# DIVERGENCE MEASURE IN INFORMATION THEORY AND RELIABILITY ANALYSIS

## DISSERTATION

**SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE**

**Of**

**Master of Philosophy**

**In statistics**

**By**

**Shabir Ahmad Chopan**

**Under the supervision of**

**Dr. M. A. K. Baig**
**Associate Professor**

**Post-graduate Department of Statistics**

**UNIVERSITY OF KASHMIR, SRINAGAR**
**(NAAC Accrediated Grade 'A')**
**(2012)**

**Post-graduate Department of Statistics**
**University of Kashmir, Srinagar**

Dr.Mirza Abdul Khalique'Baig
*(Associate Professor)*
Contact No: 09419090826
E_mail:-baigmak@gmail.com

Dated:______________

# *Certificate*

This is to certify that the scholar ***Shabir Ahmad Chopan*** has carried out the present dissertation entitled **"DIVERGENCE MEASURE IN INFORMATION THEORY AND RELIABILITY ANALYSIS"** under my supervision and the work is suitable for submission for the award of the Degree of Master of Philosophy in Statistics. It is further certified that the work has not been submitted in part or full for the award of M. Phil or any other degree.

Dr. M. A. K. Baig
(Supervisor)

# Contents

# Acknowledgement

*Praise to Allah the most gracious and merciful,*
*who bestowed upon me the courage to accomplish this work*

Knowledge gained gives new dimensions to the life of a gainer or the one who imparts. It is difficult to a researcher to complete his work without supervision and co-operation of others and I am not exception. Among those who have kindled my interest in the present endeavor is my supervisor ***Dr. M. A. K. Baig,*** Associate Professor, Department of Statistics, University of Kashmir Srinagar, whose inspiring guidance and untiring encouragement helped me to complete this work. His affectionate attitude, genuine concern and scholarly discourse always kept me lively to accomplish this work.

I am deeply grateful to ***Professor Aquil Ahmad,*** Head, Department of Statistics, University of Kashmir Srinagar, for providing the necessary facilities in the department. I express my gratefulness to my teachers, ***Dr. Tariq Rashid Jan*** and ***Dr. S. Parvaiz Ahmad*** for their help.

I am also thankful to non-teaching staff of my department and staff members of Allama Iqbal Library, University of Kashmir Srinagar, for their kind support.

Sincere thanks are due to my parents and other family members especially ***Molvu Nazir Ahmad Kashfi*** for his kind co-operation and enthusiastic love.

It would be unfair if I fail to thank my friends who extended their cooperation whenever I needed, especially Mr. Mohd Javid Dar, Mr. Raja Sultan Ahmad Reshi, Miss Humara Sultan, Miss Nusrat Mushtaq, Mr. Mohd. Shafi Malik, Mr. Aijaz Ahmad Sheikh) and Mr. Farooq Ahmad Lone.

Finally, thanks are also due to Mr. Javeed Ahmad Bhat who took pains to accomplish the present endeavour.

I owe the responsibility for any errors, omissions, typographical mistakes and other things alike that have incorporated in this manuscript.

***Shabir Ahmad Chopan***

# *Preface*

The concept of information originated when an attempt was made to create a theoretical model for the transmission of information of various kinds. Information theory is a branch of Mathematical theory of probability and is applied in a wide variety of fields: Communication Theory, Thermodynamics, Economics, Cybernetics, Operation Research and Psychology.

Much work has been done on this branch of probability and it had acquired a great currency in various research Journals of Statistics. In this light, I compiled my dissertation on the topic *DIVERGENCE MEASURE IN INFORMATION THEORY AND RELIABILITY ANALYSIS* and the chapter wise scheme is as follows:

**Chapter-I:** gives the basic concepts and preliminary results. This makes the rest of the dissertation readable.

**Chapter-II:** deals with some new generalization of entropy measure.

**Chapter-III:** throws light on the measure of discrimination between lifetime distributions.

**Chapter-IV:** discusses the cumulative residual entropy and its properties.

**Chapter- V:** deals with the measure of information and its applications.

The intent of this manuscript is to present a survey of the existing literature on divergence measures and reliability. It will be a useful document for the future researchers in this area. The area of divergence measure and reliability analysis is fertile and there is a lot of scope to work on this concept.

***Shabir Ahmad Chopan***

# Chapter – I

The word ‘information’ is very common word used in everyday language. Information transmission usually occurs through human voice (as in telephone, radio, television, etc.), books, newspapers, letters, etc. In all these cases a piece of information is transmitted from one place to another. However, one might like to quantitatively assess the quality of information contained in a piece of information. Few examples are as follows:

1. Suppose, one states, ‘It is raining’. Now the question is ‘Have we received much information?’ Here it may be concluded that if a piece of information is presented, which was already known, then, obviously, no information has been received. Again, if one states that ‘The sun will shine the whole day tomorrow’. In this case aninformation has been received without being specific. Since we have been informed that something will happen about which we did not know, therefore, we do not have to be much surprised by the statement that was made.
2. Suppose, we have come to know from the weather forecast on television that ‘The rain will continue for the next two days’. In this case, we have received information more than in the first example above because a statement has been made whose truth is not at all so surprising.

Information theory is a new branch of probability theory with extensive potential applications to the communication systems. Like several other branches of mathematics, information theory has a physical origin. It was principally originated by C.E.Shannon [122], through twooutstanding contributions to the mathematical theory of communications. These were followed by a flood of research papers speculating upon the possible applications of the newly born theory to broad spectrum of research areas such as pure mathematics, psychology, economics, biology, etc.

The first attempt to develop the mathematical measure for communication channels was made by Nyquist {[103], [104]} and Heartley[62]. The main contributions which really gave birth to the so called information theory, came shortly after the second world war from the mathematicians C. E. Shannon [122] and N. Wiener [137]. In the paper entitled, "the mathematical theory of communication" Shannon made the first attempt to deal with the new concept of the amount of information and its main consequences. Perhaps the most important theoretical result of information theory is the Shannon's fundamentaltheorem in which he first set up a mathematical model for quantitative measure of average amount of information provided by a probabilistic experiment and proved a number of interesting results which showed the importance and usefulness measure of information.

In the last 40 years, the information theory has been more precise and has grown into staggering literature. Some of its terminology even has become part of our daily language and has been brought to a point where it has found its wide applications in various fields of importance. e.g., The work of Bar-Hillel [15], B. Subrahmanyan and Siromoney[14]in Linguistic, Brillouins[23] in physics, Theil [132] in economics, Quastler [111] in Psychology, Quastler[110] in Biology and Chemistry, Wiener [137] in Cybernetics,Kerridge[78] in Statistical estimation, Kapur[73] in Operation Research, Kullback[80] in Mathematical Statistics, Zaheerudin[141] in Inference, Zadeh[139] in Fuzzy set theory, Ebrahimi[43] in Survival analysis, Rao[113] in Anthropology, Mei [87] in Genetics, Sen [118] in Political Science and Chen [27] in pattern recognition.Jaynes[68] first stated the maximum entropy principle explicitly and during last three decades, the principle has been applied with the varying degree of success in fields such as Thermodynamics and Statistical Mechanics, Design of Experiment andContingency Tables, Search theory, Reliability theory, Banking, Insurance, Accountancy and Marketing, Transportation problems, etc. We restrict ourselves only to those aspects of information theory which are closely related to our research work.

In the context of reliability and lifetime distributions, there are some measures such as the hazard rate function or the mean residual lifetime function that have been used to characterize or compare the aging process of a component. Cox [29] and Kotz and Shanbhag[79] have shown that both the functions determines the distribution function uniquely. Ebrahimi[43] proposed an alternative characterization of a lifetime distribution in terms of conditional Shannon's entropy. Based on the measure of residual entropy, Ebrahimi and Pellery[41] and Ebrahimi and Kirmani[40] have studied some ordering and aging properties of lifetime distributions. Belzunce*et al.*[18] extended some results given by Nair and Rajesh [93] and Asadi and Ebrahimi [7] to characterize a distribution from functional relationships between the residual entropy and the mean residual life or hazard rate function. Various generalizations of Ebrahimi's measure have been proposed by many researchers including Abraham and Sankaran[1], Nanda and Paul[95] and Hooda and Kumar [66]. Measure of uncertainty in past lifetime distributions have been proposed by Crescenzo and Longobadri[30] and generalized by Nanda and Paul [93].Ebrahimi and Kirmani[40] and Gupta and Nanda [58] gave an overview of some aspects of residual divergence measures and studied some characterization theorems under the assumption that the distribution function satisfy the Cox's proportional hazard rate model.

## 1.1 Information Function and Shannon's Entropy

***1.1.1 Information Function:*** Let $E_i$ be the *ith* event with probability of occurrence $p_i$ , the information function may be defined as

$$h(p_i) = -\log(p_i) \tag{1.1.1}$$

***1.1.2 Shannon's Entropy:*** Let $S$ be the sample space belongs torandom events. Compose this sample space into a finite number of mutually exclusive events $E_1, E_2, \ldots, E_n$, whose respective probabilities are $p_1, p_2, \ldots, p_n$, then the average amount of information or Shannon's entropy is defined as

$$H(P) = E(h(p_i)) = -\sum_{i=1}^{n} p_i \log p_i, \qquad 0 \le p_i \le 1 \tag{1.1.2}$$

Some important properties of Shannon's entropy is given below:

**I)** **Continuity:** $H(P)$ iscontinuous $P$, i.e. the measure should be continuous, so that changing the values of the probabilitiesby a very small amount should only change the entropy by a small amount.

**II)** **Symmetry:** The measure remains unchanged if the outcomes $x_i$ are re-ordered.

$$H(p_1, p_2) = H(p_2, p_1).$$

**III)** **Maximality:** The measure should be maximal if all the outcomes are equally likely (uncertainty is highest when all possible events are equiprobable).

$$\max H(p_1, \ldots, p_n) = H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$$

**IV)** **Additivity:** Theadditivity property states that for twoindependent probability distributions $= (p_1, p_2, \ldots, p_n)$, $Q = (q_1, q_2, \ldots, q_m)$,

$$H(p_1, p_2, \ldots, p_{n-1}, q_1, q_2, \ldots, q_m) = H(p_1, p_2, \ldots, p_{n-1}, p_n)$$

$$+ p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}, \ldots, \frac{q_m}{p_n}\right),$$

where $p_n = \sum_{k=1}^{m} q_k$ .

In addition to the above four basic properties, we have the following properties

**v)** **Expansibility:**The value of the entropy function should not change, if an impossible outcome is added to the probability scheme, i.e.

$$H_{m+1}(p_1, p_2, \ldots, p_m, 0) = H_m(p_1, p_2, \ldots, p_m).$$

**vi)** For the two independent probabilitydistributions

$$P = (p_1, p_2, \ldots, p_m), \quad Q = (q_1, q_2, \ldots, q_n),$$

where $\sum_{i=1}^{m} p_i = 1, \qquad \sum_{j}^{n} q_j = 1,$

then the uncertainty of the joint scheme should be the sum of their uncertainties, i.e.,

$$H_{mn}(P \cup Q) = H_m(P) + H_n(Q).$$

**vii)** **Normality:**The entropy becomes unity for two equally probable events, i.e.

$$H\left(\frac{1}{2}, \frac{1}{2}\right) = 1$$

***1.1.3 Unit of Information:***When the logarithm is taken with the base 2, the unit of information is called bit,when the natural logarithm is taken then the resulting unit is called Nat and if the logarithm is taken with the base 10, the unit of information is known as Heartley.

It must be noted that the definition of Shannon’s entropy though defined for a discrete random variable can be extended to situations when the random variable under consideration is continuous.

Let $X$ be a continuous random variable with the density function $f(x)$ on $I$, where $I=(-\infty,\infty)$, then the entropy is defined as

$$H(X)=-\int_{-\infty}^{+\infty} f(x)\log f(x)dx \tag{1.1.3}$$

whenever it exists. The measure (1.1.3) is also called differential entropy. It has many of the properties of discrete random entropy but unlike the entropy of the discrete random variable, the differential entropy may be infinitely large, negative or positive, Ash [12]. Also, the entropy of the discrete random variable remains invariant under a change of variable. However with a continuous random variable the entropy does not necessarily remains invariant.

## 1.2 Generalizations of Shannon's Entropy

Various generalizations of Shannon's entropy are available in the literature. Some important generalizations are given below:

***i)*** ***Renyi's Entropy:***Renyi[117] generalized the Shannon's entropy by defining the entropy of order $\alpha$ as

$$H_\alpha(P)=\frac{1}{1-\alpha}\log\left[\frac{\sum_{i=1}^{n} p_i^\alpha}{\sum_{i=1}^{n} p_i}\right], \qquad \alpha>0(\neq 1) \tag{1.2.1}$$

and in continuous case

$$H_\alpha(X)=\frac{1}{1-\alpha}\log\int_0^\infty \frac{f^\alpha(x)dx}{f(x)} \qquad \alpha>0(\neq 1) \tag{1.2.2}$$

For $\alpha \to 1$, the measure (1.2.1) and (1.2.2) reduces to (1.1.2)and (1.1.3) respectively.

***ii)*** ***Havrada-Charavat's Entropy:***Havrada-Charavat[1967] introduced the entropy as

$$H^{\alpha}(P)=\frac{\left[\sum_{i=1}^{n}p_i^{\alpha}\right]-1}{2^{1-\alpha}-1} \qquad \alpha\neq 1, \alpha>0 \qquad (1.2.3)$$

and in continuous case

$$H^{\beta}(X)=\frac{1}{1-\beta}\left(\int_0^{\infty} f^{\beta}(x)dx-1\right) \qquad \beta>0(\neq 1) \qquad (1.2.4)$$

and is called generalized entropy of type $\beta$. When $\beta\rightarrow 1$, the measure (1.2.3) and (1.2.4 ) becomes Shannon's measure (1.1.2 ) and (1.1.3) respectively.

***iii)*** ***Varma's Entropy:***Varma[136] introduced the entropies as

$$H_{\alpha}^{\beta}(P)=\frac{1}{\beta-\alpha}\log\left(\sum_{i=1}^{n}p_i^{\alpha+\beta-1}\right) \qquad \beta-1<\alpha<\beta,\quad \beta\geq 1 \qquad (1.2.5)$$

and in continuous case

$$H_{\alpha}^{\beta}(X)=\frac{1}{\beta-\alpha}\log\int_0^{\infty} f^{\alpha+\beta-1}(x)dx \qquad \beta-1<\alpha<\beta, \beta\geq 1 \qquad (1.2.6)$$

For $\beta=1,\ \alpha\rightarrow 1,$ the measure (1.2.5) and (1.2.6) reduces to (1.1.2) and (1.1.3) respectively.

***iv)*** ***Arimoto'sEntropy:***Armito[6] introduced the generalized entropy as

$$A_{\alpha}(P)=\frac{1}{2^{\alpha-1}-1}\left\{\left(\sum_{i=1}^{n}p_i^{\frac{1}{\alpha}}\right)^{\alpha}-1\right\} \qquad \alpha>0(\neq 1) \qquad (1.2.7)$$

and in continuous case

$$A_{\alpha}(X)=\frac{1}{2^{\alpha-1}-1}\left\{\left(\int_0^{\infty} f^{\frac{1}{\alpha}}(x)dx\right)^{\alpha}-1\right\} \qquad \alpha>0(\neq 1) \qquad (1.2.8)$$

For $\alpha\rightarrow 1$, (1.2.7) and (1.2.8) reduces to (1.1.2) and (1.2.3) respectively.

***v)*** ***Boekee and Lubbe's Entropy:***Boekee and Lubbe[21] introduced the generalized entropy as

$$H_R(P)=\frac{R}{R-1}\left\{1-\left(\sum_{i=1}^{n}p_i^R\right)^{\frac{1}{R}}\right\} \qquad R>0(\neq 1) \quad (1.2.9)$$

and in continuous case

$$H_R(X)=\frac{R}{R-1}\left\{1-\left(\int_0^\infty f^R(x)dx\right)^{\frac{1}{R}}\right\} \qquad R>0(\neq 1) \quad (1.2.10)$$

For $R\to 1$, (1.2.9) and (1.2.10) reduces to Shannon's entropy given in (1.1.2) and (1.1.3) respectively.

***vi)*** ***Kapur's Entropy:***Kapur[72] generalized the Shannon's entropy as

$$H_{\alpha,\beta}(P)=\frac{1}{1-\alpha}\log\frac{\sum_{i=1}^{n}p_i^{\alpha+\beta-1}}{\sum_{i=1}^{n}p_i^{\beta}} \qquad (1.2.11)$$

where $p_i\geq 0,\ \alpha>(\neq 1),\ \ \beta>\alpha,\ \ \alpha+\beta-1>0$

and in continuous case

$$H_{\alpha,\beta}(X)=\frac{1}{1-\alpha}\log\frac{\int_0^\infty f^{\alpha+\beta-1}dx}{\int_0^\infty f^{\beta}(x)dx} \qquad \alpha>(\neq 1),\ \ \beta>0 \qquad (1.2.12)$$

For $\beta=1,\ \alpha\to 1,$ the measure (1.2.11) and (1.2.12) reduces to (1.1.2) and (1.1.3) respectively.

***vii)*** ***Sharma and Mittal's Entropy:*** Sharma and Mittal [123] introduced the generalized entropies as

$$H_\alpha(P)=\frac{1}{\left(2^{1-\alpha}-1\right)}\left[\exp\left((\alpha-1)\sum_{i=1}^{n}p_i\log p_i\right)-1\right]\qquad \alpha>(\neq 1)$$

(1.2.13)

and in continuous case

$$H_\alpha(X)=\frac{1}{\left(2^{1-\alpha}-1\right)}\left[\exp\left((\alpha-1)\int_0^\infty f(x)\log f(x)dx\right)-1\right]\qquad \alpha>(\neq 1)$$

(1.2.14)

For $\alpha$ =1,(1.2.13) and (1.2.14) reduces to (1.1.2) and (1.1.3) respectively.

***viii)*** ***Sharma and Taneja's Entropy:*** Sharma and Taneja[124] introduced the generalized entropies as:

$$H_\alpha(P)=-2^{\alpha-1}\sum_{i=1}^{n}p_i^\alpha\log p_i\qquad \alpha>0\qquad (1.2.15)$$

For $\alpha$ =1, (1.2.15) reduces to (1.1.2).

$$H_{\alpha,\beta}(P)=\frac{1}{2^{1-\beta}-2^{1-\alpha}}\sum_{i=1}^{n}\left(p_i^\beta\log p_i^\alpha\right)\qquad \alpha\neq\beta,\ \alpha,\ \beta>0$$

(1.2.16)

For $\beta=1,\ \alpha\to 1,$ (1.2.16) reduces to Shannon's entropy (1.1.2).

For continuous cases, we have the following generalizations

$$H_\alpha(X)=-2^{\alpha-1}\int_0^\infty f^\alpha(x)\log f(x)dx\qquad \alpha>0\qquad (1.2.17)$$

$$H_{\alpha\beta}(X)=\frac{1}{2^{1-\beta}-2^{1-\alpha}}\int_0^\infty\left(f^\beta(x)-f^\alpha(x)\right)dx\qquad \alpha\neq\beta,\alpha,\beta>0$$

(1.2.18)

For $\alpha$ = 1, (1.2.17) reduces to (1.1.3) and for $\beta=1$, $\alpha\to 1$, (1.2.18) reduces to Shannon's entropy (1.1.3).

***ix)*** ***Kerridge Inaccuracy:*** Suppose that an experiment asserts that the probabilities of $n$ events are $Q=(q_1,q_2,\ldots,q_n)$, while their true probabilities are $P=(p_1, p_2, \ldots, p_m)$, then the Kerridge[78] has proposed the inaccuracy measure as

$$K(P;Q) = -\sum_{i=1}^{n} p_i \log q_i \tag{1.2.19}$$

When $p_i = q_i \ \forall i = 1,2,\ldots,n,$ then (1.2.19) reduces to Shannon's entropy.

In case of continuous distribution

$$K(f;g) = \int_0^\infty f(x) \log g(x) dx \tag{1.2.20}$$

## 1.3 Joint and Conditional Entropy

***1.3.1 Joint Entropy:*** The joint entropy $H(X;Y)$ of a pair of discrete random variable $(X;Y)$ with a joint distribution function $p(x;y)$ is defined as

$$H(X;Y) = -\sum_{x\in R}\sum_{y\in R} p(x;y) \log p(x;y) \tag{1.3.1}$$

and in continuous case

$$H(X;Y) = -\int_0^\infty\int_0^\infty f(x;y) \log f(x;y) dx\, dy \tag{1.3.2}$$

where $f(x;y)$ is the joint density function of the random variable $X$ and $Y$.

***1.3.2 Conditional Entropy:*** The entropy is meant to measure the uncertainty in the realization of $X$. Now, we want to quantify how much uncertainty does the realization of a random variable $X$ have if the outcome of another randomvariable $Y$ is known. This is called conditional entropy and is given by:

$$H(X \mid Y) = -\sum_{x,y} P(x,y) \log \frac{P(x,y)}{P(y)} \tag{1.3.3}$$

where $p(y)$ is the marginal distribution of $Y$.

And in continuous case

$$H(X \mid Y) = -\int_{0}^{\infty} f(x, y) \log \frac{f(x, y)}{f(y)} dx \tag{1.3.4}$$

## 1.4 The Survival Analysis

***1.4.1 Cumulative Distribution Function:*** If $X$ is a continuous random variable with the probability density function $f(x)$, then the function

$$F_X(t) = P(X \le t) = \int_{-\infty}^{t} f(x) dx, \qquad -\infty < x < t$$

is called cumulative distribution function of the random variable $X$. The distribution function has the following properties:

(i) $F(t)$ is non-decreasing function in $t$, i.e.

$$F'(t) = \frac{d}{dt} F(t) = f(t) \ge 0.$$

(ii) $F(-\infty) = 0$ and $F(+\infty) = 1$, which implies that $0 \le F(x) \le 1$.

(iii) $F(t)$ is a continuous function of $t$ on the right.

(iv) It may be noted that

$$P(a \le X \le b) = F(b) - F(a).$$

Similarly

$$P(a \le X \le b) = \int_{a}^{b} f(t) dt .$$

***1.4.2 Survival Function:*** The basic quantity employed to describe time-to-event phenomena is the survival function. This function, also known as reliability function is the probability that an individual survives beyond time $t$. If $X$ is a continuous random variable then the survival function which is usually denoted by $\bar{F}_t$, is defined by

$$\bar{F}_t = P(X \ge t) = \int_{t}^{\infty} f(x) dx \tag{1.4.1}$$

In the context of equipment or manufactured item failures, $\overline{F}_t$ is referred to as the reliability function. Note that the survival function is a non-increasing function with $\overline{F}_t(0)=1$and $\overline{F}_t(\infty)=0$.

Thus, we have the following relationship between reliability function and distribution function

$$\overline{F}_t = 1 - F(t) \tag{1.4.2}$$

Differentiating (1.4.2) both sides with respect to $t$, we have

$$\frac{d}{dx}\overline{F}_t = -f(t)$$

or

$$f(t) = -\frac{d}{dt}\overline{F}_t \tag{1.4.3}$$

***1.4.3 The Hazard Rate Function:*** It is the probability that the item will fail in the next $\delta t$ time unit given that the item is functioning properly in time '$t$'. In other words, failure rate or hazard rate function is defined as the conditional probability of failure between $(t, t + \delta t)$ given that there is no failure up to time $t$,

$$r_F(t) = \lim_{\delta t \to 0} P[t \le T \le t + \delta / T > t] \tag{1.4.4}$$

or

$$r_F(t) = \frac{f(t)}{\overline{F}(t)} \tag{1.4.5}$$

If $F(t)$ be the distribution of time to failure and $f(t)$ be the p.d.f, then we have

$$\overline{F}_t = 1 - F(t) \tag{1.4.6}$$

Therefore (1.4.5) reduces to

$$r_F(t) = \frac{f(t)}{1 - F(t)} \tag{1.4.7}$$

***1.4.4 Distribution with Increasing Failure Rate (IFR) and Decreasing Failure Rate (DFR):***Its often difficult to single out a specific model to characterized behaviour of a system or a device consequently a less conventional approach. Where in the failure behaviour is characterized merely by property of hazard rate (failure rate) is often to found to be quite useful. Such an approach has not only allevaited the task of specifying failure model but also initiated the develop of comprehensive theory of reliability. The measure of an equipment reliability in frequency with which failure occur in time. A failure distribution represents an attempt to discrete mathematically, the length of the life of the material or a device, there are many physical causes that individual or collectively may be responsible for the failure of the device at any particular instance, the hazard function describes the way in which the instance probability of death individual change with time. Let $F(t)$ be the distribution of time to failure and $f(t)$ be the p.d.f. then hazard rate is defined as

$$r_F(t) = \frac{f(t)}{\overline{F}(t)} \tag{1.4.8}$$

We have $\bar{F}(t) = 1 - F(t)$, then

$$r_F(t) = \frac{f(t)}{1 - \overline{F}(t)} \tag{1.4.9}$$

$F(t)$is called survival function denoted as $\overline{F}(t)$. The failure rate which is the function of time has probabilistic interpretation namely $r_F(t)\partial t$ represents the probability that a device of age $'t'$ will fail in interval $[t+(t+\delta t)]$ or $r_F(t) = p$ {a device of age $'t'$ will fail in interval $(t, t+\delta t)$ device of function in time $'t'$}. Now in considering a life testing model, it is often more informative to consider the properties of hazard function then characterized the model, in terms of p.d.f or c.d.f. directly the monotonically hazard rate is an important consideration. If $r_F(t)$ is Hazard Function (HF), such that

$$t_1 \le t_2, \qquad r_F(t_1) = r_F(t_2) \tag{1.4.10}$$

The considering model is said to be increasing failure rate (IFR).

If $r_F(t)$ is Failure Function, such that

$$t_1 \geq t_2, \qquad r_F(t_1) \geq r_F(t_2) \tag{1.4.11}$$

The corresponding model is said to be decreasing failure rate (DFR).

A decreasing failure rate might be interpreted an improvement of unit with age.

Several alternatives criteria for assisting whether F is (IFR) DFR distribution exists.

1. F is increasing (decreasing) failure rate distribution if $F(t+\delta t)-F(t)/1-F(t)$ is increasing (decreasing) in time '$t$' for all $\delta t > 0$.
2. F is increasing (decreasing) failure rate, if $\log[(1-F(t))]$ is concave (convex) for all $\delta t > 0$.

***1.4.5 Average Failure Rate (AFR):*** The average failure rate is defined in terms of the function

$$A(X) = \frac{1}{x}\left[\int_0^x h(t)dt\right] \tag{1.4.12}$$

A life testing model is said to have an increasing failure rate average (IFRA) if $x_1 \leq x_2$ implies $A(x_1) \geq A(x_2)$. On the other hand, a life testing model is said to have a decreasing failure rate average (DFRA) if $x_1 \leq x_2$ implies $(x_1) \geq A(x_2)$.

***1.4.6 The Mean Residual Life:*** The mean residual life (MRL) is denoted by $m_F(t)$ and is defined by

$$m_F(t) = \frac{\int_x^\infty (t-x)f(t)dt}{1-F(x)} \tag{1.4.13}$$

The MRL is the generalization of the mean life of a unit, since

$$m_F(0) = \int_x^{\infty} tf(t)dt = E(x) \qquad (1.4.14)$$

One possible interpretation of the MRL involves the conditional distribution of$X$ given $X > x$ in particular for a fixed, consider the $f(t/X > x) = \dfrac{f(t)}{1 - F(t)}$, function , if $t > x$ and zero otherwise the function $f(t/X > x)$ is the conditional p.d.f. of $x$ given $X > x$ and consequently. The MRL is the conditional expectation of $X - x$ given $X > x$, $\mu(x) = E(X/X > x) - x$. In other words MRL is the average amount of unused life of unit at age *x*. A lifetime model is said to have a decreasing mean residual life (DMRL), if $x_1 \le x_2$ implies $m_F(x_1) \ge m_F(x_2)$. On the other hand a life-testing model is said to have an increasing mean residual life (IMRL), if $x_1 \le x_2$ implies $m_F(x_1) \ge m_F(x_2)$.

### *1.4.7 Some Characterization Results*

We have from (1.4.1)

$$\overline{F}_t(x) = \exp\left\{-\int_0^t h(x)dx\right\}$$

$$= \exp\{S(t)\}.$$

Therefore

$$\overline{F}_t(x) = \frac{f(0)}{f(t)} \exp\left\{-\int_0^t \frac{1}{f(x)} dx\right\} \qquad (1.4.15)$$

But,

$$f(t) = -\frac{d}{dt} \overline{F}_t(x)$$

$$= h(t)\overline{F}_t(x).$$

Thus

$$f(t)=\left(\frac{d}{dt}f(t)+1\right)\frac{f(0)}{f^{2}(t)}\exp\left\{-\int_{0}^{t}\frac{1}{f(x)}dx\right\} \qquad (1.4.16)$$

Dividing (1.4.15) to (1.4.16), we get

$$\frac{f(t)}{\overline{F}_t(x)}=\frac{\left(\frac{d}{dt}f(t)+1\right)}{f(t)}$$

Thus

$$h(t)=\frac{f'(t)+1}{f(t)} \qquad (1.4.17)$$

where $f'(t)=\frac{d}{dt}f(t)$. Equation (1.4.11) gives the functional relationship between hazard rate function and mean residual life function. It has a pivot role in some characterization results of lifetime models by using the information theoretic approach.

### *1.4.8 Reversed Hazard Rate Function*

The concept of reversed hazard rate was initially introduced as the hazard rate in the negative direction and received the cold reception in the literature at the early stage. This was because reversed hazard rate, being the ratio of probability density function and the corresponding distribution function, was conceived as a dual measure of hazard rate.

Keilson and Sumita[77] were among the first to define reversed hazard rate and calledit the dual failure function. According to them, hazard rate ordering is the uniform stochastic ordering and the reversed hazard rate ordering is the uniform stochastic ordering in the negative direction. This has followed by Shaked and Shanthikumar[121],who have presented some nice results relating to reversed hazard rate function. Also,whatis important is the inclusion of some interestingcharacterizations based on the monotonicity of reversed hazard rate function.

Let $X$ be a continuous random variable with density function $f(x)$, cumulative distribution function $F(x)$and survival function $R(x)$. Then, the reversed hazard rate of $X$ at $t$ is denoted by $\tau(t)$ and is defined as

$$\tau(t) = \frac{d}{dt} \log F(t) = \frac{f(t)}{F(t)}$$

The following relationship can be easily obtained

$$F(t) = \exp\left\{ -\int_{t}^{\infty} \tau(x) dx \right\}.$$

### *1.4.9 Discrete Case*

Let$X$ is a discrete random variabletaking the values $x_1 < x_2 < \Lambda < x_n$ with the probability mass function, $P(j) = P(X = j)$, $j = 1,2,\mathrm{K}, n$ then the survival function is defined as

$$R(j)=\sum_{k=j}^{n}P(k) \tag{1.4.18}$$

The survival function and the probability mass function are related by

$$P(j)=R(j)-R(j+1) \tag{1.4.19}$$

The hazard function is defined as

$$h(j)=\frac{P(j)}{R(j)} \tag{1.4.20}$$

Using (1.4.20), we have

$$h(j)=1-\frac{R(j+1)}{R(j)} \tag{1.4.21}$$

The survival function is related to the hazard rate function by

$$R(j)=\prod_{k=j}^{n}[1-h(k)] \tag{1.4.22}$$

For discrete lifetimes the cumulative hazard function is defined as

$$S(j)=\sum_{k=j}^{n}h(k) \tag{1.4.23}$$

The characterization relationship between survival function, hazard rate function and mean residual lifetime can be developed as:

We have from (1.4.18)

$$R(j)=\sum_{k=j}^{n}P(k)=\prod_{k=j}^{n}[1-h(k)].$$

If $X$ is an integer valued random variable with mean residual life at time $k$ equal to $m_k$, $k=0,1,2,\ldots$ and $m_0$ is finite then we have

$$R(k)=\frac{1+m_0}{m_k}\prod_{j=0}^{k}\frac{m_j}{1+m_j} \tag{1.4.24}$$

Also, for any discrete survival function, we have

$$P(j) = R(j) - R(j+1)$$

$$= h(j)R(j).$$

Therefore,

$$h(j) = \frac{P(j)}{R(j)}.$$

## 1.5 Classes of Aging Distributions

An important characteristic of survival distribution is its aging properties. There arenumber of classes that have been suggested in the literature to categorize distributions based on their aging properties or their dual. The first class is the class of increasing hazard rate *(IHR)*distributions and the dual class of decreasing hazard rate *(DHR)* distributions. A Survival distribution is said to be in the *IHR (DHR)* class if and only if$h(t)$ is increasing (decreasing) for all $t$.

