ENTROPY AND INFORMATION INEQUALITIES

THESIS

SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

IN

STATISTICS

By

RAYEES AHMAD DAR

Under the Supervision of

Dr. M. A. K BAIG

POST GRADUATE DEPARTMENT OF STATISTICS
UNIVERSITY OF KASHMIR
SRINAGAR-190006
(2007)

Post Graduate Department of Statistics
University of Kashmir, Srinagar

No:__________
Dated:_________

Certificate

This is to certify that the work embodied in this thesis entitled "ENTROPY AND INFORMATION INEQUALITIES" is the original work carried out by Mr. Rayees Ahmad Dar under my supervision and is suitable for the award of the degree of Doctor of philosophy in Statistics.

This work has not been previously submitted for the award of any degree, fellowship, associatship or any other similar distinction.

Dr. M. A. K Baig

Supervisor.

Acknowledgement

"Praise be to Allah, creator of the universe, the most gracious, most 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 very difficult for a researcher to complete his research work with out the co-operation of others and I am no exception. Among those who first kindled my interest in the present endeavor is my supervisor Dr. M. A. K Baig, Lecturer, 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 discourses always kept me lively to accomplish this work. I consider to express my feelings and regards for him in words will be extremely formal.

I am deeply grateful to Prof. Aquil Ahmad, Head, Department of Statistics, University of Kashmir, Srinagar for encouragement and providing necessary facilities in the department. I wish to express my grateful thanks to my teachers Dr. Anwar Hassan and Dr. Tariq Rashid Jan for their help.

I am also thankful to Prof. I. J Taneja, Departmento de Matematica, Universidade Federal de santa catarina, Florianopolis, Brazil, S. S Dragomir, Victoria University of technology, Australia, Dr. A.B Khan, SKUAST, Jammu, Prof. W. Pedrycz, Editor in chief, Information sciences (An international journal) for sending relevant and valuable material regarding my topic.

I am thankful to staff members of Allama Iqbal library and browsing section, University of Kashmir, Srinagar for their kind support.

I am duty bound to pay my thanks to my parents and to my brother Bilal Ahmad for their inspiring encouragement and unflagging optimism which have always been a dynamic force behind my achievements in all walks of life.

It would be unfair if I fail to thank my friends who extended their co-operation whenever I needed, specially Mr. Ajaz Ahmad (Ph.D scholar Physics), Mr. Mumtaz Ahmad (Ph. D scholar Archeology), Irfan Ahmad (Ph.D scholar Botony), Ab. Roaf (Ph.D scholar History), Mr. Ibrahim

(Ph.D scholar Mathematics), Mr. Nisar Ahmad (Ph.D scholar Mathematics), Mr. Javid Gani (Ph.D Scholar Statistics).

Cordial thanks are due to my crony Mahnaz Shafi for her inspirations. She indeed was a continuous source of encouragement for my writing.

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

Rayees Ahmad Dar

CONTENTS

PREFERACE

The concept of information theory originated when an attempt was made to create a theoretical model for the transmission of information theory of various kinds. information theory is a branch of mathematical theory of probability and is applied in wide variety of fields: communication theory, thermodynamics, econometrics, operation research and psychology etc.

The development presented here have represented a step towards generalizing various measures of information, their characterization, application in coding theory and inference. The thesis consists of five chapters.

The first chapter gives the basic concepts and preliminary results which are used in the subsequent chapters and necessary for the thesis to be self contained one.

in chapter II, certain generalized entropy functions have been considered and their bounds have been obtained by considering their suitable codeword mean length. The coding theorems obtained in this chapter not only produces new results but also generalizes some well established results in the literature of information theory. Also, codes of variable length that are capable of error correction are studied. A lower bound on the mean length of code words for personal probability codes has been established. This generalizes the result due to Kerridge, which itself is a generalization of celebrated result due to Shannon for noiseless channel. The bounds obtained provides a measure of optimality for variable length error correcting codes.

In the chapter III, various generalized 'useful' inaccuracy measures have been considered and their bounds have been obtained for suitable generalized mean codeword lengths. Several coding theorems obtained in this chapter have been published in "International Journal of Pure and

Applied Mathematics" [8] and "Sarajevo Journal of Mathematics" [12]. The beauty of these results is that it generalizes some well established results and are suitable for the more complex distributions other than exponential.

In chapter IV, several information inequalities have been obtained by considering Csiszar f-divergence and symmetric j- divergence measure. Some particular cases are also obtained by comparing it with a number of other divergence measures arising in information theory. These results have been accepted for publication in "Indian Journal of Mathematics"[13]. Also, some generalized information inequalities have been obtained, which are not only new but also generalizes the results obtained by Dragomir [47]. This work has been published in " Journal of Mathematics and system sciences" [10].

In chapter V, some generalized inequalities have been obtained for logarithmic mapping and convex mappings by using Jensen's inequality. Dragomir [46] have obtained some information inequalities for logarithmic mappings and convex mappings, but the results in this chapter are obtained by considering the functions with independent variable 's' which gives some new information inequalities and also generalizes some results obtained by Dragomir [46]. This work has been published in " International Journal of Pure and Applied Mathematics"[11]. Also, some upper bounds for the relative Arithematic Geometric divergence measure have been obtained by using some classical inequalities like Kantorovic inequality, Diaz- Metcalf inequality.

A comprehensive bibliography is given at the end.

1.1 Introduction.

Information theory is a new branch of probability theory with extensive potential applications to communication system. The term information theory does not possess a unique definition. Broadly speaking, information theory deals with the study of problems concerning any system. This includes information processing, information storage, information retrieval and decision making. In a narrow sense, information theory studies all theoretical problems connected with the transmission of information over communication channels. This includes the study of uncertainty (information) measure and various practical and economical methods of coding information for transmission.

The first studies in this direction were undertaken by Nyquist [91,92] in 1924 and 1928 and by Hartley [55] in 1928, who recognized the logarithmic nature of the measure of information. In 1948, Shannon [107] published a remarkable paper on the properties of information sources and of the communication channels used to transmit the outputs of these sources. Around the same time, Wiener [133] also considerd the communication situations and came up, independently, with results similar to those of Shannon.

