y;\phi^{-1}(t)\right) = C^*\left(\phi(X),\phi(Y);t\right).$$

Proof

The proof follows directly from the definitions (5.3) and (5.4).

In the next theorem, we obtain an inequality concerning the measures defined in equations (5.1) and (5.3) under some mild conditions.

Theorem: 5.2

Let X and Y are two non-negative random variables with distribution functions $F(x)$ and $G(x)$ and with probability density functions $f(x)$ and $g(x)$ respectively. Denote by $h_1(x)$ and $h_2(x)$ the failure rates of F and G. Suppose that

(i) $\dfrac{h_1(x)}{h_2(x)}$ is increasing in x

and

(ii) both F and G are NBU.

Then

$C_t \geq C(F,G)$.

Proof

Let X_t and Y_t denote the truncated random variables. Denote the distribution functions of X_t and Y_t by $F_t(x)$ and $G_t(x)$ and the corresponding probability density functions $f_t(x)$ and $g_t(x)$. Denote by $C(F,G)$, the Chernoff distance between $F(x)$ and $G(x)$ defined by equation (5.1) and C_t the left truncated Chernoff distance, defined in equation (5.3) .

By taking, $x = F_t^{-1}(y)$, $\boldsymbol{y} \in (0,1)$,equation (5.3) becomes

$$C_t = -\log\left(\int_0^1 \left\{\frac{g_t(F_t^{-1}(y))}{f_t(F_t^{-1}(y))}\right\}^{1-\alpha} dy\right); 0 < \alpha < 1. \tag{5.7}$$

Using the definition of hazard rate, we have

$$f_t\left(F_t^{-1}(y)\right)=h_{F_t}\left(F_t^{-1}(y)\right)\overline{F}_t\left(F_t^{-1}(y)\right). \tag{5.8}$$

Further since $x=F_t^{-1}(y)$ we have

$$\begin{aligned}\overline{F_t}\left(F_t^{-1}(y)\right)&=\overline{F_t}(x)\\&=\frac{\overline{F}\left(t+F_t^{-1}(y)\right)}{\overline{F}(t)}.\end{aligned} \tag{5.9}$$

In view of the fact that

$$f_t\left(F_t^{-1}(y)\right)=\frac{f\left(t+F_t^{-1}(y)\right)}{\overline{F}(t)}, \tag{5.10}$$

using equations (5.9) and (5.10) in (5.8) we get

$$h_{F_t}\left(F_t^{-1}(y)\right)=\frac{f\left(t+F_t^{-1}(y)\right)}{\overline{F}\left(t+F_t^{-1}(y)\right)}=h_1\left(t+F_t^{-1}(y)\right). \tag{5.11}$$

Equation (5.8) now becomes

$$f_t\left(F_t^{-1}(y)\right)=h_1\left(t+F_t^{-1}(y)\right)\overline{F_t}\left(F_t^{-1}(y)\right). \tag{5.12}$$

From equation (5.9), we have

$$t+F_t^{-1}(y)=F^{-1}\left(1-(1-y)\overline{F}(t)\right) \tag{5.13}$$

and

$$\overline{F_t}\left(F_t^{-1}(y)\right)=1-y.$$

Then, equation (5.12) can be written as

$$f_t\left(F_t^{-1}(y)\right)=h_1\left(F^{-1}\left(1-(1-y)\overline{F}(t)\right)\right)(1-y).. \tag{5.14}$$

Similarly, we get

$$g_t\left(F_t^{-1}(y)\right)=h_{G_t}\left(F_t^{-1}(y)\right)\overline{G_t}\left(F_t^{-1}(y)\right). \tag{5.15}$$

Since $h_{G_t}\left(F_t^{-1}(y)\right)=h_2\left(t+F_t^{-1}(y)\right)$, using equation (5.13), equation (5.15) can be written as

$$g_t\left(F_t^{-1}(y)\right)=h_2\left(F^{-1}\left(1-(1-y)\overline{F}(t)\right)\right)\overline{G_t}\left(F_t^{-1}(y)\right). \tag{5.16}$$

Dividing equation (5.14) by equation (5.16), we obtain

$$\frac{f_t\left(G_1^{-1}(y)\right)}{g_t\left(G_1^{-1}(y)\right)}=\frac{h_1\left(F^{-1}\left(1-(1-y)\overline{F}(t)\right)\right)}{h_2\left(F^{-1}\left(1-(1-y)\overline{F}(t)\right)\right)}\frac{(1-y)}{\overline{G_t}\left(F_t^{-1}(y)\right)}. \tag{5.17}$$

Since $F^{-1}\left(1-(1-y)\overline{F}(t)\right)\geq F^{-1}(y)$, from the condition (i), we get

$$\frac{h_1\left(F^{-1}\left(1-(1-y)\overline{F}(t)\right)\right)}{h_2\left(F^{-1}\left(1-(1-y)\overline{F}(t)\right)\right)}\geq\frac{h_1\left(F^{-1}(y)\right)}{h_2\left(F^{-1}(y)\right)}. \tag{5.18}$$

Since F and G are NBU, we have

$$\overline{G_t}\left(F_t^{-1}(y)\right)\leq\overline{G}\left(F^{-1}(y)\right). \tag{5.19}$$

Substituting equations (5.18) and (5.19) in equation (5.17), we get

$$\frac{f_t\left(F_t^{-1}(y)\right)}{g_t\left(F_t^{-1}(y)\right)}\geq\frac{h_1\left(F^{-1}(y)\right)}{h_2\left(F^{-1}(y)\right)}\frac{(1-y)}{\overline{G}\left(F^{-1}(y)\right)}.$$

That is,

$$\frac{f_t\left(F_t^{-1}(y)\right)}{g_t\left(F_t^{-1}(y)\right)}\geq\frac{f\left(F^{-1}(y)\right)}{g\left(F^{-1}(y)\right)}. \tag{5.20}$$

Using equation (5.20), equation (5.7) becomes

$$C_t \geq -\log\left(\int_0^1 \left(\frac{g(F^{-1}(y))}{f(F^{-1}(y))}\right)^{1-\alpha} dy\right)$$

$$= C(F,G).$$

The implication of this property is that the distance between two systems of age t is never smaller than the distance when the systems were new. The following theorem provides a sufficient condition for the monotonicity of C_t.

Theorem: 5.3

For the random variables X and Y considered in Theorem 5.2, assume that

(i) $\dfrac{h_1(x)}{h_2(x)}$ is increasing in x,

and

(ii) both F and G are IFR.

Then

$C(F,G;t)$ is increasing in t.

Proof

When (ii) holds, we have

$$\frac{1}{\overline{G_{t_1}}\left(F_{t_1}^{-1}(y)\right)} \geq \frac{1}{\overline{G_{t_2}}\left(F_{t_2}^{-1}(y)\right)}, \text{ for all } 0 \leq t_2 \leq t_1,\ 0 < y < 1.$$

Assumption (i) implies

$$\frac{h_F\left(F^{-1}\{1-(1-y).\overline{F}(t)\}\right)}{h_G\left(F^{-1}\{1-(1-y).\overline{F}(t)\}\right)} \text{ is increasing in } t \geq 0,\ 0 < y < 1.$$

which implies

$$\frac{f_t\left(F_t^{-1}(y)\right)}{g_t\left(F_t^{-1}(y)\right)} \text{ is increasing.}$$

That is,

$$\frac{g_t\left(F_t^{-1}(y)\right)}{f_t\left(F_t^{-1}(y)\right)} \text{ is decreasing.}$$

Hence

$$C(F,G;t) = -\log\left(\int_0^1 \left(\frac{g_t[F_t^{-1}(y)]}{f_t[F_t^{-1}(y)]}\right)^{1-\alpha} dy\right) \text{ is increasing in } t \geq 0.$$

In the next theorem, we provide a bound for the truncated Chernoff distance in terms of the hazard rates.

Theorem: 5.4

For the random variables X and Y considered in Theorem 5.2, if X is larger (smaller) than Y in the likelihood ratio order $\left(X \geq_{lr} (\leq_{lr}) Y\right)$ then

$$C_t \geq (\leq) \alpha \log\left(\frac{h_2(t)}{h_1(t)}\right), t > 0, 0 < \alpha < 1. \tag{5.21}$$

Proof

If $\left(X \geq_{lr} (\leq_{lr}) Y\right)$, we have $\frac{f(t)}{g(t)}$ is increasing (decreasing) in t. This gives

$$\frac{f(x)}{g(x)} \leq (\geq) \frac{f(t)}{g(t)}; x \leq t, \boldsymbol{t} > 0.$$

There fore

$$\int_t^\infty \left(\frac{f(x)}{\overline{F}(t)}\right)^\alpha \left(\frac{g(x)}{\overline{G}(t)}\right)^{1-\alpha} dx \leq (\geq) \left(\frac{h_1(t)}{h_2(t)}\right)^\alpha.$$

This implies

$$\log\left(\int_t^\infty \left(\frac{f(x)}{\overline{F}(t)}\right)^\alpha \left(\frac{g(x)}{\overline{G}(t)}\right)^{1-\alpha} dx\right) \leq (\geq) \alpha \log\left(\frac{h_1(t)}{h_2(t)}\right). \tag{5.22}$$

Since

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

using equation (5.22), we get

$$C_t \geq (\leq)\, \alpha \log\left(\frac{h_2(t)}{h_1(t)}\right).$$

This completes the proof.