A second more general aging class is the class of increasing (decreasing) hazard rate on the average, *IHRA (DHRA)* distributions. A distribution is said tofall in the *IHRA (DHRA)* class if and only if

$$-\left(\frac{1}{t}\right)\log[R(t)] \tag{1.5.1}$$

is increasing (decreasing) in $t$.

The definition arises by declaring a distribution to be in the *IHRA* class when its cumulative hazard rate, $-\log[S(t)]$ is increasing faster than the cumulative hazard rate of an exponential random variable. Since the exponential distribution reflects a model with no aging, this class is one of distributions for which individuals are on the average aging.

Since (1.5.1) implies that $R^{-\frac{1}{t}}(t)$ is increasing in $t$, we have that $X$ is in *IHRA* class if and only if $R(\theta\, t) \geq R^{\theta}(t)$.

A third aging class is the class of decreasing (increasing) mean residual life, *DMRL (IMRL)* distributions. A distribution is said to be *DMRL (IMRL)* class if

$$r(t) = \frac{\int_t^{\infty} R(x)dx}{R(t)} \tag{1.5.2}$$

is increasing (decreasing) in $t$.

This aging class, which include all *IHR* models, is one where the mean remaining life of anindividual of age *t* is becoming shorter as *t* increases.

## 1.6 Some Mathematical Functions

***1.6.1 Convex Function:*** A real valued function $f(x)$ defined on $(a, b)$is said to be convex function if for every $\alpha$ such that $0 \le \alpha \le 1$ and for any two points $x_1$ and $x_2$ such that $a < x_1 < x_2 < b$, we have

$$f(\alpha x_1 + (1-\alpha)x_2) \le \alpha f(x_1) + (1-\alpha) f(x_2) \tag{1.6.1}$$

If we put = ½, then (1.6.1) reduces to

$$f\left(\frac{x_1 + x_2}{2}\right) \le \frac{f(x_1) + f(x_2)}{2} \tag{1.6.2}$$

which is taken as the definition of convexity.

**Remark 1.6.1:** If $f''(x) \ge 0$, then $f(x)$ is convex function.

***1.6.2 Strictly Convex Function:*** A real valued function $f(x)$ defined on $(a, b)$ is said to be strictly convex function if for every $\alpha$ such that $0 < \alpha < 1$ and for any two points $x_1$ and $x_2$ in $(a, b)$, we have

$$f(\alpha x_1 + (1-\alpha)x_2) < \alpha f(x_1) + (1-\alpha) f(x_2) \tag{1.6.3}$$

**Remark 1.6.2:** If $f''(x) > 0$, then $f(x)$is strictly convex function.

***1.6.3 Concave Function:*** A function $f(x)$ is said to be concave if $-f(x)$ is convex.

**Remark 1.6.3:** If $f''(x) \le 0$, then $f(x)$ isconcave function.

***1.6.4 Strictly Concave Function:*** A function $f(x)$ is said to be strictly concave if $-f(x)$ is strictly convex.

**Remark 1.6.4:** $f''(x) > 0$, then $f(x)$ is strictly concave function.

***1.6.5 IncreasingFunction*****:** Let $I$ be an open interval contained in the domain of a real function. The function $f(x)$ is an increasing function on $I$ if $x_1 < x_2$ in $I$, implies

$$f(x_1) \le f(x_2).$$

***1.6.6 Decreasing Function:*** Let $I$ be an open interval contained in the domain of a real function. The function $f(x)$ is a decreasing function on $I$ if $x_1 < x_2$ in $I$, implies

$$f(x_1) \ge f(x_2).$$

***1.6.7 Maximum of a Function:*** A function $f(x)$ is said to have a maximum value in an interval $I$ at $x_0$, if $f(x_0) \ge f(x)$ for all $x$ in $I$.

***1.6.8 Minimum of a Function:*** A function $f(x)$ is said to have a minimum value in an interval $I$ at $x_0$, if $f(x_0) \le f(x)$ for all $x$ in $I$.

The following theorems gives the working rule for finding the points of local maxima or points of local minima. The proof is simple and hence omitted.

***Theorem 1.6.1:*** (First derivative test) Let $f(x)$ be a differentiable function on $I$ and let $x_0 \in I$. Then

(a) $x_0$ is a point of local maximum of $f(x)$ if

$i)$ $f'(x) = 0$

$ii)$ $f'(x) > 0$ at every point close to the left of $x_0$ and $f'(x) < 0$ at every point close to the right of $x_0$.

(b) $x_0$ is a point of local minimum of $f(x)$ if

$i)$ $f'(x_0) = 0$

$ii)$ $f'(x) < 0$ at every point close to the left of $x_0$ and $f'(x) > 0$ at every point close to the right of $x_0$.

***Theorem 1.6.2:*** (Second derivative) Let $f(x)$ be a differential function on $I$ and let $x_0 \in I$. Let $f''(x)$ be continuous at $x_0$. Then

(i) $x_0$ is a local maximum if both $f'(x_0) = 0$ and $f''(x_0) < 0$.

(ii) $x_0$ is a local minimum if both $f'(x_0) = 0$ and $f''(x_0) > 0$.

***1.6.9 Gamma Function:*** If $n > 0$, then the integral $\int_0^\infty x^{n-1} e^{-x} dx$ which is a function of $n$, is called a Gamma function and is denoted by $\Gamma(n)$. Thus

$$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx \qquad \forall n > 0 \tag{1.6.4}$$

***1.6.10 Properties of Gamma Function:*** The gamma function has the following properties

(i) For $n > 1$,

$$\Gamma(n) = (n-1)\Gamma(n-1).$$

(ii) When $n$ is a positive integer, then

$$\Gamma(n) = (n-1)!.$$

***1.6.11 Digamma Function:*** The logarithmic derivative of the gamma function is called digamma function and is given by

$$\Psi(n) = \frac{d}{dn}\log\Gamma(n) = \frac{\Gamma'(n)}{\Gamma(n)} \tag{1.6.5}$$

***1.6.12 Beta Function*:** If*m, n>0,* then the integral $\int_0^1 x^{m-1}(1-x)^{n-1}dx$, which is a function of $m$ and $n$ is called the beta function and is denoted by

$$\beta(m;n) = \int_0^1 x^{m-1}(1-x)^{n-1}dx \qquad \forall m,\ n > 0 \tag{1.6.6}$$

***1.6.13 Properties of Beta Function:*** Following are the properties of Beta function

(i) Beta function is symmetric i.e., $\beta(m;n) = \beta(m;n)$.

(ii) If $m,\ n$ are positive integers, then

$$\beta(m;n) = \frac{(m-1)!(n-1)!}{(m+n-1)!} \tag{1.6.7}$$

Following is the relationship between Beta and Gamma functions

$$\beta(m;n) = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)} \tag{1.6.7}$$

***1.6.14 Leibniz Integration Rule:*** The Leibniz integral rule gives a formula for differentiation of a definite integral whose limits are the functions of the differential variable. It states that

$$\frac{\delta}{\delta z}\int_{z(z)}^{b(z)} f(x;z) = \int_{z(z)}^{b(z)} \frac{\delta f(x;z)}{\delta z}dx + f(b(z),z)\frac{\delta b(z)}{\delta z} - f(a(z),z)\frac{\delta a(z)}{\delta z} \tag{1.6.9}$$

It is important to note that, if $a(z)$ and $b(z)$are constants, then the last two terms of (1.6.9) vanishes.

## 1.7 Some inequalities

***i)*** ***Jensen's inequality:*** If $X$ is a random variable such that $E(X)=\mu$ exists and $f(x)$ is a convex function, then

$$E[f(X)] \geq f[E(X)] \tag{1.7.1}$$

with equality iff the random variable $X$ has a degenerate distribution at $\mu$.

***ii)*** ***Holder's Inequality:*** If $x_i, y_i > 0,\ \ i = 1,2,\ldots,n$ and $\frac{1}{p}+\frac{1}{q}=1,\ p>1,$ then the following inequality holds

$$\sum_{i=1}^{n} x_i y_i \leq \left[\sum_{i=1}^{n} x_i^p\right]^{\frac{1}{p}} \left[\sum_{i=1}^{n} y_i^q\right]^{\frac{1}{q}} \tag{1.7.2}$$

***iii)*** ***Chebychev's Inequality:*** If $X$ is a random variable with mean $\mu$ and variance $\sigma^2$, then for any positive number $k$

$$P\{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2} \tag{1.7.3}$$

or

$$P\{|X-\mu| < k\sigma\} \geq 1-\frac{1}{k^2}\ .$$

***iv)*** ***Bienayne-Chebychev's Inequality:*** Let $g(x)$ be a non-negative function of a random variable $X$, then for any $k>0$,

$$P\{g(x) \geq k\} \leq \frac{E[g(x)]}{k} \tag{1.7.4}$$

***v)*** ***Markov's Inequality:*** If we take $g(x) = |x|$ in (1.7.4), then

$$P\{|x| \geq k\} \leq \frac{E|x|}{k} \tag{1.7.5}$$

which is Markov's Inequality.

Taking, $g(x) = |x|^r$ and replacing $k$ by $k^r$ in (1.7.4), we get a more generalized form of Markov's inequality

$$P\left\{|x|^r \geq k^r\right\} \leq \frac{E|x|^r}{k^r} \tag{1.7.6}$$

***vi)*** ***Log Sum Inequality:*** For non-negative numbers $a_1, a_2, \dots, a_n$ and $b_1, b_2, \dots, b_n$, the log sum inequality is given as

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \sum_{i=1}^{n} a_i \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \tag{1.7.7}$$

with equality, iff $\frac{a_i}{b_i} = k,$ where $k$ is a constant.

## 1.8 Divergence Measures:

Kullback and Leibler introduced the idea of relative information. Sometimes it is called cross entropy, directed divergence and measure of discrimination. The entropy of a random variable is a measure of the uncertainty of the random variable; it is a measure of the amount of information required on the average to describethe random variable. The relative information is a measure of the distance between two distributions, it arises as an expected logarithm of the likelihood ratio. According to the second law of thermodynamics, for a Markov chain, the relative information decreases with time. The relationship between information theory and thermodynamics has been discussed extensively by Brillouinsand Jaynes.

There exists several divergence measure in the literature of information theory, some of them are given here:

(1) $x^2$ (chi square) divergence

Pearson introduced the measure as

$$x^2(P//Q) = \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i} = \sum_{i=1}^{n} \frac{p_i^2}{q_i} - 1 \tag{1.8.1}$$

(2) Kullback-Leibler'srelative information

Kullback-Leibler introduced the divergence measure as

$$K(P//Q)=\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \tag{1.8.2}$$

(3) Relative Jensen-Shannon divergence measure

Sibson and Lin introduced the divergence measure as

$$F(P//Q)=\sum_{i=1}^{n} p_i \log\left(\frac{2p_i}{p_i+q_i}\right) \tag{1.8.3}$$

(4) J-Divergence measure

Jeffrey, Kullback and Leibler introduced the divergence measure as

$$J(P//Q)=\sum_{i=1}^{n} (p_i-q_i)\log \frac{p_i}{q_i} \tag{1.8.4}$$

(5) Hellinger discrimination measure

Hellinger introduced the divergence measure as

$$h(P//Q)=\frac{1}{2}\sum_{i=1}^{n}\left(\sqrt{p_i}-\sqrt{q_i}\right)^2 \tag{1.8.5}$$

# Chapter II: Some NeChapter-II Generalization of

## 2.1 Introduction

The concept of entropy in communication theory was first introduced by Shannon [122] and it was then realized that entropy is a property of any stochastic system and the concept is now used widely in different disciplines. The tendency of the systems to become more disordered over time is described by the second law of thermodynamics, which states that the entropy of the system cannot spontaneously descrease. Today, information theory is still principally concerned with communications systems, but there are widespread applications in statistics, information processing and computing. A great deal of insight is obtained by considering entropy equivalent to uncertainty, the generalized theory of which has well been explained by Zadeh [140].

The uncertainty associated with probability of outcomes, known as probabilistic uncertainty, is called entropy, since this is the terminology that is well entrenched in the literature. Shannon [122] introduced the concept of information theoretic entropy by associating uncertainty with every probability distribution $P=\left(p_1, p_2, \Lambda\ , p_n\right)$ and found that there is a unique function that can measure the uncertainty, is given by

$$H(P) = -\sum_{i=1}^{n} p_i \log P_i \qquad 0 \le p_i \le 1 \qquad (2.1.1)$$

The probabilistic measure of entropy (2.1.1) possesses a number of interesting properties. Immediately, after Shannon gave his measure, research workers in many fields saw the potential of the application of this expression and a large number of other measures of information theoretic entropies were derived. Renyi [117] defined entropy of order $\alpha$ as

$$H_\alpha(P)=\frac{1}{1-\alpha}\log\left(\frac{\sum_{i=1}^{n}P_i^\alpha}{\sum_{i=1}^{n}P_i}\right), \quad \alpha\neq 1, \alpha>0 \qquad (2.1.2)$$

which includes Shannon's [122] entropy as a limiting case as $\alpha\to 1$. Zyczkowski [142] explored the relationships between the Shannon's [122] entropy and Renyi's [117] entropies of integer order.

Havrada and Charvat [63] introduced first non-additive entropy, given by

$$H^\alpha(P)=\frac{\left[\sum_{i=1}^{n}P_i^\alpha\right]-1}{2^{1-\alpha}-1}, \qquad \alpha\neq 1, \alpha>0 \qquad (2.1.3)$$

Kapur [72] generalized Renyi's [117] measure further to give a measure of entropy of order '$\alpha$' and type '$\beta$' viz.

$$H_{\alpha,\beta}(P)=\frac{1}{1-\alpha}\log\left[\frac{\sum_{i=1}^{n}P_i^{\alpha+\beta-1}}{\sum_{i=1}^{n}P_i^\beta}\right] \quad \alpha\neq 1,\ \alpha>0,\ \beta>0,\quad \alpha+\beta-1>0 \quad (2.1.4)$$

The measure (2.1.4) reduces to Renyi's [117] measure when $\beta=1,$ to Shannon's [122] measure when $\beta=1,\ \alpha\to 1.$ When $\beta=1, \alpha\to\infty,$ it gives the measures

$$H_\infty(P)=-\log P_{\max}.$$

Many other probabilistic measures of entropy have been discussed and derived by Brissaud [23], Chakrabarti [25], Chen [28], GarbacZewski [55], Herremoes [64], Lavenda [84], Nanda and Paul [96]. Rao Yunmei and Wang [115], Sergio [119], Sharma and Taneja [124] etc. The applications of the results obtained by various authors have been provided to various fields of Mathematical Sciences. In section 2.2 a new generalized probabilistic information theoretic measure have been presented.

## 2.2 New Generalized Information Theoretic Measure

In this section, a new generalized information measure for a probability distribution $P = \left\{ (p_1, p_2, \Lambda\ , p_n), p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\}$ and their essential and desirable properties have been discussed. This generalized entropy depending upon *n* real parameters $\alpha_1, \alpha_2, \Lambda\ , \alpha_n$ is given by the following mathematical expression.

$$H^{n}_{\alpha,\alpha_1,\Lambda\ ,\alpha_n}(P) = \frac{\sum_{i=1}^{n} P_i^{\alpha+\alpha_1+\Lambda\ +\alpha_n} - 1}{2^{1-\alpha-\alpha_1 \Lambda\ \alpha_n} - 1} \tag{2.2.1}$$

where

$$\alpha + \sum_{i=1}^{n} \alpha_i \neq 1,\ \alpha + \sum_{i=1}^{n} \alpha_i > 1,\ \text{and}\ \alpha \neq 1,\ \alpha > 0,\ \alpha_i \geq 0 \tag{2.2.2}$$

If $\sum_{i=1}^{n} \alpha_i = 0$ then $\alpha \neq 1, \alpha > 0$. Thus, we see that the proposed measure (2.2.1) becomes

$$H^{n}_{\alpha}(P) = \frac{\sum_{i=1}^{n} P_i^{\alpha} - 1}{2^{1-\alpha} - 1} \tag{2.2.3}$$

which is Havorada and Charvat's [63] measure of entropy of order $\alpha$. The measure (2.2.3) again reduces to Shannon's [122] measure of entropy as $\alpha \rightarrow 1$. Thus, we see that the measure proposed in equation (2.2.1) is a generalized measure of entropy. Next, we present some important properties of this generalized measure. The measure (2.2.1) satisfies the following properties:

i) It is continuous function of $p_1, p_2, \ldots, p_n$ so, that it changes by a small amount when $p_1, p_2, \ldots, p_n$ change by small amounts.

ii) It is permutationally symmetric function of $p_1, p_2, \ldots, p_n$ i.e., it does not change when $p_1, p_2, \ldots, p_n$ are permuted among themselves.

iii) $H^{n}_{\alpha,\alpha_1,\ldots,\alpha_n}(P) \geq 0.$

iv) $$H^{n+1}_{\alpha,\alpha_1,\ldots,\alpha_n}(p_1, p_2, \ldots, p_n, 0) = \frac{\sum_{i=1}^{n} p_i^{\alpha+\alpha_1+\ldots+\alpha_n} - 1}{2^{1-\alpha-\alpha_1-\ldots-\alpha_n} - 1}$$

$$= H^{n}_{\alpha,\alpha_1,\ldots,\alpha_n}(P).$$

This property says that entropy does not change by the inclusion of an impossible event with probability zero.

v) Since $H^{n}_{\alpha,\alpha_1,\ldots,\alpha_n}(P)$ is an entropy measure, its maximum value must occur. To find the maximum value, we proceed as follows:

let

$$f(P) = \frac{\sum_{i=1}^{n} p_i^{\alpha+\alpha_1+\ldots+\alpha_n} - 1}{2^{1-\alpha-\alpha_1-\ldots-\alpha_n} - 1} - \lambda\left(\sum_{i=1}^{n} p_i - 1\right)$$

then, we have

$$\frac{\partial f}{\partial P_1} = \frac{(\alpha + \alpha_1 + \ldots + \alpha_n) P_1^{\alpha+\alpha_1+\ldots+\alpha_n} - 1}{2^{1-\alpha-\alpha_1-\ldots-\alpha_n} - 1} - \lambda$$

$$\frac{\partial f}{\partial P_2} = \frac{(\alpha + \alpha_1 + \ldots + \alpha_n) P_2^{\alpha+\alpha_1+\ldots+\alpha_n} - 1}{2^{1-\alpha-\alpha_1-\ldots-\alpha_n} - 1} - \lambda$$

$$\vdots$$

$$\frac{\partial f}{\partial P_n} = \frac{(\alpha + \alpha_1 + \ldots + \alpha_n) P_n^{\alpha+\alpha_1+\ldots+\alpha_n} - 1}{2^{1-\alpha-\alpha_1-\ldots-\alpha_n} - 1} - \lambda$$

For maximum value, we take

$$\frac{\partial f}{\partial P_1} = \frac{\partial f}{\partial P_2} = \mathrm{K} = \frac{\partial f}{\partial P_n} = 0$$

which gives

$$\frac{(\alpha + \alpha_1 + \mathrm{K} + \alpha_n)P_1^{\alpha+\alpha_1+\mathrm{K}+\alpha_n} - 1}{2^{1-\alpha-\alpha_1-\mathrm{K}-\alpha_n} - 1} = \Lambda = \frac{(\alpha + \alpha_1 + \mathrm{K} + \alpha_n)P_n^{\alpha+\alpha_1+\mathrm{K}+\alpha_n} - 1}{2^{1-\alpha-\alpha_1-\mathrm{K}-\alpha_n} - 1}$$

which is possible only if $p_1 = p_2 = \Lambda = p_n$ thus

$\sum_{i=1}^{n} p_i = 1$ gives $p_1 = p_2 = \Lambda = p_n = \frac{1}{n}$.

Hence, we see that the generalized entropy measure (2.2.1) possesses maximum value and this value is subject to natural constraint $\sum_{i=1}^{n} p_i = 1$ arises when $p_1 = p_2 = \Lambda = p_n = \frac{1}{n}$ this result is most desirable.

vi) The maximum value is an increasing function of *n*. To prove this result we have

$$f(p) = \frac{n^{1-\alpha-\alpha_1-\mathrm{K}-\alpha_n} - 1}{2^{1-\alpha-\alpha_1-\mathrm{K}-\alpha_n} - 1}$$

thus

$$f'(p) = \frac{(1 - \alpha - \alpha_1 - \mathrm{K} - \alpha_n)n^{-\alpha-\alpha_1-\mathrm{K}-\alpha_n}}{2^{1-\alpha-\alpha_1-\mathrm{K}-\alpha_n} - 1}$$

$$= \frac{(\alpha + \alpha_1 + \mathrm{K} + \alpha_n - 1) \cdot 2^{\alpha+\alpha_1+\mathrm{K}+\alpha_n}}{\left(2^{\alpha+\alpha_1+\mathrm{K}+\alpha_n} - 2\right) \cdot n^{\alpha+\alpha_1+\mathrm{K}+\alpha_n}} > 0$$

Since $\alpha + \sum_{i=1}^{n} \alpha_i \neq 1, \quad \alpha + \sum_{i=1}^{n} \alpha_i > 1$

Hence maximum value is an increasing function of *n*.

vii) Recursivity property:

To prove that the measure (2.2.1) is recursive in nature, we consider

$$H^{n-1}_{\alpha,\alpha_1,\ldots,\alpha_n}\left(p_1+p_2, p_3, p_4\ldots p_n\right)$$

$$=\frac{\left(p_1+p_2\right)^{\alpha+\alpha_1+\ldots+\alpha_n}+\sum_{i=3}^{n} p_i^{\alpha+\alpha_1+\ldots+\alpha_n}-1}{2^{1-\alpha-\alpha_1-\ldots-\alpha_n}-1}$$

$$=\frac{\left(p_1+p_2\right)^{\alpha+\alpha_1+\ldots+\alpha_n}-p_1^{\alpha+\alpha_1+\ldots+\alpha_n}-p_2^{\alpha+\alpha_1+\ldots+\alpha_n}+\sum_{i=1}^{n} p_i^{\alpha+\alpha_1+\ldots+\alpha_n}-1}{2^{1-\alpha-\alpha_1-\ldots-\alpha_n}-1}$$

$$=-\left(p_1+p_2\right)^{\alpha+\alpha_1+\ldots+\alpha_n}\left[\frac{\left(\frac{p_1}{p_1+p_2}\right)^{\alpha+\alpha_1+\ldots+\alpha_n}+\left(\frac{p_2}{p_1+p_2}\right)^{\alpha+\alpha_1+\ldots+\alpha_n}-1}{2^{1-\alpha-\alpha_1-\ldots-\alpha_n}-1}\right]$$

$$+\frac{\sum_{i=1}^{n} p_i^{\alpha+\alpha_1+\ldots+\alpha_n}-1}{2^{1-\alpha-\alpha_1-\ldots-\alpha_n}-1}$$

$$=-\left(p_1+p_2\right)^{\alpha+\alpha_1+\ldots+\alpha_n} H_2\left(\frac{p_1}{p_1+p_2}+\frac{p_2}{p_1+p_2}\right)+H^{n}_{\alpha,\alpha_1,\ldots,\alpha_n}\left(p_1\ldots p_n\right)$$

thus, we have proved that

$$H^{n}_{\alpha,\alpha_1\ldots\alpha_n}\left(p_1, p_2\ldots p_n\right)=H^{n-1}_{\alpha,\alpha_1\ldots\alpha_n}\left(p_1+p_2, p_3, p_4\ldots p_n\right)$$
$$+\left(p_1+p_2\right)^{\alpha+\alpha_1+\ldots+\alpha_n} H_2\left(\frac{p_1}{p_1+p_2}+\frac{p_2}{p_1+p_2}\right)$$

this shows that the measure (2.2.1) possesses recursivity property.

viii) Additive property: To show that the measure (2.2.1) is additive, we consider

$$H^{n,m}_{\alpha,\alpha_1\ldots\alpha_n}\left(P\cup Q\right)=\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(p_i q_j\right)^{\alpha+\alpha_1+\ldots+\alpha_n}-1}{2^{1-\alpha-\alpha_1-\ldots-\alpha_n}-1}$$

$$= \left[\frac{\sum_{i=1}^{n} p_i^{\alpha+\alpha_1+\dots+\alpha_n} - 1}{2^{1-\alpha-\alpha_1-\dots-\alpha_n} - 1}\right]\left(\sum_{j=1}^{m} q_j^{\alpha+\alpha_1+\dots+\alpha_n} - 1\right)$$

$$+\left[\frac{\sum_{i=1}^{n} p_i^{\alpha+\alpha_1+\dots+\alpha_n} - 1}{2^{1-\alpha-\alpha_1-\dots-\alpha_n} - 1}\right]+\left[\frac{\sum_{j=1}^{m} q_j^{\alpha+\alpha_1+\dots+\alpha_n} - 1}{2^{1-\alpha-\alpha_1-\dots-\alpha_n} - 1}\right]$$

$$= \left(2^{1-\alpha-\alpha_1-\dots-\alpha_n} - 1\right)\left[\frac{\sum_{i=1}^{n} p_i^{\alpha+\alpha_1+\dots+\alpha_n} - 1}{2^{1-\alpha-\alpha_1-\dots-\alpha_n} - 1}\right]\left[\frac{\sum_{j=1}^{m} q_j^{\alpha+\alpha_1+\dots+\alpha_n} - 1}{2^{1-\alpha-\alpha_1-\dots-\alpha_n} - 1}\right]$$

$$+\left[\frac{\sum_{i=1}^{n} p_i^{\alpha+\alpha_1+\dots+\alpha_n} - 1}{2^{1-\alpha-\alpha_1-\dots-\alpha_n} - 1}\right]+\left[\frac{\sum_{j=1}^{m} q_j^{\alpha+\alpha_1+\dots+\alpha_n} - 1}{2^{1-\alpha-\alpha_1-\dots-\alpha_n} - 1}\right]$$

$$= \left(2^{1-\alpha-\alpha_1-\dots-\alpha_n} - 1\right)\cdot H^n(P)\cdot H^m(Q) + H^n(P) + H^m(Q)$$

which shows that the generalized entropy (2.2.1) is additive.

## 2.3 Interval Entropy

Information theory has attracted the attention of statisticians. Sunoj *et al*. [130] have explored the use of information measures for doubly truncated random variables, which plays a significant role in studying the various aspects of a system when it fails between two time points. In reliability theory and survival analysis, the residual entropy was considered by Ebrahimi and Pellerey [41], which basically measures the expected uncertainty contained in remaining lifetime of a system. The residual entropy has been used to measure the wear and tear of components and to characterize, classify and order distributions of lifetimes by Belzunce *et al*. [18] and Ebrahimi [42]. The notion of past entropy, which can be viewed as the entropy of the inactivity time of a system was introduced in Di Crescenzo and Longobardi [34].

Let $X$ be a non-negative random variable describing a system failure time. We denote the probability density function of $X$ as *f(x)*, the failure distribution as $F(x) = P(X \le x)$ and the survival function as $\bar{F}(x) = P(X > x)$. The Shannon (122) information measure of uncertainty is defined as:

$$H(X) = -E(\log f(X)) = -\int_0^\infty f(x) \log f(x) dx \tag{2.3.1}$$

where log denotes the natural logarithm. Ebrahimi and Pellerey [41] considered the residual entropy of the non-negative random variable $X$ at time $t$ as:

$$H_X(t) = -\int_t^\infty \frac{f(x)}{\bar{F}(t)} \log \frac{f(x)}{\bar{F}(t)} dx. \tag{2.3.2}$$

Given that a system has survived up to time $t, H_X(t)$ essentially measures the uncertainty represented by the remaining lifetime. The residual entropy has been used to measure the wear and tear of systems and to characterize, classify and order distributions of lifetimes, (Belzunce *et al*. [18], Ebrahimi [42] and Ebrahimi and Kirmani [40]. Di Crescenzo and Longobardi [34] introduced the notion of past entropy and motivated its use in real-life situations. They also discussed its relationship with the residual entropy. Formally, the past entropy of $X$ at time $t$ is defined as follows:

$$\bar{H}_X(t) = -\int_0^t \frac{f(x)}{F(t)} \log \frac{f(x)}{F(t)} dx. \tag{2.3.3}$$

Given that the system $X$ has failed at time $t$, $\bar{H}_X(t)$ measures the uncertainty regarding its past lifetime. Now recall that the probability density function of $(X | t_1 < X < t_2)$ for all $0 < t_1 < t_2$ is given by $f(x)/(F(t_2) - F(t_1))$. Sunoj *et al.* [130] considered the notion of interval entropy of $X$ in the interval $(t_1, t_2)$ as the uncertainty contained in $(X | t_1 < X < t_2)$ which is denoted by:

$$IH(t_1,t_2) = -\int_{t_1}^{t_2} \frac{f(x)}{F(t_2)-F(t_1)} \log \frac{f(x)}{F(t_2)-F(t_1)} dx \tag{2.3.4}$$

We can rewrite the interval entropy as

$$IH(t_1,t_2) = 1 - \frac{1}{F(t_2)-F(t_1)} \int_{t_1}^{t_2} f(x) \log r(x) dx$$

$$+ \frac{1}{F(t_2)-F(t_1)} \{ \bar{F}(t_2) \log \bar{F}(t_2) - F(t_1) \log F(t_1)$$

$$+ [\bar{F}(t_2) - \bar{F}(t_1)] \log [F(t_2) - F(t_1)] \}$$

where $r(x) = f(x)/\bar{F}(x)$ is the hazard function of *X*. Note that interval entropy can be negative and also it can be $-\infty$ or $+\infty$. Given that a system has survived up to time $t_1$ and has been found to be down at time $t_2$. $IH(t_1,t_2)$ measures the uncertainty about its lifetimes between $t_1$ and $t_2$. Misagh and Yari [90] introduced a shift-dependent version of $IH(t_1,t_2)$. The entropy (2.2.4) has been used to characterize and ordering random lifetime distributions, (Misagh and Yari [89] and Sunoj *et al.* [130].

The general characterization problem is to obtain when the interval entropy uniquely determines the distribution function. The following proposition attempts to solve this problem. We first give definition of general failure rate (GFR) functions extracted from Navarro and Ruiz [101].

**Definition 2.3.1:** The GFRs of a random variable *X* having density function $f(x)$ and cumulative distribution function $F(x)$ are given by

$$h_1^X(t_1,t_2) = \frac{f(t_1)}{F(t_2)-F(t_1)} \text{ and } h_2^X(t_1,t_2) = \frac{f(t_2)}{F(t_2)-F(t_1)}.$$

**Remark 2.3.1:** GFR functions determine distribution function uniquely (Navarro and Ruiz [101]).

**Proposition 2.3.1:** Let $X$ be a non-negative random variable, and assume $IH(t_1,t_2)$ be increasing with respect to $t_1$ and decreasing with respect to $t_2$. Then $IH(t_1,t_2)$ uniquely determines $F(x)$.

***Proof:*** By differentiating $IH(t_1,t_2)$ with respect to $t_1$, we have

$$\frac{\partial IH_X(t_1,t_2)}{\partial t_1} = h_1(t_1,t_2)[IH(t_1,t_2) - 1 + \log h_1(t_1,t_2)]$$

and

$$\frac{\partial IH_X(t_1,t_2)}{\partial t_2} = -h_2(t_1,t_2)[IH_X(t_1,t_2) - 1 + \log h_2(t_1,t_2)]$$

thus, for fixed $t_1$ and arbitrary $t_2, h_1(t_1,t_2)$ is a positive solution of the following equation.

$$g(x_{t_2}) = x_{t_2}\left[IH_X(t_1,t_2) - 1 + \log x_{t_2}\right] - \frac{\partial IH(t_1,t_2)}{\partial t_1} = 0 \qquad (2.3.5)$$

Similarly, for fixed $t_2$ and arbitrary $t_1$, we have $h_2(t_1,t_2)$ as a positive solution of the following equation:

$$\gamma(Y_{t_1}) = Y_{t_1}\left[IH_X(t_1,t_2) - 1 + \log Y_{t_1}\right] + \frac{\partial IH(t_1,t_2)}{\partial t_2} = 0 \qquad (2.3.6)$$

By differentiating $g$ and $\gamma$ with respect to $xt_2$ and $Yt_1$, we get

$$\frac{\partial g(xt_2)}{\partial xt_2} = \log xt_2 + IH(t_1,t_2)$$

and

$$\frac{\partial \gamma(Yt_1)}{\partial \gamma t_1} = \log \gamma t_1 + IH(t_1,t_2).$$

Furthermore, second-order derivatives of $g$ and $\gamma$ with respect to $xt_2$ and $Yt_1$ are $\frac{1}{xt} > 0$ and $\frac{1}{Yt} > 0$ respectively. Then the functions $g$ and $\gamma$ are minimized at points $xt_2 = e^{-IH(t_1,t_2)}$ and $Yt_1 = e^{-IH(t_1,t_2)}$ respectively. In addition, $g(0) = -\frac{\partial IH(t_1,t_2)}{dt_1} < 0, g(\infty) = \infty$ and $\gamma(0) = -\frac{\partial IH(t_1,t_2)}{\partial t_1} < 0,\ \gamma(\infty) = \infty.$ So, both functions $g$ and $\gamma$ first descrease and then increase with respect to $xt_2$ and $Yt_1$ respectively. Now, $IH(t_1,t_2)$ uniquely determines GFRs and by virtue of Remark 2.3.1 the distribution function.