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 wide applications in various fields of importance. e.g., The work of Bar-Hillel [16], Balasubrahmanyan and Siromoney [15] in linguistic, Brillouins [27], Jaynes [64] in Physics, Kullback [82], Kerridge [73] in Statistical estimation, Theil [126] in Economics, Quastler [98] in Psychology, Quastler [97] in Biology and Chemistry, Wiener [133] in Cybernetics, Renyi [104], Zaheeruddin [135] in inference, Kapur [71] in Operation Research, Kullback [82] in Mathematical Statistics, Zadeh [134] in Fuzzy set theory, Rao [99] in anthropology, Mei [88] in Genetics, Sen [106] in Finance, Theil [127] in Poltical Science, Pielou [96] in Biology, Gokhale and Kullback [50] in analysis of contingency tables, Chow and Lin [32] and Kazakos and Cotsidas [72] in approximation in probability distributions, Kadota and Shepp [66] and Kailath [67] in signal processing and Beth Bassat [21] and Chen [31] in pattern recognition.

A key feature of Shannon information theory is that the term information can often be given a Mathematical meaning as a numerically measurable quantity, on the basis of a probabilistic model, in such a way that the solutions of many important problems of information storage and transmission can be formulated in terms of this measure of the amount of information. This important measure has a very concrete operational interpretation: it roughly equals the minimum number of binary digits needed, on the average, to encode the message in question. The coding theorems of information theory provide such a overwhelming evidence for the adequacy of the Shannon information measure that to look for essentially different measures of information might appear to make no sense at all. Moreover, it has been shown by several authors, starting with Shannon [107] that the measure of the amount of information is uniquely determined by some rather natural postulates. Still, all the evidence that the Shannon information measure is the only possible one is valid only within the restricted scope of coding problems considerd by Shannon. As pointed out by Renyi [104] in his paper on generalized information measures, in other sorts of problems other quantities may serve just as well, or even better, as measures of information. This should be supported either by their operational significance or by a set of natural postulates characterizing them, or, preferably, by both. Thus the idea of generalized entropies arises in the literature. It found its birth in Renyi [104], who characterizing a scalar parametric entropy as an entropy of order α, which includes Shannon entropy as a limiting case.

1.2 Shannon's entropy

Let X be a discrete random variable taking on a finite number of possible values $x_1, x_2, ..., x_n$ happening with probabilities $P = (p_1, p_2, ..., p_n)\,, p_i \geq 0, i = 1, 2, ..., n, \sum_{i=1}^{n} p_i = 1$.We denote

(1.2.1) $$X = \begin{bmatrix} x_1 \; x_2 \dots\dots \; x_n \\ p_1 \; p_2 \dots \dots \; p_n \end{bmatrix}$$

and call the scheme (1.2.1) as the information scheme. Shannon [107] proposed the following measure of information for the information scheme (1.2.1) and call it entropy.

(1.2.2) $$H(P) = H(p_1, p_2, ..., p_n) = -\sum_{i=1}^{n} p_i \log p_i$$

Generally, the base of logarithm is taken as '2' and it is assumed $0 \log 0 = 0$. When the logarithm is taken as base '2' the unit of information is called a 'bit'. When the natural logarithim is taken, then the resulting unit is called a ‘nit’. If the logarithim is taken with base 10, the unit of information is known as ‘Hartley’.

The information measure (1.2.2) satisfies the following properties.

(1) Non-negativity

$H(p_1, p_2, ..., p_n) \geq 0$

The entropy is always non -negative.

(2) Symmetry

$H(p_1, p_2, ..., p_n) = H\left(p_{k(1)}, p_{k(2)}, ..., p_{k(n)}\right) \quad \forall \quad (p_1, p_2, ..., p_n) \in P$

where $\left(k_{(1)}, k_{(2)}, ..., k_{(n)}\right)$ is an arbitrary permutation on $(1, 2, ..., n)$.

$H(p_1, p_2, ..., p_n)$ is a symmetric function on every p_i , $i = 1, 2, ..., n$

(3) Normality

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

The entropy becomes unity for two equally probable events.

(4) Expansibilty

$$H_n(p_1, p_2, ..., p_n) = H_{n+1}(0, p_1, p_2, ..., p_n)$$
$$= H_{n+1}(p_1, p_2, ..., p_i, 0, p_{i+1}, ..., p_n)$$
$$= ...$$
$$= H_{n+1}(p_1, p_2, ..., p_n, 0)$$

(5) Recursivity

$$H_n(p_1, p_2, ..., p_n) = H_{n-1}(p_1 + p_2, p_3, ..., p_n) + (p_1 + p_2)\, H_2\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right)$$

where $p_i \geq 0, \ \sum_{i=1}^{n} p_i = 1$ with $p_1 + p_2 > 0$.

(6) Decisivity

$$H_2(1,0) = H_2(0,1) = 0$$

If one of the event is sure to occur then the entropy is zero in the scheme.

(7) Maximality

$$H(p_1, p_2, ..., p_n) \leq H\left(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n}\right) = \log n$$

The entropy is maximum when all the events have equal probabilities.

(8) Additivity

$$H_{nl}(PQ) = H_{nl}(p_1q_1, p_1q_2, ..., p_1q_l, p_2q_1, p_2q_2, ..., p_2q_l, ..., p_nq_1, p_nq_2, ..., p_nq_l)$$

$$= H_n(p_1, p_2, ..., p_n) + H_l(q_1, q_2, ..., q_l)$$

for all $(p_1, p_2, ..., p_n) \in P$ and for all $(q_1, q_2, ..., q_l) \in Q$.

If the two experiments are independent then the entropy contained in the experiment is equal to the entropy in the first experiment plus entropy in the second experiment.

(9) Strong Additivity

$$H_{nl}(PQ) = H_{nl}(p_1q_1, p_1q_2, ..., p_1q_l, p_2q_1, p_2q_2, ..., p_2q_l, ..., p_nq_1, p_nq_2, ..., p_nq_l)$$

$$= H_n(p_1, p_2, ..., p_n) + \sum_{i=1}^{n} p_i H_l(q_{i1}, q_{i2}, ..., q_{in})$$

for all $(p_1, p_2, ..., p_n) \in P$ and for all $(q_1, q_2, ..., q_l) \in Q$ and q_{ij} are the conditional probabilities i.e., entropy contained in the two experiments is equal to the entropy in the first plus the conditional entropy in the second experiment given that the first experiment has occured.

The Shannon's entropy (1.2.2) was characterized by Shannon assuming a set of postualtes. There exists several other characterization of the measure (1.2.2) using different set of postulates. Notably amongst them are those of Khinchin [79], Tverberg [130], Chaundy and Mcloed [30], Renyi [104], Lee [84] , Daroczy [37], Rathie [102], Guiasu [52], Kapur [70] etc.