An analogous bound can be obtained in the right truncated situation. This is given as Theorem 5.5 below.

Theorem: 5.5

If the random variables X and Y as defined as in Theorem 5.2, and if X is larger (smaller) than Y in the likelihood ratio order $\left(X \geq_{lr} (\leq_{lr}) Y\right)$, then

$$C_t^* \geq (\leq) \alpha \log\left(\frac{\lambda_2(t)}{\lambda_1(t)}\right), t > 0, 0 < \alpha < 1.$$

The proof of the result is similar to that of Theorem 5.4 and hence omitted.

Remark: 5.1

X is smaller than Y in the likelihood ratio order $\left(X \leq_{lr} Y\right)$ implies that $h_1(t) \leq h_2(t)$ for all $t > 0$. Thus the right side of the expression (5.21) is non-negative for all $t > 0$. If C_t is increasing,

$C_t^{'} \geq 0$, where $C_t^{'}$ is the derivative of C_t.

From equation (5.5), it follows that

$$C_t \geq \log\left(\alpha\left(\frac{h_1(t)}{h_2(t)}\right)^{1-\alpha} + (1-\alpha)\left(\frac{h_2(t)}{h_1(t)}\right)^{\alpha}\right), t > 0.$$

Remark: 5.2

Denote by $D(F,G;t)$ is the modified Kullback-Leibler divergence measure defined in equation (2.48). It is immediate that there exist the following relationship between C_t and Kullback-Leibler divergence measures, namely

$$\lim_{\alpha\to 0}\left(\frac{C_t}{\alpha}\right)=D(G,F;t)$$

and

$$\lim_{\alpha\to 1}\left(\frac{C_t}{1-\alpha}\right)=D(F,G;t).$$

It may be observed that the following functional relationship exists between $H_E(t)$ and C_t, namely

$$C_t=-\log\left(1-H_E(t)\right),$$

where $H_E(t)=\int_t^\infty\left(\sqrt{\frac{f(x)}{\overline{F}(t)}}-\sqrt{\frac{g(x)}{\overline{G}(t)}}\right)^2 dx$ is the Hellinger's distance for truncated random variable.

5.3 Characterization theorems

In this Section, we look in to the situation where the truncated Chernoff distance is independent of t.

Theorem: 5.6

Let X and Y be two non negative random variables admitting absolutely continuous distribution functions and let C_t defined as in equation (5.3). C_t is independent of t if and only if $\left(Y,\overline{G}\right)$ is the proportional hazards model of $\left(X,\overline{F}\right)$.

Proof

The truncated Chernoff distance C_t is related to the Renyi divergence of order α defined by Renyi (1961) through the relationship

$$C_t = p.K_\alpha\left(f, g; t\right),$$

where, $p = 1-\alpha$. Asadi et al. (2005) has proved that $K_\alpha\left(f, g; t\right)$ is independent of t if and only if F and G satisfy the condition for being a proportional hazards model. The proof of the theorem is immediate from the above observations.

The following example describes an application of the above theorem in the context of series systems.

Example: 5.1

Let X_i, $i = 1, 2, ...n$ denote the life times of the components in series system. Assume that the probability density function of life times is $f\left(x\right)$ and that survival function is $\overline{F}\left(x\right)$.The lifetime of the system is then $Y = Min\left\{X_1, X_2, ...X_n\right\}$ with probability density function $g\left(x\right) = n\left[\overline{F}\left(x\right)\right]^{n-1} f\left(x\right)$ and with survival function $\overline{G}\left(x\right) = \left[\overline{F}\left(x\right)\right]^n$. Observe that X_i, $i = 1, 2, ...n$ and Y satisfies the condition for being a proportional hazards model. Further in view of equation (5.3), we have

$$e^{-C_t} = \int_t^\infty \left(\frac{f\left(x\right)}{\overline{F}\left(t\right)}\right)^\alpha \left(\frac{n\left[\overline{F}\left(x\right)\right]^{n-1} f\left(x\right)}{\left(\overline{F}\left(t\right)\right)^n}\right)^{1-\alpha} dx$$

$$= \frac{n^{1-\alpha}}{\left[\overline{F}\left(t\right)\right]^{n(1-\alpha)+\alpha}} \int_t^\infty f\left(x\right)\left(\overline{F}\left(x\right)\right)^{(n-1)(1-\alpha)} dx. \tag{5.23}$$

On simplifying equation (5.23), we get

$C_t = \dfrac{n^{1-\alpha}}{n(1-\alpha)+\alpha}$, which is independent of t as claimed in Theorem: 5.6.

The implication of the above result is that when a system of components with life distribution $F(x)$ are in series the Chernoff distance between the distribution of minimum and the original distribution $F(x)$ depends only on the number of components and the parameter α.

Theorem: 5.7

Let X and Y be two non-negative random variables with distribution functions $F(x)$ and $G(x)$ and with probability density functions $f(x)$ and $g(x)$ respectively. Then C_t^* defined in equation (5.4) is independent of t if and only if the relation

$$F(x) = (G(x))^{\theta};\theta > 0,$$

holds.

Proof

This result follows from Theorem: 2 of Asadi et al. (2005), who has considered the Renyi divergence of order α between two distributions for past life time namely

$$K_\alpha^*(f,g;t) = -\frac{1}{1-\alpha}\log\int_0^t \left(\frac{f(x)}{F(t)}\right)^{\alpha}\left(\frac{g(x)}{G(t)}\right)^{1-\alpha} dx .$$

It is established that $K_\alpha^*(f,g;t)$ is independent of t if and only if F and G have proportional reversed hazards rates. In view of the fact that C_t^* and $K_\alpha^*(f,g;t)$ is functionally related through the relationship

$$C_t^* = pK_\alpha^*(f,g;t),$$

where $p = 1-\alpha$, the theorem follows.

The following example shows an instance involving application of the above theorem in the case of a parallel system.

Example: 5.2

Let $\{X_1, X_2 ... X_n\}$ be independent and identically distributed random variables representing the lifetime of the components in a parallel system with probability density function $f(x)$ and distribution function $F(x)$. The lifetime of the system is then $Y = Max\{X_1, X_2, ... X_n\}$ with probability density function $g(x) = n(F(x))^{n-1} f(x)$ and distribution function $G(x) = (F(x))^n$. Here X_i and Y satisfy the condition for the proportional reversed hazards model. By direct calculation using equation (5.4), one can conclude that the truncated Chernoff distance C_t^*, is independent of t.

In the next section, we investigate the behavior of the Chernoff distance between the original and weighted distributions.

5.4 Chernoff distance between original and weighted distributions

Chernoff distance between the original random variable X and weighted random variable X_w, reviewed in Section 2.4, takes the form

$$C_w(x) = -\log\left(\int_0^\infty f^\alpha(x) f_w^{1-\alpha}(x) dx\right); 0 < \alpha < 1, \tag{5.24}$$

where $f_w(x) = \dfrac{w(x) f(x)}{E(w(X))}$, $E(w(X)) < \infty$, (5.25)

is the weighted distribution.

In the left truncated situation, equation (5.24) becomes

$$C_w(t) = -\log\left(\int_t^\infty \left(\frac{f(x)}{\overline{F}(t)}\right)^\alpha \left(\frac{f_w(x)}{\overline{F_w}(t)}\right)^{1-\alpha} dx\right); 0 < \alpha < 1 . \tag{5.26}$$

When we consider the length-biased model, given in equation (2.24), the above equation becomes

$$C_L(t) = (1-\alpha)\log\left(\frac{\mu \overline{F_L}(t)}{\overline{F}(t)}\right) - \log\left(\int_t^\infty x^{1-\alpha} \frac{f(x)}{\overline{F}(t)} dx\right), \tag{5.27}$$

where, $\overline{F_L}(t)$ is the survival function of the length biased random variable X_L. Further we have

$$\frac{\mu \overline{F_L}(t)}{\overline{F}(t)} = v(t) \tag{5.28}$$

$$= E(X \mid X \geq t) \text{ , is the vitality function}$$

and

$$\int_t^\infty x^{1-\alpha} \frac{f(x)}{\overline{F}(t)} dx = E\left(X^{1-\alpha} \mid X > t\right)$$

$$= v^{1-\alpha}(t). \tag{5.29}$$

Using (5.28) and (5.29) in (5.27), we get

$$C_L(t) = (1-\alpha)\log v(t) - \log\left(v^{1-\alpha}(t)\right) \tag{5.30}$$

The identity (5.30) enables one to find the non parametric estimator of $C_L(t)$ from the non parametric estimate of $v(t)$.