The effect of monotone transformations on the residual and past entropy has been discussed in Ebrahimi and Kirmani [46] and Di Crescenzo and Longobardi [34] respectively. Following proposition gives similar results for interval entropy.

**Proposition 2.3.2**

Suppose $X$ be a non-negative random variable with cumulative distribution function $F$ and survival function $\overline{F}$; let $Y = \phi(X)$, with $\phi$, strictly increasing and differentiable function, then for all $0 < t_1 < t_2 < \infty$

$$
\begin{aligned}
IH_Y(t_1,t_2) = IH_X\left(\phi^{-1}(t_1), \phi^{-1}(t_2)\right) + \frac{1}{F\left(\phi^{-1}(t_2)\right) - F\left(\phi^{-1}(t_1)\right)}\{E(\log \phi'(X)) \\
- F\left(\phi^{-1}(t_1)\right)E\left(\log \phi'(X) \mid X < \phi^{-1}(t_1)\right) \\
- \overline{F}\left(\phi^{-1}(t_2)\right)E\left(\log \phi'(X) \mid X > \phi^{-1}(t_2)\right)\}
\end{aligned}
$$

***Proof:*** Recalling (2.3.1), the Shannon information of $X$ and $Y$ can be expressed as:

$$H(Y) = H(X) + E(\log \phi'(X)) \tag{2.3.7}$$

From Ebrahimi and Kirmani [46] and from Di Crescenzo and Longobardi [34]. We have

$$H_Y(t_1) = H_X\left(\phi^{-1}(t_2)\right) + E\left(\log \phi'(X) \middle| X > \phi^{-1}(t_1)\right) \tag{2.3.8}$$

and $$\overline{H}_Y(t_1) = \overline{H}_X\left(\phi^{-1}(t_1)\right) + E\left(\log\phi'(X) \middle| X < \phi^{-1}(t_1)\right) \tag{2.3.9}$$

Due to Sunoj *et al*. [131], there holds:

$$H(Y) = H\left(G(t_1), \overline{G}(t_2), 1 - G(t_1) - \overline{G}(t_2)\right) + G(t_1)\overline{H}_Y(t_1) + \overline{G}(t_2)H_Y(t_2)$$

$$+\left[1 - G(t_1) - \overline{G}(t_2)\right]IH_Y(t_2, t_2) \tag{2.3.10}$$

where $G$ and $\overline{G}$ denote distribution and survival functions of y respectively. Substituting $H(Y), \overline{H}_Y(t_1)$ and $H_Y(t_1)$ in (2.3.7), (2.3.8) and (2.3.9) into terms of (2.3.10), we get:

$$H(X) + E(\log\phi'(X)) = \left(F\left(\phi^{-1}(t_2)\right) - F\left(\phi^{-1}(t_1)\right)\right)IH_X(t_1, t_2)$$

$$- F\left(\phi^{-1}(t_1)\right)E\left(\log\phi'(X)\middle|X < \phi^{-1}(t_1)\right)$$

$$- \overline{F}\left(\phi^{-1}(t_2)\right)E\left(\log\phi'(X)\middle|X > \phi^{-1}(t_2)\right)$$

$$+ F\left(\phi^{-1}(t_1)\right)\overline{H}_X\left(\phi^{-1}(t_1)\right) + \overline{F}\left(\phi^{-1}(t_2)\right)H_X\left(\phi^{-1}(t_2)\right)$$

$$+ H\left(F\left(\phi^{-1}(t_1)\right), \overline{F}\left(\phi^{-1}(t_2)\right), F\left(\phi^{-1}(t_2)\right) - F\left(\phi^{-1}(t_1)\right)\right)$$

(2.3.11)

Three terms of the right hand side of (2.3.11) are equal to:

$$\left[F\left(\phi^{-1}(t_2)\right) - F\left(\phi^{-1}(t_1)\right)\right]IH_X\left(\phi^{-1}(t_1), \phi^{-1}(t_2)\right) -$$

and the proof is complete.

**Remark 2.3.2:** Suppose $\phi(X) = F(X)$, then the function $\phi$ satisfies the assumptions of proposition 2.3.2 and uniformly distributed over (0, 1), then:

$$IH_{F(X)}(t_1, t_2) = IH_X\left(F^{-1}(t_1), F^{-1}(t_2)\right) - \frac{1}{t_2 - t_1}\{H(X)$$

$$+t_1 E\left(\log f(X)\middle|X < F^{-1}(t_1)\right) + t_2 E\left(\log f(X)\middle|X > F^{-1}(t_2)\right)\}$$

**Remark 2.3.3:** For all $0 < \theta < t_1$, we get $IH_{X+\theta}(t_1 + t_2) = IH_X(t_1 - \theta, t_2 - \theta)$

**Remark 2.3.4:** Let $Y = aX$ where $a > 0$, then, we have

$$IH_{aX}(t_1, t_2) = IH_X\left(\frac{t_1}{a}, \frac{t_2}{a}\right) + \log a.$$

## 2.4 Informative Distance

In this section, we review some basic definitions and facts for measures of discrimination between two residual and past lifetime distributions. Misargh and G. Yari [91] measure of discrepancy between two random variables at an interval time.

Let *X* and *Y* are two non-negative random variables describing times to failure of two systems. We denote the probability density functions of *X* and *Y* as $f(x)$ and $g(y)$, failure distributions as $F(x) = P(X \le x)$and $G(y) = G(Y \le y)$ and survival functions as $\overline{F}(x) = P(X > x)$ and $\overline{G}(y) = G(Y > y)$ respectively, with $F(0) = G(0) = 1.$ Kullback–Leibler [82] informative distance between *F* and *G* is defined by:

$$I_{X,Y} = \int_0^\infty f(x) \log \frac{f(x)}{g(x)} dx \tag{2.4.1}$$

where log denotes natural logarithm, $I_{X,Y}$ is known as relative entropy and it is shift and scale invariant. However, it is not metric, since symmetrization and triangle inequality does not hold. We point out the Jensen-Shannon divergence (JSD) which is based on Kullback-Leibler divergence, with the notable differences that is always a finite value and its square root is a metric. (Nielsen [102] and Amari *et al*. [5]). The application of $I_{X,Y}$ as an informative distance in residual and past lifetimes has increasingly studied in recent years. In particular, Ebrahimi and Kirmani [46] considered the residual Kullback-Leibler discrimination information of non-negative lifetimes of the systems *X* and *Y* at time *t* as:

$$I_{X,Y}(t)=\int_t^{\infty}\frac{f(x)}{\bar{F}(t)}\log\frac{f(x)/\bar{F}(t)}{g(x)/\bar{G}(t)}dx \tag{2.4.2}$$

Given that both systems have survived up to time t, $I_{X,Y}(t)$ identifies with the relative entropy of remaining lifetimes $(X|X\geq t)$ and $(Y|Y>t)$. Furthermore, the Kullback-Leibler distance for two past lifetimes was studied in Di Crescenzo and Longobardi [35] which is dual to (2.4.2) in the sense that it is an informative distance between past lifetimes $(X|X<t)$ and $(Y|Y<t)$. Formally, the past Kullback-Leibler distance of non-negative random lifetimes of the systems $X$ and $Y$ at time $t$ is defined as:

$$\bar{I}_{X,Y}(t)=\int_0^{t}\frac{f(x)}{\bar{F}(t)}\log\frac{f(x)/F(t)}{g(x)/G(t)}dx \tag{2.4.3}$$

Given that at time $t$, both systems have been found to be down, $\bar{I}_{X,Y}(t)$ measures the informative distance between their past lives.

Along a similar line, Misagh and Yari [91] define a new discrepancy measure that completes studying informative distance between two random lifetimes.

**Definition 2.4.1**

The interval distance between random lifetimes $X$ and $Y$ at interval $(t_1,t_2)$ is the Kullback-Leibler discrimination measure between the Truncated lives $(X|t_1<X<t_2)$ and $(Y|t_1<Y<t_2)$:

$$ID_{X,Y}(t_1,t_2)=\int_{t_1}^{t_2}\frac{f(x)}{F(t_2)-F(t_1)}\log\frac{f(x)/[F(t_2)-F(t_1)]}{g(x)/[G(t_2)-G(t_1)]}\partial x \tag{2.4.4}$$

**Remark 2.4.1:**

Clearly $ID_{X,Y}(o,t)=\bar{I}_{X,Y}(t), ID_{X,Y}(t,\infty)=I_{X,Y}(t)$ and $ID_{X,Y}(o,\infty)=I_{X,Y}$

Given that both systems $X$ and $Y$ have survived up to time $t_1$ and have seen to be down at time $t_2$ $ID_{X,Y}(t_1,t_2)$ measures the discrepancy between their unknown failure times in the interval $(t_1,t_2)$. $ID_{X,Y}(t_1,t_2)$ satisfies all properties of Kullback-Leibler discrimination measure and can be rewritten as:

$$ID_{X,Y}(t_1,t_2) = -\int_{t_1}^{t_2} \frac{f(x)}{F(t_2)-F(t_1)} \log \frac{g(x)}{G(t_2)-G(t_1)} dx - IH_X(t_1,t_2) \qquad (2.4.5)$$

where $IH_X(t_1,t_2)$ is the interval entropy of $X$ in (2.3.4).

An alternative way of writing (2.4.5) is the following:

$$ID_{X,Y}(t_1,t_2) = \log \frac{G(t_2)-G(t_1)}{F(t_2)-F(t_1)} + \frac{1}{F(t_2)-F(t_1)} \int_{t_1}^{t_2} f(x) \log \frac{f(x)}{g(x)} dx \qquad (2.4.6)$$

The following example clarifies the effectiveness of the interval discrimination measure.

**Example 2.4.1:** Suppose $X$ and $Y$ be random lifetimes of two systems with joint density function:

$f(x,y) = 1/4,\ 0 < x < 2,\ 0 < y < 4-2x$ and that the marginal densities of $X$ and $Y$ are $f(x) = 1/2(2-x);\ 0 < x < 2$ and $g(y) = 1/8(4-y),\ \ 0 < y < 4$ respectively. Because $X$ and $Y$, belongs to different domains, using relative entropy to measure the informative distance between $X$ and $Y$ is not interpretable. The interval distance between $X$ and $Y$ in the intervals (0, 1.5) and (1.5, 2) are 0.01 and 0.16 respectively. Hence the informative distance between $X$ and $Y$ in the interval (1.5, 2) is greater than of it in the interval (0, 1.5). In the following proposition we decompose the Kullback-Leibler discrimination function in terms of residual, past and interval discrepancy measure.

**Proposition 2.4.1**

Let $X$ and $Y$ are two non-negative random lifetimes of two systems. For all $0 \le t_1 < t_2 < \infty$, the Kullback-Leibler discrimination measure is decomposed as follows:

$$I_{X,Y} = [F(t_2) - F(t_1)] ID_{X,Y}(t_1, t_2) + F(t_1)\bar{I}_{X,Y}(t_1) + \bar{F}(t_2) I_{X,Y}(t_2) + I_{U,V}(t_1, t_2) \tag{2.4.7}$$

where:

$$I_{U,V}(t_1, t_2) = F(t_1) \log \frac{F(t_1)}{G(t_1)} + \bar{F}(t_2) \log \frac{\bar{F}(t_2)}{\bar{G}(t_2)} + [F(t_2) - F(t_1)] \log \frac{F(t_2) - F(t_1)}{G(t_2) - G(t_1)}$$

is the Kullback-Leibler distance between two Trivalent discrete random variables. Proposition 2.4.1 admits the following interpretation: the Kullback-Leibler discrepancy measure between random lifetimes of systems $X$ and $Y$ is composed from four parts:

i) the discrepancy between the past lives of two systems at time $t_1$;

ii) the discrepancy between residual lifetimes of $X$ and $Y$ that have both survived up to time $t_2$;

iii) the discrepancy between the lifetimes of both systems in the interval $(t_1, t_2)$;

iv) the discrepancy between two random variables which determines if the systems have been found to be failing before $t_1$, between $t_1$ and $t_2$ or after $t_2$.

## 2.5 Results on Interval Based Measures

In this section, we study the properties of $ID(t_1, t_2)$ and point out certain similarities with those of $I_{X,Y}(t)$ and $\bar{I}_{X,Y}(t)$. The following proposition gives lower and upper bounds for the interval distance. We first give definition of likelihood ratio ordering.

**Definition 2.5.1**

$X$ is said to be larger than $Y$ in likelihood ratio $\left(X \overset{LR}{\geq} Y\right)$ if $\dfrac{f(x)}{g(x)}$ is increasing in $x$ over the union of the supports of $X$ and $Y$.

**Proposition 2.5.1**

Let $X$ and $Y$ are random variables with common support $(0,\infty)$ then:

i) $X \overset{LR}{\geq} Y$ implies:

$$\log\frac{h_1^X(t_1,t_2)}{h_1^Y(t_1,t_2)} \leq ID_{X,Y}(t_1,t_2) \leq \log\frac{h_2^X(t_1,t_2)}{h_2^Y(t_1,t_2)} \qquad (2.5.1)$$

when $f(x)/g(x)$ is decreasing in $x>0$, then the inequalities in (2.5.1)are reversed.

ii) Descreasing $g(x)$ in $x>0$, implies:

$$\log\frac{1}{h_1^Y(t_1,t_2)} \leq ID_{X,Y}(t_1,t_2) + IH_X(t_1,t_2) \leq \log\frac{1}{h_2^Y(t_1,t_2)} \qquad (2.5.2)$$

for increasing $g(x)$ then the inequalities in (2.5.2) are reversed.

***Proof***: Because of increasing $\dfrac{f(x)}{g(x)}$, $n>0$ from (2.4.4), we have

$$ID_{X,Y}(t_1,t_2) \leq \int_{t_1}^{t_2} \frac{f(x)}{F(t_2)-F(t_1)} \log\frac{f(t_2)/[F(t_2)-F(t_1)]}{g(t_2)/[G(t_2)-G(t_1)]}dx$$

$$= \log\frac{h_2^X(t_1,t_2)}{h_2^Y(t_1,t_2)}$$

And $$ID_{X,Y}(t_1,t_2) \geq \int_{t_1}^{t_2} \frac{f(x)}{F(t_2)-F(t_1)} \log\frac{f(t_1)/[F(t_2)-F(t_1)]}{g(t_1)/[G(t_2)-G(t_1)]}dx$$

$$= \log\frac{h_1^X(t_1,t_2)}{h_1^Y(t_1,t_2)}$$

which gives (2.5.1) when $\frac{f(x)}{g(x)}$ is descreasing, the proof is similar. Furthermore, for all $t_1 < x < t_2$ descreasing $g(x)$ in $x > 0$ implies $g(t_2) < g(x) < g(t_1)$. Then from (2.4.5) we get:

$$ID_{X,Y}(t_1,t_2) \le -\log h_2^Y(t_1,t_2) - IH_X(t_1,t_2)$$

and $$ID_{X,Y}(t_1,t_2) \ge -\log h_1^Y(t_1,t_2) - IH_X(t_1,t_2)$$

so that (2.5.2) holds. When $g(x)$ is increasing the proof in similar.

**Remark 2.5.1**

Consider $X$ and $Y$ are two non-negative random variables corresponding to weighted exponential distributions with positive rates $\lambda$ and $\mu$ respectively and with common positive real weight function $\omega(\cdot)$. The densities of $X$ and $Y$ are $f(x) = \frac{\omega(x)e^{-\lambda x}}{h(\lambda)}$ and $g(x) = \frac{\omega(x)e^{-\mu x}}{h(\mu)}$ respectively, where $h(\cdot)$ denotes the Laplace transform of $\omega(\cdot)$ given by $h(\theta) = \int_0^\infty \omega(x)e^{-\theta x}dx, \theta > 0$, therefore, for $\lambda \ne \mu$ the interval distance between $X$ and $Y$ at interval $(t_1,t_2)$ is the following:

$$ID_{X,Y}(t_1,t_2) = \log\frac{G(t_2) - G(t_1)}{F(t_2) - F(t_1)} + \log\frac{h(\mu)}{h(\lambda)} - (\lambda - \mu)E(X|t_1 < X < t_2) \tag{2.5.3}$$

**Remark 2.5.2**

Let $X$ be a non-negative random lifetime with density function $f(x)$ and cumulative distribution function $F(t) = P(X \le t)$. Then the density function and cumulative distribution function for the weighted random variable $X_\omega$ associated to a positive real function $\omega(\cdot)$ are $f_\omega(x) = \frac{\omega(x)}{E(\omega(X))} f(x)$ and

$F_\omega(t) = \frac{E(\omega(X)|X \le t)}{E(\omega(X))} F(t)$, respectively, where $E(\omega(X)) = \int_0^\infty \omega(x) f(x)dx.$

Then, from (2.4.6) we have:

$$ID_{X,X_\omega}(t_1,t_2) = \log E(\omega(X)|t_1 < X < t_2) - E(\log(\omega(X))|t_1 < X < t_2) \quad (2.5.4)$$

A similar expression is available in Maya and Sunoj [85] for past lifetime. Due to (2.5.4) and from non-negativity of $ID_{X,X_\omega}(t_1,t_2)$ we have:

$$\log E(\omega(X)|t_1 < X < t_2) \geq E(\log(\omega(X))|t_1 < X < t_2)$$

which is a direct result of Markov inequality for concave functions.

**Example 2.5.1:** For $\omega(x) = x^{n-1}$ and $h(\theta) = \frac{(n-1)!}{\theta^n}$ the distributions of random variables in Remark 2.5.1 called Erlang distributions with scale parameters $\lambda$ and $\mu$ and with common shape paramter *n*. The conditional mean of $(X|t_1 < X < t_2)$ is the following:

$$E(X|t_1 < X < t_2) = \frac{1}{F(t_2) - F(t_1)} \int_{t_1}^{t_2} x \frac{x^{n-1}\lambda^n e^{-\lambda x}}{(n-1)!} dx$$

$$= \frac{Y(n+1, \lambda t_2) - Y(n+1, \lambda t_1)}{\lambda (n-1)!\,(F(t_2) - F(t_1))}$$

where $Y(a,x) = \int_0^x e^{-\mu} \cdot \mu^{\alpha - 1} \partial\mu$ is the incomplete Gamma function. From (2.5.3) we obtain:

$$ID_{X,Y}(t_1,t_2) = \frac{\log Y(n, \mu t_2) - Y(n, \mu t_1)}{Y(n, \lambda t_2) - Y(n, \lambda t_1)} - n \log \frac{\lambda}{\mu}$$

$$+ (\lambda - \mu) \frac{Y(n+1, \lambda t_2) - Y(n+1, \lambda t_1)}{\lambda (n-1)!\,(F(t_2) - F(t_1))}$$

In the following proposition, sufficient condition for $ID_{X_1,Y}(t_1,t_2)$ to be smaller than $ID_{X_2,Y}(t_1,t_2)$ is provided.

**Proposition 2.5.2**

Consider three non-negative random variables $X_1, X_2$ and $Y$ with probability density functions $f_1, f_2$ and $g$ respectively. $X_1 \overset{LR}{\geq} Y$ implies

$$ID_{X_1,Y}(t_1,t_2) \le ID_{X_2,Y}(t_1,t_2)$$

***Proof:*** From (2.4.6) we have:

$$ID_{X_1,Y}(t_1,t_2) - ID_{X_2,Y}(t_1,t_2)$$

$$= -ID_{X_2,X_1}(t_1,t_2) + \int_{t_1}^{t_2} \left( \frac{f_1(x)}{F_1(t_2) - F_1(t_1)} - \frac{f_2(x)}{F_2(t_2) - F_2(t_1)} \right) \log \frac{f_1(x)}{g(x)} dx.$$

$$\le \int_{t_1}^{t_2} \left( \frac{f_1(x)}{F_1(t_2) - F_1(t_1)} - \frac{f_2(x)}{F_2(t_2) - F_2(t_1)} \right) \log \frac{f_1(x)}{g(x)} dx$$

$$\le \log \frac{f_1(t_2)}{g(t_2)} \int_{t_1}^{t_2} \left( \frac{f_1(x)}{F_1(t_2) - F_1(t_1)} - \frac{f_2(x)}{F_2(t_2) - F_2(t_1)} \right) dx = 0$$

where the first inequality comes from the fact that $ID_{X_2,X_1}(t_1,t_2) \ge 0$ and the second one follows from the increasing $\frac{f_1(x)}{g(x)}$ in $x > 0$.

**Example 2.5.2:**

Let $\{N(t), t \ge 0\}$ be a non-homogeneous Poisson process with a differentiable mean function $M(t) = E(N(t))$ such that *M(t)* tends to $\infty$ as $t \to \infty$. Let $R_n$, $n = 1, 2, 3$ denote the time of the occurrence of the *n-th* event in such a process. Then $R_n$ has the following density function:

$$f_n(x) = \frac{(M(t))^{n-1}}{(n-1)!} f_1(x), \qquad x > 0$$

where

$$f_1(x) = -\frac{\partial}{\partial x} \exp(-M(x)), \qquad x > 0$$

clearly $f_n(x)/f_1(x)$ is increasing in $x > 0$. It follows from proposition (2.5.2) that for all $m \le n$.

$$ID_{X_n,X_1}(t_1,t_2) \le ID_{X_m,X_1}(t_1,t_2)$$

**Remark 2.5.3**

For all $0 \le \theta < t_1$, we get

$$ID_{X+\theta,Y+\theta}(t_1,t_2) = ID_{X,Y}(t_1-\theta, t_2-\theta)$$

**Remark 2.5.4**

Let $Y = aX$ where $a > 0$,

$$ID_{aX,aY}(t_1,t_2) = ID_{X,Y}\left(\frac{t_1}{a}, \frac{t_2}{a}\right)$$

# Chapter III
# Measure of Discrimination Between Lifetime Distributions

## 3.1 Introduction

After the creation of Shannon [122], a number of research papers and monographs discussing and extending Shannon's original work have appeared. Among them Dragomir [38], Kagan, Linnik and Rao [71] and Kullback [80] are using and extending results due to information measures. Asad *et al.*[10] proposed a dynamic measure based on differential geometry applicable to residual lifetime. This measure has been used for the classification and ordering of survival function. Ebrahimi and Kirmani [40], Nanda *et al.* [96] and Asadi *et al.* [10] gave an overview of some aspects of residual Renyi divergence and residual Kullback-Leibler information and residual entropy. Further implications and properties of the dynamic measures such as above and the uncertainty ordering, proportional hazard model through a measure of discrimination between two residual life distributions on the basis of the measures that are mentioned in the literature are obtained. In this chapter characterization results for residual entropy residual information measures are obtained.

In Ebrahimi and Pellery [41] and in Ebrahimi [42] it has been pointed out the potentiality of the classical machinery of Shannon information theory as suitable tool to measure the uncertainty related to random lifetimes and their reliability.

## 3.2 Measure of Discrepancy

Let $X$ and $Y$ be absolutely continuous non-negative honest random variables that describe the lifetimes of two items. We denote by $f(t)$, $F(t)$ and $\overline{F}(t)=1-F(t)$ the probability density function (p.d.f), the cumulative distribution function (c.d.f) and the survival function of $X$, respectively and by $g(t)$, $G(t)$ and $\overline{G}(t)$ the corresponding functions of $Y$. Moreover, let $\lambda_X(x)=f(x)/\overline{F}(x)$ and $\lambda_Y(x)=g(x)/\overline{G}(x)$ be the hazard rate functions of $X$ and $Y$, respectively, whereas $\tau_X(x)=f(x)/F(x)$ and $\tau_Y(x)=g(x)/G(x)$ will denote their reversed hazard rate functions. Without loss of generality we assume that densities $f(t)$ and $g(t)$ have support $(0,+\infty)$.

As an information distance between *F* and *G*, Kullback and Leibler [82] proposed the following discrimination measure, also known as relative entropy of $X$ and $Y$:

$$I_{X,Y}=\int_0^{+\infty} f(x)\log\frac{f(x)}{g(x)}dx \tag{3.2.1}$$

where 'log' denotes natural logarithm. Distance (3.2.1) is shift and scale invariant. Furthermore, $I_{X,Y}\geq 0$ with equality if and only if $f(x)=g(x)$ a. e. However, $I_{X,Y}$ is not a metric, as it does not satisfy the triangle inequality and is asymmetric. A symmetrized version of $I_{X,Y}$, introduced in Kullback and Leibler [82] is defined by $\hat{I}_{X,Y}=I_{X,Y}+I_{Y,X.}$

Ebrahimi and Kirmani [40] have defined the Kullback-Leibler discrimination information of $X$ and $Y$ at time $t$ by

$$I_{X,Y}(t) = \int_{t}^{+\infty} \frac{f(x)}{\overline{F}(t)} \log \frac{f(x)/\overline{F}(t)}{g(x)/\overline{G}(t)} dx, \qquad t > 0 \tag{3.2.2}$$

According to (3.2.1), $I_{X,Y}(t)$ identifies with the relative entropy of $[X - t|X > t]$ and $[Y - t|Y > t]$, whereas customary, $[X|B]$ denotes a random variable whose distribution is the same as the conditional distribution of $X$ given $B$. Information measure (3.2.2) is thus useful to compare the residual lifetimes of two items that have both survived up to time $t$.

Along a similar line, hereafter we shall define a new information measure $\bar{I}_{X,Y}(t)$. This is dual to (3.2.2) in the sense that it is an information distance between the past lives $[X|X \le t]$ and $[Y|Y \le t]$.

**Definition 3.2.1**

The discrimination measure between the past lives $[X|X \le t]$ and $[Y|Y \le t]$ is

$$\bar{I}_{X,Y}(t) = \int_{0}^{t} \frac{f(x)}{F(t)} \log \frac{f(x)/F(t)}{g(x)/G(t)} dx, \qquad t > 0 \tag{3.2.3}$$

Given that at time $t$ two items have been found to be failing, $\bar{I}_{X,Y}(t)$ measures the discrepancy between their past lives. In analogy with the Kullback-Leibler discrimination information, it is $\bar{I}_{X,Y}(t) \ge 0$, with equality holding if and only if $f(x) = g(x)$ a. e. From (3.2.3) we have

$$\bar{I}_{X,Y}(t) = -\int_{0}^{t} \frac{f(x)}{F(t)} \log \frac{g(x)}{G(t)} dx - \overline{H}_X(t), \qquad t > 0 \tag{3.2.4}$$

where $\overline{H}_X(t) = -\int_{0}^{t} \frac{f(x)}{F(t)} \log \frac{f(x)}{F(t)} dx$

is the past entropy at time $t$ of $X$ introduced by Di Creacenzo and Longobardi [34]. Another way of representing $\bar{I}_{X,Y}(t)$ is the following:

$$\bar{I}_{X,Y}(t) = \log \frac{G(t)}{F(t)} + \frac{1}{F(t)} \int_{0}^{t} f(x) \log \frac{f(x)}{g(x)} dx, \qquad t > 0 \tag{3.2.5}$$

**Remark 3.2.1**

From Equations (3.2.1–3.2.3), we obtain the following relation between the three discrimination measures considered above:

$$I_{X,Y} = I_{X,Y}(t)\bar{F}(t) + \bar{I}_{X,Y}(t)F(t) + F(t)\log\frac{F(t)}{G(t)} + \bar{F}(t)\log\frac{\bar{F}(t)}{\bar{G}(t)}, \qquad t>0$$

In example 3.2.1, we pinpoint the role of the discrimination measure between the past lives for the comparison of random lifetimes.

**Example 3.2.1:** Assume that $X$ and $Y_\alpha$ are the random lifetimes of two items, where $X$ is uniformly distributed on (0,1) and $Y_\alpha$ has probability density

$$g_\alpha(t) = \alpha(t-1/2)+1 \qquad 0<t<t_1.$$

With $-2 \le \alpha \le 2$. Due to the symmetry of $g_\alpha(t)$ with respect to $t=1/2$, and due to (3.2.1) it follows that

$$I_{X,Y_\alpha} = I_{X,Y_{-\alpha}}, \quad \text{for all} \quad \alpha \in [-2,2],$$

so, that the information distance of $X$ and $Y_\alpha$ equals the information distance existing between $X$ and $Y_{-\alpha}$. Moreover for $\alpha \neq 0$, from (3.2.5) we have

$$\bar{I}_{X,Y_\alpha}(t) = 1 + \frac{1}{\alpha t}\left\{\left(1-\frac{\alpha}{2}\right)\log\left(1-\frac{\alpha}{2}\right) - \left(1-\frac{\alpha}{2}+\alpha t\right)\log\left(1-\frac{\alpha}{2}+\alpha t\right)\right\}$$

$$+\log\left(1-\frac{\alpha}{2}+\frac{\alpha t}{2}\right), \quad 0<t<1 \qquad (3.2.6)$$

equation (3.2.6) implies that in general $\bar{I}_{X,Y_\alpha}(t)$ is not equal to $\bar{I}_{X,Y_{-\alpha}}(t)$ for all $t \in (0,\ 1)$, as is shown for instance in Fig 3.2.1

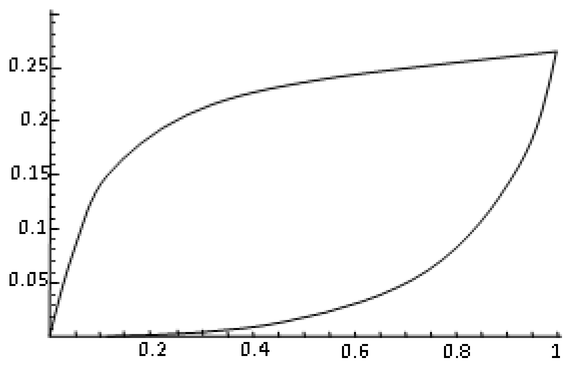



**Fig. 3.2.1:** The discrimination measure is given for $t \in (0,1)$, with $x=1.95$ (top) and $x = -1.95$ (bottom)

Hence, even though $I_{X,Y\alpha} = I_{X,Y-\alpha}$ the information distance between $[X|X \le t]$ and $[Y_\alpha|Y_\alpha \le t]$ is in general different from the information distance between $[X|X \le t]$ and $[Y_{-\alpha}|Y_{-\alpha} \le t]$.

### 3.3 Properties of $\bar{I}_{X,Y}(t)$

In this section, we study some properties of $\bar{I}_{X,Y}(t)$ and point out certain similarities with those of $I_{X,Y}(t)$. First of all, Equations (3.2.1 – 3.2.3), we observe that

$$\underset{t\to\infty}{Lim}\ \bar{I}_{X,Y}(t) = I_{X,Y} = \underset{t\to 0^+}{Lim}\ I_{X,Y}(t)$$

Let us now obtain some bounds for $\bar{I}_{X,Y}(t)$. In the following, the terms 'increasing' and 'decreasing' are used in non-strict sense. We shall also make use of some nations of stochastic orders;

**Theorem 3.3.1**

i) If $f(t)/g(t)$ is increasing in t > 0, i.e. $X \ge_{lr} Y$, then

$$\bar{I}_{X,Y}(t) \le \log \tau_x(t)/\tau_Y(t), \quad t > 0 \tag{3.3.1}$$

when $f(t)/g(t)$ is decreasing in $t > 0$, i.e. $X \le_{lr} Y$, then the inequality in (3.3.1) is reversed.

ii) If $g(t)$ is decreasing in $t > 0$, then

$$\bar{I}_{X,Y}(t) \le -\log \tau_Y(t) - \bar{H}_X(t), \quad t > 0 \tag{3.3.2}$$

***Proof:*** Due to increasingness of $f(t)/g(t)$, from (3.2.3) for all $t > 0$ we have

$$\bar{I}_{X,Y}(t) = \int_0^t \frac{f(x)}{F(t)} \log \frac{f(x)/F(t)}{g(x)/G(t)} dx \le \int_0^t \frac{f(x)}{F(t)} \log \frac{f(t)/F(t)}{g(t)/G(t)} dx$$
$$= \log \frac{\tau_X(t)}{\tau_Y(t)}$$

which gives (3.3.1). When $f(t)/g(t)$ is descreasing the proof is similar. Furthermore, if $g(x) \ge g(t)$ for all $t > x > 0$, then from (3.2.4) we obtain

$$\bar{I}_{X,Y}(t) = -\int_0^t \frac{f(x)}{F(t)} \log \frac{g(x)}{G(t)} dx - \overline{H}_X(t) \le -\log \frac{g(t)}{G(t)} - \overline{H}_X(t)$$

so that (3.3.2) holds.