1.3 Generalizations of Shannon's entropy

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

(1) **Renyi's entropy**

Renyi [104] generalized the Shannon's entropy for an incomplete probability distribution as

(1.3.1) $$H_1\left(P\right)=\frac{-\sum\limits_{i=1}^{n}p_i\log p_i}{\sum\limits_{i=1}^{n}P_i}$$

and the entropy of order α as

(1.3.2) $$H_\alpha\left(P\right)=\frac{1}{1-\alpha}\log\frac{\sum\limits_{i=1}^{n}P_i^\alpha}{\sum\limits_{i=1}^{n}P_i}\quad ,p_i\geq 0,\quad \sum\limits_{i=1}^{n}P_i\leq 1,\ \ \alpha>0\,(\neq 1)$$

For $\alpha\rightarrow 1$, (1.3.2) reduces to (1.3.1).

For a complete probability distribution, (1.3.2) reduces to

(1.3.3) $$H_\alpha\left(P\right)=\frac{1}{1-\alpha}\log\sum\limits_{i=1}^{n}p_i^\alpha\quad ,\alpha>0\,(\neq 1)$$

and is called Renyi's [104] entropy of order α , as $\alpha\rightarrow 1$, the measure (1.3.3) reduces to Shannon's entropy (1.2.2).

(2) Kapur's entropy

Kapur [69] generalized the Shannon's entropy for an incomplete probability distribution as

(1.3.4) $$H_1^\beta(P) = -\frac{\sum\limits_{i=1}^{n} P_i^\beta \log p_i}{\sum\limits_{i=1}^{n} p_i^\beta}$$

and

(1.3.5) $$H_\alpha^\beta(P) = \frac{1}{1-\alpha}\log\frac{\sum\limits_{i=1}^{n} p_i^{\alpha+\beta-1}}{\sum\limits_{i=1}^{n} p_i^\beta}$$

where $p_i \geq 0,\ \sum\limits_{i=1}^{n} p_i \leq 1, \alpha > 0\,(\neq 1)\,, \beta > 0$

For a complete probability distribution and for $\alpha \to 1, \beta = 1$, the measure (1.3.5) reduces to Shannon's entropy (1.2.2).

(3) Havrada- Charavt's entropy

Havrada- Charvat [57] introduced non- additive entropy as

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

and called it generalized entropy of type β. When $\beta \to 1$, the measure (1.3.6) becomes Shannon's entropy (1.2.2).

(4) Aczel and Daroczy's entropy

Aczel and Daroczy [3] introduced the entropy of order α and type β

(1.3.7) $$H_\alpha^\beta(P) = \frac{1}{\beta-\alpha}\log\frac{\sum\limits_{i=1}^{n} p_i^\alpha}{\sum\limits_{i=1}^{n} p_i^\beta}, \qquad \alpha \neq \beta, \quad \alpha, \beta > 0$$

For $\beta = 1$, (1.3.7) becomes Renyi's [104] entropy and for $\alpha \to 1, \beta = 1$, (1.3.7) reduces to Shannon's entropy (1.2.2).

(5) Varma's entropy

Varma [132] introduced the entropies as

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

and

(1.3.9) $$H_\alpha^\beta(P) = \frac{1}{\beta(\beta-\alpha)} \log\left(\sum_{i=1}^{n} p_i^{\frac{\alpha}{\beta}}\right) \quad , \; 0 < \alpha < \beta, \;\; \beta \geq 1$$

For $\beta = 1$ and $\alpha \to 1$, (1.3.8) and (1.3.9) becomes Shannon's entropy (1.2.2).

(6) Rathie's entropy

Rathie [103] introduced the generalized entropy

(1.3.10) $$H_\alpha^{\beta_i} = \frac{1}{1-\alpha} \log \frac{\sum_{i=1}^{n} p_i^{\alpha+\beta_i-1}}{\sum_{i=1}^{n} p_i^{\beta_i}}, \qquad \alpha > 0\,(\neq 1)\,, \quad \beta_i > 0$$

For $\beta_i = \beta \;\; \forall \;\; i = 1, 2, ..., n$, (1.3.10) reduces to (1.3.5). Also when $\beta_i = 1 \;\; \forall \; i = 1, 2, ..., n$ and $\alpha \to 1$ then (1.3.10) reduces to Shannon's entropy (1.2.2).

(7) Armito's entropy

Armito [5] introduced the generalized entropy as

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

For $\alpha \to 1$, (1.3.11) reduces to Shannon's entropy (1.2.2).

(8) Sharma and Mittal's entropy

Sharma and Mittal [111] introduced the generalized entropies as

(1.3.12) $$H_\alpha(P) = \frac{1}{(2^{1-\alpha}-1)} \left[\exp\left((\alpha - 1) \sum_{i=1}^{n} p_i \log p_i \right) - 1 \right] \;, \alpha > 0(\neq 1)$$

For $\alpha \to 1$,(1.3.2) reduces to Shannon's entropy (1.2.2).

and

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

For $\alpha \to 1, \beta \to 1$. (1.3.13) reduces to Shannon's entropy (1.2.2).

(9) Sharma and Taneja's entropy

Sharma and Taneja [112] introduced the generalized entropies as

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

For $\alpha = 1$, (1.3.14) reduces to Shannon's entropy (1.2.2).

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

For $\beta = 1$ and $\alpha \to 1$, (1.3.15) reduces to Shannon's entropy (1.2.2).

(1.3.16) $$H_s^{(\alpha,\beta)}(P) = -\frac{2^{\alpha-1}}{\sin\beta}\sum_{i=1}^{n} p_i^\alpha \sin(\beta \log p_i), \quad \alpha > 0, \beta \neq k\pi, k = 0, 1, ...$$

For $\beta \to 0, \alpha = 1$, (1.3.16) reduces to Shannon's entropy (1.2.2).

(10) Picard's entropy

Picard [95] introduced the generalized entropies as

(1.3.17) $$H_{\nu_i}(P) = -\frac{\sum_{i=1}^{n}\nu_i \log p_i}{\sum_{i=1}^{n}\nu_i}, \quad \nu_i > 0, \quad i = 1, 2, ..., n$$

For $\nu_i = p_i \;\; \forall \; i = 1, 2, ..., n$, (1.3.17) reduces to Shannon's entropy (1.2.2).