In the case of equilibrium model namely

$$f_E(x) = \frac{\overline{F}(x)}{\mu},\ X_E > 0,\ \mu = E\left(w(X)\right) < \infty.$$

Equation (5.26) becomes

$$C_E'(t) = \frac{\left(h(t)m(t)\right)^\alpha e^{C_E(t)} - \left(\alpha m'(t) + 1\right)}{m(t)},$$

where, $h(t)$ and $m(t)$ are the hazard rate and mean residual life function of original random variable and $C_E(t)$ is the left truncated Chernoff distance between original and equilibrium random variable.

Theorem: 5.8

Let X and Y are two non-negative random variables with distribution functions $F(x)$ and $G(x)$. Then $C_w(t)$ is independent of t if and only if the weight function $w(t)$ has the form

$$w(t) = \left(\overline{F}(t)\right)^{\theta-1}, \ \theta > 0, \ \theta \neq 1. \tag{5.31}$$

Proof

Suppose that

$C_w(t) = k$, a constant.

Then equation (5.26) becomes

$$-\log\left(\int_t^{\infty}\left(\frac{f(x)}{\overline{F}(t)}\right)^{\alpha}\left(\frac{f_w(x)}{\overline{F_w}(t)}\right)^{1-\alpha} dx\right) = k\ . \tag{5.32}$$

Differentiating (5.32) with respect to t and assuming that

$$\lim_{x\to\infty}\left(\left(f(x)\right)^{\alpha}\left(f_w(x)\right)^{1-\alpha}\right) = 0\,,$$

we get

$$\alpha\frac{h(t)}{h_w(t)} - e^k\left(\frac{h(t)}{h_w(t)}\right)^{\alpha} = \alpha - 1\,. \tag{5.33}$$

Substituting $\dfrac{h(t)}{h_w(t)} = c(t)$ in (5.33) and differentiating with respect to t, we get

$$\alpha c'(t)\left(1-\alpha\left(c(t)\right)^{\alpha-1}e^k\right) = 0\,.$$

The above equation gives $c^{'}(t)=0$ or $c(t)=\left(\frac{e^{-k}}{\alpha}\right)^{\frac{1}{\alpha-1}}$. In either case $c(t)$ is constant, say θ.

This gives

$$h_w(t)=\theta h(t),\ \theta>0,\ \theta\neq 1. \tag{5.34}$$

But we have the relation

$$h_w(t)=\frac{w(t)}{m_w(t)}h(t), \tag{5.35}$$

where

$$m_w(t)=E\left(w(X)\mid X>t\right). \tag{5.36}$$

From (5.34) and (5.35), we have

$$m_w(t)=\frac{w(t)}{\theta}\ . \tag{5.37}$$

Using (5.36) and (5.37), we get

$$\frac{1}{\overline{F}(t)}\int_t^{\infty}w(x)\,f(x)\,dx=\frac{w(t)}{\theta}. \tag{5.38}$$

Differentiating (5.38) with respect to t and assuming

$$\lim_{x\to\infty}\left(w(x)\,f(x)\right)=0,$$

we get (5.31).

Conversely assume that

$$w(t)=\left(\overline{F}(t)\right)^{\theta-1}.$$

We have the relation

$$\overline{F_w}(t) = \frac{m_w(t)}{\mu_w}\overline{F}(t), \tag{5.39}$$

where $\mu_w = E\left(w(t)\right) = \dfrac{1}{\theta}$.

Using equation (5.39) in equation (5.26), we get

$C_w(t) = \dfrac{2\sqrt{\theta}}{\theta+1}$, which is independent of t.

Theorem: 5.9

For the random variables X and Y considered in Theorem 5.8, the weight function $w(t)$ of model (5.25) is increasing (decreasing) in t, then

$$C_w(t) \le (\ge) \alpha \log\left(\frac{h_w(t)}{h(t)}\right), \; 0 < \alpha < 1, \tag{5.40}$$

where $h(t)$ and $h_w(t)$ are the hazard rates of X and X_w respectively.

Proof

Suppose that $w(t)$ is increasing (decreasing) in t. From (5.25), we get $\dfrac{f(t)}{f_w(t)}$ is increasing (decreasing) in t.This gives

$$\frac{f(t)}{f_w(t)} \le (\ge) \frac{f(x)}{f_w(x)}, \; x \le t.$$

Then

$$\int_t^\infty \left(\frac{f(x)}{\overline{F}(t)}\right)^{\alpha} \left(\frac{f_w(x)}{\overline{F_w}(t)}\right)^{1-\alpha} dx \ge (\le) \left(\frac{h(t)}{h_w(t)}\right)^{\alpha}.$$

Using the above expression, (5.26) becomes

$$C_w(t) \le (\ge) \alpha \log\left(\frac{h_w(t)}{h(t)}\right).$$

Corollary: 5.1

When $w(t)=t$ (length biased model), we have

$$C_L(t)\le \alpha\log\left(\frac{t}{t+m(t)}\right), t>0.$$

The result follows from Theorem 5.9 and the relationship $\frac{h(t)}{h_L(t)}=\frac{t+m(t)}{t}$, where $m(t)$ is the mean residual life function.

Corollary: 5.2

When $w(t)=\frac{1}{h(t)}$ (equilibrium distribution) is increasing (decreasing),

$$C_E(t)\le(\ge)-\alpha\log\big(h(t)m(t)\big).$$

5.5 Affinity for truncated distributions

As pointed out in Chapter 2 the concept of affinity defined by equation (2.50), is extensively used as a useful tool for discrimination among distributions. The measure of affinity given in equation (2.50) is a special case of the Chernoff distance defined in equation (5.1). In fact, when $\alpha=\frac{1}{2}$ equation (5.1) reduces to $-\log\rho$, where ρ is the measure of affinity given in equation (2.50). Affinity finds application in several practical situations. In the reliability context, the concept of affinity helps an experimenter to decide whether the distribution of life times for two components differ or are closer. Chitty Babu (1973) has used the concept for the extraction of effective features from imperfectly labeled patterns. Comaniciu et al. (2000) used this measure to examine the similarity in images or section of images in communication networking. There are several practical instances where complete data are not available to the experimenter. For instance, in life testing experiments the data on failure times are usually truncated. Motivated by this, in the present Section we extend the definition of affinity to the truncated situation. It may be noted that the proposed measure is an extension of the Bhattacharyya measure given in Thacker et al. (1997).

In reliability studies, if X and Y represents the lifetime of two systems, then X_t and Y_t , defined in equation (5.2) represent the remaining life of the system. The affinity between X_t and Y_t is a measure of similarity between the distribution of the residual lifetime of the systems. For instance, if X and Y represents the amount of profit of two firms and t is the tax exempt level, then the affinity between X_t and Y_t represents the similarity between the taxable incomes of the two firms.

Using the definition for affinity, given in equation (2.50), the affinity between X_t and Y_t takes the form

$$A_l\left(F,G;t\right)=\int_0^\infty \sqrt{f_t\left(y\right)g_t\left(y\right)}dy\,, \tag{5.41}$$

where

$f_t\left(y\right)=\dfrac{f\left(t+y\right)}{\overline{F}\left(t\right)}$ and $g_t\left(y\right)=\dfrac{g\left(t+y\right)}{\overline{G}\left(t\right)}$ are the probability density functions of X_t and Y_t and $\overline{F}\left(t\right)=P\left(X>t\right)$ and $\overline{G}\left(t\right)=P\left(Y>t\right)$ are the survival functions of X and Y.

Equation (5.41) can also be written as

$$A_l\left(F,G;t\right)=A_l=\frac{1}{\sqrt{\overline{F}\left(t\right)\overline{G}\left(t\right)}}\int_t^\infty \sqrt{f\left(x\right)g\left(x\right)}\,dx\ . \tag{5.42}$$

In the case of right truncated random variables, the variables under consideration are $X_t^*=t-X\mid X<t$ and $Y_t^*=t-Y\,|\,Y<t$ are the measure of affinity turns out to be

$$A_r\left(F,G;t\right)=A_r=\frac{1}{\sqrt{F\left(t\right)G\left(t\right)}}\int_0^t \sqrt{f\left(x\right)g\left(x\right)}\,dx\,. \tag{5.43}$$

In view of the fact that the measure of affinity, defined in equation (2.50) is a special case of general Chernoff distance, when $\alpha = \frac{1}{2}$, we get $C_t = -\log A_t$, the properties and characterizations based on the Chernoff distance can be suitably reformulated in the context of affinity. Since affinity is more often used in literature as a potential measure of discrimination, the formulation of characterization results in this frame work seems to be in order. In the sequel, we state some important characterization results using the concept of affinity. The proof of the results are similar to that of Chernoff distance discussed above.

Theorem: 5.10

Let X and Y be two non-negative random variables admitting absolutely continuous distribution functions $F(x)$ and $G(x)$ and probability density functions $f(x)$ and $g(x)$ respectively. A_t, defined by equation (5.42) is independent of 't', if and only if $\left(Y, \overline{G}\right)$ is the proportional hazards model of $\left(X, \overline{F}\right)$.