The assumption that $f(t)/g(t)$ is increasing implies $\tau_X(t) \ge \tau_Y(t)$ for all $t > 0$ i.e. $X \ge_{rh} Y$. So that the right-hand side of (3.3.1) is non-negative for all $t > 0$.

From (3.2.3), it is not hard to see that

$$\frac{\partial}{\partial t} \bar{I}_{X,Y}(t) = \tau_X(t) \left[ \log \frac{\tau_X(t)}{\tau_Y(t)} - \bar{I}_{X,Y}(t) \right] + \tau_Y(t) - \tau_X(t), \quad t > 0 \qquad (3.3.3)$$

equation (3.3.3) implies that if $\bar{I}_{X,Y}(t)$ is increasing, then

$$\bar{I}_{X,Y}(t) \le \frac{\tau_Y(t)}{\tau_X(t)} - 1 + \log \frac{\tau_X(t)}{\tau_Y(t)}, \quad t > 0$$

**Remark 3.3.1**

Let $X$ and $Y$ have weighted distributions corresponding to exponential distributions with rates $\lambda$ and $\mu$, respectively and with weighted function $\phi(t)$. Their densities are then given by

$$f(t) = \frac{\phi(t) e^{-\lambda t}}{h(\lambda)}, \; g(t) = \frac{\phi(t) e^{-\mu t}}{h(\mu)} \qquad t > 0 \qquad (3.3.4)$$

where $h(\cdot)$ denotes the Laplace Transform of $\phi(t)$:

$$h(\xi) = \gamma_\xi[\phi(t)] = \int_0^\infty e^{-\xi t} \phi(t) dt, \; \xi > 0$$

Hence from (3.2.5) and (3.3.4) it follows that for $\lambda \ne \mu$ the discrimination measure of the past lives can be expressed as

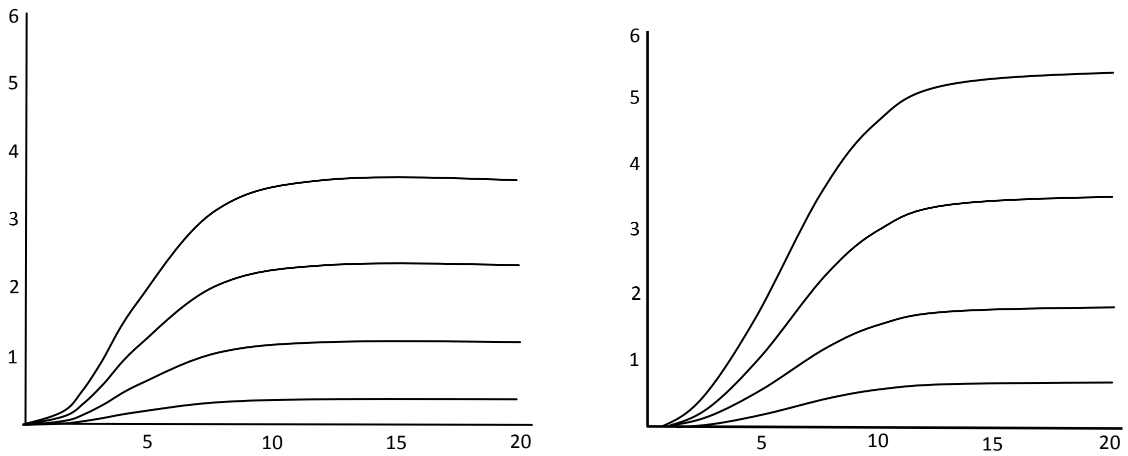
$$\bar{I}_{X,Y}(t) = \log \frac{G(t)}{F(t)} + \log \frac{h(\mu)}{h(\lambda)} + (\mu - \lambda) E[X | X \le t] \quad t > 0 \qquad (3.3.5)$$

Due to (3.3.4) if $\lambda \ge \mu$ $[\lambda \le \mu]$ then $f(t)/g(t)$ is decreasing [increasing] in $t > 0$ i.e. $X_{lr} \le Y [X \ge_{lr} Y]$, which implies $X \le_{rh} Y [X \ge_{rh} Y]$ so that $\log\{G(t)/F(t)\}$ is increasing [decreasing] in $t > 0$. Hence noting that $E[X | X \le t]$ is increasing for $t > 0$, the first and the third terms at the right-hand side of (3.3.5) exhibit different behaviours for all $\lambda \neq \mu$. Indeed, when $\lambda > \mu$, the first term is increasing whereas the third one is decreasing, the opposite occurring when $\lambda < \mu$.

**Example 3.3.1:** Let $X$ and $Y$ be Erlang distributed, with scale parameters $\lambda$ and $\mu$, respectively and with common shape parameter $n$. Hence, they possesses densities (3.3.4) where $\phi(t) = t^{n-1}$ and $h(\xi) = (n-1)!/\xi^n$. Hence for $\lambda \neq \mu$, equation (3.3.5) becomes

$$\bar{I}_{X,Y}(t) = \log\frac{G(t)}{F(t)} + n\log\frac{\lambda}{\mu} + (\mu - \lambda)\frac{\gamma(n+1, \lambda t)}{F(t)\lambda(n-1)!} \quad t > 0 \qquad (3.3.6)$$

where $\gamma(a,x) = \int\limits_0^x \bar{e}^{\gamma}.\gamma^{a-1}dy$ denotes the incomplete Gamma function. Fig 3.3.1 shows some plots of the discrimination measure (3.3.6)




(a) (b)

**Fig. 3.3.1:** The discrimination measure is plotted for $t \in (0,20)$, $\lambda = 1$, and $\mu = 1.5, 2.0, 2.5, 3.0$ (bottom to top), with $n=4$ in (a) and $n=6$ in (b)

The following example shows that the discrimination measure $\bar{I}_{X,Y}(t)$ is not necessarily monotone.

**Example 3.3.2:** Let

$$F(x)=\begin{cases}\exp\left\{-\frac{1}{2}-\frac{1}{x}\right\} & If \quad 0<x\le 1\\ \exp\left\{-2+\frac{x^2}{2}\right\} & If \quad 1<x\le 2.\\ 1 & If \quad x\ge 2.\end{cases}$$

$$G(x)=\begin{cases}\frac{x^2}{4} & If \quad 0<x\le 2\\ 1 & If \quad x\ge 2.\end{cases}$$

be the c.d.f's of $X$ and $Y$. From Fig 3.3.2 we see that $\bar{I}_{X,Y}(t)$ is not monotone for all $t\in(0,\,2)$

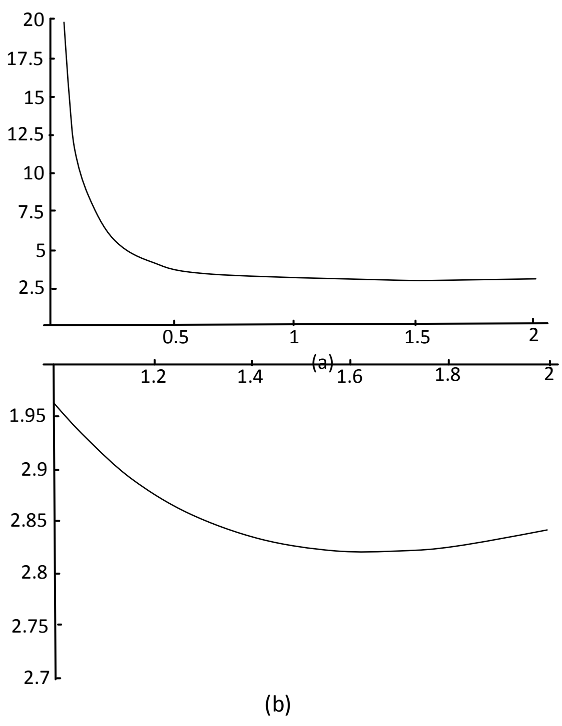



**Fig. 3.3.2:** The discrimination measure of the past lives of Example 3.3.1 is sketched for $t\in(0,20)$. (b) An enlargement of (a) close to $t=2$, showing the lack of monotonicity.

## Theorem 3.3.2

Let us consider three random lifetimes $X_1, X_2$ and $Y$, with p.d.f.'s $f_1, f_2$ and $g$ and with reversed hazard functions $\tau_1,\ \tau_2$ and $\tau_Y$ respectively, if

i) $f_1(t)/g(t)$ is decreasing for all $t>0$, i.e. $X_1 \ge_{lr} Y$ and

ii) $\tau_1(t) \le \tau_2(t)$ for all $t > 0$ i.e. $X_1 \le_{rh} X_2$ then

$$\bar{I}_{X_1,Y}(t) \le \bar{I}_{X_2,Y}(t), \quad t > 0$$

***Proof:*** From definition (3.2.1) and since

$$\bar{I}_{X_2,X_1}(t) \ge 0,$$

we have

$$\bar{I}_{x_1,Y}(t) - \bar{I}_{X_2,Y}(t) = \int_0^t \frac{f_1(x)}{F_1(t)} \log \frac{f_1(x)/F_1(t)}{g(x)/G(t)} dx$$

$$- \int_0^t \frac{f_2(x)}{F_2(t)} \log \frac{f_1(x)/F_1(t)}{g(x)/G(t)} dx - I_{X_2,X_1}(t)$$

$$\bar{I}_{X_1,Y}(t) - \bar{I}_{X_2,Y} \le \int_0^t \frac{f_1(x)}{F_1(t)} \log \frac{f_1(x)}{g(x)} dx - \int_0^t \frac{f_2(x)}{F_2(t)} \log \frac{f_1(x)}{g(x)} dx$$

Next, denoting the p.d.f of $X_{i,t} = [X_t | X_i \le t]$ by $f_{i,t}(x) := \dfrac{f_i(x)}{F_i(t)}, \quad 0 < x < t, \quad i = 1,2$ we are led to

$$\bar{I}_{x_1,Y}(t) - \bar{I}_{X_2,Y}(t) \le \int_0^t f_{1,t}(x) \log \frac{f_1(x)}{g(x)} dx - \int_0^t f_{2,t}(x) \log \frac{f_1(x)}{g(x)} dx$$

$$= E[\phi(X_{1,t})] - E[\phi(X_{2,t})] \tag{3.3.7}$$

where $\phi(x) = \log[f_1(x)/g(x)]$ is increasing in $x$ due to assumption (i). Hence due to Remark (3.3.2) the right-hand side of (3.3.7) is non-positive if $X_{1,t} \le_{St} X_{2,t}$ i.e. if $X_1 \le_{rh} X_2$.

**Theorem 3.3.3**

Let us consider three random lifetimes $X, Y_1$ and $Y_2$ with p.d.f's $f, g_1$ and $g_2$ respectively. If $g_1(t)/g_2(t)$ is increasing for all $t > 0$, i.e. $Y_1 \ge_{lr} Y_2$ then

$\bar{I}_{X,Y_1}(t) \ge \bar{I}_{X,Y_2}(t) - \log \dfrac{\tau_1(t)}{\tau_2(t)}, \quad t > 0$. Where $\tau_1$ and $\tau_2$ are the reversed hazard rate functions of $Y_1$ and $Y_2$.

***Proof:*** Making use of (3.2.5) we have

$$\bar{I}_{X,Y_1}(t)-\bar{I}_{X,Y_2}(t)=\log\frac{G_1(t)}{G_2(t)}+\frac{1}{F(t)}\int_0^t f(x)\log\frac{g_2(x)}{g_1(x)}dx,\qquad t>0$$

since $g_2(x)/g_1(x)\geq g_2(t)/g_1(t)$ for all $x\in(0,t)$. The proof immediately follows on account that $\tau_i(t)=g_i(t)/G_i(t),\quad i=1,\ 2.$

### 3.4 Residual Information Measures for Weighted Distributions

Let $(\Omega,\beta,\mu)$ be a measure space and $f$ be a measurable function from $\Omega$ to $[0,\infty)$, such that

$$\int_\Omega f d\mu=1.$$

The Shannon entropy (or simply the entropy) of $f$ relative to $\mu$, is defined by

$$H(f,\mu)=-\int_\Omega f\log f d\mu,\qquad (with\ f\log f=0\ if\ f=0)\qquad (3.4.1)$$

and assumed to be defined for which $f\log f$ is integrable. If $X$ is any random variable with pdf $f$, then we refer to $H$ as the entropy of $X$ and denotes also it by the notation $H_X$. In the case $\mu$ is a version of counting measure, (3.4.1) leads us to a specialized version that introduced by Shannon [122] as

$$H_X=-\sum_{i=1}^{n}p_i\log p_i$$

where $p_i\geq 0$ and $\sum_{i=1}^{n}p_i=1$. One of the important issues in many applications of probability theory is finding an appropriate measure of distance between two probability distributions. A number of divergence measure for this purpose are available in the literature related to various type of information measure. These measures have applied in a variety of fields. Consider $F$ and $G$ be two distributions which are absolutely continuous w.r.t measure $\mu$ and $\frac{\partial F}{\partial\mu}=f$ and $\frac{\partial G}{\partial\mu}=g$. Some of the divergence measures are given below:

**Kullback-Leibler information:**

$$D_{KL}(F,G)=\int_{\chi}\log\frac{f(x)}{g(x)}.f(x)d\mu \tag{3.4.2}$$

**$\chi^2-$ Divergence:**

$$D_{\chi^2}(F,G)=\int_{\chi}\frac{[f(x)-g(x)]^2}{f(x)}d\mu \tag{3.4.3}$$

**Bhattacharyya Distance and Hellinger Distance:**

$$D_{BH}(F,G)=\int_{\chi}\sqrt{g(x)f(x)d\mu}\cdot D_H(F,G)=2[1-D_{Bh}(F,G)] \tag{3.4.4}$$

**$\alpha-$Divergence:**

$$D_{\alpha}(F,G)=\frac{1}{1-\alpha^2}\int_{\chi}\left\{1-\frac{g^{\frac{1+\alpha}{2}}(x)}{f^{\frac{1+\alpha}{2}}(x)}\right\}^2 f(x)d\mu \tag{3.4.5}$$

**Relative Jensen-Shannon divergence measure**

Sibsen in (1969) and Lin in (1991) introduced the divergence measure as

$$D_J(F,G)=\int_{\chi}f(x)\log\left(\frac{2f(x)}{f(x)+g(x)}\right)d\mu \tag{3.4.6}$$

**J-Divergence measure**

Jeffrey in (1946) introduced the divergence measure

$$D_J(F,G)=\int_{\chi}(f(x)-g(x))\log\frac{f(x)}{g(x)}d\mu \tag{3.4.7}$$

**Hellinger Divergence Measure**

Hellinger in (1909) introduced the divergence measure as

$$D_h(F,G)=\tfrac{1}{2}\int_{\chi}\left(\sqrt{f(x)}-\sqrt{g(x)}\right)^2 d\mu \tag{3.4.8}$$

The following results are related to the above measures:

i) It is easy to see that $D_H(F,\ G)\le 2$ via Taylor expansion and approximation, we can get,

$$D_{KL}(F,G) \approx \frac{1}{2} D_{\chi^2}(F,G)$$

$$D_J(F,G) \approx \frac{1}{2}\left[D_{\chi 2}(G,F) + D_{\chi 2}(G,F)\right]$$

$$D_{\chi^2}(F,G) \approx 4D_H(F,G) \text{ and } D_{\chi^2}(F,G) \geq D_H(F,G)$$

ii) Sometimes we are interested in the distances that is introduced in (3.4.2) to (3.4.8), between the distributions $F = F_{\theta 1}$ and $G = F_{\theta 2}$. Between the corresponding sample distributions which we denote $F_{\theta 1}^{n}$ and $F_{\theta 2}^{n}$, the distances are meaningful for arbitrary distributions and have no relation to the nature of spaces.

iii) The chi-squared divergence $D_{\chi 2}(f,G) = 2D_{\alpha}(F,G)$ on taking $\alpha = -3$ in (3.4.5). Also, the Hellinger distance $D_H(F,G) = \frac{1}{2} D_{\alpha}(F,G)$ on taking $\alpha = 0$m (3.4.5). The Hellinger distance and Bhattacharyya distance are symmetric and have all properties of metric. The relative information generating function of $f$ given the reference measure $g$ is defined as

$$R(F,G,\gamma) = \int_{\chi} \left(\frac{f}{g}(x)\right)^{\gamma-1} f(x)dx \qquad (3.4.9)$$

where $\gamma \geq 1$ and the integral is convergent on noting that $R(F,G,1) = 1$. In particular, $R'(F,G,1)$ is just Kullback-Leilber information and $-R'(F.1,1)$ and $R(F,1,2)$ are Shannon entropy and second order entropy respectively. The power divergence measure (PWD) which gathers most of the interesting specification is indexed by

$$PWD(F,G,\lambda) = \frac{1}{\lambda(\lambda+1)} \int_{\chi} \left\{ \left[\frac{f(x)}{g(x)}\right]^{\lambda} - 1 \right\} f(x)dx \qquad (3.4.10)$$

the power divergence family implies different well-known divergence measures for different values of $\lambda$. PWD for $\lambda = (-2,-1,-0.5,\, 0,1)$ implies Neyman Chi-square, Kullback-Leibler, Squared Hellinger distance, Likelhihood disparity and Pearson Chi-Square divergence respectively. Note that

$$PWD(F,G,\lambda)=\frac{1}{\lambda(\lambda+1)}[R(F,G,\lambda+1)-1]$$

### 3.5 Weighted Residual Entropy

If a unit is known to have survived upto an age $t$, Ebrahimi [42] defined residual entropy of the non-negative continuous random variable $X$ as

$$H(F,t)=-\int_t^\infty \log\left[\frac{f(x)}{\overline{F}(t)}\right]\log\frac{f(x)}{\overline{F}(t)}dx \tag{3.5.1}$$

where $\overline{F}(t)$ is the survival function of *X*. If we put $t=0$, then we get $H(X,0)$ is the Shannon entropy. Nanda *et al.* [96] defined

$$H_1^\beta(F,t)=\frac{1}{\beta-1}\left[1-\int_t^\infty\left[\frac{f(x)}{\overline{F}(t)}\right]^\beta\right]dx \tag{3.5.2}$$

and

$$H_2^\beta(F,t)=\frac{1}{1-\beta}\log\left[\int_t^\infty\left(\frac{f(x)}{\overline{F}(t)}\right)^\beta dx\right] \tag{3.5.3}$$

where $H_1^\beta(F,t)$ and $H_2^\beta(F,t)$ are first kind residual entropy of order $\beta$ and second kind residual entropy of order $\beta$ of the random variable $X$ respectively. It can be noted that as $\beta\to1$, then (3.5.2) and (3.5.3) reduce to residual entropy that defined in (3.5.1) $H_1^\beta(F,t)$ and $H_2^\beta(F,t)$ can always be made non-negative by choosing appropriate $\beta$. In the following results we consider $g(x)=\frac{\omega(x)f(x)}{E_\omega(X)}$ as a weighted version of $f$ and $A(t)=\int_t^\infty\frac{\omega(x)f(x)dx}{\overline{F}(t)}$

i) Let $X$ and $Y$ be two non-negative random variables having densities $f$ and $g$ and distribution functions $F$ and $G$ and survival functions $\overline{F}$ and $\overline{G}$ respectively as defined in previously then $X$ is said to have less uncertainty than $Y$ if $H(F,t) \le H(G,t)$ for all $t \ge 0$. We write $X \le^{LU} Y$.

$X$ is said to be less than $Y$ in (first kind residual entropy of order $\beta$ $\left(written\ X \le^{\beta(1)} Y\right)\Big)$ if $H_1^{\beta}(F,t) \le H_1^{\beta}(G,t)$ for all $t > 0$. $X$ is said to be less than $Y$ in (first kind residual entropy of order $\beta$ $\left(written\ X \le^{\beta(2)} Y\right)\Big)$ if $H_2^{\beta}(F,t) \le H_2^{\beta}(G,t)$ for all $t > 0$. Let $\omega(x) \le A(t)$ for $\forall x > t$, then $H_i^{\beta}(F,t) \le H_i^{\beta}(G,t)$ for all i=1,2. That $g$ is a weighted version of distribution $F$.

ii) $X$ is said to be larger than $Y$ in likelihood ratio ordering $\left(X \ge^{LR} Y\right)$ if $\dfrac{f(x)}{g(x)}$ is a non-decreasing function of $x \ge 0 . \omega(x)$ non-increasing in $x$ implies that $X \ge^{LR} Y$.

iii) Let $X \le^{LR} Y$ and $\lambda_F(x)$ or $\lambda_G(x)$ be non-decreasing in $x$. Then it follows that $X \le^{LU} Y$. So, let $\omega(x)$ be non-decreasing and $\lambda_F(x)$ or $\lambda_G(x)$ be non-decreasing in $x$, the $X \le^{LU} Y$. Also let $A(x)/\omega(x) - \lambda_F(x)$ and $A(x)$ be non-decreasing function of $x$, $X \le^{LU} Y$.

iv) If $\lambda_F(x) \ge E(\omega(X))\left[\log \lambda_F(x) + \log \dfrac{\omega(x)}{A(t)}\right]$ for $\forall t \ge 0, \forall x > t$ and $\lambda_F(x)$ be non-decreasing in . then $X \le^{LU} Y$.

v) A natural question whether residual entropy like mean residual life and hazard rate characterizes survival function or distribution function. Ebrahimi [42] proved that $H(X,t) < \infty, t \ge 0,$ uniquely determine the distribution function $H_i^{\beta}(X,t)$ is increasing in $t$, then $H_i^{\beta}(X,t)$ uniquely determines $\overline{F}(t)$ for i = 1, 2 (Nanda *et al*. [96]).

vi) A non-negative random variable is said to have decreasing (increasing) uncertainty in residual life (DURL (IURL)) if $H(X,t)$ is decreasing (increasing). A non-negative random variable is said to have DURL (IURL) of first kind of order $\beta(DURLF(\beta))$ if $H_1^{\beta}(X,t)$is decreasing in $t \ge 0$. A non-negative random variable is said to have DURL (IURL) of second kind of order $\beta$ (DURL SF( $\beta$ )) if $H_2^{\beta}(X,t)$

is decreasing in $t \geq 0$. In the above definition if we replace the word "decreasing" by "increasing" then we call them IURLF $(\beta)$ and IURLS $(\beta)$ respectively.

vii) The first system is very strongly better than second system if $X \leq^{LU} Y$ and $X \geq^{LR} Y$. So let $\omega(x)$ and $\omega(x)/A(x)$ be non-decreasing functions of $x$, then $H(F,t)-H(G,t)$ is increasing in $t$.

viii) $X$ is said to be stochastically than $Y\left(X \leq^{ST} Y\right)$ if $\bar{F}(x) \leq \bar{G}(x)$ for all $x \geq 0$. Hence for weighted case, if $A(x) \geq E(\omega(x))$ for all $x$, then $X \leq^{ST} Y$.

## 3.6 Weighted Residual Information Measures

In view of Ebrahimi [42], we defined the above measures for the case that after the unit has survived for time $t$ so, assume that the set [t,∞) be the suitable support of distributions and $F$ and $G$ be two distributions which are absolutely continuous w.r.t measure $\mu$ and $\frac{\partial F}{\partial \mu} = f$ and $\frac{\partial G}{\partial \mu} = g$. We have the following definitions:

***Residual Kullback Leibler Information:***

$$D_{KL}(F,G,t) = \int_t^\infty \log \frac{f(x)/\bar{F}(t)}{g(x)/\bar{G}(t)} \frac{f(x)}{\bar{F}(t)} d\mu \tag{3.6.1}$$

***Residual $\chi^2$ – Divergence:***

$$D_{\chi^2}(F,G,t) = \int_t^\infty \frac{\left[\frac{f(x)}{\bar{F}(t)} - \frac{g(x)}{\bar{G}(t)}\right]}{f(x)/\bar{F}(t)} d\mu \tag{3.6.2}$$

***Residual Bhattacharyya Distance and Residual Hellinger Distance:***

$$D_{\beta h}(F,G,t) = \int_t^\infty \sqrt{\frac{g(x)}{\bar{G}(t)} \frac{f(x)}{\bar{F}(t)}} d\mu, \quad D_H(F,G,t) = 2\left[1 - D_{\beta h}(F,G,t)\right] \tag{3.6.3}$$

***Residual α – Divergence:***

$$D_\alpha(F,G,t) = \frac{1}{1-\alpha^2} \int_t^\infty \left\{ 1 - \frac{\left[g(x)/\bar{G}(t)\right]^{\frac{1+\alpha}{2}}}{\left[\frac{f(x)}{\bar{F}(t)}\right]^{\frac{1+\alpha}{2}}} \right\} \frac{f(x)}{\bar{F}(t)} d\mu \tag{3.6.4}$$

The following results are related to the above measures:

- It is easy to see that $D_H(F,G,t) \leq 2$. via Taylor expansion and approximation. We can get $D_{Kl}(F,G,t) \approx D_{\chi^2}(F,G,t), D_J(F,G,t \approx 1/2)$ $\lfloor D_{\chi^2}(F,G,t) + D_{\chi^2}(G,F,t) \rfloor$, $D_{\chi^2}(F,G,t) \approx 4D_H(F,G,t)$
- Sometimes we are interested in the residual distances that is introduced in (3.6.1) to (3.6.4), between the distributions $F = F_{\theta_1}$ and $G = F_{\theta_2}$. Between the corresponding samples distributions which we denote $F^n_{\theta_1}$ and $F^n_{\theta_2}$, the distances are meaningful for arbitrary distributions and have no relation to the nature of spaces. Hence, results can be applicable similar the case that $t = 0$ for any $t$, but not easier than the case $t = 0$.
- The residual chi-squared divergence $D_{\chi^2}(F,G,t) = 2D_\alpha(F,G,t)$ on taking $\alpha = -3$ in (3.4.5) also, the residual Hellinger distance $D_H(F,G,t) = 1/2D_\alpha(F,G,t)$ on taking $\alpha = 0$ in (3.4.5) the relative information generating function of $f$ given the reference measure $g$ is defined as

$$R(F,G,\gamma,t) = \int_t^\infty \left( \frac{f(x)/\bar{F}(t)}{g(x)/\bar{G}(t)} \right)^{\gamma-1} \frac{f(x)}{\bar{F}(t)} d\mu \qquad (3.6.5)$$

where $\gamma \geq 1$ and the integral is convergent on noting that $R(F,G,1,t) = 1$. In particular, $R'(F,G,1,t)$ which is just residual Kullback-Leibler information and $R'(F,G,1,0) + R'(G,F,1,0)$ is residual J–divergence between $F$ and $G$, $-R'(F,1,1,t)$ and $R(F,1,2,t)$ are residual Shannon [122] entropy and residual second order entropy respectively.

The residual power divergence measure (PWD) is indexed by

$$PWD(F,G,t) = \frac{1}{\lambda(\lambda+1)} \int_t^\infty \left\{ \left[ \frac{f(x)/\bar{F}(t)}{g(x)/\bar{G}(t)} \right]^\lambda - 1 \right\} f(x) d\mu \qquad (3.6.6)$$

The power divergence family implies different well-known divergence measures for different values of $\lambda$ PWD for $\lambda = -2,-1,-0.5,0,1,$ implies residual Neyman Chi-square, residual Kullback-Leibler, residual squared Hellinger distance, residual likelihood disparity and residual Pearson Chi-square divergence respectively. Not that

$$PWD(F,G,t) = \frac{1}{\lambda(\lambda+1)}\left[R(F,G,\lambda+1,t)-1\right]$$

- Let ϕ be an invertible increasing function, then $D_\alpha\left(F_1,G_1,\phi^{-1}(t)\right) = D_\alpha\left(\phi(F_1),\ \phi(G_1)t\right)$ because

$$D_\alpha(\phi(F),\phi(G),t) = \frac{1}{1-\alpha^2}\int_{\phi^{-1}(t)}^{\infty}\left\{1-\frac{\left[g(y)/\overline{G}\left(\phi^{-1}(t)\right]^{\frac{1+\alpha}{2}}}{\left[f(y)/\overline{F}\left(\phi^{-1}(t)\right]^{\frac{1+\alpha}{2}}}\right\}^2\frac{f(y)}{\overline{F}\left(\phi^{-1}(t)\right)}dy$$
$$= D_\alpha\left(F,G,\phi^{-1}(t)\right). \qquad (3.6.7)$$

It is clear that for residual Kullback-Leibler information (Ebrahimi *et al* [40]), residual $\chi^2$ divergence, residual Bhattachryya distance, residual Hellinger distance as special cases of (3.6.7) the result holds

- $D_\alpha(F_1,G_1,t)$ is independent of $t$ and only if $F_1$ and $G_1$ have proportional hazard rate. On noting that the only if is easy but for if case, assume that $D_\alpha(F_1,G_1,t) = b$ that $b$ is constant. It is clear that for residual Kullback-Leibler information (Ebrahimi *et al.* [40]) residual $\chi^2$ divergence, residual Bhattacharyya distance, residual Hellinger distance as special cases of the above result. Note that the results achieved via the technique that applied in the Asadi *et al*. [10].

We can consider $f_1(x) = \dfrac{\omega_1(x)f(x)}{E\omega_1(X)}$ and $g_1(x) = \dfrac{\omega_2(x)f(x)}{E\omega_2(X)}$ as the weighted distributions of $f$. Then, the above measures are expressed as:

Residual Kullback-Leibler information:

$$D_{KL}(F_1,G_1,t)=\frac{1}{A_1(t)\bar{F}(t)}\int_t^{\infty}\log\left[\frac{\omega_1(x)}{\omega_2(x)}\right]\omega_1(x)f(x)dx+\log\left[\frac{A_1(t)}{A_2(t)}\right] \quad (3.6.8)$$

where $A_1(t)\bar{F}(t)=\int_t^{\infty}\omega_i(x)f(x)dx$ for i = 1, 2,

**Residual $\chi^2$ – Divergence:**

$$D_{\chi^2}(F_1,G_1,t)=\frac{1}{A_2(t)\bar{F}(t)}\int_t^{\infty}\left[\frac{[\omega_2(x)]^2}{\omega_1(x)}\frac{A_1(t)}{A_2(t)}\right]f(x)dx \quad (3.6.9)$$

***Residual Bhattacharya Distance and Residual Hellinger Distance:***

$$D_{\beta h}(F_1,G_1,t)=\frac{1}{\bar{F}(t)}\int_t^{\infty}\sqrt{\frac{\omega_1(x)\omega_2(x)}{A_1(t)A_2(t)}}f(x)dx\,,$$

$$D_H(F_1,G_1,t)=2-2D_{\beta h}(F_1,G_1,t) \quad (3.6.10)$$

***Residual $\alpha-$divergence:***

$$D_\alpha(F_1,G_1,t)=\frac{1}{(1-\alpha^2)A_1(t)\bar{F}(t)}\int_t^{\infty}\left\{1-\left[\frac{\omega_2(x)A_1(t)}{\omega_1(x)A_2(t)}\right]^{\frac{1+\alpha}{2}}\right\}^2\times\omega_1(x)f(x)dx$$

(3.6.11)

- For the case that $\omega_1(x)=1,\ A_1(t)=1$ statements in (3.6.8) to (3.6.11). Change to simple statements that their calculation is easier than the previous statements.
- For the above residual information measures, for weights like, order statistics, record value, proportional hazard rate, reversed proportional hazard rate, hazard rate, selection samples, we can find the values of these residual measures and some properties of them in special cases some of them lead us to calculating the integrals via incomplete gamma and incomplete beta functions.

## 3.7 Generalized Measures of Discrimination between Past Lifetime Distributions

The measure (3.2.3) can be generalized in many ways. Here we study parametric generalization by introducing a parameter $\beta$. Thus we get a class of discrimination measures of which (3.2.3) is a particular case. New generalized measure has more flexibility application due to infinite values of $\beta$. However, here we consider the following generalization.

$$\bar{I}_{X,Y}^{\beta}(t)=\frac{1}{\beta-1}\left[\int_0^t\left(\frac{f(x)}{F(t)}\right)^{\beta}\left(\frac{g(x)}{G(t)}\right)^{1-\beta}dx-1\right],\ \beta\neq 1,\ \beta>0 \qquad (3.7.1)$$

It may be noted that (3.7.1) reduces to (3.1.3) when $\beta\to 1$ so we may call (3.7.1) the discrimination information measure of degree $\beta$.

Consequently, the measure (3.2.4) can also be generalized as follows:

$$\bar{I}_{X,Y}^{\beta}(t)=-\int_0^t\frac{f(x)}{F(t)}\left(\frac{g(x)/G(t)}{f(x)/F(t)}\right)^{1-\beta}\log\frac{g(x)}{G(t)}dx-\bar{H}_X^{\beta}(t) \qquad (3.7.2)$$

where

$$\bar{H}_X^{\beta}(t)=-\int_0^t\frac{f(x)}{F(t)}\left(\frac{g(x)/G(t)}{f(x)/F(t)}\right)^{1-\beta}\log\frac{f(x)}{F(t)}dx.$$

Next we obtain some bounds of $\bar{I}_{X,Y}^{\beta}(t)$ using the term “increasing” and “decreasing” in non-strict sense.