(1.3.18) $$H_{\nu_i}^\alpha(P) = \frac{1}{1-\alpha}\log\left(\frac{\sum_{i=1}^{n} p_i^{\alpha-1}\nu_i}{\sum_{i=1}^{n}\nu_i}\right)$$

where $\alpha > 0\,(\neq 1), \nu_i > 0, i = 1, 2, ..., n$

For $\nu_i = p_i \;\; \forall \; i = 1, 2, ..., n$ and $\alpha \to 1$, (1.3.18) reduces to Shannon's entropy (1.2.2).

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

where $\alpha > 0\,(\neq 1), \nu_i > 0, i = 1, 2, ..., n$

For $\nu_i = p_i \;\; \forall \; i = 1, 2, ..., n$ and $\alpha \to 1$, (1.3.19) reduces to Shannon's entropy (1.2.2).

(1.3.20) $$H_{\nu_i}^{(\alpha,\beta)}(P) = \frac{1}{(2^{1-\alpha}-1)}\left[\left(\frac{\sum_{i=1}^{n} p_i^{\beta-1}\nu_i}{\sum_{i=1}^{n}\nu_i}\right)^{\frac{\alpha-1}{\beta-1}} - 1\right]$$

where $\alpha \neq 1, \beta \neq 1, \alpha, \beta > 0, \nu_i > 0, i = 1, 2, ..., n$

For $\nu_i = p_i \;\; \forall \; i = 1, 2, ..., n$ and $\alpha \to 1, \beta \to 1$, (1.3.20) reduces to Shannon's entropy (1.2.2).

(11) Boekee and Lubbe's entropy

Boekee and Lubbe [26] introduced the generalized entropy as

$$(1.3.21) \qquad H_R(P) = \frac{R}{R-1}\left[1 - \left(\sum_{i=1}^{n} p_i^R\right)^{\frac{1}{R}}\right], \quad R > 0\,(\neq 1)$$

For $R \to 1$, (1.3.21) reduces to Shannon's entropy (1.2.2).

(12) Kerridge's inaccuracy

Suppose that an experiment asserts that the probabilities of n events are $Q = (q_1, q_2, ..., q_n)$ while their true probabilities are $P = (p_1, p_2, ..., p_n)$, then Kerridge [73] has proposed a measure of inaccuracy as

$$(1.3.22) \qquad H(P;Q) = -\sum_{i=1}^{n} p_i \log q_i$$

when $p_i = q_i \ \ \forall\, i = 1, 2, ..., n$, then (1.3.22) reduces to Shannon's entropy (1.2.2).

(13) Khan and Autar's inaccuracy

Khan and Autar [75] gave a generalized non- additive measure of inaccuracy as

$$(1.3.23) \qquad H(P^\nu, Q : \alpha, \beta) = \frac{1}{\left(D^{1-\beta}-1\right)}\left[\left(\frac{\sum_{i=1}^{n} p_i^\nu q_i^{\alpha-1}}{\sum_{i=1}^{n} p_i^\alpha}\right)^{\frac{\beta-1}{\alpha-1}} - 1\right]$$

where $\alpha \neq 1, \beta \neq 1, \sum_{i=1}^{n} p_i \leq 1, \sum_{i=1}^{n} q_i \leq 1$

D being the size of the code alphabet. For $\beta \to 1$, (1.3.23) tends to inaccuracy of order α and type ν due to Sharma [114], which further for $p_i = q_i \ \forall \ i = 1, 2, ..., n$ gives Kapur's [69] entropy of order α and type ν.

(14) Rathie's inaccuracy

Rathie [103] generalized the non-additive measure of inaccuracy as

$$(1.3.24) \qquad H(P;Q : \alpha, \beta_i) = \frac{1}{1-\alpha} \log\left(\frac{\sum_{i=1}^{n} p_i^{\beta_i} q_i^{\alpha-1}}{\sum_{i=1}^{n} p_i^{\beta_i}}\right)$$

where $\alpha \neq 1, \beta_i \geq 0, \sum_{i=1}^{n} p_i \leq 1, \sum_{i=1}^{n} q_i \leq 1$

For $\beta_i = 1 \ \forall \ i = 1, 2, ..., n, \ \ \alpha \to 1$ and the distribution is complete then (1.3.24) reduces to Kerridge [73] measure of inaccuracy.

(15) Nath's inaccuracy

Nath [90] introduced the generalized inaccuracy measure

(1.3.25) $$I_\alpha (P; Q) = \frac{1}{1-\alpha} \log \sum_{i=1}^{n} p_i q_i^{\alpha-1}, \quad \alpha > 0 \, (\neq 1)$$

For $\alpha \to 1$, (1.3.25) reduces to Kerridge [73] measure of inaccuracy.

(16) Rathie and Kannapan's measure of inaccuracy

Rathie and Kannapan [101] introduced the generalized inaccuracy measure as

(1.3.26) $$I^\alpha (P; Q) = \frac{1}{1-\alpha} \left(\sum_{i=1}^{n} p_i q_i^{\alpha-1} - 1 \right), \quad \alpha > 0 \, (\neq 1)$$

For $\alpha \to 1$, (1.3.26) reduces to Kerridge [73] inaccuracy measure.

(17) Sharma and Gupta'a inaccuracy measure

Sharma and Gupta [110] introduced inaccuracy measure as

(a) Log measure

(1.3.27) $$I^L_{\alpha,\beta} (P; Q) = -2^\beta \sum_{i=1}^{n} p_i^\alpha q_i^\beta \log q_i \ , \quad \alpha > 0, \beta \geq 0$$

For $\alpha \to 1, \beta \to 0$, (1.3.27) reduces to Kerridge [73] inaccuracy measure.

(b) Power measure

(1.3.28) $$I^P_{\alpha,\beta,\gamma} (P; Q) = \frac{1}{2^{-\beta} - 2^{-\gamma}} \sum_{i=1}^{n} p_i^\alpha \left(q_i^\beta - q_i^\gamma \right), \quad \alpha > 0, \beta, \gamma \geq 0, \beta \neq \gamma$$

For $\alpha \to 1, \beta, \gamma \to 0$, (1.3.28) reduces to Kerridge [73] inaccuracy measure.