Proof

When A_t is independent of t, we have from equation (5.42)

$$\int_t^\infty \sqrt{f(x)g(x)}dx = c\sqrt{\overline{F}(t)\overline{G}(t)},$$

where $0 < c < \sqrt{2}$ is a constant , not depending on t.

Differentiating the above equation with respect to t and using the condition

$$\lim_{x\to\infty}\left(f(x)g(x)\right) = 0,$$

we obtain

$$-\sqrt{f(t)g(t)} = \frac{c}{2\sqrt{\overline{F}(t)\overline{G}(t)}}\left(-g(t)\overline{F}(t) - f(t)\overline{G}(t)\right)$$

or

$$\frac{2}{c} = \sqrt{\frac{g(t)\overline{F}(t)}{f(t)\overline{G}(t)}} + \sqrt{\frac{f(t)\overline{G}(t)}{g(t)\overline{F}(t)}}.$$

The above equation can be written as

$$\frac{2}{c}=\sqrt{\frac{h_2(t)}{h_1(t)}}+\sqrt{\frac{h_1(t)}{h_2(t)}}. \tag{5.44}$$

where $h_1(t)$ and $h_2(t)$ are the hazard rates of F and G respectively.

Denoting by

$$k(t)=\frac{h_2(t)}{h_1(t)},$$

equation (5.44) takes the form

$$\left(1+k(t)\right)^2=\frac{4}{c^2}k(t).$$

This gives $k(t)$ is a constant (say θ), independent of t

Thus

$$h_2(t)=\theta h_1(t),$$

or equivalently

$$\overline{G}(x)=\left(\overline{F}(x)\right)^{\theta}\ \theta>0, \tag{5.45}$$

as claimed in theorem.

Conversely, when equation (5.45) holds, we have

$$g(t)=\theta f(t)\left(\overline{F}(t)\right)^{\theta-1}. \tag{5.46}$$

Using equation (5.46) in equation (5.42), we get

$$A_l=\frac{\sqrt{\theta}}{\sqrt{\overline{F}(t)\overline{G}(t)}}\int_t^{\infty}\left(\overline{F}(x)\right)^{\frac{\theta-1}{2}}f(x)\,dx. \tag{5.47}$$

$$A_l = \frac{2\sqrt{\theta}}{\theta+1},$$

which is independent of t, and the sufficiency part follows.

In the right truncated situation, the property that A_r is constant is characteristic to the proportional reversed hazards model. This result is stated as Theorem 5.11 below.

Theorem: 5. 11

Under the conditions of the above theorem, A_r defined in equation (5.43), is independent of 't' if and only if the relationship

$$G(t) = (F(t))^{\phi} \ ; \ \phi > 0, \tag{5.48}$$

holds for all $t > 0$. That is when (Y, G) is the proportional reversed hazards model of (X, F).

Proof

When equation (5.48) holds, we have

$$g(t) = \phi f(t)(F(t))^{\phi-1}. \tag{5.49}$$

Using equation (5.49) in equation (5.43), we get

$$\sqrt{F(t)G(t)}A_r = \sqrt{\phi}\left\{(F(t))^{\frac{\phi+1}{2}} - \left(\frac{\phi-1}{2}\right)\frac{1}{\sqrt{\phi}}\sqrt{F(t)G(t)}A_r\right\}.$$

The solution of the above equation is

$$A_r = \frac{2\sqrt{\phi}}{\phi+1}, \qquad \text{a constant.}$$

The proof of the converse part is similar to that of Theorem: 5. 10 and hence omitted.

Note:

Instead of assuming the condition that A_l is independent of t, if we assume that A_l is linear in t, say $A_l = at+b$, where a and b are constants, the following relationship between the hazard rates of F and G is immediate.

$$\left(\frac{h_1(t)+h_2(t)}{2}\right)(at+b)-\sqrt{h_1(t)h_2(t)}=a,$$

where $h_1(t)$ and $h_2(t)$ are the hazard rates. Similarly, in the case of A_r, we get

$$\left(\frac{\lambda_1(t)+\lambda_2(t)}{2}\right)(at+b)-\sqrt{\lambda_1(t)\lambda_2(t)}=-a,$$

where $\lambda_1(t)$ and $\lambda_2(t)$ are the reversed hazard rates of F and G respectively.

In certain cases, the dependence structure may be such that $G(x)$ a weighted distribution obtained from $F(x)$. Denote by $\overline{F}_w(t)$ and $f_w(t)$, the survival and probability density functions of X_w, the weighted random variable. The affinity between the original and weighted random variables, namely X and X_w, takes the form

$$A_w(t)=\int_t^{\infty}\sqrt{\frac{f(x)}{\overline{F}(t)}\frac{f_w(x)}{\overline{F}_w(t)}}dx, \tag{5.50}$$

where $f_w(x)=\dfrac{w(x)f(x)}{E(w(X))}, E\left[w(X)\right]<\infty.$ (5.51)

The relationship connecting $A_w(t)$ and hazard rates are immediate from equation (5.50), and is given by

$$A_w'(t)=\left(\frac{h(t)+h_w(t)}{2}\right)A_w(t)-\sqrt{h(t)h_w(t)}\,, \tag{5.52}$$

where $A_w'(t)=\frac{d}{dt}A_w(t)$, $h(t)$ and $h_w(t)$ are the hazard rates of the random variables X and X_w respectively.

5.6 Relationship with other discrimination measures

(i) Bhattacharyya distance

First, we discuss the relationship between the Bhattacharyya distance [Kailath (1967)] and the affinity in the truncated situation. Equation (5.42) can be written as

$$\sqrt{\overline{F}(t)\overline{G}(t)}A_l=\int_t^\infty\sqrt{f(x)g(x)}\,dx.$$

This is equivalent to

$$\sqrt{\overline{F}(t)\overline{G}(t)}\,A_l=\int_0^\infty\sqrt{f(x)g(x)}\,dx-\int_0^t\sqrt{f(x)g(x)}\,dx\,. \tag{5.53}$$

Using the equations (2.50) and (5.43), equation (5.53) can be written as

$$\sqrt{\overline{F}(t)\overline{G}(t)}\,A_l=\rho-\sqrt{F(t)G(t)}\,A_r\,. \tag{5.54}$$

But we have the relationship

$$\rho=e^{-B},$$

where B is the Bhattacharyya distance [Kailath (1967)]. Then equation (5.54) now becomes

$$B=-\ln\left\{\sqrt{\overline{F}(t)\overline{G}(t)}\,A_l+\sqrt{F(t)G(t)}\,A_r\right\}.$$

(ii) Modified Kullback- Leibler divergence measure

Here we consider the modified Kullback-Leibler divergence measure defined in equation (2.48) namely

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

$D(F,G,t)$ can also be written as

$$D(F,G,t)=-2\int_t^{\infty}\frac{f(x)}{\bar{F}(t)}\log\left(\sqrt{\frac{g(x)/\bar{G}(t)}{f(x)/\bar{F}(t)}}\right)dx\ . \tag{5.55}$$

Using Jensen's inequality, equation (5.55) becomes

$$D(F,G,t)\geq -2\log\int_t^{\infty}\frac{f(x)}{\bar{F}(t)}\sqrt{\frac{g(x)\bar{F}(t)}{f(x)\bar{G}(t)}}dx.$$

That is,

$$D(F,G,t)\geq -2\log\int_t^{\infty}\sqrt{\frac{f(x)g(x)}{\bar{F}(t)\bar{G}(t)}}dx.$$

From equation (5.42), the above expression becomes

$$D(F,G,t)\geq -2\log A_l. \tag{5.56}$$

However, modified Kullback-Leibler divergence measure is the difference between the residual inaccuracy measure and residual entropy function, so from equation (2.53), the expression (5.56) can be read as

$$I(F,G,t)-H(F,t)\geq -2\log A_l,$$

where $I(F,G,t)$ is the inaccuracy measure in truncated setup and $H(F,t)$ is the residual entropy.

(iii) Hellinger's distance

Hellinger's distance for truncated random variable is

$$H_E(t) = \int_t^\infty \left(\sqrt{\frac{f(x)}{\overline{F}(t)}} - \sqrt{\frac{g(x)}{\overline{G}(t)}} \right)^2 dx . \tag{5.57}$$

On simplifying, we get equation (5.57) as

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

$$= 2(1 - A_l) .$$

Chapter **6**

RESIDUAL INACCURACY MEASURE AND RELATED CONCEPTS IN DISCRETE TIME DOMAIN

6.1 Introduction

Compared to the volume of literature available in the continuous case, only little work seems to have been done in the analysis of lifetime data in discrete time. However, discrete model provides a good approximation for their continuous counterparts. Xekalaki (1983) provides examples of situations where discrete models are appropriate. Gupta and Gupta (1983), Hitha and Nair (1989) and Roy and Gupta (1999) have characterized probability distributions using reliability concepts in discrete time. Rajesh and Nair (1998) has defined the residual entropy function in the discrete set up and has obtained characterization results for the geometric distribution using the functional form of the residual entropy function. Nanda and Paul (2006) have extended the definition to residual entropies of order β and has studied their properties. Recently, an alternate definition for the generalized residual entropy in discrete case has been proposed by Baig and Dar (2009). Motivated by this, in the present chapter we extend the concept of residual inaccuracy and affinity to the discrete time domain and examine its properties.