**Theorem 3.7.1:**

If $f(t)/g(t)$ is increasing in $t>0$ i.e. $X\geq_{lr}Y$, then

$$\bar{I}_{X,Y}^{\beta}(t)\leq\left(\frac{\tau_X(t)}{\tau_Y(t)}\right)^{\beta-1}\log\frac{\tau_X(t)}{\tau_Y(t)}.$$

***Proof:*** From the measure (3.7.1) we have

$$\bar{I}_{X,Y}^{\beta}(t)=\frac{1}{\beta-1}\left[\int_0^t\left(\frac{f(x)}{F(t)}\right)^{\beta}\left(\frac{g(x)}{G(t)}\right)^{1-\beta}dx-1\right]$$

$$=\frac{1}{\beta-1}\left[\int_0^t\left(\frac{f(x)}{F(t)}\right)\left(\frac{G(t)}{g(x)}\right)^{\beta-1}\left(\frac{f(x)}{F(t)}\right)^{\beta-1}dx-1\right]$$

Due to increasing of $f(t)g(t)$, from (3.2.3) for all $t>0$ we have

$$=\frac{1}{\beta-1}\left[\int_0^t\left(\frac{f(x)}{F(t)}\right)\left(\frac{G(t)}{g(x)}\right)^{\beta-1}\left(\frac{f(x)}{F(t)}\right)^{\beta-1}dx-1\right.$$

$$\le\frac{1}{\beta-1}\left[\frac{1}{F(t)}\int_0^t f(x)\left(\frac{f(t)/F(t)}{g(t)/G(t)}\right)^{\beta-1}dx-1\right]$$

$$=\frac{1}{F(t)}\int_0^t f(x)\left(\frac{f(t)/F(t)}{g(t)/G(t)}\right)\log\left(\frac{f(t)/F(t)}{g(t)/G(t)}\right)dx$$

$$=\frac{1}{F(t)}\left(\frac{f(t)/F(t)}{g(t)/G(t)}\right)^{\beta-1}\log\left(\frac{f(t)/F(t)}{g(t)/G(t)}\right)\int_0^t f(x)dx$$

$$=\left(\frac{f(t)/F(t)}{g(t)/G(t)}\right)^{\beta-1}\log\left(\frac{f(t)/F(t)}{g(t)/G(t)}\right)$$

$$\bar{I}^{\beta}_{X,Y}(t)=\left(\frac{\tau_X(t)}{\tau_\gamma(t)}\right)^{\beta-1}\log\frac{\tau_X(t)}{\tau_\gamma(t)}.$$

Hence proved.

When $f(t)/F(t)$ is decreasing the proof is similar.

## 3.8 Applications

Let $X$ and $Y$ have weighted exponential distribution with rates $\lambda$ and $\mu$ respectively and with weight function $\phi(t)$. Their densities are given by

$$f(t)=\frac{\phi(t)e^{-\lambda t}}{h(\lambda)},\ g(t)=\frac{\phi(t)e^{-\mu t}}{h(\mu)}\qquad t>0 \tag{3.8.1}$$

which $h(\cdot)$ denotes the Laplace transform of $\phi(t)$

$$h(\xi)=L_\xi[\phi(t)]=\int_0^\infty e^{-\xi t}\phi(t)dt\qquad \xi>0$$

solving the measure (3.2.5) and (3.8.1) Crescenzo and Longobadi [35] expressed the discrimination measure of the past lives for $\lambda \neq \mu$ as follows:

$$\bar{I}_{X,Y}(t) = \log\frac{G(t)}{F(t)} + \log\frac{h(\mu)}{h(\lambda)} + (\mu-\lambda)E[X|X \le t] \qquad t > 0 \qquad (3.8.2)$$

Next we generalize measure (3.8.1) in the following way:

$$\bar{I}^{\beta}_{X,Y}(t) = \frac{1}{F(t)}\left[\left(\frac{G(t)}{F(t)}\right)^{\beta-1} \log\left(\frac{G(t)}{F(t)}\right)\int_0^t f(x)\left(e^{(\mu-\lambda)x}\frac{h(\mu)}{h(\lambda)}\right)^{\beta} dx\right]$$

$$+\left(\frac{h(\mu)}{h(\lambda)}\right)^{\beta-1}(\mu-\lambda)\left(\frac{G(t)}{F(t)}\right)^{\beta-1}\int_0^t xf(x)\left(e^{(\mu-\lambda)x}\right)^{\beta-1}dx$$

$$+\left(\frac{G(t)}{F(t)}\right)^{\beta-1}\log\frac{h(\mu)}{h(\lambda)}\left(\frac{h(\mu)}{h(\lambda)}\right)^{\beta-1}\int_0^t f(x)\left(e^{(\mu-\lambda)}\right)^{\beta-1}dx \quad (3.8.3)$$

It may be noted that (3.8.3) reduces to (3.8.2) when $\beta \to 1$.

**Example 3.7.1:** Let $X$ and $Y$ be Erlang distributed, with scale parameter $\lambda$ and $\mu$ respectively. With the common shape parameter *n* hence they possess densities given by (3.8.1), where $\phi(t) = t^{n-1}$ and $h(\xi) = (n-1)!\xi^n$.

If we put $f(x) = \dfrac{\phi(x)\cdot e^{-\lambda x}}{h(\lambda)}$, where $h(\lambda)\dfrac{(n-1)!}{\lambda^n}$ and $\phi(x) = x^{n-1}$ in (3.8.3) then

$$\bar{I}^{\beta}_{X,Y}(t) = \frac{1}{F(t)}[\cdot\left(\frac{G(t)}{F(t)}\right)^{\beta-1} \log\left(\frac{G(t)}{F(t)}\right)\int_0^t f(x)\left(e^{(\mu-\lambda)^x}\frac{h(\mu)}{h(\lambda)}\right)^{\beta-1} dx$$

$$+\left(\frac{h(\mu)}{h(\lambda)}\right)^{\beta-1}(\mu-\lambda)\left(\frac{G(t)}{F(t)}\right)^{\beta-1}\int_0^t \left(xe^{-\lambda x}x^{n-1}\right)\lambda^n dx$$

$$+\left(\frac{G(t)}{F(t)}\right)^{\beta-1}\cdot n\log\frac{\lambda}{\mu}\left(\frac{h(\mu)}{h(\lambda)}\right)^{\beta-1}\int_0^t f(x)\left(e^{(\mu-\lambda)}\right)^x dx \quad (3.8.4)$$

In case $\beta \to 1$, (3.8.4) reduces to

$$\bar{I}_{X,Y}(t) = \log\frac{G(t)}{F(t)} + \frac{(\mu-\lambda)}{F(t)}\cdot\frac{\gamma((n+1),\lambda t)}{(n-1)!\lambda} + n\log\frac{\lambda}{\mu}$$

This is clearly indicated in the Fig. 3.8.1 & 3.8.2

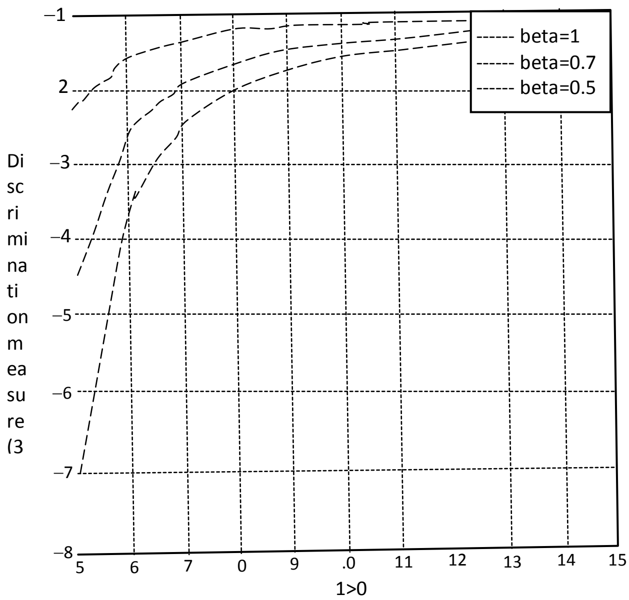



**Fig. 3.8.1:** The discrimination measure (3.8.4) for $\mu > \lambda$ (from top to bottom)

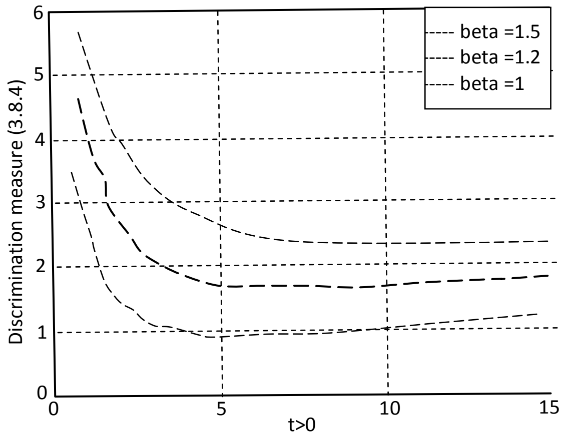



**Fig. 3.8.2:** The discrimination measure (3.8.4) for $\mu > \lambda$

The following example shows that discrimination measure $\bar{I}^{\beta}_{X,Y}(t)$ is not necessarily monotone.

**Example 3.8.2:** Let

$$F(x)=\begin{cases}\exp\left\{-\dfrac{1}{2}-\dfrac{1}{x}\right\} & if \quad 0<x\le 1\\[2ex] \exp\left\{-2+\dfrac{x^2}{2}\right\} & if \;\; 1<x\le 2`\end{cases}$$

$$G(x)=\begin{cases}\dfrac{x^2}{4} & if \quad 0<x\le 2\\ 1 & if \quad x\ge 2\end{cases}$$

be the c.d.f's of $X$ and $Y$. From figure 3.8.3 we see that $\bar{I}^{\beta}_{X,Y}(t)$ is not monotone for all $t\in(0,2)$.

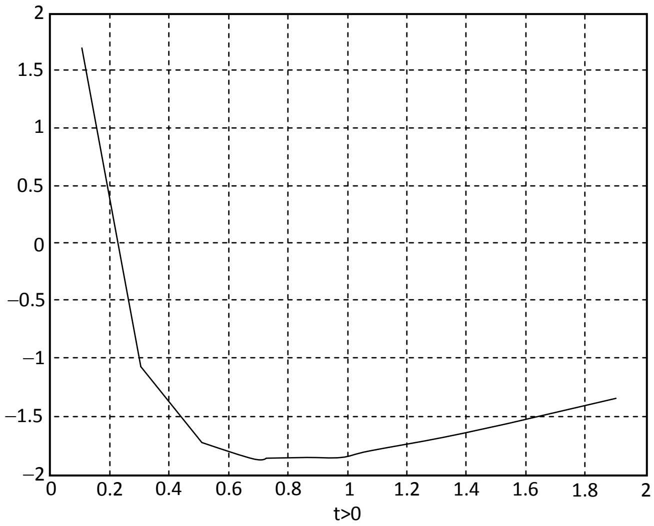



**Fig. 3.8.3:** The discrimination measure of past lives of example (3.8.2) is plotted for $t > 0$.

# Chapter-IV: Cumulative Residual Entropy and its Properties

## 4.1: Introduction

Let $X$ be a non-negative random variable with distribution function $F$ and density function $f$ (in the case when $X$ is continuous) respectively. Shannon [122] introduced a measure of uncertainty, associated with the probability distribution $F$, which is also known as the Shannon entropy (or 'differential entropy' in the continuous case). The differential entropy of a non-negative continuous random variable $X$, is defined as:

$$H(X) = -\int_0^\infty f(x) \log f(x) dx \tag{4.1.1}$$

The differential entropy plays the central role in information theory and a large number of research work are available in the literature in both theory and applications.

Rao *et al.* [114] introduced an alternative measure of uncertainty called Cumulative Residual Entropy (CRE). This measure is based on the cumulative distribution function $F$ and is defined in the univariate case and for non-negative variables as follows:

$$\xi(X) = -\int_0^\infty \overline{F}(x) \log \overline{F}(x) dx \tag{4.1.2}$$

They have obtained several properties of this measure and provided some applications of it in reliability engineering and computer vision. Rao [115] developed some new mathematical properties of CRE and gave an alternative formula for it.

The aim of the present chapter is to show that CRE is connected to some well-known reliability measures. In particular in section 4.2, CRE is infact the expectation of the mean residual lifetime of $X$ which plays an important role in reliability and survival analysis. In section 4.3 we define a dynamic (Time-dependent) CRE (DCRE) and show that DCRE is connected to the mean residual life function (MRL). In section 4.4 we study some properties of DCRE. We give the necessary and sufficient conditions under which DCRE in monotone. Further, we show that under some conditions DCRE has a one-to-one relation with the underlying distribution function.

## 4.2 Relation between CRE and mean residual life function

The MRL and HR plays important roles in reliability and survival analysis to model and analyze the data let $X$ be a continuous random variable with survival function $\overline{F} = 1 - F$ and density function f. The MRL of $X$, which we denote by $m_F$, is defined as

$$m_F(t) = E(X - t | X \geq t) = \frac{\int_t^\infty \overline{F}(x)dx}{\overline{F}(t)}$$

for $t$ such that $\overline{F}(t) > 0$. It is well-known that MRL $m_F'(t)$ determines the distribution function $F$ uniquely. The HR of $X$, which we denote by $r_F(t)$, is defined as

$$r_F(t) = \frac{f(t)}{\overline{F}(t)} \tag{4.2.1}$$

for $t$ such that $\overline{F}(t) > 0$. It is also well-known that there is a one-to-one relation between $r_F(t)$ and the distribution function $F$. The relation between the MRL $m_F(t)$ and HR $r_F(t)$ is as follows:

$$r_F(t) = \frac{m_F'(t)+1}{m_F(t)} \tag{4.2.2}$$

where $m_F'(t)$ denotes the derivative of $m_F(t)$ with respect to $t$.

Teitler et al [131] have shown that a useful representation of the differential entropy in terms of HR is as follows:

$$H(X) = 1 - \int_0^\infty f(x) \log r_F(x) dx = 1 - E(\log r_F(X)).$$

In the following theorem we show that, in the case where the underlying distribution $F$ is absolutely continuous the CRE has a direct relation to the MRL $m_F(t)$.

**Theorem 4.2.1:** Let $X$ be a non-negative continuous random variable with MRL function $m_F$ and CRE $\xi(X)$, such that $\xi(X) < \infty$ then

$$\xi(X) = E(m_F(X)) \tag{4.2.3}$$

***Proof:*** We know that

$$E(m_F(X)) = \int_0^\infty \frac{\int_t^\infty \overline{F}(x)dx}{\overline{F}(t)} f(t)dt$$

$$= \int_0^\infty \left( \int_t^\infty \overline{F}(x)dx \right) \frac{f(t)}{\overline{F}(t)} dt$$

$$= \int_0^\infty \left( \int_0^x \frac{f(t)}{\overline{F}(t)} dt \right) \overline{F}(x)dx$$

$$= \int_0^\infty \left( \int_0^x r_F(t)dt \right) \overline{F}(x)dx$$

$$= \int_0^\infty \left( -\log \overline{F}(x) \right) \overline{F}(x)dx \tag{4.2.4}$$

$$= \xi(X).$$

Hence the proof is complete.

**Example 4.2.1:**

a) If $X$ is distributed as exponential with mean $\lambda$, then it is well-known that the MRL function of $X$ is $\lambda$ (which does not depend on $t$). Hence, we have $\xi(X)=\lambda$.

b) If $X$ is distributed uniformly on (0, a), a > 0, then it can be easily seen that $m_F(t)=(a-t)/2$. From this it can be shown that $\xi(X)=E((a-X)/2)=\frac{1}{4}a.$

c) If $X$ has Pareto distribution with density function

$$f(x)=\begin{cases}\frac{\alpha\beta^{\alpha}}{(x+\beta)^{\alpha+1}}, & 0\le x,\quad \alpha>1,\quad \beta>0\\ 0 & otherwise\end{cases}$$

then it can be easily seen that in this case

$$m_F(t)=\frac{1}{(\alpha-1)(t+\beta)}$$

Hence

$$\xi(X)=\frac{1}{\alpha-1}E(X+\beta)=\frac{\alpha\beta}{(\alpha-1)^2}$$

**Remark 4.2.1:** Classification of distributions with respect to ageing properties is a popular theme in reliability theory. A class of distributions which arises in the study of replacement and maintenance policies is the class of new better (worse) than used in expectation (NBUE) (NWUE) distributions. Let $X$ be the lifetime of a component with a continuous distribution function and the MRL function $m_F$. $F$ is said to be a new better (worse) than used in expectation (NBUE) (NWUE) distribution if

$$m_F(t)\le(\ge)m(0)=\mu \qquad t\ge 0.$$

If we assume that $F$ is NBUE (NWUE) then based on representation (4.2.3)

$$\xi(X)=E(m_F(X))\le(\ge)\mu \tag{4.2.5}$$

This gives an upper (lower) bound for the CRE of the statistical models which are in the class of NBUE (NWUE) distributions. That is, when $F$ is NBUE one can easily conclude that the amount of uncertainty of that, based on the measure CRE, is at most equal to the mean of $F$. Also when $F$ is in the class of NWUE distributions one can conclude that the amount of uncertainty of $F$, based on measure CRE, is at least equal to the mean of $F$.

Rao *et al*. [114], shown that when $X$ is a non-negative random variable, then

$$\xi(X) \leq \frac{E(X^2)}{2E(X)} \tag{4.2.6}$$

It is shown by Hall and Wellner [60] that for a NBUE (NWUE) distribution the coefficient of variation (which is defined as the ratio of the standard deviation and the mean) is always less than (greater than) unity. This implies that

$$\frac{E(X^2)}{2E(X)} < E(X).$$

Hence the upper bound given by Rao *et al*. [114] is sharper than the upper bound in (4.2.5) in the NBUE case. However, it should be pointed out that to have the upper bound (4.2.5) one needs only the knowledge of $E(X)$ while for using (4.2.6) as a upper bound one needs to have both $E(X)$ and $E(X^2)$. To illustrate the above results let us consider the following examples.

**Example 4.2.2:** The mixture of distributions arise naturally in many branches of statistics and applied probability. Let $X$ be distributed as the mixture of two exponential distributions with mean $\lambda_1$ and $\lambda_2$ respectively. Then the survival function of $X$ is given by

$$\overline{F}(X) = Pe^{-x/\lambda_1} + (1-P)e^{-x/\lambda_2}$$

where $P \in (0,1)$. It is well-known that, the class of NWUE distributions includes the class of distributions with decreasing HR. On the other hand $X$ has a decreasing HR and hence it is in the class of NWUE distributions. Barlow and Proschan [16]. Thus based on (4.2.5) we obtain

$$\xi(X) \geq p\lambda_1 + (1-p)\lambda_2 \tag{4.2.7}$$

where the right-hand side is the mean of $X$. This shows that for the mixture of two exponential distributions, the CRE is at least $p\lambda_1 + (1-p)\lambda_2$. Based on the result (4.2.6) the upper bound of the CRE for this model is given as follows:

$$\xi(X) \leq \frac{p\lambda_1^2 + (1-P)\lambda_2^2}{P\lambda_1 + (1-P)\lambda_2} \tag{4.2.8}$$

**Example 4.2.3:** Let $X$ be distributed as Gamma $(\alpha, \beta)$ with probability density function

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \quad x > 0, \quad \alpha > 0, \quad \beta > 0 \tag{4.2.9}$$

for $\alpha \in (0, 1)$. The Gamma distribution has decreasing HR (Barlow and Proschan, [16]). Hence for $\alpha \in (0, 1)$. It is also NWUE and the lower bound (4.2.5) for the uncertainty CRE is given by

$$\xi(X) \geq \alpha\beta.$$

Based on the result (4.2.6) the upper bound for CRE in this case is

$$\xi(X) \leq \frac{\beta(1+\alpha)}{2} \tag{4.2.10}$$

Note, on the other hand, that for α >1, the Gamma distribution is NBUE. Also it is trival that for $\alpha > 1$, $(\beta(1+\alpha))/2 \leq \alpha\beta$. This means that the upper bound (4.2.6) is sharper than the upper bound (4.2.5). Note also that the upper bound in (4.2.10) holds for all values of $\alpha \in (0, \infty)$.

## 4.3 Dynamic Cumulative Residual Entropy

Study of duration is a subject of interest in many branches of science such as reliability, survival analysis, actuary, economics, business and many other fields. Let $X$ be a non-negative random variable denoting a duration such as a lifetime where we assume that it has the distribution function $F$, and the probability density function $f$. Capturing effects of the age $t$ of an individual or an item under study on the information about the residual lifetime is important in many applications. e.g., in reliability when a component or a system of components is working at time $t$, one is interested in the study of the lifetime of component or system beyond $t$. In such case, the set of interest is the residual lifetime

$$\gamma_t = \{x : x > t\}.$$

Hence the distribution of interest for computing uncertainty and information is the residual distribution with survival function

$$\overline{F}_t(x) = \begin{cases} \dfrac{\overline{F}(x)}{\overline{F}(t)}, & x \in \gamma_t \\ 1 & otherwise \end{cases} \tag{4.3.1}$$

Where $\overline{F}$ denotes the survival function of $X$.

Ebrahimi [42] defined the concept of dynamic Shannon entropy and obtained some properties of that since then several attempts have been made to study and extend the concept of dynamic Shannon's entropy. Among others, Asadi *et al*. [8] studied the maximum dynamic entropy models. Asadi *et al*. [9] introduced the minimum dynamic discrimination information approach to probability modelling. Asadi *et al.* [10] developed dynamic information divergence and entropy of order $\alpha$, which is also known as Renyi information and entropy, respectively.

Now the CRE for the residual lifetime distribution with survival function $\overline{F}_t(x)$ is

$$\xi(X;t) = -\int_t^\infty \overline{F}_t(x) \log \overline{F}_t(x) dx$$

$$= -\int_t^\infty \frac{\overline{F}(x)}{\overline{F}(t)} \log \frac{\overline{F}(x)}{\overline{F}(t)} dx$$

$$= -\frac{1}{\overline{F}(t)}\int_t^{\infty} \overline{F}(x)\log \overline{F}(x)dx + m_F(t)\log \overline{F}(t) \qquad (4.3.2)$$

Asadi and Zohrevand [11] introduced a new measure of uncertainty, the CRE for the residual lifetime distribution that is, CRE of $X_t$. This function, called the dynamic cumulative residual entropy (DCRE) is defined by

$$\xi(X;t) = -\int_t^{\infty} \frac{\overline{F}_X(X)}{\overline{F}_X(t)}\log \frac{\overline{F}_X(X)}{\overline{F}_x(t)}dx \qquad (4.3.3)$$

It is clear that $\xi(X;0) = \xi(0)$. From (4.3.3) the DCRE can be rewritten as

$$\xi(X;t) = e_X(t)\log \overline{F}_X(t) - \frac{1}{\overline{F}_X(t)}\int_t^{\infty} \overline{F}_X(X)\log \overline{F}_X(X)dx \qquad (4.3.4)$$

It can be easily seen that for each $t, t \geq 0, \xi(X;t)$ possesses all the properties of the CRE (4.1.2). It is worth noting that $\xi(X;t)$ provides a dynamic information measure for measuring the information of the residual life distribution. We call this as dynamic cumulative residual entropy (DCRE). It is clear that $\xi(X;0) = \xi(X)$.

**Theorem 4.2.2** Let $F$ be an absolutely continuous distribution function with DCRE $\xi(X;t)$ and the MRL $m_F(t)$, such that $\xi(X;\ t) < \infty$ for all $t \geq 0$. Then

$$\xi(X;\ t) = E(m_F(X) |\ X \geq t)$$

***Proof:*** The result follows easily using the same steps as used to prove theorem 4.2.1.

**Example 4.2.3:** Again Consider Example 4.2.2, then

**a)** For the exponential distribution as we have $m_F(t) = \lambda$, we get $\xi(X;t) = \lambda$, which does not depend on $t$.

**b)** The DCRE for a uniform distribution is

$$\xi(X;t) = E(m_F(x) | X \geq t) = E\left(\frac{\alpha - X}{2} | X > t\right) = \frac{\alpha - t}{4}$$

This shows that the DCRE for uniform distribution is a decreasing function of $t$. Hence as $t$ gets larger the uncertainty gets smaller.

**c)** For Pareto distribution we have

$$\xi(X;t)=E\left(m_F(X)\middle|X\geq t\right)=E\left(\frac{X+\beta}{\alpha-1}\middle|X>t\right)=\frac{\alpha(t+\beta)}{(\alpha-1)^2}.$$

This shows that the DCRE of the Pareto distribution is an increasing function of $t$. Hence as $t$ gets larger the uncertainty gets larger.

## 4.4 Properties of Dynamic Cumulative Residual Entropy

**Definition 4.4.1:** The distribution function $F$ is said to be increasing (decreasing) DCRE of $\xi(X;t)$ is an increasing (decreasing) function of $t$.

The following theorem gives the necessary and sufficient conditions for $\xi(X;t)$ to be increasing (decreasing) DCRE.

**Theorem 4.4.1:** The distribution function *F* is increasing (decreasing) DCRE. If and only if for $t\geq 0$.

$$\xi(X;t)\geq(\leq)m_F(t) \tag{4.4.1}$$

***Proof:*** Differentiation of $\xi(X;t)$in (4.3.2), where we assume that the derivative exists with respect to $t$, implies that

$$\frac{\partial\xi(X;t)}{\partial t}=r_F(t)(\xi(X;t)-m_F(t)) \tag{4.4.2}$$

Which in turn implies the assertion of the theorem.

**Corollary 4.4.1:** The distribution function $F$ has constant DCRE if and only if it is an exponential distribution.

***Proof:*** An absolutely continuous distribution function $F$ is both increasing and decreasing DCRE if and only if $\partial\xi(X;t)/\partial t=0$. Using theorem 4.4.1. This is equivalent to saying that for $t,\, t\geq 0$

$$\xi(X;t)=m_F(t)$$

Now the result follows from example 4.2.1

**Corollary 4.4.2:** An absolute continuous distribution function $F$ has decreasing (increasing) DCRE if and only if the corresponding MRL is decreasing (increasing).

***Proof*:** To prove the 'if' part, assume that the MRL $m_F(t)$ is decreasing (increasing) then one can easily see that

$$\xi(X;t)=\frac{\int_t^\infty m_F(t)f(x)dx}{\bar{F}(t)}\le(\ge)m_F(t).$$

Hence we have

$$\frac{\partial\xi(X;t)}{\partial t}=r_F(t)(\xi(X;t)-m_F(t))\le(\ge)r_F(t)(m_F(t)-m_F(t))=0$$

This shows that $\xi(X;t)$ is decreasing (increasing) the proof of the 'only if' part is obtained based on theorem 4.4.1.

Table 4.4.1 gives some applications of corollary 4.4.2. In the table we have presented some well-known statistical models such as Pareto Weibull, Gamma, Rayleigh and Half logistic. These models have monotone MRLs and consequently from corollary 4.4.2 they have also monotone DCREs.

**Table 4.4.1**

| Model | MRL | DCRE |
|---|---|---|
| Pareto<br>$f(x)=\frac{x\beta^2}{(\alpha+\beta)^{\alpha+1}},\qquad \alpha>1,\beta>0$ | Increasing | Increasing |
| Weibull<br>$f(x)=\frac{\alpha x^{\alpha-1}}{\beta^\alpha}e,\qquad \beta>0$<br>$0<\alpha<1$<br>$1<\alpha$ | Increasing<br>Decreasing | Increasing<br>Decreasing |
| Gamma<br>$f(x)=\frac{1}{I^*(\alpha)\beta^\alpha}x^{\alpha-1}e^{-\alpha\beta},\qquad \beta>0$<br>$0<\alpha<1$<br>$1<\alpha$ | Increasing<br>Decreasing | Increasing<br>Decreasing |
| Rayleigh<br>$f(x)=(\alpha+\beta x)e^{-(\alpha x+\beta/2x^2)},\qquad \alpha,\beta>0$ | Decreasing | Decreasing |
| Half Logistic<br>$f(x)=\frac{2\alpha e^{\alpha x}}{(1+e^{\alpha x})^2},\qquad \alpha>0$ | Decreasing | Decreasing |

**Theorem 4.4.2:** Let $X$ and $Y$ be two non-negative continuous random variables with survival functions $\bar{F}(t)$ and $\bar{G}(t)$, MRL functions $m_F(t)$ and $m_G(t)$ and HRs $r_F(t)$ and $r_G(t)$, respectively. Let for $t \geq 0$, $r_F(t) \leq r_G(t)$ and $m_F(t)$ be an increasing function of $t$. Then

$$\xi(X;t) \geq \xi(Y:t)$$

***Proof*:** The assumption that $r_F(t) \leq r_G(t)$ implies that $m_F(t) \geq m_G(t)$, for all $t$, $t \geq 0$. Hence

$$\xi(Y:t) = \frac{\int_t^\infty m_G(x)g(x)dx}{\bar{G}(t)} \leq \frac{\int_t^\infty m_F(x)g(x)dx}{\bar{G}(t)} \leq \frac{\int_t^\infty m_F(x)f(x)dx}{\bar{F}(t)}$$

$$= \xi(X;t) \tag{4.4.3}$$

where the last inequality is obtained from the fact that $r_F(t) \leq m_G(t)$ implies that $E(h(X)|X \geq t) \geq E(h(Y)|Y \geq t)$ for all increasing function $h$.

The following example shows an application of this theorem.

**Example 4.4.2:** Let $X_1, \ldots, X_n$, $n \geq 2$, denote the lifetimes of $n$ independent and identically distributed components which are connected in a series system. Assume that the $X_i's$ are distributed as Weibull with survival function

$$\bar{F}(t) = e^{-(t/\beta)^\alpha}, \; t \geq 0, \quad \beta > 0, \quad 0 < \alpha < 1$$

Assume $r_F(t)$ denotes the common HR of the components and $r_n(t)$ denotes the HR of the system. Then using the fact that the lifetime of the system is $Y = \min(X_1, \ldots, X_n)$ we get

$$\bar{F}_n(t) = e^{-n(t/\beta)^\alpha}, \; t \geq 0, \quad \beta > 0, \quad 0 < \alpha$$

where $\bar{F}_n$ denotes the survival function of the system.

Hence

$$r_n(t) = n\alpha \frac{t^{\alpha-1}}{\beta^{\alpha}} \geq \alpha \frac{t^{\alpha-1}}{\beta^{\alpha}} = r_F(t)$$

for all $t \geq 0$. On the other hand since $r_F(t)$, in this case, is a decreasing function of $t$, the common MRL of the components is an increasing function of $t$. Hence based on theorem (4.4.2), we get that

$$\xi(X;t) \geq \xi(Y:t).$$

An important question regarding the DCRE is whether it characterizes the underlying distribution function uniquely, in the following theorem we show that when DCRE $\xi(X;t)$ is an increasing function of $t$ then the relation between the distribution function $F$ and $\xi(X;t)$ is one-to-one.

**Theorem 4.4.3:** Let $X$ and $Y$ be two non-negative absolutely continuous random variables with survival functions $\bar{F}(t)$ and $\bar{G}(t)$, MRL functions $m_F(t)$ and $m_G(t)$ and HRs $r_F(t)$ and $r_G(t)$, respectively. Let the DCREs $\xi(X;t)$ and $\xi(Y;t)$ corresponding to $X$ and $Y$ be increasing functions of $t$. If for all $t \geq 0$; $\xi(X;t) = \xi(X;t)$ then $\bar{F}(t) = \bar{G}(t)$.

***Proof:*** Under the assumptions that $\xi(X;t)$ and $\xi(Y;t)$ are differentiable with respect to $t$. We get by differentiating both sides of $\xi(X;t) = \xi(Y;t)$

$$r_F(t)(\xi(X;t) - m_F(t)) = r_G(t)(\xi(Y;t) - m_G(t)) \tag{4.4.4}$$

If for all values of $t$, $t \geq 0$, $r_F(t) = r_G(t)$ then $\bar{F}(t) = \bar{G}(t)$and the proof is complete. Hence, we assume that there exists some $t_\theta$ such that $r_F(t_\theta) \neq r_G(t_\theta)$. Without loss of generality, we assume that $r_G(t_\theta) > r_F(t_\theta)$. Using this and equation (4.4.4) with $t = t_\theta$, we get

$$\xi(X;t_\theta) - m_F(t_\theta) > \xi(Y;t_\theta) - m_G(t_\theta). \tag{4.4.5}$$

which implies that

$$m_F(t_\theta) < m_G(t_\theta)$$

this is a contradiction with the fact that $r_G(t_\theta) > r_F(t_\theta)$ implies that $m_F(t_\theta) > m_G(t_\theta)$. Therefore for all values of $t$ we have $r_F(t) = r_G(t)$ or equivalently $\overline{F}(t) = \overline{G}(t)$. Hence the proof is complete. In the following theorem we give characterization of some well-known distributions in terms of DCRE.