(18) Sharma's inaccuracy

Sharma [114] introduced the inaccuracy measure of order α and type β

(1.3.29) $$H^\beta_\alpha (P; Q) = \frac{1}{1-\alpha} \log \left(\frac{\sum_{i=1}^{n} p_i^\beta q_i^{\alpha-1}}{\sum_{i=1}^{n} p_i^\beta} \right)$$

where $\alpha > 0 \, (\neq 1) \, , \beta > 0, \sum_{i=1}^{n} p_i \leq 1, \sum_{i=1}^{n} q_i \leq 1$

For $\alpha \to 1, \beta = 1$ and the distributions are complete then (1.3.29) reduces to Kerridge [73] measure of inaccuracy.

(19) Parkash's inaccuracy measure

Parkash [93] introduced the generalized inaccuracy measure

(1.3.30) $$I_n^{\beta}(P;Q) = \frac{1}{\left(2^{-\beta}-1\right)}\left[\sum_{i=1}^{n} p_i q_i^{\beta} - 1\right], \quad \beta \neq 0$$

For $\beta \to 0,$ (1.3.30) reduces to Kerridge [73] inaccuracy measure.

1.4 Divergence measures

The concept of entropy was first introduced in Thermodynamics, where it was used to provide a statement of the second law of thermodynamics, which states that the entropy of an isolated system is non- decreasing. In statistical thermodynamics, entropy is often defined as the log of the number of microstates in the system. Boltzman, who has the logarithmic equation inscibed as the epitaph on his gravestone, carried out this work. In 1928, Hartley [55] introduced a logarithmic measure of information for communication. Shannon [107] was the first to define entropy and mutual information from the statistical point of view for communication. Kullback and Leibler [81] introduced the idea of relative information. Some times 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 describe the random variable.The relative information is a measure of the distance between two distributions. In statistics, 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 Brillouins [27] and Jaynes [64].

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

(1) χ^2 (chi square) divergence

Pearson [94] introduced the measure as

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

(2) Kullback-Leibler's relative information

Kullback -Leibler [81] introduced the divergence measure as

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

(3) Relative Jensen-Shannon divergence measure

Sibson [117] and Lin [85] introduced the divergence measure as

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

(4) Relative Arithmetic- geometric divergence measure

Taneja [125] introduced the divergence measure as

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

(5) J- Divergence measure

Jeffrey [65] and Kullback and Leibler [81] introduced the divergence measure as

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

(6) Arithmetic- geometric divergence measure

Taneja [123] introduced the measure as

(1.4.6) $$T\left(P//Q\right)=\sum_{i=1}^{n}\left(\frac{p_i+q_i}{2}\right)\log\left(\frac{p_i+q_i}{2\sqrt{p_iq_i}}\right)$$

(7) Hellinger discrimination measure

Hellinger [58] introduced the divergence measure as

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

(8) Triangular discrimination measure

Topsoe [128] introduced the divergence measure as

(1.4.8) $$\Delta\left(P//Q\right)=\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{p_i+q_i}$$

(9) Variational distance measure

Ben-tal et al [18] introduced the divergence measure as

(1.4.9) $$V\left(P//Q\right)=\sum_{i=1}^{n}\left|p_i-q_i\right|$$

Some important generalizations of divergence measures in information theory are:

(1) Relative information of order s

Renyi [104] introduced the generalized divergence measure as

(1.4.10) $$K^s(P//Q) = \frac{1}{s-1}\log\left(\sum_{i=1}^{n} p_i^s q_i^{1-s}\right), \quad s>0\,(\neq 1)$$

For $s \to 1$, (1.4.10) reduces to Kullback - Leibler divergence measure (1.4.2).

(2) Relative information of type s

Sharma and Autar [109] and Taneja [122] introduced the measure as

(1.4.11) $$K_s(P//Q) = \frac{1}{s-1}\left[\sum_{i=1}^{n} p_i^s q_i^{1-s} - 1\right], \quad s>0\,(\neq 1)$$

(3) Relative information of type s

Vajda [131] introduced the generalized measure as

(1.4.12) $$K_s(P//Q) = \frac{1}{s(s-1)}\left[\sum_{i=1}^{n} p_i^s q_i^{1-s} - 1\right], \quad s \neq 0,1$$

(4) Relative JS and AG divergence of type s

Pranesh and Taneja [83] introduced the generalized measure as

(1.4.13) $$FG_s(P//Q) = \frac{1}{s(s-1)}\left[\sum_{i=1}^{n} p_i\left(\frac{p_i+q_i}{2p_i}\right)^s - 1\right], \; s \neq 0,1$$

(5) J -Divergence of type s

Taneja [124] introduced the generalized measure as

(1.4.14) $$J_s(P//Q) = \frac{1}{s(s-1)}\left[\sum_{i=1}^{n}\left(p_i^s q_i^{1-s} + p_i^{1-s} q_i^s\right) - 2\right], s \neq 0,1$$

(6) AG and JS divergence of type s

Taneja [124] introduced the generalized measure as

(1.4.15) $$I_s(P//Q) = \frac{1}{s(s-1)}\left[\sum_{i=1}^{n}\left(\frac{p_i^{1-s}+q_i^{1-s}}{2}\right)\left(\frac{p_i+q_i}{2}\right)^s - 1\right], s \neq 0,1$$

(7) Csiszar’s f- divergence

Given a convex function $f:[0,\infty)\to\Re$, the f- divergence measure introduced by Csiszar [36] is given by

(1.4.16) $$C_f(p,q) = \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right)$$

where $p, q \in \Re_+^n$.

The following two important concepts are due to Csiszar and Korner [34].

Joint convexity: Let $f : [0,\infty) \to \Re$ be a convex, then $C_f(p,q)$ is jointly convex in p and q, where $p, q \in \Re_+^n$.

Jensen's inequality: Let $f : [0,\infty) \to \Re$ be a convex function. Then for any $p, q \in \Re_+^n$ with $P_n = \sum_{i=1}^{n} p_i > 0,\ Q_n = \sum_{i=1}^{n} q_i > 0,$

we have the inequality

(1.4.17) $$C_f(p,q) \geq Q_n f\left(\frac{P_n}{Q_n}\right)$$

The equality sign holds iff

(1.4.18) $$\frac{p_1}{q_1} = \frac{p_2}{q_2} = \ldots = \frac{p_n}{q_n}$$

In particular, for all $P, Q \in \Delta_n$, we have

$$C_f(p,q) \geq f(1)$$

with equality iff $P = Q$.