6.2 Residual inaccuracy in discrete time

Let X and Y be two random variables in the support of the set of non-negative integers and with probability mass functions $f(x)$ and $g(x)$. Denote the survival functions of X and Y by $\overline{F}(x)$ and $\overline{G}(x)$ respectively. The inter relationships between the reliability concepts in discrete case is reviewed in Section 2.1.

If

$$X_t = X - t \mid X > t \text{ and } Y_t = Y - t \mid Y > t,\ t \geq 0, \tag{6.1}$$

the discrete analogue of the inaccuracy measure, defined in equation (2.54), associated with X_t and Y_t can be defined as

$$I(F,G;t) = -\sum_{x=t+1}^{\infty} \frac{f(x)}{\overline{F}(t)} \log\left(\frac{g(x)}{\overline{G}(t)}\right). \tag{6.2}$$

For convenience in notation, we denote $I(F,G;t)$ by $I(t)$.

Equation (6.2) can also be written as

$$I(t) = \log \overline{G}(t) - \frac{1}{\overline{F}(t)} \sum_{x=t+1}^{\infty} f(x) \log g(x). \tag{6.3}$$

Further, the residual inaccuracy function, defined in equation (6.2) can be expressed in terms of the hazard rate, $h(x) = \dfrac{f(x)}{\overline{F}(x-1)}$.

Observing that equation (6.2) can be written as

$$I(t) = -\log h_2(t+1) - \sum_{x=t+1}^{\infty} \frac{f(x)}{\overline{F}(t)} \log\left(\frac{g(x)}{g(t+1)}\right), \tag{6.4}$$

where

$$h_2(t+1) = \frac{g(t+1)}{\overline{G}(t)},$$

we have

$$I(t) = -\log h_2(t+1) - \sum_{x=t+1}^{\infty} \frac{f(x)}{\overline{F}(t)} \log\left(\frac{h_2(x)}{h_2(x+1)(1-h_2(x))} \frac{g(x+1)}{g(t+1)}\right). \tag{6.5}$$

Using the relationship

$$f(x) = \overline{F}(x-1) - \overline{F}(x),$$

equation (6.5) becomes

$$I(t) = -\log h_2(t+1) + \sum_{x=t+1}^{\infty} \frac{\overline{F}(x)}{\overline{F}(t)} \log\left(\frac{h_2(x)}{h_2(x+1)(1-h_2(x))} \right)$$

$$\sum_{x=t+1}^{\infty} \frac{\overline{F}(x-1)}{\overline{F}(t)} \log\left(\frac{g(x)}{g(t+1)} \right) + \sum_{x=t+1}^{\infty} \frac{\overline{F}(x)}{\overline{F}(t)} \log\left(\frac{g(x+1)}{g(t+1)} \right). \tag{6.6}$$

In view of the fact that

$$\sum_{x=t+1}^{\infty} \frac{\overline{F}(x)}{\overline{F}(t)} \log\left(\frac{g(x+1)}{g(t+1)} \right) - \sum_{x=t+1}^{\infty} \frac{\overline{F}(x-1)}{\overline{F}(t)} \log\left(\frac{g(x)}{g(t+1)} \right) = 0, \tag{6.7}$$

using equation (6.7) in equation (6.6), we get

$$I(t) = -\log h_2(t+1) + \sum_{x=t+1}^{\infty} \frac{\overline{F}(x)}{\overline{F}(t)} \log\left(\frac{h_2(x)}{h_2(x+1)(1-h_2(x))} \right). \tag{6.8}$$

The above equation provides a useful relation connecting the residual inaccuracy measure and the hazard rate.

We now establish a recurrence relation satisfied by $I(t)$.

Theorem: 6.1

Let X and Y be two discrete random variables in the support of non-negative integers with probability mass functions $f(x)$ and $g(x)$, failure rates $h_1(x)$ and $h_2(x)$ and residual inaccuracy function $I(t)$. Then $I(t)$ satisfies the relationship

$$I(t)=\frac{1}{1-h_1(t)}\left\{I(t-1)+h_1(t)\log h_2(t)+\left(1-h_1(t)\right)\log\left(1-h_2(t)\right)\right\}.\ t=1,2,... \quad (6.9)$$

Proof

Equation (6.3) can be written as

$$\overline{F}(t)I(t)=\overline{F}(t)\log\overline{G}(t)-\sum_{x=t+1}^{\infty}f(x)\log g(x)\ . \quad (6.10)$$

Changing t to $t+1$ in equation (6.10), we get

$$\overline{F}(t+1)I(t+1)=\overline{F}(t+1)\log\overline{G}(t+1)-\sum_{x=t+2}^{\infty}f(x)\log g(x). \quad (6.11)$$

Subtracting equation (6.10) from equation (6.11), we get

$$I(t+1)=\frac{\overline{F}(t)}{\overline{F}(t+1)}I(t)+\log\overline{G}(t+1)+\frac{f(t+1)}{\overline{F}(t+1)}\log g(t+1)-\frac{\overline{F}(t)}{\overline{F}(t+1)}\log\overline{G}(t). \quad (6.12)$$

Using the relationship

$$f(t+1)=\overline{F}(t)-\overline{F}(t+1),$$

equation (6.12) becomes

$$I(t+1)=\frac{\overline{F}(t)}{\overline{F}(t+1)}I(t)+\log\overline{G}(t+1)+\left(\frac{\overline{F}(t)}{\overline{F}(t+1)}-1\right)\log g(t+1)-\frac{\overline{F}(t)}{\overline{F}(t+1)}\log\overline{G}(t). \quad (6.13)$$

Equation (6.13) can be written as

$$I(t+1)=\frac{\overline{F}(t)}{\overline{F}(t+1)}\left(I(t)+\log h_2(t+1)\right)-\log\left(\frac{g(t+1)}{\overline{G}(t+1)}\right). \quad (6.14)$$

Since $\frac{g(t+1)}{\overline{G}(t+1)} = \left(\frac{h_2(t+1)}{1-h_2(t+1)}\right)$ and $\frac{\overline{F}(t)}{\overline{F}(t+1)} = \frac{1}{1-h_1(t+1)}$, equation (6.14) becomes

$$I(t+1) = \frac{1}{1-h_1(t+1)}\left(I(t) + \log h_2(t+1)\right) - \log\left(\frac{h_2(t+1)}{1-h_2(t+1)}\right). \tag{6.15}$$

Rearranging the terms in equation (6.15), we get

$$I(t+1) = \frac{1}{1-h_1(t+1)}\left\{I(t) + h_1(t+1)\log h_2(t+1) + \left(1-h_1(t+1)\right)\log\left(1-h_2(t+1)\right)\right\}. \tag{6.16}$$

Taking t in place of $t+1$ in equation (6.16), we get the relationship given in equation (6.9), as claimed.

Corollary: 6.1

$I(t)$ can be expressed uniquely in terms of the inaccuracy measure $I(F,G)$, defined in equation (2.51) and the hazard rates associated with F and G.

Proof

Substituting for $I(t-1)$ in equation (6.9), we get

$$\begin{aligned}I(t) = {} & \frac{1}{\left[1-h_1(t-1)\right]\left[1-h_1(t)\right]} I(t-2) \\ & + \frac{1}{\left[1-h_1(t-1)\right]\left[1-h_1(t)\right]}\left\{h_1(t-1)\log h_2(t-1) + \left(1-h_1(t-1)\right)\log\left(1-h_2(t-1)\right)\right\} \\ & + \frac{\left\{h_1(t)\log h_2(t) + \left(1-h_1(t)\right)\log\left(1-h_2(t)\right)\right\}}{\left[1-h_1(t)\right]}.\end{aligned}$$

Proceeding like this, we get

$$I(t)=\frac{I(-1)}{[1-h_1(t)][1-h_1(t-1)]\ldots[1-h_1(0)]}$$

$$+\frac{\{h_1(0)\log h_2(0)+(1-h_1(0))\log(1-h_2(0))\}}{[1-h_1(t)][1-h_1(t-1)]\ldots[1-h_1(0)]}$$

$$+\frac{\{h_1(1)\log h_2(1)+(1-h_1(1))\log(1-h_2(1))\}}{[1-h_1(t)][1-h_1(t-1)]\ldots[1-h_1(1)]}$$

$$+\ldots+\frac{h_1(t)\log h_2(t)+(1-h_1(t))\log(1-h_2(t))}{[1-h_1(t)]}.$$

$$=\frac{I(F,G)}{\prod_{x=0}^{t}[1-h_1(x)]}+\frac{\{h_1(0)\log h_2(0)+(1-h_1(0))\log(1-h_2(0))\}}{\prod_{x=0}^{t}[1-h_1(x)]}$$

$$+\frac{\{h_1(1)\log h_2(1)+(1-h_1(1))\log(1-h_2(1))\}}{\prod_{x=0}^{t}[1-h_1(x)]} \tag{6.17}$$

$$+\ldots+\frac{h_1(t)\log h_2(t)+(1-h_1(t))\log(1-h_2(t))}{[1-h_1(t)]}.$$

where $I(F,G)$ is the Kerridge's inaccuracy measure associated with the random variables X and Y. Equation (6.17) expresses the truncated inaccuracy measure in terms of the hazard rates and the inaccuracy measure, defined in equation (2.51).