**Theorem 4.4.4:** Let $X$ be a non-negative absolutely continuous random variable with survival function $\overline{F}(t)$, MRL $m_F(t)$, HR $r_F(t)$, and DCRE $\xi(X;t)$. Then

$$\xi(X;t) = cm_F(t),\ \ C > 0. \tag{4.4.6}$$

If and only if $X$ distributed as

i) Exponential where c=1

ii) Power where 0<c<1

iii) Pareto where c>1.

***Proof:*** The 'if' part of the theorem is straight forward to prove. To prove the 'only if' part of equation (4.4.6) hold. That is

$$m_F(t)\log \overline{F}(t) - \frac{1}{\overline{F}(t)}\int_t^\infty \overline{F}(x)\log \overline{F}(x)dx = cm_F(t)$$

Differentiating both sides of this with respect to $t$ gives

$$cm'_F(t) = m'_F(t)\log \overline{F}(t) - r_F(t)m_F(t) + \log \overline{F}(t) - r_F(t)\frac{1}{\overline{F}(t)}\int_t^\infty \overline{F}(x)\log \overline{F}(x)dx \tag{4.4.7}$$

$$= m'_F(t)\log \overline{F}(t) - r_F(t)m_F(t) + \log \overline{F}(t) + r_F(t)\left[cm_F(t) - m_F(t)\log \overline{F}(t)\right] \tag{4.4.8}$$

on the other hand using equation (4.2.2) we have

$$m'_F(t) = m_F(t)r_F(t) - 1$$

substituting $m'(t)$ from this last equation in (4.4.8) we get

$$c(m_F(t)r_F(t) - 1) = (m_F(t)r_F(t) - 1)\log \overline{F}(t) - r_F(t)m_F(t)$$
$$+ \log \overline{F}(x) + cr_F(t)m_F(t) - r_F(t)m_F(t)\log \overline{F}(t)$$

from this we obtain for any $x > 0$

$$m_F(x) \cdot r_F(x) = c$$

or again using (4.2.2)

$$m'_F(x) = C - 1$$

Integrating both sides of this last equation with respect to $x$ on $(\theta, t)$ yields the following linear form for $m_F(t)$

$$m_F(t) = (C-1)t + m_F(\theta). \tag{4.4.9}$$

Hall and Wellner [60] proved that the MRL function of a continuous random variable $X$ is linear of the form (4.4.9), if and only if the underlying distribution is exponential (C=1), Pareto (C>1) or power (0<C<1) this completes the theorem.

**Theorem 4.4.5:** Let $X$ and $Y$ be two non-negative absolutely continuous random variables with survival functions $\overline{F}(t)$ and $\overline{G}(t)$, MRL functions $m_F(t)$ and $m_G(t)$, HR $r_F(t)$ and $r_G(t)$, and DRCEs $\xi(X;t)$ and $\xi(Y;t)$, respectively assume that for all $t > 0$

$$m_G(t) = c m_F(t)$$

where $c > 1$ is real valued. If $\xi(X;t)$ is an increasing function of $t$, then $\xi(Y;t)$ is also an increasing function of $t$ provided that

$$\lim_{t \to \infty} \frac{\overline{G}(t)}{\overline{F}(t)} < \infty$$

***Proof:*** Using theorem (4.4.1) the assumption that $\xi(X;t)$ is increasing implies that

$$\xi(X;t) \geq m_F(t).$$

Hence to show that $\xi(Y:t)$ is increasing in $t$ we need to show that

$$\xi(Y;t) \geq m_G(t) = c m_F(t)$$

To this end we show that

$$\frac{1}{c} \xi(Y;t) \geq \xi(X;t)$$

Define the function $\beta(t)$ as follows:

$$\beta(t) = \overline{G}(t)\left( \xi(X;t) - \frac{\xi(Y;t)}{C} \right)$$

we show that $\beta(t) < 0$. Differentiation of $\beta(t)$ with respect to $t$ gives

$$\begin{aligned} \beta'(t) &= -g(t)\xi(X;t) + r_F(t)[\xi(X;t) - m_F(t)]\overline{G}(t) + \frac{1}{c} m_G(t) g(t) \\ &= \overline{G}(t)[-r_G(t)\xi(X;t) + r_F(t)(\xi(X;t) - m_F(t)) + m_F(t) r_G(t)] \\ &= \overline{G}(t)[r_F(t) - r_G(t)(\xi(X;t) - m_F(t))] \end{aligned} \quad (4.4.10)$$

since we assume that $\xi(X;t)$ is increasing from theorem (4.4.1) we obtain $\xi(X;t) \geq m_F(t)$. On the other hand from the equality $m_G(t) = c m_F(t)$ one can easily see that

$$r_G(t) = r_F(t) + \frac{1 - C}{c m_F(t)}$$

since $C > 1$. This equation implies that for all $t$, $t \geq 0$, $r_F(t) > r_G(t)$. Therefore $\beta'(t) > 0$ that is $\beta(t)$ is an increasing function of $t$. Now using the fact $\xi(X;t) \geq 0$ and the assumption $\underset{t\to\infty}{Lt} \frac{\overline{G}(t)}{\overline{F}(t)} < \infty$ we have

$$\underset{t\to\infty}{Lim}\, \beta(t) = \underset{t\to\infty}{Lim} \left( \frac{\overline{G}(t)}{\overline{F}(t)} \int_t^\infty m_F(x) f(x) dx - \frac{1}{c} \int_t^\infty m_G(x) g(x) dx \right) = 0$$

this implies that $\beta(t) < 0$. Hence

$$\frac{1}{c}\xi(y:t) \geq \xi(X:t)$$

this completes the proof.

Asadi and Zohrevand [11] showed that for any non-negative random variable $X$, the CRE of $X$ is the expectation of the mean residual lifetime of $X$ i.e. $\xi(X) = E(e_X(X))$. They also proved that

$$\xi(X;t) = E(e_x(X) | X > t). \quad (4.4.11)$$

Moreover, applying theorem 8 in Rao *et al*. [115] to $X_t$, we have

$$\xi(X;t) \geq c\ \exp(H(X;t))$$

where c $\cong$ 0.2065.

Asadi and Zohrevand [11] proposed the following two classes of lifetime distributions based on the DCRE function

**Proposition 4.4.1:** If $X$ and $Y$ are two non-negative random variables with finite means $E(X)$ and $E(Y)$, respectively and such that $X \leq_{ST} Y$ then

$$\xi(X) \leq \xi(Y) - E(X)\log\frac{E(X)}{E(Y)} \tag{4.4.12}$$

***Proof:*** Using the log-sum inequality we have

$$\int_0^\infty \overline{F}(X)\log\frac{\overline{F}(X)}{\overline{G}(X)}dx \geq \int_0^\infty \overline{F}(X)dx\ \log\frac{\int_0^\infty \overline{F}(X)dx}{\int_0^\infty \overline{G}(X)dx} = E(X)\log\frac{E(X)}{E(Y)}$$

Hence we obtain

$$\begin{aligned}\xi(X) &= -\int_0^\infty \overline{F}(X)\log\overline{F}(X)dx \\ &\leq -\int_0^\infty \overline{F}(X)\log\overline{G}(X)dx - E(X)\log\frac{E(X)}{E(Y)} \end{aligned} \tag{4.4.13}$$

Finally, using that $\overline{F} \leq \overline{G}$, we obtain

$$\begin{aligned}\xi(X) &\leq -\int_0^\infty \overline{F}(X)\log\overline{G}(X)dx - E(X)\log\frac{E(X)}{E(Y)} \\ &\leq -\int_0^\infty \overline{G}(X)\log\overline{G}(X)dx - E(X)\log\frac{E(X)}{E(Y)} \\ &= \xi(Y) - E(X)\log\frac{E(X)}{E(Y)}.\end{aligned}$$

Now using the Weibull distribution we obtain an upper bound for the CRE similar to that obtained by Rao *et al*. [114].

**Proposition 4.4.2:** If $X$ is a non-negative random variable then

$$\xi(X) \leq \frac{E\left(X^{\beta+1}\right)\Gamma^{\beta}\left(1+\frac{1}{\beta}\right)}{(\beta+1)E^{\beta}(X)} \qquad \text{for all } \beta > 0 \tag{4.4.14}$$

***Proof:*** If $Y$ is a random variable with a Weibull distribution and reliability function $\overline{G}(t) = e^{-(\lambda t)^{\beta}}$ from (4.4.13) we obtain

$$
\begin{aligned}
-\xi(X) &\geq \int_0^{\infty} \overline{F}(X)\log \overline{G}(X)dx + E(X)\log\frac{E(X)}{E(Y)} \\
&= E(X)\log\frac{E(X)}{E(Y)} - \int_0^{\infty}(\lambda X)^{\beta}\,\overline{F}(X)dx \\
&= E(X)\log\frac{E(X)}{E(Y)} - \lambda^{\beta}\,\frac{E\left(X^{\beta+1}\right)}{\beta+1}
\end{aligned}
$$

where

$$
\mu = E(Y) = \int_0^{\infty} e^{-(\lambda X)^{\beta}}\,dx = \frac{\Gamma\left(1+\frac{1}{\beta}\right)}{\lambda}
$$

Hence

$$
-\xi(X) \geq E(X)\log\frac{E(X)}{\mu} - \frac{E\left(X^{\beta+1}\right)\Gamma^{\beta}\left(1+\frac{1}{\beta}\right)}{\beta+1}\,\mu^{-\beta}
$$

which is maximized for a fixed $\beta$ at

$$
\mu_{\beta} = \left(\frac{\beta E\left(X^{\beta+1}\right)\Gamma^{\beta}\left(1+\frac{1}{\beta}\right)}{(\beta+1)E(X)}\right)^{1/\beta}
$$

Substituting this value we obtain

$$
\begin{aligned}
-\xi(X) &\geq E(X)\log\frac{E(X)}{\mu_{\beta}} - \frac{E\left(X^{\beta+1}\right)\Gamma^{\beta}\left(1+\frac{1}{\beta}\right)}{\beta+1}\,\mu_{\beta}^{-\beta} \\
&= -\frac{E(X)}{\beta}\log\left(\frac{\beta E\left(X^{\beta+1}\right)\Gamma^{\beta}\left(1+\frac{1}{\beta}\right)}{(\beta+1)E^{\beta+1}(X)}\right) - \frac{E(X)}{\beta},
\end{aligned}
$$

and using that $-\log X \geq 1 - X$ we get

$$-\xi(X)\geq \frac{E(X)}{\beta}\left(1-\frac{\beta E\left(X^{\beta+1}\right)\Gamma^{\beta}\left(1+\frac{1}{\beta}\right)}{(\beta+1)E^{\beta+1}(X)}\right)-\frac{E(X)}{\beta}$$

$$=-\frac{E\left(X^{\beta+1}\right)\Gamma^{\beta}\left(1+\frac{1}{\beta}\right)}{(\beta+1)E^{\beta}(X)}$$

In particular, if we take $\beta=1$ then we obtain expression similar that obtained by Rao *et al*. [115].

**Definition 4.4.2:** A random variable $X$ is said to be increasing (decreasing) DCRE, denoted by IDCRE (DDCRE) if $\xi(X;t)$ is an increasing (decreasing) function of $t$.

**Definition 4.4.3:** A random variable $X$ is said to be increasing (decreasing) failure rate, denoted by IFR (DFR) if $r_X(t)$ is increasing (decreasing) in $t$.

**Definition 4.4.4:** A random variable $X$ is said to be increasing (decreasing) mean residual life, denoted by IMRL (DMRL), if $e_X(t)$ is increasing (decreasing) in $t$.

Asadi and Zohrevand [11] obtained characterizations for the exponential, power and Pareto distributions from the following relationship between the DCRE and the MRL

$$\xi(X;t)=ce_X(t)$$

where $c$ is a non-negative real constant. In the following theorem, we extend this result to the more general case where $c$ is a function of $t$.

**Theorem 4.4.6:** Let $X$ be a non-negative absolutely continuous random variable such that $\xi(X;t)=c(t)e_X(t)$ for $t\geq 0$, then

$$e_X(t)=\left(K-\int_0^t\left(1-c(X)e^{c(X)}dx\right)\right)e^{-c(t)} \tag{4.4.15}$$

With $K=\mu e^{c(0)}$ and $\mu=E(X)$

***Proof:*** If $X$ has survival function $\overline{F}_X(t)$, from $\xi(X;t)=-\int_t^\infty \frac{\overline{F}_X(X)}{\overline{F}_X(t)}\log\frac{\overline{F}_X(X)}{\overline{F}_X(t)}dx$

we have

$$\xi(X;t)\overline{F}_X(t)=\log\overline{F}_X(t)\int_t^\infty \overline{F}_X(X)dx-\int_t^\infty \overline{F}_X(X)\log\overline{F}_X(X)dx$$

Differentiating with respect to $t$, we obtain

$$\xi'(X;t)=r_X(t)(\xi(X;t)-e_X(t)) \tag{4.4.16}$$

which jointly with $\xi(X;t)=c(t)e_x(t)$ and $r_X(t)=\dfrac{e_X(t)+1}{e_X(t)}$ give

$$e'_X(t)+c'(t)e_X(t)=c(t)-1$$

solving this linear differential equation we obtain (4.4.14).

In the next example we show how to use this general result to obtain new characterization results.

**Example 4.4.3:** If $c(t)=at+b$ for $t>0$ and $a>0$, from theorem (4.4.6) we obtain the general model with mean residual life function

$$e_X(t)=\frac{b-2at}{a}-\frac{(b-2)e^{-at}}{a}+Ke^{-at-b}$$

If a=0 from theorem (4.4.6) we obtain the characterization results given by Asadi and Zohrevand [11] for the exponential, power and Pareto distributions. Other characterization results can be obtained from our general result by solving the corresponding differential equation.

The following example shows that DDCRE does not imply DMRL.

**Example 4.4.4:** Let $X$ be a continuous random variable with survival function

$$\overline{F}_X(t)=e^{-t^5-(2t)^{1/2}} \quad \text{for} \quad t\geq 0$$

the mean residual life function $e_X(t)$ of $X$ has a maximum at $t = 0.120786$ (see fig 4.4.1). Hence, the function $e_X(t)$ is increasing to the maximum and it is decreasing later. Therefore, $X$ is not DMRL. However, it can be checked that the DCRE at time $t$ of $X$ is a decreasing function in $t$ i.e. $X$ is DDCRE (see Fig. 4.4.1)

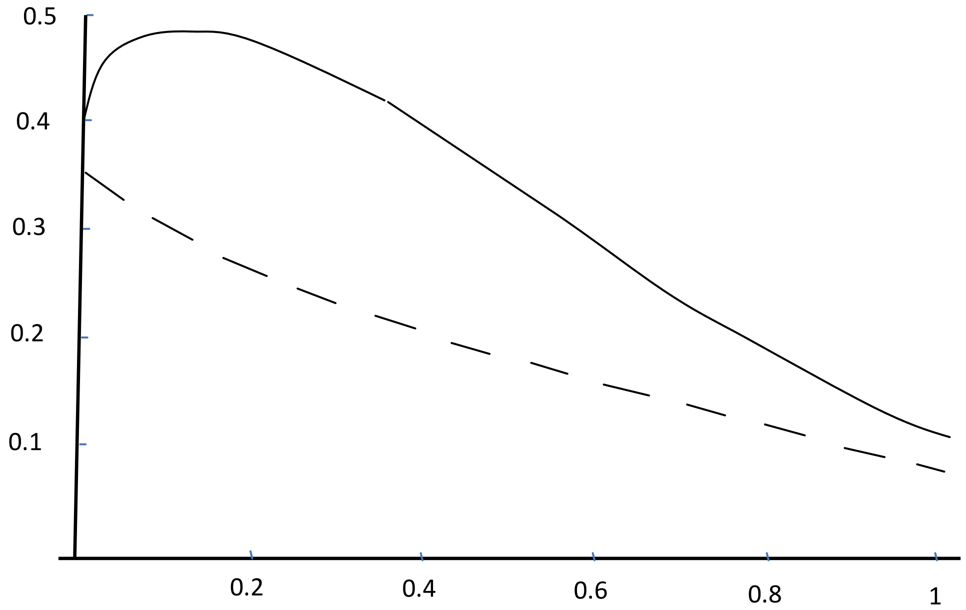



**Fig.4.4.1:** MRI (continuous line) and DCRE (dashed line) functions of the survival function

The following example shows that we do not know if the DCRE function uniquely determines the distribution when $\xi(X;t)$ is an increasing function of $t$.

**Example 4.4.5:** Let $X$ and $Y$ be two random variables with survival functions

$$\bar{F}_X(t) = \begin{cases} 1 - 0.5t & for \quad 0 \le t \le 1 \\ 0.5e^{-(t-1)} & for \quad t > 1 \end{cases}.$$

and

$$\bar{F}_Y(t) = \begin{cases} 1 - 0.5t^2 & for \quad 0 \le t \le 1 \\ 0.5e^{-(t-1)} & for \quad t > 1 \end{cases}.$$

The failure rate functions of $X$ are $Y$ given by

$$r_X(t)=\begin{cases}\dfrac{0.5}{1-0.5t} & for \quad 0\le t\le 1\\ 1 & for \quad t>1\end{cases}$$

and

$$r_Y(t)=\begin{cases}\dfrac{t}{1-0.5t^2} & for \quad 0\le t\le 1\\ 1 & for \quad t>1\end{cases}.$$

Note that $r_X(t)>r_Y(t)$ in $t\in[0,0.585786]$ and $r_X(t)<r_Y(t)$ in $t\in[0.585786,1]$

The mean residual life functions of $X$ and $Y$ are given by

$$e_X(t)=\begin{cases}\dfrac{1.25-t+0.25t^2}{1-0.5t} & for \quad 0\le t\le 1\\ 1 & for \quad t>1\end{cases}$$

and

$$e_Y(t)=\begin{cases}\dfrac{\frac{4}{3}-t+\frac{1}{6}t^3}{1-0.5t^2} & for \quad 0\le t\le 1\\ 1 & for \quad t>1\end{cases}$$

Note that $e_X(t)<e_Y(t)$ in $t\in(0,0.154534)$ and $e_X(t)>e_Y(t)$ in $t\in(0.154534,1)$fig 4.4.2. shows the difference of failure rate functions $(r_X(t)-r_Y(t))$ and the difference of mean residual life functions $(e_X(t)-e_Y(t))$. Thus we have that for all $t\in(0.154534,\ 0.585786)$ $r_X(t)>r_Y(t)$ and $e_X(t)>e_X(t)$.

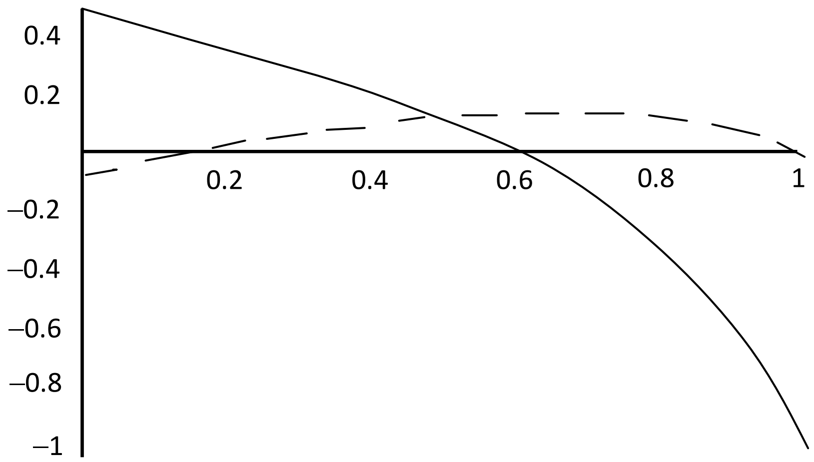



**Fig.4.4.2:** Difference of failure rate function (continuous line) and mean residual life functions (dashed line) of the survival functions

Let $X$ be a random variable with support $S$ and probability density function $f_X$ and let $\omega$ be a non-negative real function in $S$ such that $0 < E(w(X)) < \infty$ then, a random variable $Y$ is said to the weighted distribution associated to $X$ and $\omega$ if its probability density function is given by

$$f_Y(t) = \frac{\omega(t) f_X(t)}{E(\omega(X))}$$

for all $t \in S$. In this case, $Y$ is called the weighted random variable associated to $X$ and $\omega$. The concept of weighted distribution was formulated by Rao [112] to model various situations in which the sampling probabilities are proportional to a 'weighted' function $\omega$.

The usefulness of weighted distributions can be seen in Patil and Rao [108] and Navarro *et al*. [97, 98, 99, 100] and in the references therein. The reliability (survival) function for the weighted random variable $Y$ associated to $X$ and $\omega$ is given by

$$\overline{F}_Y(t) = \frac{E(\omega(X)|X > t)}{E(\omega(X))} \overline{F}_X(t) \tag{4.4.17}$$

From (4.4.17) the dynamic cumulative residual entropy DCRE for the weighted random variable $Y$ can be expressed as

$$\xi(Y:t) = -\int_t^\infty \frac{\overline{F}_Y(X)}{\overline{F}_Y(t)} \log \frac{\overline{F}_Y(X)}{\overline{F}_Y(t)} dx$$

$$= -\int_t^\infty \frac{E(\omega(X)|X > \chi)\overline{F}_X(\chi)}{E(\omega(X)|X > t)\overline{F}_X(t)} \log \frac{E(\omega(X)|X > \chi)\overline{F}_X(X)}{E(\omega(X)|X > t)\overline{F}_X(t)} dx$$

Rao *et al.* [114] proved that if $Y$ is the equilibrium random variable associated to $X$, then

$$H(Y) = \log \mu + \frac{\xi(X)}{\mu}$$

**Proposition 4.4.3:** Let $X$ be a non-negative continuous random variable and let $Y$ be the equilibrium random variable associated to $X$. Then

1. X DDCRE $\Leftrightarrow$ Y DURL
2. X IDCRE $\Leftrightarrow$ Y IURL

***Proof:*** We know that if $X$ be a non-negative continuous random variable and let $Y$ be the equilibrium random variable associated to $X$ then

$$H(Y;t) = \log e_X(t) + \frac{\xi(X;t)}{e_X(t)}$$ we also obtain from this

$$H(Y;t) = \frac{e'_X(t)}{e_X(t)} + \frac{\xi'(X;t)e_X(t) - \xi(X;t)e'_X(t)}{e_X^2(t)}$$

using now (4.4.15) we get

$$H'(Y;t) = \frac{e'_X(t)}{e_X(t)} + \frac{r_X(t)e_X(t)(\xi(X;t) - e_X(t)) - \xi(X;t)e'_X(t)}{e_X^2(t)}$$

then a straightforward calculation gives

$$H'(Y;t) = \frac{r_X(t)e_X(t) - e'_X(t)}{e_X^2(t)}(\xi(X;t) - e_X(t))$$

Hence using " $X$ is DDCRE (IDDCRE) if and only if

$\xi(X;t) \le e_X(t)(\ge)$ for all $t$. The Proof is complete.

## 4.5 Results on the Dynamic Cumulative Past Entropy

Let $X$ be a non-negative random variable with absolutely continuous distribution function $F_X(t) = \Pr(X \le t)$ and probability density function $f_X(t)$. In reliability and survival analysis one usually works with the conditional random variable ${}_tX = [t - X | X < t]$ which is usually known as 'inactivity time'. To illustrate the importance of the random variables of the form ${}_tX$ we give two examples here. First, let us assume that, at time $t$, one has undergone a medical test to check for a certain disease. Let us assume that the test is positive. If we denote by $X$ the age when the patient was infected, then it is known that $X < t$.

Now, the question is, how much time has elapsed since the patient had been infected by this disease. The second example arises naturally in life testing. Suppose that an item has been put under test by an engineer at time t = 0. Usually, in life testing, the items under test are not monitored continuously. Assume that when the engineer checks the items at time $t$, some of them have already failed. Then the same question of that of the first example arises naturally here.

The reliability function ${}_tX$ is given by

$$\bar{F}_{{}_tX}(X) = P({}_tX > x) = P(t - X | X < t) = \frac{F_X(t - X)}{F_X(t)}$$

for $0 \le \chi \le t$. The random variable ${}_tX$ is related with two relevant ageing functions, the reversed failure (or hazard) rate defined by

$$r_X(t) = \frac{f_X(t)}{F_X(t)}$$

and the mean inactivity time (MIT) function defined by

$$K_X(t) = E(t - X | X < t) = \frac{1}{F_X(t)} \int_0^t F_X(\chi) d\chi$$

for $t$ such that $F_X(t) > 0$. The reversed failure rate of $X$ is related with $K_X(t)$ by

$$r_X(t) = \frac{1 - K'_X(t)}{K_X(t)} \tag{4.5.1}$$

these functions are used to define the following classes.

**Definition 4.5.1:** A random variable $X$ is said to be increasing (Decreasing) reversed failure rate, denoted by IRFR (DRFR) if $r_X(t)$ is increasing (decreasing) in $t$.

**Definition 4.5.2:** A random variable $X$ is said to be increasing (Decreasing) mean inactivity time, denoted by IMIT (DMIT), if $K_X(t)$ is increasing (decreasing) in it.

The CRE of ${}_tX$ denoted by $\bar{\xi}(x;t)$ can be called as the dynamic cumulative past entropy (DCPE) of $X$ and it is given by

$$\bar{\xi}(X;t)=-\int_0^t \frac{F_X(\chi)}{F_X(t)}\log\frac{F_X(\chi)}{F_X(t)}dx \tag{4.5.2}$$

From (4.5.2), the DCPE function can be rewritten as

$$\bar{\xi}(X;t)=K_X(t)\log F_X(t)-\frac{1}{F_X(t)}\int_0^t F_X(\chi)\log F_X(\chi)d\chi$$

where $K_X(t)$ is the MIT of $X$.

**Example 4.5.1:** Let $X$ be a continuous random variable with distribution function

$$F_X(t)=\begin{cases}\frac{t^2}{16} & for \quad 0\le t<1\\ \frac{t^4-2t+2}{16} & for \;\; 1\le t<a\\ 1 & for \;\; t\ge a\end{cases}$$

where $a\cong 2.06338$ is the unique positive root of the equation $a^4-2a-14=0$. This distribution was given by Nanda *et al.* [94] the mean inactivity time of $X$, $K_X(t)$ is

$$K_X(t)=\begin{cases}\frac{t}{3} & for \quad 0\le t<1\\ \frac{1}{t^4-2t+2}\left(\frac{t^5}{5}-t^2+2t-\frac{13}{15}\right) & for \;\; 1\le t<a\\ 0.359524 & for \;\; t\ge a\end{cases}$$

It is easy to see that $K_X(t)$ is not monotone (see fig 4.5.1) However, it can be checked that the DCPE of $X$ is an increasing function, i.e. $X$ is IDCPE (See fig 4.5.1).

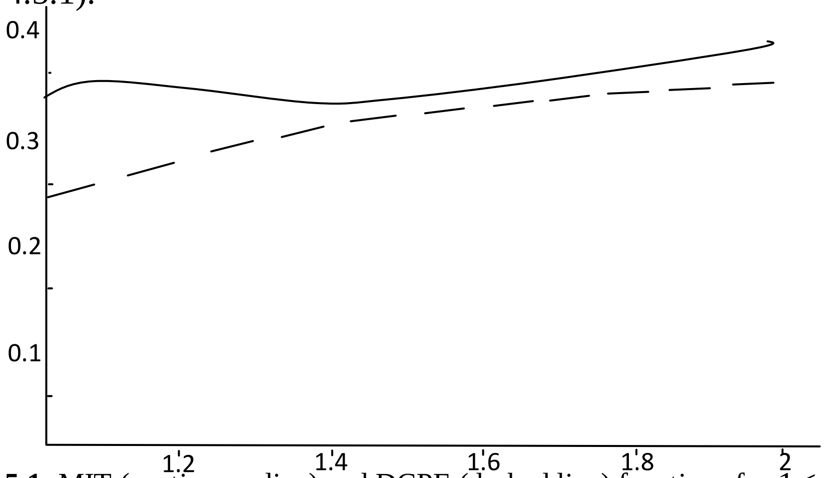



**Fig.4.5.1:** MIT (continuous line) and DCPE (dashed line) functions for $1\le t\le\alpha$ of the distribution as in example 4.5.1

## 4.6 Testing Goodness-of-Fit for Exponential Distribution Based on Cumulative Residual Entropy

The notion of entropy is of fundamental importance in different areas such as physics, probability and statistics, communication theory and economics. In information theory, entropy is a measure of the uncertainty associated with the random variable. This concept was introduced by Shannon [122] Rao *et al*. [114] introduced a new measure of information that extends the Shannon entropy to continuous random variables, and called it cumulative residual entropy (CRE). Its definition is valid for both continuous and discrete cases. It can easily be computed from sample data and its estimation asymptotically converges to the true value. CRE has applications in reliability engineering and computer vision. This measure is based on the cumulative distribution function (cdf) $F$ and is defined as follows:

$$CRE(X) = -\int_{R_+^N} P(|X| > \lambda) \log P(|X| > \lambda) d\lambda ,$$

where $X = (X_1, \ldots, X_N)$ and $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)$ and $|X| > \lambda$ means that, for every $i$, $|X_i| > \lambda_i$, and $R_+^N = \{(\lambda_1, \ldots, \lambda_N);\ \lambda_i \geq 0,\ 1 \leq i \leq N\}$. In reliability theory, CRE is based on survival function $\bar{F}(x) = 1 - F(x)$ and is defined as

$$CRE(X) = -\int_0^\infty \bar{F}(x) \log \bar{F}(x) dx .$$

Testing for exponentiality still attracts considerable attention and is the topic of a good amount of recent research. Many authors provide test statistics for detecting departures from the hypothesis of exponentiality against specific or general alternatives. Alwasel [4] and Ahmad and Alwasel [2] used the lack of memory property of the exponential distribution. Grzegorzewski and Wieczorkowski [57] and Ebrahimi and Habibullah [44] make use of the maximum entropy principle. Also, since early work of Sukhatme [127] and later work by Epstein and Sobel [50, 51, 52] and Epstein [48, 49] considerable attention has been given to testing the hypothesis of exponentiality. Park and Park [107] established the entropy-based goodness of fit test statistics based on the non-parametric distribution functions of the sample entropy and modified sample entropy and compare their performances for the exponential and normal distributions.

### 4.6.1: Test Statistics and its Properties

Suppose $X$ and $Y$ be two non-negative and absolutely continuous random variables with cdf $F$ and $G$ and pdf $f$ and $g$, respectively. As an information distance between two distribution function $F$ and $G$, Kullback and Leibler [81] proposed the following discrimination measure, also known as relative entropy of $X$ and $Y$

$$I_{X,Y} = \int_0^{\infty} f(x)\log\frac{f(x)}{g(x)}dx.$$

To construct a goodness of fit test for exponentiality, we first define a new measure of distance between two distributions that is similar to Kullback-Leibler divergence $(KL)$, but using the distribution function rather than density function and called it cumulative Kullback-Leibler $(CKL)$ divergence.

**Definition 4.6.1**: If $X$ and $Y$ be two non-negative and absolutely continuous random variables with, respectively cdfs $F$ and $G$, then $CKL$ between these distributions is defined as

$$CKL(F:G) = \int_0^{\infty} \overline{F}(x)\log\frac{\overline{F}(x)}{\overline{G}(x)}dx - [E(X) - E(Y)]$$

where $\overline{F}(x) = 1 - F(x)$ and $\overline{G}(x) = 1 - G(x)$ are respectively, cumulative residual distributions.

**Lemma 4.6.1:** $CKL(F:G) \geq 0$, and equality holds if and only if and only if $F = G, a.e$

***Proof:*** By the log-sum inequality, we have

$$\int_0^\infty \overline{F}(x)\log\frac{\overline{F}(x)}{\overline{G}(x)}dx \ge \int_0^\infty \overline{F}(x)dx\log\frac{\int_0^\infty \overline{F}(x)dx}{\int_0^\infty \overline{G}(x)dx} = E(X)\log\frac{E(X)}{E(Y)}$$

The proof is complete if we use the inequality $x\log\frac{x}{y} \ge x-y,\ \forall\ x>0$ and $\forall\ y>0$ and note that in the log-sum inequality, equality holds if and only if $\overline{F}(x)=\overline{G}(x)$.