1.5 Generalized 'useful' information measures

Shannon's entropy is indeed a measure of uncertainty in the scheme and is treated as information supplied by a probabilistic experiment. This formula gives us the measure of information as a function of the probabilities with which various events occur without considering the effectiveness or importance of the events. The possible events of a given experiment is represented by the relevance or the utility of the information they carry with respect to a goal. These utilities may be either of objective or subjective character. We shall suppose that these qualitative (usefulness) are non- negative, finite, real numbers as the utility in decision theory. Also, if one event is more useful than another one, the utility of the first event will be greater than that of the second one. We now try to evaluate how the amount of information supplied by a probabilistic experiment, whose elementary events are characterized both by their probabilities and by utility of the events. Motivated with this idea Belis and Guiasu [17] introduced a utility distribution

$U = (u_1, u_2, ..., u_n)$ where each u_i is a non -negative real numbers accounting for the utility of the occurrence of the i^{th} event. If X is a discrete random variable taking on a finite number of possible values $x_1, x_2, ..., x_n$. We define the utility information scheme by

(1.5.1) $$X = \begin{bmatrix} x_1 & x_2 \dots .. & x_n \\ p_1 & p_2 \dots . & p_n \\ u_1 & u_2 \dots . & u_n \end{bmatrix}$$

The utility u_i of an event may be either independent, or dependent on its objective probability. The measure of information for the utility information scheme (1.5.1) given by Belis and Guiasu [17] is

(1.5.2) $$H(U;P) = -\sum_{i=1}^{n} u_i p_i \log p_i$$

where $P = (p_1, p_2, ..., p_n), p_i \geq 0, \sum_{i=1}^{n} p_i = 1$ and $U = (u_1, u_2, ..., u_n), u_i > 0$.

It satisfies many algebraic and analytical properties (Guiasu [52]). It reduces to Shannon's entropy (1.2.2) when utility aspect of the scheme is ignored by taking $u_i = 1 \ \forall\, i = 1, 2, ..., n$.

The generalized additive 'useful' information of order α given by Gurdial and Pessoa [53] is

(1.5.3) $$H_\alpha(U;P) = \frac{1}{1-\alpha} \log \left(\frac{\sum_{i=1}^{n} u_i p_i^\alpha}{\sum_{i=1}^{n} u_i p_i} \right), \quad \alpha > 0\, (\neq 1)$$

and used it in studying noiseless coding theorems for source having utilities. The measure $H_\alpha(U;P)$ resembles to Renyi's [104] entropy of order α when utility is ignored by taking $u_i = 1 \ \forall \ i = 1, 2..., n$.

The non- additive 'useful' information of degree β was first introduced and characterized by Sharma, Mittal and Mohan [113] which is given by

(1.5.4) $$I^\beta(P;U) = \frac{\sum_{i=1}^{n} u_i p_i \left(p_i^{\beta-1} - 1\right)}{\left(2^{1-\beta} - 1\right) \sum_{i=1}^{n} p_i}, \quad \beta \neq 1$$

The measure (1.5.4) was further generalized by Hooda and Tuteja [62] as

(1.5.5) $$H_\alpha^\beta(P;U) = \frac{\sum_{i=1}^{n} u_i^\alpha p_i^\alpha \left(p_i^{\beta-1} - 1\right)}{\left(2^{1-\beta} - 2^{1-\alpha}\right) \sum_{i=1}^{n} p_i}, \quad \alpha \neq 0, \beta \neq 1, \alpha \neq \beta$$

The measure involving utilities given by Hooda and Tuteja [62] as

(1.5.6) $$H_\alpha(P;U) = -2^{\alpha-1}\sum_{i=1}^{n}(u_i p_i)^\alpha \log p_i \ , \quad \alpha \neq 0$$

For $\alpha = 1$, (1.5.6) reduces to measure of 'useful' information given by Belis and Guiasu [17]. For $u_i = 1 \ \forall \ \ i = 1, 2, ..., n$ in (1.5.6), it reduces to entropy studied by Sharma and Taneja [112].

Autar and Khan [6] introduced the 'useful' information measure for incomplete probability distribution as

(1.5.7) $$_\alpha H(U;P) = \frac{1}{\left(D^{\frac{\alpha}{\alpha-1}}-1\right)}\left[1-\left(\frac{\sum_{i=1}^{n} u_i p_i^\alpha}{\sum_{i=1}^{n} u_i p_i}\right)^{\frac{1}{\alpha}}\right], \quad \alpha > 0\,(\neq 1)\,, \sum_{i=1}^{n} p_i \leq 1$$

For $\alpha \to 1$, (1.5.7) reduces to measure given by Belis and Guiasu [17] for incomplete probability distribution.

Hooda and Singh [61] introduced the 'useful' information measure of order α and type β as

(1.5.8) $$H_\alpha^\beta(P;U) = \frac{1}{1-\alpha}\log\frac{\sum_{i=1}^{n} u_i^\beta p_i^{\alpha+\beta-1}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta}, \quad \alpha, \beta > 0, \quad \alpha \neq 1$$

Singh et al [118] introduced the two 'useful' information measures as

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

and

(1.5.10) $$H_R(P;U) = \frac{R}{R-1}\left[1-\left(\frac{\sum_{i=1}^{n} u_i p_i^R}{\sum_{i=1}^{n} u_i p_i}\right)^{\frac{1}{R}}\right],$$

where $R \in \Re, \sum_{i=1}^{n} p_i = 1, i = 1, 2, ..., n$

Khan et al [78] introduced the generalized 'useful' information measure for the incomplete probability distribution as

(1.5.11) $$_{\alpha\beta}H(U;P) = \frac{\alpha}{\alpha-1}\left[1-\left(\frac{\sum_{i=1}^{n} u_i p_i^{\alpha\beta}}{\sum_{i=1}^{n} u_i p_i^\beta}\right)^{\frac{1}{\alpha}}\right]$$

where $\alpha > 0\,(\neq 1)\,, \beta > 0, p_i \geq 0, \sum_{i=1}^{n} p_i \leq 1, i = 1, 2, ..., n$

For $\beta = 1, \alpha \to 1$, (1.5.11) reduces to a measure of 'useful' information for the incomplete probability distribution due to Belis and Guiasu [17].