Now we look into the situation where the residual inaccuracy measure defined in equation (6.2) is a constant, independent of t.

Theorem: 6.2

Let X and Y be two discrete random variables in the support of non-negative integers. Assume that the relationship

$$\overline{G}(x) = \left(\overline{F}(x)\right)^{\theta} , \tag{6.18}$$

holds for all $x > 0$. The residual inaccuracy measure is independent of t if X follows geometric distribution.

Proof

Assume that equation (6.18) holds.

This gives

$$\overline{G}(x-1) - \overline{G}(x) = \left(\overline{F}(x-1)\right)^{\theta} - \left(\overline{F}(x)\right)^{\theta} .$$

Since $\overline{G}(x-1) - \overline{G}(x) = g(x)$, the above expression becomes

$$g(x) = \left(\overline{F}(x-1)\right)^{\theta} - \left(\overline{F}(x)\right)^{\theta} . \tag{6.19}$$

Using equations (6.18) and (6.19) in equation (6.3), we get

$$I(t) = \theta \log \overline{F}(t) - \frac{1}{\overline{F}(t)} \sum_{x=t+1}^{\infty} f(x) \log\left\{\left(\overline{F}(x-1)\right)^{\theta} - \left(\overline{F}(x)\right)^{\theta}\right\}. \tag{6.20}$$

When X follows geometric distribution with survival function

$$\overline{F}(x) = q^{x+1}, x = 0,1,2... \tag{6.21}$$

using equation (6.21), equation (6.20) can be written as

$$I(t) = \theta(t+1)\log q - q^{-(t+1)} \sum_{x=t+1}^{\infty} pq^{x} \log\left(q^{x\theta} - q^{(x+1)\theta}\right).$$

That is,

$$I(t)=\theta(t+1)\log q-q^{-(t+1)}\left(\theta p\log q\sum_{x=t+1}^{\infty}xq^{x}+p\log\left(1-q^{\theta}\right)\sum_{x=t+1}^{\infty}q^{x}\right). \qquad (6.22)$$

Equation (6.22) can be read rewritten as

$$I(t)=\theta(t+1)\log q-q^{-(t+1)}\left[\theta tq^{t+1}\log q+\frac{\theta q^{t+1}}{p}\log q-q^{t+1}\log\left(1-q^{\theta}\right)\right]. \qquad (6.23)$$

Simplifying equation (6.23), we get

$$I(t)=\theta\log q-\frac{\theta}{p}\log q-\log\left(1-q^{\theta}\right),$$

which is independent of t.

In this sequel we consider the discrete analogue of the generalized inaccuracy measure considered in Nath (1968) in the truncated context. Let X and Y be the random variables in the support of the set of non-negative integers with probability mass functions $f(x)$ and $g(x)$ respectively. Nath (1968) defines the generalized inaccuracy measure as

$$H_r(F,G)=\frac{1}{1-r}\log\left(\sum_{x=1}^{\infty}f(x)g^{r-1}(x)\right),\ r\neq 1, r>0. \qquad (6.24)$$

When $f(x)=g(x)$, equation (6.24) reduces to generalized Renyi entropy considered in Renyi (1961). Further $\lim_{r\to 1}H_r(F,G)$ becomes the Kerridge's inaccuracy measure. Notice that the measure defined in equation (6.24) is a monotonically decreasing function of r.

For the random variables X_t and Y_t, defined in equation (6.1), with probability mass functions $f_t(x)$ and $g_t(x)$ and survival functions $\overline{F_t}(x)$ and $\overline{G_t}(x)$, equation (6.24) turns out to be

$$H_r(F,G;t)=\frac{1}{1-r}\log\left(\sum_{x=t+1}^{\infty}\left(\frac{f(x)}{\overline{F}(t)}\right)\left(\frac{g(x)}{\overline{G}(t)}\right)^{r-1}\right). \tag{6.25}$$

The following relationship between the measure defined in the above equation and hazard rates is immediate.

$$e^{(1-r)H_r(t-1)}-e^{(1-r)H_r(t)}=h_1(t)\left(h_2(t)\right)^{r-1},$$

where $h_1(t)=\frac{f(t)}{\overline{F}(t)}$ and $h_2(t)=\frac{g(t)}{\overline{G}(t)}$.

6.3 Inaccuracy in the context of length biased distributions

Gupta (1979) has defined the length-biased distribution of a non-negative discrete random variable X, analogous to the continuous case, as

$$f_L(x)=\frac{xf(x)}{\mu},\ x=1,2,... \tag{6.26}$$

where $\mu=E(X)<\infty$.

The survival function, $\overline{F_L}(x)$ corresponding to equation (6.26) is

$$\overline{F}_L(x)=\frac{\overline{F}(x)v(x)}{\mu}, \tag{6.27}$$

where $v(x)=E(X\mid X>x)$ is the vitality function of X.

Now for the random variables X and X_L, the measure (6.2) is defined as

$$I_L(t)=-\sum_{x=t+1}^{\infty}\frac{f(x)}{\overline{F}(t)}\log\left(\frac{f_L(x)}{\overline{F_L}(t)}\right). \tag{6.28}$$

The measure (6.28) gives the measure of inaccuracy when $f_L\left(x\right)$ is assumed instead of $f\left(x\right)$. In the context of length-biased distributions, the following relationship exists between $I_L\left(t\right)$, the residual entropy function, geometric vitality function and vitality function.

Using equations (6.26) and (6.27), equation (6.28) can be written as

$$I_L\left(t\right)=H\left(F,t\right)-\log G\left(t\right)+\log v\left(t\right), \tag{6.29}$$

where

$H\left(F;t\right)=-\sum_{x=t+1}^{\infty}\frac{f\left(x\right)}{\overline{F}\left(t\right)}\log\left(\frac{f\left(x\right)}{\overline{F}\left(t\right)}\right)$ is the residual entropy measure defined in Rajesh and Nair (1998) and

$\log G\left(t\right)=\frac{1}{\overline{F}\left(t\right)}\sum_{x=t+1}^{\infty}f\left(x\right)\log x$ is the geometric vitality function discussed in Nair and Rajesh (2000).

6.4 Affinity between two residual life distributions in the discrete case

Let X and Y be two discrete random variables in the support of the set of non-negative integers. Denote the probability mass function of X and Y by $f\left(x\right)$ and $g\left(x\right)$ and the survival functions by $\overline{F}\left(x\right)$ and $\overline{G}\left(x\right)$ respectively. We define the affinity between the residual life distributions X_t and Y_t, defined in equation (6.1),as

$$\rho_t=\sum_{x=t+1}^{\infty}\sqrt{\frac{f\left(x\right)g\left(x\right)}{\overline{F}\left(t\right)\overline{G}\left(t\right)}}\ . \tag{6.30}$$

Equation (6.30) provides a measure of similarity between the distribution functions associated with the truncated random variables X_t and Y_t.

In the next theorem, we establish a recurrence relation satisfied by ρ_t .

Theorem: 6.3

Let ρ_t be as defined in equation (6.30) and $h_1(x)$ and $h_2(x)$ be the hazard rates associated with F and G respectively. Then ρ_t satisfies the recurrence relation

$$\rho_t = \frac{\rho_{t-1} - \sqrt{h_1(t)h_2(t)}}{\sqrt{(1-h_1(t))(1-h_2(t))}}, \ t = 1,2,3... \tag{6.31}$$

Proof

From equation (6.30), we have

$$\sqrt{\overline{F}(t)\overline{G}(t)}\rho_t = \sum_{x=t+1}^{\infty} \sqrt{f(x)g(x)}. \tag{6.32}$$

Replace t by $t-1$ in equation (6.32), we get

$$\sqrt{\overline{F}(t-1)\overline{G}(t-1)}\rho_{t-1} = \sum_{x=t}^{\infty} \sqrt{f(x)g(x)}. \tag{6.33}$$

Subtracting equation (6.33) from equation (6.32), and simplifying, we get

$$\rho_{t-1} - \sqrt{\frac{\overline{F}(t)\overline{G}(t)}{\overline{F}(t-1)\overline{G}(t-1)}}\rho_t = \sqrt{h_1(t)h_2(t)}. \tag{6.34}$$

Since $\frac{\overline{F}(t)}{\overline{F}(t-1)} = 1 - h_1(t)$, equation (6.34) can be written as

$$\rho_{t-1} - \sqrt{[1-h_1(t)][1-h_2(t)]}\rho_t = \sqrt{h_1(t)h_2(t)}.$$

Rearranging the terms, we get equation (6.31), as claimed.