Let $X_1, X_2, \ldots, X_n$ be non-negative, independent and indentically distributed (iid) random variables from an absolutely continuous cdf $F$ with order statistics $X_{(1)} \le \ldots \le X_{(n)}$, and with finite $\lambda = \frac{E(X_1^2)}{2E(X_1)}$. Let $F_\theta(x,\lambda) = 1-e^{-x/\lambda,},\ \lambda>0,\ x>0$, denote an exponential cdf, where $\lambda$ is the unknown mean parameter. The aim of this section is testing the hypothesis

$$H_0 : F(x) = F_\theta(x,\lambda),\ vs\ H_a : F(x) \ne F_\theta(x,\lambda).$$

Under the null hypothesis $CKL(F:F_0)=0$ and large value of $CKL(F:F_\theta)$ leads us to reject the null hypothesis $H_0$ in favour of the alternative hypothesis $H_a$. Since evaluation of the integral in $CKL(F:F_\theta)$ requires complete knowledge of $F$ and $F_\theta$ then $CKL(F:F_\theta)$ is not operational. We operationalize $CKL(F:F_\theta)$ by developing a discrimination information statistics. Toward this end, $CKL(F:F_\theta)$ is written as

$$\begin{aligned} CKL(F:F_\theta) &= -CRE(F) - \int_0^\infty \overline{F}(x)\log\overline{F}_\theta(x;\lambda)dx - E(X) + \lambda \\ &= -CRE(F) + \frac{1}{\lambda}\int_0^\infty x\overline{F}(x)dx - E(X) + \lambda \\ &= -CRE(F) + \frac{1}{2\lambda}E(X^2) - E(X) + \lambda \\ &= -CRE(F) + \lambda \qquad (4.6.1) \end{aligned}$$

The last equality is obtained by noting that $\lambda = \dfrac{E\left(X_1^2\right)}{2E\left(X_1\right)}$. An estimator of $CRE(F)$ is the $CRE$ of the empirical distribution $F_n(x) = \sum_{i=0}^{n-1} \frac{i}{n} I_{\left[x_{(i)}, x_{(i+1)}\right]}$. Thus

$$\widehat{CRE}(F) = -\int_0^{\infty} \overline{F}_n(x) \log\left(\overline{F}_n(x)\right) dx$$

$$= -\sum_{i=1}^{n-1} \frac{n-i}{n}\left(\log \frac{n-i}{n}\right)\left(X_{(i+1)} - X_i\right).$$

where $\overline{F}_n(x) = 1 - F_n(x)$. By replacing $CRE(F)$ by $CRE(F)$ and $\lambda$ by $\hat{\lambda} = \dfrac{\sum_{i=1}^{n} X_i^2}{2\sum_{i=1}^{n} X_i}$ in (4.6.1), an estimator of $CKL(F : F_\theta)$ is obtained as follows:

$$\widehat{CKL}(F : F_\theta) = \sum_{i=1}^{n-1} \frac{n-i}{n}\left(\log \frac{n-i}{n}\right)\left(X_{(i+1)} - X_{(i)}\right) + \frac{\sum_{i=1}^{n} X_i^2}{2\sum_{i=1}^{n} X_i}$$

Thus the test statistics is defined as

$$T_n = \frac{\sum_{i=1}^{n-1} \frac{n-i}{n}\left(\log \frac{n-i}{n}\right)\left(X_{(i=1)} - X_{(i)}\right) + \dfrac{\sum_{i=1}^{n} X_i^2}{2\sum_{i=1}^{n} X_i}}{\dfrac{\sum_{i=1}^{n} X_i^2}{2\sum_{i=1}^{n} X_i}}$$

We reject $H_0$ at the significance level $\alpha$ and favour $H_a$ if $T_n \geq T_{n,1-\alpha}$, where $T_{n,1-\alpha}$ is $100(1-\alpha)$- percentile of $T_n$ under $H_0$.

The type I error control using the 0.95 percentiles of the $T_n$ statistics was evaluated by simulating random samples from a spectrum of exponential populations. A selection of the result is presented in table 4.6.1. It can be seen that the empirical percentiles given in table 4.6.1 provide an excellent type I error control.

**Table 4.6.1**

**Type I error control of $T_n$: $\alpha = 0.05$**

**(Simulation estimates based on 100000 replications)**

| $\text{Exp}(\lambda)$ | ***N*** | | |
|---|---|---|---|
| | 5 | 15 | 25 |
| λ=2 | 0.04954 | 0.05015 | 0.05020 |
| λ=3 | 0.05026 | 0.04877 | 0.04887 |
| λ=4 | 0.04962 | 0.05059 | 0.04942 |
| λ=5 | 0.05041 | 0.04995 | 0.04925 |

# Chapter –V

# Measure of information and its application

## 5.1: Introduction

Measures of information appear everywhere in probability and statistics. They also play a fundamental role in communication theory. They have a long history since the research work of Fisher, Shannon and Kullback. There are many measures each claiming to capture the concept of information or simply being measures of (directed) divergence or distance between two probability distributions. Also there exist many generalizations of these measures. One may mention here the research work of Lindley and Jaynes who introduced entropy based Bayesian information and the maximum entropy principle for determining probability models respectively.

Broadly speaking there are three classes of measures of information and divergence. Fisher-type, divergence–type and entropy (discrete and differential) type measures. Some of them have been developed axiomatically (Shannon's entropy and its generalizations) but most of them have been established operationally in the sense that they have been introduced on the basis of their properties.

There have been several phases in the history of information theory. Initially we have (i) the development of generalizations of measures of information and divergence ($f$-divergence, ($h$ – $f$)-divergence, hypo-entropy, etc); (ii) the synthesis (collection) of properties they ought to satisfy, and (iii) attempts to unify them. All this work refers to populations and distributions. Later on we have the emergence of information or divergence statistics based on data or samples and their use in statistical inference primarily in minimum "distance" estimation and for the development of asymptotic tests of goodness of fit or model selection criteria. Lately we have a resurgence of interest on measures of information and divergence which are used in many places, in several contexts and in new sampling situations.

The measures of information and divergence enjoy several properties such a non-negativity, maximal information, sufficiency etc. and statisticians do not agree on all of them. There is a body of knowledge known as statistical information theory which has made many advances but not-achieved a wide acceptance and application. This approach is more operational rather than axiomatic as it is the case with Shannon's entropy.

## 5.2: Classes of Measures

As it was mentioned earlier there are three classes of measures of information and divergence, Fisher-type, divergence-type and entropy-type measures. In what follows assume that $f(x,\theta)$ is a Probability Density Function (pdf) corresponding to a random variable $X$ and depending on a parameter $\theta$. At other places $X$ will follow a distribution with pdf $f_1$ or $f_2$.

### *5.2.1: Fisher-Type Measures*

The Fisher's measure of information introduced in 1925 is given by

$$I_X^F(\theta)=\begin{cases} E\left[\dfrac{\partial}{\partial\theta}\log f(X,\theta)\right]^2 = -E\left[\dfrac{\partial^2}{\partial\theta^2}\log f(X,\theta)\right], \theta \text{ univariate} \\ \left\| E\left[\dfrac{\partial}{\partial\theta_i}\log f(X,\theta)\dfrac{\partial}{\partial\theta_j}\log f(X,\theta)\right]\right\|, \qquad \theta \;\; k-\text{variate} \end{cases}$$

where $\|\cdot\|$ denotes the norm of a matrix. The above is the classical or expected information while the observed Fisher information where $\hat{\theta}$ an estimate of $\theta$ is given by:

$$\hat{I}_X^F(\theta)=\begin{cases} -\dfrac{\partial^2 \log f\left(X,\hat{\theta}\right)}{\partial\theta^2}, & 1 \quad \text{observation} \\ -\dfrac{\partial^2 \log f\left(\hat{\theta}/x_1,\ldots,x_n\right)}{\partial\theta^2}, & n \quad \text{observations} \end{cases}$$

Finally the Fisher information measure is given by:

$$I_X^F = E\left[\frac{\partial \log f(X)}{\partial x}\right] \quad \text{or equivalently by}$$

$$J_X^F(\theta) = -E\left[\frac{\partial^2 \log f(X)}{\partial x^2}\right] = E\left[\frac{\partial \log f(X)}{\partial x}\right]^2 - [f'(b) - f'(a)]$$

where $a$ and $b$ are the end points of the interval of support of $X$.

Vajda [134] extended the above definition by raising the score function to a power $a, a \geq 1$ for the purpose of generalizing inference with loss function other than the squared one which leads to the variance and mean squared error criteria. The corresponding measure for a univariate parameter $\theta$ is given by:

$$I_X^V(\theta) = E\left|\frac{\partial}{\partial \theta} \log f(X,\theta)\right|^a, \qquad a \geq 1.$$

In case of a vector parameter $\theta$, Ferentinos and Papaioannou [54] proposed as a measure of information $I_X^{FP}(\theta)$ any eigenvalue or special functions of the eigenvalues of Fisher's information matrix, such as the trace or its determinant.

Finally Tukey [133] and Chandrasekar and Balakrishnan [26] discussed the following measure of information

$$I_X^{T\beta}(\theta) = \begin{cases} \dfrac{(\partial\mu/\partial\theta)^2}{\sigma^2} & X \quad univariate \quad f(x,\theta), \quad \theta \quad scalar \\ , & \\ \left(\dfrac{\partial\mu}{\partial\theta}\right)' \sum^{-1} \left(\dfrac{\partial\mu}{\partial\theta}\right) & X \quad Vector \end{cases}$$

where $\mu$ and $\sigma^2$ (matrix $\sum$ for the vector case) are the mean and variance of the random variable $X$.

### *5.5.2: Measures of Divergence*

A measure of divergence is used as a way to evaluate the distance (divergence) between any two populations or functions. Let $f_1$ and $f_2$ be two probability density functions which may depend or not on an unknown parameter of fixed finite dimension. The most well-known measure of (directed) divergence is the Kullback-Leibler divergence which is given by

$$I_X^{KL}(f_1, f_2) = \int f_1 \log \frac{f_1}{f_2} d\mu$$

for a measure $\mu$. If $f_1$ is the density of $X = (U, V)$ and $f_2$ is the product of the marginal densities of $U$ and $V$, $I_X^{KL}$ is the well-known mutual or relative information in coding theory.

The additive and non-additive directed divergences of order $\alpha$ were introduced in the 60's and the 70's (Renyi [117], CSiczar [32] and Rathie and Kannappan [116]). The so called order α information measure of Renyi [117] is given by

$$I_X^R(f_1, f_2) = \frac{1}{\alpha - 1} \log \int f_1^\alpha f_2^{1-\alpha} d\mu, \ \alpha > 0, \ \alpha \neq 1$$

It should be noted that for $\alpha$ tending to 1 the above measure becomes the Kullback-Leibler divergence. Another measure of divergence is the measure of Kagan [71] which is given by

$$I_X^{K\alpha}(f_1, f_2) = \int (1 - f_1 / f_2)^2 f_2 d\mu$$

Csiszar's measure of information (Csiszar [32]) is a general divergence-type measure, known also as $\phi -$divergence based on a convex function $\phi$. Csiszar's measure is defined by

$$I_X^C(f_1, f_2) = \int \phi(f_1 / f_2) f_2 d\mu$$

where $\phi$ is a convex function in $[\theta, \infty)$ such that $\theta\phi(\theta/\theta) = \theta, \quad \phi(\mu) \to \theta$ and $\theta\phi(\mu/\theta) = \mu\phi_\infty$ with $\phi_\infty = \lim_{\mu \to \infty} [\phi(\mu)/\mu]$. Observe that Csiszar's measure reduces to

Kullback-Leibler divergence if $\phi(\mu)=\mu\log\mu$. If $\phi(\mu)=(1-\mu)^2$ or $\phi(\mu)=\mathrm{sgn}(\alpha-1)\mu^{\alpha}$, $\alpha>\theta$, $\alpha\neq 1$. Csiszar's measure yields the Kagan (Pearson's $X^2$) and Renyi's divergence respectively.

Another generalization of measures of divergence is the family of power divergences introduced by Cressie and Read [31] which is given by

$$I_X^{CR}(f_1, f_2)=\frac{1}{\lambda(\lambda+1)}\int f_1(z)\left[\left(\frac{f_1(z)}{f_2(z)}\right)^{\lambda}-1\right]\mathrm{dz} \quad \lambda\in R$$

where for $\lambda=0,1$ is defined by continuity. Note that the Kullback-Leibler divergence is obtained for $\lambda$ tending to 0.

One of the most recently proposed measures of divergence is the BHHJ power divergence between $f_1$ and $f_2$ (Basu *et al.* [17]) which is denoted by BHHJ, indexed by a positive parameter $\alpha$ and defined as

$$I_X^{BHHJ}(f_1, f_2)=\int\left\{f_2^{1+\alpha}(z)-\left(1+\frac{1}{\alpha}\right)f_1(z)f_2^{\alpha}+\frac{1}{\alpha}f_1^{1+\alpha}(z)\right\}dz \qquad \alpha>0.$$

Note that the above family which is also referred to as a family of power divergence is loosely related to the Cressie and Read power divergence. It should be also noted that the BHHJ family reduces to the Kullback-Leibler divergence for $\alpha$ tending to 0 and to the standard $L_2$ distance between $f_1$ and $f_2$ for $\alpha$ =1. The above measures can be defined also for discrete settings let $P=(p_1, p_2, \ldots, p_m)$ and $Q=(q_1, q_2, \ldots, q_m)$ be two discrete finite probability distributions. Then the discrete version of Csiszar's measure is given by

$I_X^{C}(P,Q)=\sum_{i=1}^{m}q_i\phi(p_i/q_i)$ while the Creasie and Read divergence is given by

$$I_X^{CR}(P,Q)=\frac{1}{\lambda(\lambda+1)}\sum_{i=1}^{m}p_i\left[\left(\frac{p_i}{q_i}\right)^{\lambda}-1\right] \qquad \lambda\in R$$

The discrete version of the BHHJ measure can be defined in a similar fashion.

### 5.2.3 Entropy-Type Measures

Let $P=(p_1, p_2, \ldots, p_m)$ be a discrete finite probability distribution associated with a random variable $X$. Shannon's entropy is defined by

$$H_X^S = -\sum p_i \log p_i$$

It was later generalized by Renyi [117] as entropy of order $\alpha$:

$$H_X^R = \frac{1}{1-\alpha} \log \sum p_i^\alpha, \quad \alpha > 0, \quad \alpha \neq 1.$$

A further generalization along the lines of Csiszar's measure based on a convex function $\phi$ known as $\phi-$entropy was proposed by Burbea and Rao [24] and is given by $H_X^\phi = -\sum_{i=1}^{K} \phi(p_i)$. Finally, it is worth mentioning the entropy measure of Havrda and Charvat [63]:

$$H_X^C = \frac{1-\sum p_i^\alpha}{\alpha - 1}, \; \alpha > 0, \; \alpha \neq 1$$

which for $\alpha$ = 2 it becomes the Gini-Simpson index. Other entropy-type measures include the $\gamma-$entropy given by

$$H_X^\gamma = \frac{1-\left(\sum p_i^{1/\gamma}\right)^\gamma}{1-2^{\gamma-1}} \;, \quad \gamma > 0, \quad \gamma \neq 1$$

and the paired entropy given by

$$H_X^P = -\sum p_i \log p_i - \sum (1-p_i) \log (1-p_i)$$

where pairing is in the sense of $(p_i, 1-p_i)$(cf Burbea and Rao [24]).

### 5.2.4 Properties of Information Measures

The measures of divergence are not formal distance functions. Any bivariate function $I_X(\cdot,\cdot)$ that satisfies the non-negativity property, namely $I_X(\cdot,\cdot) \geq 0$ with equality iff its two arguments are equal can possibly be used as a measure of information or divergence. The three types of measures of information and divergence share similar statistical properties. Several properties have been investigated some of which are of axiomatic character and others of operational. Here we will briefly mention some of these properties. In what follows we shall use $I_X$ for either $I_X(\theta_1,\ldots,\theta_k)$, $K \geq 1$. The information about $(\theta_1,\ldots,\theta_k)$ based on the random variable $X$ or $I_X(f_1, f_2)$, a measure of divergence between $f_1$ and $f_2$. One of the most distinctive properties is the additivity property. The weak additivity property is defined as

$I_{X,Y} = I_X + I_Y$, *If* $X$ is independent of *Y*.

While the strong additivity is defined by

$I_{X,Y} = I_X + I_{Y|X}$ .

Where $I_{Y|X} = E(I_{Y|X=x})$ is the conditional information or divergence of $Y|X$. The sub-additivity and super-additivity properties are defined through the weak additivity when the equal sign is replaced with an inequality.

$I_{X,Y} \leq I_X + I_Y$ (sub-additivity)

and

$I_{X,Y} \geq I_X + I_Y$ (super-additivity).

Observe that super and sub-additivty are contradictory. Sub-additivity is not satisfied for any known measure except Shannon’s entropy [cf. Papaioannou (105)]. Super-additivity coupled with equality iff $X$ and $Y$ independent is satisfied by Fisher’s information number (Fisher’s shift-invariant information) and mutual information (cf. Papaioannou and Ferentinos [106] and Micheas and Zografos [88]). Super-addivity generates measures of dependence or correlation while sub-additivity stems from the conditional inequality (entropy).

Three important inequality properties are the conditional inequality given by

$$I_{X|Y} \leq I_X$$

The nuisance parameter property given by

$$I_X(\theta_1, \theta_2) \leq I_X(\theta_1)$$

where $\theta_1$ the parameter of interest and $\theta_2$ a nuisance parameter and the monotonicity property (maximal information property) given by

$$I_{T(X)} \leq I_X$$

for any statistic $T(X)$. Not that if $T(X)$ is sufficient then the monotonicity property holds as equality which shows the invariance property of the measure under sufficient transformations.

Let positive numbers $\alpha_1$ and $\alpha_2$ such that $\alpha_1 + \alpha_2 = 1$ . Also let $f_1$ and $f_2$ be two probability density functions. The convexity property is defined as

$$I_X(\alpha_1 f_1 + \alpha_2 f_2) \leq \alpha_1 I_X(f_1) + \alpha_2 I_X(f_2)$$

the order preserving property has been introduced by Shiva, Ahmed and Georganas [125] and shows that the relation between the amount of information contained in a random variable $X_1$ and that contained in another random variable $X_2$ remains intact irrespectively of the measure of information used. In particular, if the superscripts 1 and 2 represent two different measures of information then

$$I_{X_1}^1 \leq I_{X_2}^1 \rightarrow I_{X_1}^2 \leq I_{X_2}^2$$

the limiting property is defined by

$$f_n \rightarrow f \ \ iff \ \ I_X(f_n) \rightarrow I(f) \ \ or \ \ I_X(f_n, f) \rightarrow 0 .$$

Where $f_n$ is a sequence of probability density functions, $f$ is the limiting probability density function and $I(f_n)$ and $I(f_n, f)$ are measures of information based on one or two pdfs respectively.

We finally mention the Ali-Silvey property. If $f(x,\theta)$ or (simply $f_\theta$) has a monotone likelihood ratio in $x$ then $\theta_1 < \theta_2 < \theta_3 \rightarrow I_X(f_{\theta_1}, f_{\theta_2}) < I_X(f_{\theta_1}, f_{\theta_3})$.

## 5.3: Information Measures for Generalized Gamma Family

The generalized gamma $(GG)$ distribution offers a flexible family in the varieties of shapes and hazard functions for modeling duration. It was introduced by Stacy [126] difficulties with convergence of algorithms for maximum likelihood estimation Hager and Ban [59] inhibited applications of the $GG$ model. Prentice [109] resolved the convergence problem using a nonlinear transformation of $GG$ model. However, despite its long history and growing use in various applications, the $GG$ family has been remarkably absent in the information theoretic literature. Thus for a maximum entropy $(ME)$ derivation of $GG$ is given in Kapur [74] where it is referred to as generalized Weibull distribution and the entropy of $GG$ has appeared in the context of flexible families of distributions Nadarajah and Zografos [92].

Analysis of duration data is increasingly used in various areas of economics and related fields Keifer [76]. In labor economics, examples include studies of the duration of unemployment (Lancaster [83], Kiefer [75], McDonald and Butler [86], Yamaguchi [138]). Examples in other areas include studies of firms survival (Audretsch and Mahmoud [13]), duration that a property is on the market (Genesove and Mayer [56]), duration of schooling at higher education (Diaz [36]) duration of stages of oilfield exploration (Favero *et al*. [53]) household inter purchase time (Vakratsas and Bass [135]) inter purchase time in financial markets (Allenby *et al*. [3]) and length of the time that new movies stay on screens (Blumenthal [20]).

Distributions that are used in duration analysis in economics include exponential, lognormal, gamma and Weibull. The $GG$ family which encompasses exponential, gamma, and Weibull as subfamilies and lognormal as a limiting distribution has been used in economics by Jaggia. Yamaguchi [138] and Allenby *et al*. [3]. Some authors have argued that the flexibility of $GG$ makes it suitable for duration analysis while others have advocated use of simpler models because of estimation difficulties caused by the complexity of $GG$ parameter structure. Obviously there would be no need to endure the costs associated with the application of a complex $GG$ model if the data do not discriminate between the $GG$ and members of its subfamilies or if the fit of a simpler model to the data is as good as that for the complex $GG$ model. Despite its long history and growing use in various applications, the $GG$ family and its properties has been remarkably presented in different papers. Maximum likelihood estimation of the parameters and quasi maximum likelihood estimators for its subfamily (two-parameter gamma distribution) can be found in [61,65,127,128] some concepts of this family in information theory has introduced by Dadpay *et al*. [33].

### *5.3.1: Information Properties of GG Family*

The probability density function of the $GG$ distribution, $GG(\alpha,\tau,\lambda)$ is

$$f_{GG}(Y|\alpha,\tau,\lambda)=\frac{\tau}{\lambda^{\alpha\tau}\Gamma(\alpha)}\gamma^{\alpha\tau-1}e^{-(\gamma/\lambda)^{\tau}},\ \gamma\geq 0,\ \ \alpha,\tau,\lambda>0 \tag{5.3.1}$$

where $\Gamma(\cdot)$ is the gamma function, $\alpha$ and $\tau$ are shape parameters and $\lambda$ is the scale parameter.

The $GG$ family is flexible in that it includes several well-known models as subfamilies (Johnson *et al*. [70]). The subfamilies of $GG$ thus for considered in the literature are exponential $(\alpha=\tau=1)$ gamma for $(\tau=1)$, and Weibull for $(\alpha=1)$. The lognormal distribution is also obtained as a limiting distribution when $\alpha \to \infty$. By letting $\tau = 2$ we obtain a subfamily of $GG$ which is known as the generalized normal distribution, $GN(2\alpha, \lambda)$. The $GN$ is itself a flexible family and includes Half-normal $(\alpha = 1/2)$, Rayleigh $(\alpha = 1)$. Maxwell-Bottzmann $(\alpha = 3/2)$ and Chi $\alpha = K/2,\ K = 1,\ 2...)$. Moreover, the $GG$ family is more flexible than gamma and Weibull distributions in terms of hazard rate function. It allows for nonmonotonicity in the form of single-peaked hazard functions (but that it would not be able to 'handle' multi-peaked hazard functions). An important property of $GG$ family for information analysis is that the family is closed under power transformation. That is, if $Y \sim GG(\alpha, \tau, \lambda)$, then

$$Z=Y^S \sim GG(\alpha, \tau/S, \lambda^S), \quad S > 0 \tag{5.3.2}$$

in particular,

$$X=Y^t \sim G(\alpha, \lambda^\tau) \tag{5.3.3}$$

where $G(\alpha, \lambda^\tau)$ denotes the gamma density with shape parameter $\alpha$ and scale $\lambda^\tau$.

The power and logarithmic moments of $GG$ distribution are given by

$$\mu_S(\alpha, \tau, \lambda) = E_{GG}(Y^S | \alpha, \tau, \lambda) = \frac{\lambda^S \Gamma(\alpha + S/t)}{\Gamma(\alpha)}, S > 0$$

$$\upsilon_S(\alpha, \tau, \lambda) = E_{GG}(\log Y^S) = \log \lambda^S + \frac{S}{\tau}\psi(\alpha) \tag{5.3.4}$$

where $\psi(\alpha) = \dfrac{\partial \log \Gamma(\alpha)}{\partial \alpha}$ is the digamma function.

For studying the information properties of $GG$ family, we consider the class of distribution functions

$$\Omega_\theta = \{F(\gamma|\theta) : E_f[T_j(\gamma|\theta)] = \theta_j, \quad j = 0,1,2\} \tag{5.3.5}$$

Where $\theta = (\theta_0, \theta_1, \theta_2)$ and $\theta_0 = T_0(\gamma) = 1$ normalizes the density. For a given $\tau$, $T_1(\gamma) = \gamma^{\tau}$,

$$\theta_1 = \mu_{\tau}(\alpha, \tau, \lambda) = E_{GG}\left(\gamma^{\tau} \middle| \alpha, \tau, \lambda\right) = \lambda^{\tau}\alpha \tag{5.3.6}$$

$T_2(\gamma) = \log\gamma$ and

$$\theta_2 = \upsilon(\alpha, \tau, \lambda) = E_{GG}(\log\gamma) = \log\lambda + \frac{1}{\tau}\psi(\alpha) \tag{5.3.7}$$

is the geometric mean.

### *5.3.2: Entropy Properties*

The entropy of a distribution $F$ in $\Omega_{\theta}$ is given by

$$H(F) = -\int_0^{\infty} f(\gamma|\alpha, \tau, \lambda)\log f(\gamma|\alpha, \tau, \lambda)d\gamma .$$

The ME model in (5.3.5) is $F^* = GG^* = GG(\alpha, \tau, \lambda)$ with density (5.3.1).

Kapur [74] gives a proof for a different parameterization of (5.3.1) refers to it as generalized Weibull distribution

The $GG$ entropy is

$$H\left(GG^*\right) = \max_{F \in \Omega_{\theta}}[H(F)] = \log\lambda + \log\Gamma(\alpha) + \alpha - \log\tau + \left(\frac{1}{\tau} - \alpha\right)\psi(\alpha) \tag{5.3.8}$$

Nadarajah and Zografos [92] includes $H(GG)$ among their list of entropies of flexible classes of distributions. For specific values of the paramaters, (5.3.8) gives entropy expressions for gamma, Weibull, exponential and half-normal distributions.

Entropy ordering of distributions within many parametric families is studied in Ebrahimi *et al*. [45], but $GG$ is not included. It is clear that the entropy of $GG$ family is ordered by scale parameter $\lambda$. For the entropy orderings in terms of the shape parameters, we have

$$\frac{\partial H(GG)}{\partial \alpha} \geq 0 \quad \text{for} \quad (\tau\alpha - 1)\psi'(\alpha) \leq \tau$$

$$\frac{\partial H(GG)}{\partial \tau} \geq 0 \quad \text{for} \quad \tau \leq -\psi(\alpha)$$

The first inequality holds for all values of $\tau$, and hence $H(GG)$ is increasing in $\alpha$ since $\psi(\alpha) < 0$ for $\alpha < 1.5$ approximately. $H(GG)$ can be increasing in $\tau$ only when $\alpha < 1.5$.

### *5.3.3: Discrimination Information Properties*

Suppose that we wish to examine if a distribution $F \in \Omega_\theta$ can be approximated by a given model $F_\theta$. The measure of information discrepancy between $F$ and $F_\theta$ is the Kullback-Leibler discrimination information function

$$K(F : F_\theta) = \int f(y) \log \frac{f(y)}{f_\theta(y)} dy \tag{5.3.9}$$

It is well-known that $K(F : F_\theta) \geq \theta$; the equality holds if and only if $f(y) = f_\theta(Y)$ for all $\gamma$ in the support of the distributions $K(F : F_\theta)$ is not symmetric and is a measure of directed divergence between $F$ and $F_\theta$, where $F_\theta$ is referred to as the reference distribution. Symmetric versions of $K(F : F_\theta)$ include Jeffreys divergence, $J(F, F_\theta) = K(K : F_\theta) + K(F_\theta : F)$, (Jeffreys [69]), and $\min\{K(F : F_\theta), K(F_\theta : F)\}$ referred to as the intrinsic information by Bernardo and Rueda [19].

Let $F_\theta = GG_\theta = GG(\alpha_\theta, \tau_\theta, \lambda_\theta)$ be a given $GG$ distribution. It can be shown that the discrimination information function between $F = GG(\alpha, \tau, \lambda)$ and $F_\theta$ is given by

$$K[GG:GG_\theta] = \log\frac{\phi_\tau}{\phi_\lambda^{\alpha\phi_\tau}} - \log\frac{\Gamma(\alpha)}{\Gamma(\alpha_\theta)} - \alpha + \mu(\alpha.\phi_\tau, \phi_\lambda) + (\alpha\phi_\tau - \alpha_\theta)\nu(\alpha, \phi_\tau, \phi_\lambda)$$

(5.3.10)

where $\phi_\tau = \tau/\tau_\theta$, $\phi_\lambda(\lambda/\lambda_\theta)^{\tau_\theta}$, $\mu(\alpha, \phi_\tau, \phi_\lambda)$ is the first moment and $\nu(\alpha, \phi_\tau, \phi_\lambda)$ is the geometric mean of a $GG$ distribution with parameters $(\alpha, \phi_\tau, \phi_\lambda)$.

Although $K(GG:GG_\theta)$ is a complicated function of the parameters (5.3.10) is a general representation that encompasses discrimination information functions between the $GG$ and its subfamilies, between distributions within each subfamily, and between members of different subfamilies. The discrimination information between $GG(\alpha,\tau,\lambda)$ and a gamma $G(\alpha_\theta, \lambda_\theta)$ is given by (5.3.10) with $\phi_\tau = \tau$. The discrimination information between $GG(\alpha,\tau,\lambda)$ and Weibull $\omega(\tau_\theta, \lambda_\theta)$ is given by (5.3.10) with $\alpha_\theta = 1$. The discrimination information between $GG(\alpha,\tau,\lambda)$ and exponential $\xi(\lambda_\theta)$ is given by (5.3.10) with $\phi_\tau = \tau$ and $\alpha_\theta = 1$.The discrimination information between $GG(\alpha,\tau,\lambda)$ and generalized normal $GN(\alpha_\theta, \lambda_\theta)$ is given by (5.3.10) with $\phi_\tau = \tau/2$ and $\alpha_\theta = 2\alpha$.

### *5.3.4: Data Transformation*

Information analysis of the $GG$ family provides some interesting measures in terms of data transformation. Since the $GG$ family is closed under power transformation, by (5.3.2) we can assess the effect of power transformation $Z = Y^S$ by the discrimination information between $Y \sim GG(\alpha,\tau,\lambda)$ and $Y^S \sim GG_S(\alpha_1\tau|S_1\lambda^S)$. In this case, $\phi_\tau = S$ and $\phi_\lambda = \lambda^{\tau/S-\tau}$ in (5.3.10). After some simplifications, we find that the information effect of transformation is given by

$$K_{GG}(Y:Y^S) = K(GG:GG_S) = \log s + \alpha\left[\frac{\mu_{\tau/s}(\alpha,\tau,\lambda)}{\mu_\tau(\alpha,\tau,\lambda)} - \left[\nu_{\tau/s}(\alpha,\tau,\lambda) - \nu_\tau(\alpha,\tau,\lambda)\right] - 1\right]$$

(5.3.11)

Thus, the effect of power transformation is captured through the ratio of the power means and the difference between the geometric means of the transformed and original variables. The information function (5.3.11) is a general representation of some important power transformation information measures for the *GG* family and subfamilies.

As a measure of information disparity between the distributions of the real data (prior to transformation) and the transformed data, $K_{GG}\left(Y:Y^{S}\right)$ may be interpreted as the loss of information due transformation. A large $K_{GG}\left(Y:Y^{S}\right)$ indicates the effect of transformation on the distribution in pronounced.

The information function $K_{GG}\left(Y:Y^{\tau}\right)$ measures the effect of transformation (5.3.3) i.e. discrepancy between $GG(\alpha,\tau,\lambda)$ and gamma $G\left(\alpha,\lambda^{\tau}\right)$

A $GN$ variable $Z$ can be obtained from a $GG$ variable $Y$ by $Z=Y^{\tau/2}$. For $S=\tau/2$, (5.3.11) gives the effect of this transformation. A gamma variable $X$ can be obtained by the square transformation of a $GN$ variable $Z$. The effect of this transformation is measured by (5.3.11) with $\tau=2$ and $S=2$. However, there is no simple relationship between $K_{GG}\left(Y:Y^{\tau/2}\right)$, $K_{GG}\left(Y^{\tau/2}:Y^{\tau}\right)$, and $K_{GG}\left(Y:Y^{\tau}\right)$. The simplest information theoretic model in the $GG$ family is the exponential distribution $\xi$. The exponential model can be obtained from $GG$ sequentially in two ways.

### *5.3.5: Exponential Distribution*

The exponential distribution occurs naturally when describing the lengths of the interval times in a homogeneous Poisson process. Exponential variables can also be used to model situations where certain events occur with a constant probability per unit length, such as the distance between mutations on a DNA strand, or between road kills on a given road. In queuing theory the service times of agents in a system (e.g. how long it takes for a bank teller etc to serve a customer) are often modeled as exponentially distributed variables. Because of the memoryless property of this distribution, it is well-suited to model the constant hazard rate portion of the bathtub curve used in reliability theory.