Hooda and Ram [60] introduced the 'useful' information measure as

$$(1.5.12) \qquad H_\alpha(P;U) = \frac{1}{2^{1-\alpha}-1}\left[\frac{\sum_{i=1}^{n} u_i p_i^\alpha}{\sum_{i=1}^{n} u_i p_i} - 1\right], \quad \alpha > 0\,(\neq 1)$$

1.6 Useful inaccuracy measures and their generalizations

Let $P = (p_1, p_2, ..., p_n)\,, p_i \geq 0, \sum_{i=1}^{n} p_i = 1$ be the probability distribution associated with a finite system of events $X = (x_1, x_2, ..., x_n)$ representing the realization of some experiment. The different x_i depend upon the experiments's goal or upon some qualitative characteristic of the physical system taken into account; ascribe to each event x_i a non- negative number $u_i\,(> 0)$ directly proportional to its importance and call u_i the utility of the event x_i. Then the weighted entropy [17] of the experiment X is defined as

$$(1.6.1) \qquad I(P;U) = -\sum_{i=1}^{n} u_i p_i \log p_i$$

Now let us suppose that the experimenter asserts that the probability of the i^{th} outcome x_i is q_i, whereas the true probability is p_i, with $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1$. Thus, we have two utility information schemes

$$(1.6.2) \qquad S = \begin{bmatrix} x_1 & x_2 \ldots\ldots & x_n \\ p_1 & p_2 \ldots\ldots & p_n \\ u_1 & u_2 \ldots\ldots & u_n \end{bmatrix}, \quad p_i \geq 0, u_i > 0, \sum_{i=1}^{n} p_i = 1$$

of a set of n events after an experiment, and

$$(1.6.3) \qquad S^* = \begin{bmatrix} x_1 & x_2 \ldots\ldots & x_n \\ q_1 & q_2 \ldots\ldots & q_n \\ u_1 & u_2 \ldots\ldots & u_n \end{bmatrix}, \quad q_i \geq 0, u_i > 0, \sum_{i=1}^{n} q_i = 1$$

of the same set of n events before the experiment.

In both the schemes (1.6.2) and (1.6.3) the utility distribution is the same because we assume that the utility u_i of an outcome x_i is independent of its probability of occurrence p_i , or predicted probability q_i; u_i is only a 'utility' or value of the outcome x_i for an observer relative to some specified goal (refer to [87]).

The quantitative- qualitative measure of inaccuracy [119] associated with the statement of an experimenter as

$$(1.6.4) \qquad I(P;Q;U) = -\sum_{i=1}^{n} u_i p_i \log q_i$$

when $u_i = 1 \quad \forall \quad i = 1, 2, ..., n$, the measure (1.6.4) reduces to Kerridge's [73] inaccuracy.

Bhatia [23] generalized the 'useful' inaccuracy measure as

$$(1.6.5) \qquad I_\alpha\left(P; Q; U\right) = \frac{1}{1-\alpha} \log \frac{\sum_{i=1}^{n} u_i p_i q_i^{\alpha-1}}{\sum_{i=1}^{n} u_i p_i}, \quad \alpha > 0 \left(\neq 1\right)$$

Further, Bhatia [22] introduced the 'useful' inaccuracy of order α for incomplete probability distribution, which is given by

$$(1.6.6) \qquad I_\alpha\left(P; Q; U\right) = \frac{1}{D^{\frac{\alpha-1}{\alpha}}-1} \left[1 - \left(\frac{\sum_{i=1}^{n} u_i p_i q_i^{\alpha-1}}{\sum_{i=1}^{n} u_i p_i}\right)^{\frac{1}{\alpha}}\right]$$

where $\alpha > 0 \left(\neq 1\right), p_i \geq 0, \sum_{i=1}^{n} p_i \leq 1, i = 1, 2, ..., n$. D is the size of the code alphabet.

Tuteja and Bhaker [129] introduced the 'useful' inaccuracy measure of order α and type β as

$$(1.6.7) \qquad I_\alpha^\beta\left(P/Q; U\right) = \frac{\sum_{i=1}^{n} u_i^\beta p_i^\beta \left(q_i^{\alpha-1} - 1\right)}{\left(2^{1-\alpha}-1\right) \sum_{i=1}^{n} p_i^\beta}, \quad \alpha \neq 1, \beta > 0$$

They also studied non- additive 'useful' inaccuracy of order α and type (β, γ) as

$$(1.6.8) \qquad I_\alpha^{(\beta,\gamma)}\left(P/Q; U\right) = \frac{\sum_{i=1}^{n} u_i^\beta p_i^\beta \left(p_i^{\gamma-1} q_i^{\alpha-\gamma} - 1\right)}{\left(2^{1-\alpha}-1\right) \sum_{i=1}^{n} p_i^\beta}, \quad \alpha, \beta, \gamma > 0, \alpha \neq 1$$

Parkash [93] introduced the 'useful' inaccuracy measure as

$$(1.6.9) \qquad I_n^\beta\left(P; Q; U\right) = \frac{1}{2^{-\beta}-1} \left[\sum_{i=1}^{n} u_i p_i \left(q_i^\beta - 1\right)\right], \quad \beta \neq 0$$

1.7 Coding theorems

The elements of a finite set of n input symbols $X = (x_1, x_2, ..., x_n)$ be encoded using alphabet of D symbols. The number of symbols in a codeword is called the length of the codeword. It becomes clear that some restriction must be placed on the assignment of codewords. One of the restrictions may be that the sequence may be decoded accurately. A code is uniquely decipherable if every finite sequence of code character corresponds to at most one message. In other words, we can say uniquely decipherability is to require that no code be a prefix of another codeword. We mean by prefix that a sequence 'A' of code character is prefix of a sequence 'B', if 'B' may be written as 'AC' for some sequence 'C'.

A code having the property that no codeword is prefix of another codeword is said to be instantaneous code. Feinstein [49] proved that instantaneous/uniquely decipherable code with lengths $l_1, l_2, ..., l_n$ is possible iff

(1.7.1) $$\sum_{i=1}^{n} D^{-l_i} \leq 1$$

where D is the size of the code alphabet. The inequality (1.7.1) is known as Kraft [80] inequality. Also if

(1.7.2) $$L = \sum_{i=1}^{n} l_i p_i$$

is the average codeword length, where p_i is the probability of the i^{th} input symbol to a noiseless channel then for a code which satisfy (1.7.1), the following inequality holds

(1.7.3) $$L \geq H(P)$$

where $D\,(D > 1)$ is an arbitrary base. Equality in (1.7.3) holds iff

(1.7.4) $$l_i = -\log p_i\ , \quad i = 1, 2, ..., n$$

by suitable encoding the message, the average code length can be arbitrarly close to $H(P)$.

Shannon's [107] and Renyi's [104] entropies have been studied by several research workers. The study has been carried out from essentially two different point of view. The first is an axiomatic approach and the second is a pragmatic approach. However these approaches have little connection with the coding theorem of information theory.