Corollary: 6.2

From the recurrence relation (6.31) by successive iteration, one can express ρ_t in terms of $\rho(F,G)$ and hazard rates, similar to that of inaccuracy, described in Corollary 6.1.

Plan for future study

The present study has unfolded several problems which needs further investigation. Compared to the volume of work done on inaccuracy measure in the continuous case, only very little work seems to have been done in discrete domain. Characterizations of distributions based on the functional form of the discrete residual inaccuracy measure as well as its generalization can be done analogous to the continuous case. The proposed affinity measure based on truncated observations would be useful to decide whether two populations differ or is consistent with respect to their distributions. To apply this measure in practical situations, one has to develop a reasonable estimator for this measure. Non parametric estimation procedures can be utilized by suitably extending the work by Ahamad (1980). These problems shall be taken up in a future work.

REFERENCES

[1] Abraham, B. and Sankaran, P.G. (2005). Renyi's entropy for residual lifetime distribution. *Statistical Papers*, 46, 17-30.

[2] Aczel, J. and Daroczy, Z. (1975). *On measures of information and their characterizations*. Academic Press, New York.

[3] Ahmad, I. A. (1980). Nonparametric estimation of Matusita's measure of affinity between absolutely continuous distributions. *Annals of the Institute of Statistical Mathematics*, 32, 223-240.

[4] Asadi, M. and Ebrahimi, N. (2000). Residual entropy and its characterizations in terms of hazard function ande mean residual life function. *Statistics and Probability Letters*, 49, 263-269.

[5] Asadi, M., Ebrahimi, N. and Soofi, E. (2005). Dynamic generalized Information measures. *Statistics and Probability and Letters*, 71, 85-98.

[6] Akahira, M. (1996). Loss of information of a statistic for a family of non-regular distributions. *Annals of the Institute of Statistical Mathematics*, 48, 349-364.

[7] Azlarov, T.A. and Volodin, N.A. (1986). *Characterization Problems Associated with the Exponential Distribution,* Springer, New York.

[8] Baig, M.A.K. and Dar, J.G. (2009). A generalized residual information measure and its properties. *Journal Of inequalities in pure and applied mathematics,* 10 (3), 1-8.

[9] Barlow, R. E., Marshall, A. W. and Proschan, F. (1963). Properties of probability distributions and monotone hazard rate. *The Annals of Mathematical Statistics*, 34 (2), 375-389.

[10] Belzunce, F., Navarro, J., Ruiz J.M. and Aguila, Y. (2004). Some results on residual entropy function. *Metrika* , 59 , 147–161.

[11] Bhattacharyya, A. (1946). On a measure of divergence between two multinomial populations. *Sankhya* , 7, 401–406.

[12] Block, H.W., Savits, T.H. and Singh, H. (1998). The reversed hazard rate function. *Probability in Engineering and Informational Sciences*, 12, 69-90.

[13] Chaundy, T.W., and Mcleod, T.W. (1960). On a functional equation. *Proceedings of Edinburgh Mathematical Society*, Edinburgh Mathematical Notes, 43, 7-8.

[14] Chitti Babu, C. (1973). On the application of probabilistic distance measures for the extraction of features from imperfectly labeled patterns. *International Journal of Computer and Information Sciences*, 2(2), 103-114.

[15] Comaniciu, D., Ramesh,V . and Meer, P. (2000). Real-time tracking of non-rigid objects using mean shift. *IEEE Conference on Computer Vision and Pattern Recognitio.* Hilton HeadIsland, South Carolina, 142 - 149.

[16] Cox, D.R. (1962). *Renewal Theory*. John Wiley, New York.

[17] Cox, D.R. (1972). Regression models and life tables. *Journal of Royal statistical society*, Series B (34), 187-220.

[18] Csiszar, I. (2008). Axiomatic characterizations of information measures. *Entropy*, 10(3), 261-273.

[19] Di Crescenzo, A. (2000). Some results on the proportional reversed hazard model. *Statistics Probability Letters*, 50, 313-321.

[20] Di Crescenzo, A. and Longobardi, M. (2002). Entropy-based measure of uncertainty in past lifetime distributions. *Journal of Applied Probability*, 39, 434-440.

[21] Di Crescenzo, A. and Longobardi, M. (2004). A measure of discrimination between past lifetime distributions, *Statistics and Probability Letters*, 67, 173-182.

[22] Ebanks, B., Sahoo, P. and Sander, W. (1998). *Characterizations of information measures,* World Scientific, Singapore.

[23] Ebrahimi, N. and Pellery, F. (1995). New partial ordering of survival functions based on the notion of uncertainty. *Journal of applied probability,* 23, 202-211.

[24] Ebrahimi, N. (1996). How to measure uncertainty about residual life time. *Sankhya,* Series A 58, 48–57.

[25] Ebrahimi, N. and Kirmani, S.N.U.A. (1996,a). A measure of discrimination between two residual life-time distributions and its applications. *Annals of the Institute of Statistical Mathematics,* 48, 257-265.

[26] Ebrahimi, N. and Kirmani, S.N.U.A. (1996,b). A characterization of the proportional hazards model through a measure of discrimination between two residual life distributions. *Biomertika,* 83, 233-235.

[27] El Barmi, H. and Nelson, P.I. (2002). A note on estimating a non-increasing density in the presence of selection bias. *Journal of Statistical Planning and Inference,* 107, 353-364.

[28] Finkelstein, M.S. (2002). On the reversed hazard rate. *Reliability Engineering and System Safety,* 78, 71-75.

[29] Galambos, J. and Kotz, S. (1978). *Characterizations of Probability Distributions.* Springer-Verlag. New York.

[30] George, A. and Mathai, A.M. (1974). Applications of the concepts of affinity and distance to population problems. *Journal of Biosocial Sciences (London),* 6, 347-356.

[31] Gupta, R. C. (1979). Waiting time paradox and size biased sampling. *Communications in Statistics- Theory and Methods,* A8 (6), 601–607.

[32] Gupta, P. L. and Gupta, R. C. (1983). On the moments of residual life in reliability and some characterization results. *Communications in Statistics - Theory and Methods,* 12, 449-461.

[33] Gupta, R.C. and Keating, J.P. (1986). Relations for reliability measures under length-biased sampling. *Scandavian Journal of Statistics*, 13, 49-56.

[34] Gupta, R.C. and Kirmani, S.N.U.A. (1990). The role of weighted distributions in stochastic modelling. *Communications in Statistics-Theory and Methods*, 19 (9), 3147-3162.

[35] Gupta, R.C. and Kirmani, S.N.U.A. (2004). Some characterization of distributions by functions of failure rates and mean residual life. *Communications in Statistics- Theory and Methods*, 33, 3115-3131.

[36] Gupta , R.C, Gupta, P.L. and Gupta, R.D. (1998). Modeling failure time data with Lehman alternative. *Communications in Statistics-Theory and methods*, 27, 887-904.

[37] Gupta, R.C. and Han, W. (2001). Analyzing survival data by PRH models. *International Journal of Reliability and Applications*, 2, 203-216.

[38] Hall, W.J. and Wellner, J.A. (1981). Mean residual life. *In Statistics and Related Topics*, 169-184, North Holland, Amster dam, The Netharlands.

[39] Hayashi,M., (2007). Limiting behavior of relative Rényi entropy in a non-regular location shift family. *Annals of the Institute of Statistical Mathematics*, online sample.

[40] Hitha, N. and Nair, N.U. (1989). Characterization of some discrete models by properties of residual life function. *Calcutta Statistical Association Bulletin*, 38, 219–225.

[41] Ikeda, S. (1963). Asymptotic equivalence of probability distributions with applications to some problems of asymptotic independence. *Annals of the Institute of Statistical Mathematics*, 15, 87-116.

[42] Jain, K., Singh, H. and Bagai, I. (1989). Relations for reliability measures of weighted distributions. *Communications in Statistics-Theory and Methods*, 18 (2), 4393-4412.

[43] Kailath, T. (1967). The divergence and Bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology*, 15(1), 52-60.

[44] Kannappan, P.I. and Rathie, P.N. (1973). On a Characterization of Directed Divergence. *Information and Control*, 22(2), 163-171.

[45] Kapur, J.N. (1968). On information of order α and β. *Proceedings of the Indian Academic Sciences*, 48 (A), 65-75.

[46] Kapur, J. N. (1989). *Maximum Entropy Models in Science and Engineering*, New Delhi, India, Wiley Eastern Limited.

[47] Kapur, J.N. (1994). *Measures of information and their applications*, Wiley Eastern Limited, New Delhi.

[48] Keilson, J. and Sumita, V. (1982). Uniform stochastic ordering and related inequalities. *The Canadian Journal of Statistics*, 10 (3), 181-191.