Failure rate is the frequency with which an engineered system or component fails expressed for example in failures per hour. It is important in reliability engineering. By calculating the failure rate for smaller and smaller intervals of time $< t$, the interval becomes infinitely small. This results in the hazard function, which is the instantaneous failure rate at any point in time.

$$h(t) = \lim_{\Delta t \to 0} \frac{R(t) - R(t + \Delta t)}{\Delta t . R(t)}$$

Continuous failure rate depends on a failure distribution $F(t)$, which is a cumulative distribution function that describes the probability of failure prior to time $t$

$$P(T \le t) = F(t) = 1 - R(t) \qquad t \ge 0 .$$

The hazard function can be defined now as

$$h(t) = \frac{f(t)}{R(t)}$$

Many probability distribution can be used to model the failure distribution. A common model is the exponential failure distribution,

$$h(t) = \frac{f(t)}{R(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda .$$

For an exponential failure distribution the hazard rate is a constant with respect to time. (i.e. the distribution is 'memoryless'). For other distributions, such as Weibull distribution or a log-normal distribution, the hazard function may not be constant with respect to time.

### *5.3.6: Gamma Distribution*

In probability theory and statistics, the gamma distribution is a two parameter family of continuous probability distributions. It has a scale parameter $\lambda$ and a shape parameter $\alpha$. If $\alpha$ is an integer then the distribution represents the sum of $\alpha$ independent exponentially distributed random variables each of which has a mean of $\lambda$ (which is equivalent to a rate parameter of $\lambda^{-1}$). The gamma distribution is frequently a probability model for waiting times; for instance in life testing, the waiting time until death is a random variable that is frequently modeled with a gamma distribution.

### *5.3.7: Weibull Distribution*

The Weibull distribution is a continuous probability distribution. It is named after Waloddi Weibull who described it in detail in 1951, although it was first identified by Frehet in 1927 and first applied by Rosin and Rammler in 1933 to describe the size distribution of particles. The Weibull distribution is often used in the field of life data analysis due to its ability to fit the exponential distribution and the normal distribution and interpolate a range of shapes in between them.

### *5.3.8: Generalized Normal Distribution*

The generalized normal distribution or generalized Gaussion distribution is either of parametric continuous probability distributions on the real line. The *GN* family includes the below well-known models as subfamilies.

### *5.3.9: Half Normal Distribution*

The half-normal distribution is the probability distribution of the absolute value of a random variable that is normally distributed with expected value 0 and variance $\alpha^2$, i.e. if $X$ is normally distributed with mean 0 then $Y = |X|$ is half-normally distributed.

### *5.3.10: Reyleigh Distribution*

In statistic literature, the Reyleigh distribution is a continuous probability distribution. As an example of how it arises, the wind speed will have a Rayleigh distribution if the components of the two-dimensional wind velocity vector are uncorrelated and normally distributed with equal variance. The distribution is named after Lord Rayleigh.

### *5.3.11: Maxwell-Baltzmann Distribution*

The Maxwell-Baltzmann distribution applies to ideal gases close to thermodynamic equilibrium, negligible quantum effects and non-relativistic speeds. It forms the basis of the Kinetic theory of gases, which explains many fundamental gas properties, including pressure and diffusion. The Maxwell-Baltzmann distribution is usually thought of as the distribution for molecular speeds, but it can also refer to the distribution for velocities, momenta and magnitude of the momenta of the molecules, each of which will have a different probability distribution function, all of which are related. The Maxwell-Baltzmann distribution can now most readily be derived from the Baltzmann distribution for energies.

### *5.3.12: Chi-Square Distribution*

In probability theory, the chi-square distribution ($\lambda = 1$ in the chi distribution) with $K$ degrees of freedom is the distribution of a sum of squares of $K$ independent standard normal random variables. It is one of the most widely used probability distributions in inferential statistics, e.g. in hypothesis testing, goodness of fit tests, independence of two criteria of classification of qualitative data, Friedman's analysis of variance by Ranks, estimating variances, estimating the slope of a regression line via, its role in students t-distribution, analysis of variance problems via its role in the F-distribution and so on. The sum of squares of statistically independent unit-variance Gaussian variables which do not have mean zero yields a generalization of the chi-square distribution called the non-central chi-square distribution.

## 5.4: Entropy and its Estimation for Generalized Gamma Distribution

The concept of Shannon's entropy [122] is the central role of information theory, sometimes referred as measure of uncertainty. The entropy of a random variable is defined in terms of its probability distribution and can be shown to be a good measure of randomness or uncertainty. Henceforth we assume that log is to the base 2 and entropy is expressed in bits. For deriving entropy of the generalized gamma distribution, we need the following two definitions.

### *5.4.1: Definition*

The entropy of a discrete alphabet random variable $f$ defined on the probability space $(\Omega, \beta, P)$ is defined by

$$H_P(f) = -\sum_{\alpha \in A} P(f = a) \log(P(f = a)) \tag{5.4.1}$$

It is obvious that $H_P(f) \geq 0$.

***5.4.2: Definition***

The obvious generalization of the definition of entropy for a probability density function $f$ defined on the real line is

$$H(f) = -\int_{-\infty}^{+\infty} f(x) \log f(x) dx = E(-\log f(x)) \tag{5.4.2}$$

provided this integral exists

**Theorem 5.4.1**

Let $X \sim GG(\alpha, \tau, \lambda)$ then

$$H(GG) = \log \lambda + \log \Gamma(\alpha) + \alpha - \log \tau + \left(\frac{1}{\tau} - \alpha\right)\psi(\alpha) \tag{5.4.3}$$

***Proof:*** By definition (5.4.2) we can write

$$\begin{aligned} H(GG) &= -E(\log f(y|\alpha, \tau, \lambda)) = -\log \tau + \alpha\tau \log \Gamma(\alpha) - \alpha\tau\, E(\log Y) \\ &\quad + E(\log Y) + \frac{1}{\lambda^{\tau}} E(Y^{\tau}) \end{aligned} \tag{5.4.4}$$

We also know that if $X \sim GG(\alpha, \tau, \lambda)$ then

$$E(X^s) = \frac{\lambda^S \Gamma\left(\frac{s}{\tau} + \alpha\right)}{\Gamma(\alpha)} \tag{5.4.5}$$

and

$$E(\log(X^S)) = s \log \lambda + \frac{s}{\tau}\psi(\alpha) \tag{5.4.6}$$

where $\psi(\alpha) = \dfrac{\partial \log \Gamma(\alpha)}{\partial \alpha}$ is the digamma function.

Further, from (5.4.5) and (5.4.6) we have

i) $E\left(Y^{\tau}\right)=\dfrac{\lambda^{\tau}\Gamma(1+\alpha)}{F(\alpha)}$

ii) $E(\log(Y))=\log\lambda+\dfrac{1}{\tau}\psi(\alpha)$

Then by substitute these relations in (5.4.4) the theorem is provided

**Corollary 5.4.1:** For all values of $\tau, H(GG)$ is increasing in $\alpha$ .

**Corollary 5.4.2:** For values of $\alpha<1.5$, $H(GG)$ is increasing in $\tau$ .

We can summarizes the entropy of subfamilies of $GG$ distribution as below table:

| Distribution name | $\alpha$ | $\tau$ | $\lambda$ | Entropy |
|---|---|---|---|---|
| Exponential | 1 | 1 | $\lambda$ | $\log\lambda+1$ |
| Gamma | $\alpha$ | 1 | $\lambda$ | $\log\lambda+\log\Gamma(\alpha)+\alpha+(1-\alpha)\psi(\alpha)$ |
| Weibull | 1 | T | $\lambda$ | $\log\lambda+1-\log\tau+\left(\frac{1}{\tau}-1\right)\psi(1)$ |
| Generalized normal | $\alpha$ | 2 | $\lambda$ | $\log\lambda+\log\Gamma(\alpha)+\alpha-1+\left(\frac{1}{2}-\alpha\right)\psi(\alpha)$ |
| Half normal | 0.5 | 2 | $\sqrt{2\sigma^2}$ | $\log\sigma+\log\sqrt{\pi}$ |
| Rayleigh | 1 | 2 | $\sqrt{2\sigma^2}$ | $1/2+\log\sigma-1/2\psi(1)$ |
| Maxwell Boltzmann | 3/2 | 2 | $\lambda$ | $\log\lambda+\log\frac{\sqrt{\pi}}{2}-\frac{1}{2}-\psi\left(\frac{3}{2}\right)$ |
| Chi | k/2 | 2 | $\lambda$ | $\log\lambda+\log\Gamma\left(\frac{k}{2}\right)+\frac{K-2}{2}+\left(\frac{1-K}{2}\right)\psi\left(\frac{K}{2}\right)$ |

### 5.4.2: Entropy Estimation

Consider another form of (5.2.1) as

$$f(y|\alpha,\tau,\lambda)=\frac{\tau}{\lambda^{\alpha\tau}\Gamma(\alpha)}e^{(\alpha\tau-1)\log y-(y/\lambda)^{\tau}} \qquad y\geq 0,\ \ \tau,\alpha,\lambda>0 \qquad (5.4.7)$$

Then, the likelihood function is given by

$$L(y_1,\mathrm{K}\ ,y_n|\alpha,\tau,\lambda)=\left(\frac{\tau}{\lambda^{\alpha\tau}\Gamma(\alpha)}\right)^n e^{(\alpha\tau-1)}\sum_{i=1}^{n}\log y-\sum_{i=1}^{n}\left(\frac{y}{\lambda}\right)^{\tau} \quad y\geq 0,\ \tau,\alpha,\lambda>0 \ (5.4.8)$$

Consequently,

$$l(\alpha,\tau,\lambda)=\log L(y_1,\mathrm{K}\ ,y_n|\alpha,\tau,\lambda)=n(\log\tau-\alpha\tau\log\lambda-\log\Gamma(\alpha)$$

$$+(\alpha\tau-1)\overline{\log y}-\overline{y^{\tau}}/\lambda^{\tau}) \qquad (5.4.9)$$

where, $\overline{\log}y=\dfrac{\sum_{i=1}^{n}\log y}{n}$ and $\overline{y^{\tau}}=\dfrac{\sum_{i=1}^{n}y^{\tau}}{n}$

By taking derivative to parameters we have

$$\begin{cases}\dfrac{d\lambda(\alpha,\tau,\lambda)}{\partial\alpha}=-n\tau\log\lambda-n\psi(\alpha)+n\tau)\overline{\log y}=0\\[2ex] \dfrac{d\lambda(\alpha,\tau,\lambda)}{\partial\lambda}=-\dfrac{n\alpha\tau}{\lambda}+\dfrac{n\tau y^{\bar{\tau}}}{\lambda^{\tau+1}}=0\end{cases} \qquad (5.4.10)$$

By solving this equation and from (5.4.5) and (5.4.6) we have

$$\overline{\log y}=\log\lambda+\frac{1}{\tau}\psi(\alpha)=E(\log Y) \ \text{ and } \ \overline{y^{\tau}}=\lambda^{\tau}\alpha=E(Y^{\tau}) \qquad (5.4.11)$$

then by replacement (5.4.11) and (5.4.4) we get

$$\hat{H}(GG)=-\left(\log\hat{t}-\hat{\alpha}\hat{t}\log\hat{\lambda}-\log\Gamma(\hat{\alpha})+(\hat{\alpha}\hat{t}-1)\overline{\log y}-\frac{y^{\hat{t}}}{\hat{\lambda}^{\hat{\tau}}}\right) \qquad (5.4.12)$$

From (5.4.9) and (5.4.12) we can write

$$\hat{H}(GG)=l(\hat{\alpha},\hat{\tau},\hat{\lambda})/n \qquad (5.4.13)$$

## 5.5: Kullback-Leibler Discrimination

In information theory, the Kullback-Leibler (KL) divergence (also information divergence, discrimination information, or relative entropy) is a non-symmetric measure of the difference between two probability distributions $P$ and $Q$. *KL* divergence is a special case of a broader class of divergence called *f*-divergences. Originally introduced by soloman Kullback and Richard Leibler [81] as the directed divergence between two distributions, it is not the same as a divergence in calculus. Although it is often intuited as a distance metric, the *KL* divergence is not a true metric-for example, the *KL* from $P$ to $Q$ is not necessarily the same as the *KL* from $Q$ to $P$. For probability distributions $P$ and $Q$ of a discrete random variable the *KL* divergence of $Q$ from $P$ is defined to be

$$K(P:Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}.$$

For distributions $P$ and $Q$ of a continuous random variable the summations given way to integrals, so that

$$K(P:Q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx$$

where $p$ and $q$ denote the densities of $P$ and $Q$.

Let $GG_\theta = GG(\alpha_\theta, \tau_\theta, \lambda_\theta)$ be a given $GG$ distribution. Authors showed that the discrimination information function between $GG_\theta$ and $GG$ is given by

$$K(GG:GG_\theta) = \log \frac{\phi_\tau}{\phi_\lambda^{\alpha\phi_\tau}} - \log \frac{\Gamma(\alpha)}{\Gamma(\alpha_\theta)} - \alpha + \mu(\alpha, \phi_\tau, \phi_\lambda) + (\alpha\phi_\tau - \alpha_\theta)\nu(\alpha, \phi_\tau, \phi_\lambda) \qquad (5.4.14)$$

where $\phi_\tau = \dfrac{\tau}{\tau_\theta}, \phi_\lambda = \left(\dfrac{\lambda}{\lambda_\theta}\right)^{\tau_\theta}$, $\mu(\alpha, \phi_\tau, \phi_\lambda)$ is the first moment and $\nu(\alpha, \phi_\tau, \phi_\lambda)$ is the geometric mean of a $GG$ distribution with parameters $(\alpha, \phi_\tau, \phi_\lambda)$. The discrimination information $K(GG:GG_\theta)$ is a complicated function of the parameters, (5.4.14) is a general representation that encompasses discrimination information functions between the $GG$ and its subfamilies, between distributions within each subfamily, and between distributions from different subfamilies. The discrimination information between $GG(\alpha,\tau,\lambda)$ and Gamma $(\alpha_\theta,\lambda_\theta)$ is given by (5.4.14) with $\phi_\tau = \tau$. The discrimination information between $GG(\alpha,\tau,\lambda)$ and Weibull $(\tau_\theta,\lambda_\theta)$ is given by (5.4.14) with $\alpha_\theta = 1$. The discrimination information between $GG(\alpha,\tau,\lambda)$ and exponential $(\lambda_\theta)$ is given by (5.4.14) with $\phi_\tau = \tau$ and $\alpha_\theta = 1$. The discrimination information between $GG(\alpha,\tau,\lambda)$ and $GN(\alpha_\theta,\lambda_\theta)$ is given by (5.4.14) with $\phi_\tau = \frac{\tau}{2}$ and $\alpha_\theta = 2\alpha$.

# REFERENCES


Aczel, J. [1975]: *On Shannon's inequality, optimal coding and characterization of Shannon's and Renyi's entropies,* Institute Novonal De Alta Mathematics, *Symposia Maths* Vol.15, pp.153-179.

Ash, B.R. [1990]: Information theory. Dover, New York.

Atanassov, [1983]: Intuitionistic fuzzy sets, VII ITKR Session, Sofia (Deposed in Central Library of the Bulgarian Academy Of Sciences, (1697/84), (in Bulgarian).

Bhandari and N. R. Pal, [1993]: Some new information measures for fuzzy sets, *Information Sciences*, Vol. 67, No. 3, pp. 209–228.

Bhandari, D. and N. R. Pal [1993]: Some new information measures for fuzzy sets, Information Science, Vol. 67, pp. 204-228.

Bhandari, D., N.R. Pal and D.D. Majumdar, [1992,1993]: Fuzzy Divergence, Probability Measure of Fuzzy Events and Image Thresholding, Pattern Recognition Letters , pp. 857-868.

Campbell, L.L. [1965]: *A coding theorem and Renyi's entropy*, Information and Control, Vol. 8, pp. 423-429.

Chaudary, B.B. and Rosenfeld, A. [1996]: On a metric distance between fuzzy sets, Pattern Recognition Letters, Vol.17, No.1, pp.1157-1160.

Cisaszer, I. [1967b]: On Topological Properties of Divergence, Studia Math. Hungarica, Vol.2, pp.329-339.

Csiszar, I. [1967a]: Information type measures of difference of probabililty distribution and indirect observation, *Studia Math. Hungarica*, Vol.2, pp.229-318.

D. E. F. Kerridge [1961]: *Inaccuracy and Inference*, J. Royal Statist. Society, Vol.23(A), pp.184-194.

D. S. Hooda [2004]: *On Generalized Measures of Fuzzy Entropy*, Mathematica Slovaca, Vol.54, pp. 315 - 325.

D. S. Hooda and R. K. Bajaj [2008]: *On Generalized R-norm Measures of Fuzzy Information*, Journal of Applied Mathematics, Statistics and Informatics, Vol.4, No.2, pp.199-212.

D. Sharma and D.P. Mittal [1975]: *New non-additive measures of entropy for discrete probability distribution*, J. Math. Sci. (Calcutta), Vol.10, pp.28-40.

De Luca and S. Termini [1972]: *A Definition of a Non-probabilistic Entropy in the Setting of fuzzy sets theory*, Information and Control, Vol.20, pp.301-312.

De Luca, A. and Termini, S. [1972]: A definition of non-probabilistic entropy in setting of fuzzy set theory. *Information and Control*, Vol.20, pp.301-312.

Dubois, D. and Prade, H. [1980]: Fuzzy sets and systems, Theory and application, Academic Press, New York.

Dubois, D. and Prade, H. [1987]: Properties of information in Evidence and Possibility Theories, Fuzzy sets and systems, Vol.24, pp.161-182.

E. Shannon [1948]: *The Mathematical theory of Communication*, Bell Syst. Tech. Journal, Vol.27, pp. 423 - 467.

E. Straszecka [2007]: Combining uncertainty and imprecision in models of medical diagnosis, *Information Sciences*, Vol. 176, No. 20, pp. 3026–3059.

E. Szmidt and J. Kacprzyk [2000]: Distances between intuitionistic fuzzy sets, *Fuzzy Sets and Systems*, Vol. 114, No. 3, pp. 505–518.

E. Szmidt and J. Kacprzyk [2001]: Intuitionistic fuzzy sets in intelligent data analysis for medical diagnosis, in *Proceedings of the Computational Science (ICCS '01)*, V. N. Alexandrov, J. J. Dongarra, B. A. Juliano, R. S. Renner, and C. J. K. Tan, Eds., Vol. 2074, pp. 263–271. Springer, Berlin, Germany.

E. Szmidt and J. Kacprzyk [2001]: Intuitionistic fuzzy sets in some medical applications, in *Proceedings of the 7th Fuzzy Days Dortmund*, B. Reusch, Ed., Vol. 2206 of *Computational intelligence: theory and applications*, pp. 148–151. Springer, Berlin, Germany.

E. Szmidt and J. Kacprzyk, [2004]: A similarity measure for intuitionistic fuzzy sets and its application in supporting medical diagnostic reasoning, in *Proceedings of the Lecture Notes in Computer Science (ICAISC '04)*, L. Rutkowski, J. H. Siekmann, R. Tadeusiewicz, and L. A. Zadeh, Eds., Vol. 3070: pp. 388–393, Springer, Berlin, Germany.

Ebanks, B.R. [1983]: On measures of fuzziness and their representations, Journal of Math. Anal and Appl., Vol. 94, pp. 24-37.

F. Steimann and K. P. Adlassnig [2001]: Fuzzy Medical Diagnosis.

F. Yao and J. S. Yao [2001]: Fuzzy decision making for medical diagnosis based on fuzzy number and compositional rule of inference, *Fuzzy Sets and Systems*, Vol. 120, No. 3, pp.351-366.

Feinstin, A. [1958]: *Foundations of information theory*, McGraw-Hill, New York.

Gadaras and L. Mikhailov [2009]: An interpretable fuzzy rule-based classification methodology for medical diagnosis, *Artificial Intelligence in Medicine*, Vol. 47, No. 1, pp. 25–41.

Guiasu, S. [1971]: Weighted Entropy, Reports on Math., Physics, Vol. 2, pp.165-179.

Havrada, J. H. and Charvat, F. [1967]: *Quantification methods of classificatory processes*, the concepts of structural $\alpha$ entropy, Kybernetika, Vol.3, pp. 30-35.

Hooda, D. S. [2004]: On Generalized Measures of Fuzzy Entropy, Mathematica Slovaca, Vol.54, pp. 315-325.

Hooda, D. S. and R. K. Bajaj [2008]: On Generalized R-norm Measures of Fuzzy Information, Journal of Applied Mathematics, Statistics and Informatics, Vol.4, No.2, pp. 199-212.

Hu, Q. and Yu, D. [2004]: Entropies of fuzzy indiscernibility relation and its operations. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,* Vol.12, pp. 575-589.

J. Lin [1991]: Divergence measures based on the Shannon entropy, *IEEE Transactions on Information Theory*, Vol. 37, No. 1, pp.145-151.

J. N. Kapur [1997]: *Measures of Fuzzy Information*, Mathematical Science Trust Society, New Delhi.

J. Palma, J. M. Juarez, M. Campos, and R. Marin [2006]: Fuzzy theory approach for temporal model-based diagnosis, an application to medical domains, *Artificial Intelligence in Medicine*, Vol. 38, No. 2, pp.197-218.

J. Y. Ahn., Y. H. Kim, and S. K. Kim [2003]: A fuzzy differential diagnosis of headache applying linear regression method and fuzzy classification, *IEICE Transactions on Information and Systems*, Vol. E86-D, No. 12, pp.2790-2793.

J. Ye [2011]: Cosine similarity measures for intuitionistic fuzzy sets and their applications, *Mathematical and Computer Modelling*, Vol. 53, No.1-2, pp. 91-97.

Jaynes, E. T. [1957]: Information Theory and statistical Mechanics, Physical Review, Vol.106, pp.620-630, 108, pp.171-193.

K. Atanassov [1999]: *Intuitionistic Fuzzy Sets: Theory and Applications*, Physica, Heidelberg, Germany.

K. T. Atanassov [1989]: More on intuitionistic fuzzy sets, *Fuzzy Sets and Systems*, Vol. 33, No.1, pp.37-45.

K. Vlachos and G. D. Sergiadis [2007]: Intuitionistic fuzzy information—applications to pattern recognition, *Pattern Recognition Letters*, Vol.28, No. 2, pp.197-206.

Kapur, J. N. [1986]: *A generalization of Campbell's noiseless coding theorem*, Jour. Bihar Math, Society, Vol.10, pp.1-10.

Kapur, J. N. [1990]: *Maximum entropy models in Science and Engineering*, Wiley East, New Delhi.

Kapur, J. N. [1997]: *Inequalities Theory, Applications and Measurements*, Mathematical Science Trust Society, New Delhi.

Kapur, J. N. [1997]: *Measures of Fuzzy Information*, Mathematical Science Trust Society, New Delhi, India.

Kapur, J. N. [1997]: *Measures of Fuzzy Information*. Mathematical Sciences Trust Society, New Delhi.

Kapur, J. N. [1998]: *Entropy and Coding*, Mathematical Science Trust Society, New Delhi.

Kaufmann, A. [1957]: *Introduction to Theory of Fuzzy Subsets*, New York, Academic.

Kerridge, D. E. F. [1961]: Inaccuracy and Inference, J. Royal Statist. Society, Vol.23(A), pp.184-194.

Khan, A.B., Autar, R. and Ahmad, H. [1981]: *Noiseless Coding Theorems for Generalized Non-additive Entropy*, Tam Kong J. Math, Vol.1211, pp.15-20.

Kopka and P. W. Daly [1999]: *A Guide to LATEX*, 3rd ed. Harlow, England Addison-Wesley.

Kulback-Leibler, R.A. [1951]: On Fuzzy symmetric Divergence AFSS, pp. 904-908

Kullback, S. [1959]: *Information Theory and Statistics*, Willey and Sons, New Delhi, India.

Kullback, S. and Leibler, R. A. [1951]: On Information and sufficiency, *Ann Math. Stats.,* Vol.22, pp. 77-89 .

Kullback, S. and R.A. Leibler [1951]: On Information and Sufficiency, *Annals of Mathematical Statistics*, Vol.22: pp.79-86.

Kumar, A. [2011]: Some Generalized Measures of Fuzzy Entropy, Vol. 1, No.2.

L. A. Zadeh, *Fuzzy Sets* [1965]: Information and Control, Vol.8, pp. 338 - 353.

L. Todorova, K. Atanassov, S. Hadjitodorov, and P. Vassilev [2007]: On an Intuitionistic Fuzzy Approach for Decision Making in Medicine, part 1, *International Journal Bioautomation*, Vol.6, pp.92-101.

L. Todorova, K. Atanassov, S. Hadjitodorov, and P. Vassilev [2007]: On an intuitionistic fuzzy approach for decision making in medicine, part 2, *International Journal Bioautomation*, Vol.7, pp.64-69.

Lowen, R. [1996]: *Fuzzy Set Theorem – Basic Concepts*, Techniques and Bibliography, Kluwer Academic Publication.

M. Own [2009]: Switching between type-2 fuzzy sets and intuitionistic fuzzy sets: an application in medical diagnosis, *Applied Intelligence*, Vol. 31, No. 3, pp.283-291.

Mathai, A.M. and Rathie, P.N. [1975]: *Basic Concept in Information Theory and Statistics*. Wiley Eastern Limited, New Delhi.

Nguyen, H.T. and Walker, E.A. [1997]: *A First Course in Fuzzy Logic*, C.R.C Press, Inc.

P. Burillo and H. Bustince [1996]: Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets, *Fuzzy Sets and Systems*, Vol. 78, No. 3, pp.305-316.

P. R. Innocent and R. I. John [2004]: Computer aided fuzzy medical diagnosis, *Information Sciences*, Vol. 162, No.2, pp.81-104.

Pal and Bezdek [1994]: Measuring Fuzzy Uncertainty, IEEE Trans. of fuzzy systems, Vol. 2, No. 2, pp.107-108.

Parkash, O. [1998]: A new parametric measure of fuzzy entropy. *Information Processing and Management of Uncertainty* Vol.2, pp.1732-1737.

Parkash, O. and Sharma, P.K. [2004]: Measures of fuzzy entropy and their relations. *Inernationa. Journal of Management & Systems*, Vol.20, pp.65-72.

Parkash, O., Sharma, P. K. and Kumar, J. [2008]: Characterization of fuzzy measures via concavity and recursivity. *Oriental Journal of Mathematical Sciences*. Vol.1, pp.107-117.

Parkash. O, And Sharma, P. K. [2004]: *Noiseless coding theorems corresponding to fuzzy entropies, Asian Bulletine of Mathematical Society (Accepted)*.

Prakash, O. [2000]: On Fuzzy Symmetric Divergence, AFSS, pp. 904-908.

Q. S. Zhang and S. Y. Jiang [2008]: A note on information entropy measures for vague sets and its applications, *Information Sciences*, Vol. 178, No. 21, pp.4184-4191.

Quteishat and C. P. Lim [2008]: Application of the fuzzy min-max neural networks to medical diagnosis, *Lecture Notes in Computer Science*, Vol. 5179, part 3, pp.548-555.

R. Seising [2006]: From vagueness in medical thought to the foundations of fuzzy reasoning in medical diagnosis, *Artificial Intelligence in Medicine*, Vol. 38, No. 3, pp. 237-256.

Rao, C.R. [1965]: *Linear Statistical Inference and its Applications*, J. Willy and sons, Inc., New York.

Renyi, A. [1961]: On measures of entropy and information. *Proceedings 4th Berkeley Symposium on Mathematical Statistics and Probability*, Vol.1, pp.547-561.

Rudas, I. J. [2001]: Measures of fuzziness: theory and applications. *Advances in fuzzy systems and evolutionary computation*. pp.187-192. World Scientific and Engineering Society Press, Athens.

S. Badaloni and M. Falda [2010]: Temporal-based medical diagnoses using a fuzzy temporal reasoning system, *Journal of Intelligent Manufacturing*, Vol. 21, No. 1, pp.145-153.

S. K. De, R. Biswas, and A. R. Roy [2001]: An application of intuitionistic fuzzy sets in medical diagnosis, *Fuzzy Sets and Systems*, Vol. 117, No. 2, pp. 209-213.

S. Kullback [1959]: *Information Theory and Statistics*, Willey and Sons, New Delhi.

S. Kullback [1968]: *Information Theory and Statistics*, Dover, New York, NY, USA.

S. Kullback and R.A. Leibler [1951]: *On Information and Sufficiency*, Annals of Mathematical Statistics, Vol.22, pp.79-86.

S. Lekkas and L. Mikhailov [2010]: Evolving fuzzy medical diagnosis of Pima Indians diabetes and of dermatological diseases, *Artificial Intelligence in Medicine*, Vol.50, No.2, pp.117-126.

Sanchez [1979]: Medical diagnosis and composite fuzzy relations, in *Advances in Fuzzy Set Theory and Applications*, M. M. Gupta, R. K. Ragade, and R. R. Yager, Eds., North-Holland, Amsterdam, The Netherlands.

Shannon, C. E. [1948]: A mathematical theory of communication. *Bell System Technical Journal*, Vol.27, pp.379-423, 623-659.

Shannon, C.E. [1948]: The Mathematical theory of Communication, *Bell Syst. Tech. Journal*, Vol.27, pp.423-467.

Shannon, C.E. and Weaver, W. [1949]: *The Mathematical Theory of Communication*, The University of Illinoise Press, Urbana.

Shannon, S. Kim, Y. Kim, J. Sorsich, K. Atanassov, and P. Georgiev [1997]: A possibility for implementation of elements of the intuitionistic fuzzy logic in decision making in medicine, in *Proceedings of the First International Conference on Intuitionistic Fuzzy Sets*, J. Kacprzyk and K. Atanassov, Eds., Vol. 3 of *Notes on intuitionistic fuzzy sets*, No. 4, pp. 40-43.

Sharma, B.D. and D.P. Mittal [1975]: New non-additive measures of entropy for discrete probability distribution, J. Math. Sci. (Calcutta, India), Vol.10, pp.28-40.

Sharma, B.D. and Taneja, I. J.: *Entropies of type* $\alpha$*,* $\beta$ *and other generalized measures of information theory*, Mathematatika, Vol.22, pp. 205-215.

Sing and Tomar [2009]: On Fuzzy Measures of Divergences and their Inequalities, Proceeding 10th Annual Confrence of Indian Society of Information Theory and Applications (ISITA), pp.31-43.

Singpurwalla, N. D. and Booker, J.M. [2004]: Membership functions and probability measures of fuzzy sets. *Journal of the American Statistical Association,* Vol.99, pp. 867-889.

Smets, P. [1981]: Medical diagnosis: Fuzzy sets and degrees of belief, *Fuzzy sets and systems,* Vol.5, pp. 259-266.

Teneja, I. J. [2004]: Generalized Symmetric Divergence Measures and Inequalities. RGMIA Research Report Collection, Art 9.

Theil, H. [1967]: Economics and Information Theory, North-Holland Publishing Co., Amesterdom.

V. Khatibi and G. A. Montazer [2009]: Intuitionistic fuzzy set vs. fuzzy set application in medical pattern recognition, *Artificial Intelligence in Medicine*, Vol. 47, No. 1, pp. 43-52.

Van Der Lubbe, J.C.A. [1978]: *On certain coding theorems for the information of order* $\alpha$ *and type* $\beta$*, In. Trans, 8th Prague Conf.* Inf. Theory, D. Reidell, Dordrecht, pp.253-256.

Wang, G.J. [1999]: *On the logic foundation of fuzzy reasoning,* Inf. Sci., Vol.117, pp. 47-48.

X. G. Shang and W. S. Jiang [1997]: A note on fuzzy information measures, *Pattern Recognition Letters,* Vol. 18, No. 5, pp. 425–432.

Yager, R. R. [1979]: On measures of fuzziness and negation, Part-I: membership in the unit interval, *International Journal of General Systems,* Vol.5, pp. 221-229.

Yager, R.R. [1980]: On Measures of Fuzziness and Negation, Part II.Latices, Inform & Control, Vol.44, pp. 236-260.

Z. S. Xu [2006]: On correlation measures of intuitionistic fuzzy sets, *Lecture Notes in Computer Science*, Vol. 4224, pp.16-24.

Zadeh [1965]: Fuzzy sets, *Information and Control*, Vol. 8, No. 3, pp. 338-353.

Zadeh, L. A. [1965]: Fuzzy sets. *Information and Control,* Vol.8, pp.338-353.

Zadeh, L. A. [1966]: *Fuzzy Sets,* Inform, and Control, Vol.8, pp.94-102.

Zadeh, L. A. [1968]: Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications,* Vol.23, pp.421-427.