[49] Kerridge, D.F. (1961). Inaccuracy and inference. *Journal of the Royal Statistical Society*, Series B, 23, 184-194.

[50] Khinchin, A.I. (1957). *Mathematical Foundation of Information Theory*, New York, Dover.

[51] Khinchin, A.J. (1953). The concept of entropy in the theory of probability. *Uspekhi. Mat. Nauk.*, 8, 3-20.

[52] Kirmani, S.N.U.A. (1968). Some results on Matusita's measure of distance. *Journal of the Indian Statistical Association*, 6, 89-98.

[53] Kotz, S. and Shanbhag, D.N. (1980). Some new approaches to probability distributions. *Advances in Applied Probability*, 12, 903-921.

[54] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematics Statistics*, 22, 76-86.

[55] Kundu, D. and Gupta, R.D. (2004). Characterizations of the proportional (Reversed) hazard model. *Communications in statistics-Theory and methods*, 33, 3095-3103.

[56] Kupka, J. and Loo, S. (1989). The hazard and vitality measures of ageing. *Journal of Applied Probability,* 26, 532-542.

[57] Lee, P.M. (1964). On the axioms of information theory. *Annals of Mathematical Statistics*, 35, 415-418.

[58] Lie, C.D. and Xie, M. (2006). *Stochastic ageing and dependence for reliability*. Springer, New York.

[59] Majernik, V. (2004). A dissimilarity measure for an arbitrary number of probability distributions. *International Journal of General Systems*, 33, (6), 673-678.

[60] Mathai, A. M. and Rathie, P. N. (1975). *Basic concepts in information theory and statistics axiomatic foundations and applications*, Wiley, New York.

[61] Matusita, K. (1954). On the estimation by the minimum distance method. *Annals of the Institute of Statistical Mathematics,* 5, 59-65.

[62] Matusita, K. (1955). Decision rules based on the distance, for problem of fit, two samples and estimation. *Annals of Mathematical Statistics*, 26, 631-640.

[63] Matusita, K. (1957). Decision rules based on the distance for the classification problem. *Annals of the Institute of Statistical Mathematics*, 8, 67-77.

[64] Matusita, K. (1961). Interval estimation based on the notion of affinity. *Bulletin of the International Statistical Institute,* 38 (4), 241-244.

[65] Matusita, K. (1967). On the notion of affinity of several distributions and some of its applications. *Annals of the Institute of Statistical Mathematics*, 19, 181-192.

[66] Mukherjee, S.P and Roy, D. (1986). Some characterizations of the exponential and related life distributions. *Calcutta Statistical Association Bulletin*, 36, 189-197.

[67] Nair, K.R.M. and Rajesh, G. (1998). Characterization of probability distributions using the residual entropy function. *Journal of Indian Statistical Association,* 36, 157-166.

[68] Nair, K.R.M. and Rajesh, G. (2000). Geometric vitality function and its applications to reliability, *IAPQR Transactions*, 25 (1), 1-8.

[69] Nair, K.R.M., Sankaran, P.G. and Smitha**,** S. (2009). Chernoff distance for truncated distributions. *Statistical Papers,* 50 (4), 761 – 778.

[70] Nair, N.U. and Sankaran, P.G. (1991). Characterization of the Pearson system of distributions. *IEEE Transactions on Reliability*, 40, 75-77.

[71] Nair, N. U., Sankaran, P. G. and Asha, G. (2005). Characterizations using reliability concepts. *Journal of Applied Statistical Science*, 14, 237-242.

[72] Nair, N.U. and Gupta, R.P. (2007). Characterization of proportional hazard models by properties of information measures. *International Journal of Statistical Sciences*, 6, Special issue, 223-231.

[73] Nanda, A. and Jain, K. (1999). Some weighted distribution result on univariate and bivariate cases. *Journal of Statistical Planning and Inference,* 77, 169-180.

[74] Nanda, A. K. and Paul, P. (2006,a). Some properties of past entropy and their applications. *Metrika,* 64 (1), 47-61.

[75] Nanda, A. K. and Paul, P. (2006). Some Results on Generalized Residual Entropy. *Information Sciences*, 176(1), 27-47.

[76] Nath, P. (1968). Inaccuracy and Coding Theory. *Metrika*, 13, 123-135.

[77] Oluyede, B.O. (1999). On inequalities and selection of experiments for length-biased distributions. *Probability in Engineering and Informational Sciences,* 13, 169-185.

[78] Oluyede, B.O. (2000). On some length-biased inequalities for reliability measure. *Journal of Inequality and Applications,* 5, 447-466.

[79] Oluyede, B.O. (2002). Some inequalities and bounds for weighted reliability measures. *Journal of Inequalities in Pure and Applied Mathematics*, 3 (4), Article 60.

[80] Patil, G.P. and Rao, C.R. (1977). Weighted distribution: A survey of their applications. In *Applications of Statistics* (Krishnaiah, P.R. ed.), North Holland Publishing Company, 384-405.

[81] Rajesh, G. (2001). Characterization of Some probability distributions using the residual entropy function. Doctoral dissertation, Cochin University of Science and Technology.

[82] Rajesh, G. and Nair, K.R.M. (1998). Residual entropy function in discrete time. *The Far East Journal of Theoretical Statistics*, 2(1), 1-10.

[83] Ramkumar, R. (1975). Religious Taxonomy in India, Demography, India.

[84] Rao, C.R. (1965). *Linear Statistical Inference and its applications*. Wiley, New York

[85] Rathie, P.N. (1971). On some measures of uncertainty, inaccuracy and information and their characterizations. *Kybernetika*, 7, 394-403.

[86] Renyi, A. (1961). On measures of entropy and information. *Proceeding of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, university of California Press, Berkeley, 547-561.

[87] Roy, D. and Gupta, R.P. (1999). Characterizations and model selections through reliability measures in the discrete case. *Statistics and Probability Letters*, 43, 197-206.

[88] Roy, D and Mukharjee, S.P. (1989). Characterization based on arithmetic, geometric, and harmonic means of failure rates. *Contribution to Stochastic*, 178-185.

[89] Ruiz, J.M. and Navarro, J. (1994). Characterization of distributions by relationships between failure rate and mean residual life. *IEEE Transactions on Reliability* , 43 (4), 640-644.

[90] Sankaran, P.G. and Nair, N.U. (1993). On form-invariant length-biased models from Pearson family. *Journal of the Indian Statistical Association,* 31, 35-89.

[91] Sankaran, P.G. and Gupta, R.P.(1999). Characterization of lifetime distributions using measure of uncertainty, *Calcutta Statistical Association Bullettin*, 154–166.

[92] Sengupta, D., Singh, H. and Nanda, A. K. (1999). The Proportional Reversed Hazards Model, Technical Report, Indian Statistical Institute, Calcutta.

[93] Sen, A. and Khattree, R. (1996). Length-biased distribution, equilibrium distribution and characterization of probability laws. *Journal of Applied Statistical Science*, 3, 239-252.

[94] Shaked, M. and Shanthikumar, J. G. (1994). *Stochastic orders and their applications*. New York, Academic Press.

[95] Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 279-423 and 623-656.

[96] Smitha, S., Nair, K.R.M. and Sankaran, P.G. (2007). On measures of affinity for truncated distributions. *Calcutta Statistical Association Bulletin*, 59, 151-162.

[97] Sullo, P. and Rutherford, D.K. (1977). Characterizations of the power distribution by conditional excedence. *The American Statistical Association Proceedings of the Business and Economics*, 787-792.

[98] Sunoj, S. M., Sankaran, P. G., and Maya, S. S. (2009). Characterizations of Life Distributions Using Conditional Expectations of Doubly (Interval) Truncated Random Variables. *Communications in Statistics-Theory and Methods*, 38, 9, 1441-1452.

[99] Taneja, H. C. Vikas Kumar and Srivastava, R. (2009). A Dynamic Measure of Inaccuracy between Two Residual Lifetime Distributions, *International Mathematical Forum*, 25, 1213-1220.

[100] Thacker, N. A. , Aherne, F. J. and Rockett, P. I. (1997). The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data. *Kybernetika*, 34, 363-368.

[101] Tverberg, H. (1957). A new derivation of the information function. *Math. Stand.* 6, 297-298.

[102] Van Es, B., Klaassen, C. A. J. and Ondshoorn, K. (2000). Survival analysis under cross-sectional sampling: length-bias and multiplicative censoring. *Journal of Statistical Planning and Inference*, 91, 295-312.

[103] Xie, M., Gaudoin, O. and Bracquemond,C. (2002). Redefining failure rate function for discrete distributions. International Journal of Reliability, *Quality and Safety Engineering,* 9 (3), 275-286.

[104] Xekalaki, E. (1983*)*. Hazard functions and life distributions in discrete time. *Communications in Statistics- Theory and Methods,* 12 , 2503-2509.

[105] Yeung, R.W. (2002). *A First Course in Information Theory*. Kluwer Academic /Plenum Publishers, New York.

[106] Weiner, N. (1948), *Cybernetics*. New York, Wiley.