Campbell [29] defined a codeword length of order t as

(1.7.5) $$L(t) = \frac{1}{t} \log \left(\sum_{i=1}^{n} p_i D^{tl_i} \right), \quad -1 < t < \infty,\ t \neq 0$$

and developed a noiseless coding theorem for Renyi's [104] entropy of order α which is quite similar to the noiseless coding theorem for Shannon's [107] entropy.

By means of prefix code Longo [87], Gurdial and Pessoa [53], Sharma et al [113] , Bernard and Sharma [20], Blak [25], Autar and Soni [7], Khan and Autar [75], Khan and Haseen [74], Khan, Autar and Haseen [77], Jain and Tuteja [63], Autar and Khan [6], Baig and Zaheeruddin [14], Bhatia [22,23], Hooda and Bhaker [59], Singh, Kumar and Tuteja [118], Khan et al [78] etc have established coding theorems for various information measures.

Theorem (1.7.1): (Kraft [80]): A necessary and sufficient condition for the existance of a instantaneous code $S\left(x_i\right)$ such that the length of each word $S\left(x_i\right)$, should be l_i , $i = 1, 2, .., n$ is that the Kraft inequality

(1.7.6) $$\sum_{i=1}^{n} D^{-l_i} \leq 1$$

should hold. Where D is the number of symbols in the code alphabet.

Proof: Necessary part:

First suppose that there exists a code $S\left(x_i\right)$ with the word length l_i , $i = 1, 2, .., n$. Define $m = \max\left\{l_i\ , i = 1, 2, .., n\right\}$ and let $u_j, j = 1, 2, ..., n$ be the number of codewords with length j (some u_j may be zero). Thus the number of codewords with only one letter can not be larger than D

(1.7.7) $$u_1 \leq D$$

The number of codewords of length 2, can use only of the remaining $(D - u_1)$ symbols in their first place, because of prefix property of our codes, while any of the D symbols can be used in the second place, thus

(1.7.8) $$u_2 \leq \left(D - u_1\right) D = D^2 - u_1 D$$

Similarly,

(1.7.9) $$u_3 \leq \left(D^2 - u_1 D - u_2\right) D = D^3 - u_1 D^2 - u_2 D$$

Finally, If m is the maximum length of the encoded words, one concludes that

(1.7.10) $$u_m \leq D^m - u_1 D^{m-1} - u_2 D^{m-2} - ... - u_{m-1} D$$

Dividing (1.7.10) by D^m, we get

(1.7.11) $$0 \leq 1 - u_1 D^{-1} - u_2 D^{-2} - ... - u_{m-1} D^{1-m} - u_m D^{-m}$$

or

(1.7.12) $$\sum_{i=1}^{m} u_i D^{-i} \leq 1$$

It may not be obvious that this condition is identical with (1.7.6) but note that $m \geq l_i, i = 1, 2, ..., n$ and $\sum_{i=1}^{m} u_i D^{-i} \leq 1$ means the sum of "the numbers of all sequences of length i multiplied by D^{-i}", where the summation extends from 1 to m. The left hand side of (1.7.12) can be written as

(1.7.13) $$\sum_{i=1}^{m} u_i D^{-i} = \underbrace{D^{-1} + D^{-1} + ... + D^{-1}}_{u_1 times} + \underbrace{D^{-2} + D^{-2} + ... + D^{-2}}_{u_2\ times}$$

$$+ ... + \underbrace{D^{-m} + D^{-m} + ... + D^{-m}}_{u_m \ times}$$

each bracketed expression corresponds to a message x_i, and therefore the total number of terms in n.

$$\underbrace{1, 1, ..., 1}_{u_1}, \underbrace{2, 2, ..., 2}_{u_2}, ..., \underbrace{m, m, ..., m}_{u_m}$$

$$u_1 + u_2 + ... + u_m = n$$

The terms in u_k coressponds to the encoded message of length k. These latter terms can be considerd as $\sum D^{-l_i}$ when the summation takes place over all those terms with $l_i = k$. Therefore, by a simple re- assignment of terms, we may equivalently write

(1.7.14) $$\sum_{i=1}^{m} u_i D^{-i} = \sum_{i=1}^{n} D^{-l_i}$$

Thus

$$\sum_{i=1}^{m} u_i D^{-i} = \sum_{i=1}^{n} D^{-l_i} \leq 1$$

The desired set of positive integers $[l_1, l_2, ..., l_n]$ must satisfy the inequality (1.7.6). This proves the necessity requirement of the theorem.

Sufficient part:

Suppose now that inequality (1.7.6) is satisfied for $[l_1, l_2, ..., l_n]$. Then every summand of the left hand side of (1.7.6) being non negative, the partial sums are also at most 1.

$u_1 D^{-1} \leq 1$ or $u_1 \leq D$

$u_1 D^{-1} + u_2 D^{-2} \leq 1$ or $u_2 \leq D^2 - u_1 D$

.

.

.

$u_1 D^{-1} + u_2 D^{-2} + ... + u_n D^{-n} \leq 1$

or $u_n \leq D^n - u_1 D^{n-1} - u_2 D^{n-2} - ... - u_{n-1} D$

but these are exactly the conditions that we have to satisfy in order to guarantee that no encoded message can be obtained from any other by the addition of a sequence of letters of the encoding alphabet, therefore, which implies the existance of the instantaneous code.

Remark (1.7.1): For binary case the Kraft inequality tells us that the length l_i must satisfy the equation

(1.7.15) $$\sum_{i=1}^{n} 2^{-l_i} \leq 1$$

where the summation is over all the words of the block code.

Lemma (1.7.1): (Aczel and Daroczy [2]): For a probability distribution $P = (p_1, p_2, ..., p_n), p_i \geq 0, \sum_{i=1}^{n} p_i = 1$ and incomplete distribution $Q = (q_1, q_2, ..., q_n), q_i \geq 0, \sum_{i=1}^{n} q_i \leq 1$, The following inequality holds

(1.7.16) $$-\sum_{i=1}^{n} p_i \log p_i \leq -\sum_{i=1}^{n} p_i \log q_i$$

Proof: Before proving Lemma (1.7.1), we state the following Lemma.

Lemma (1.7.2): (Aczel and Daroczy [2]): If ψ is differentiable concave function in (a,b), then for all $x_i \in (a,b), i = 1, 2, ..., n$ and for all $(q_1, q_2, ..., q_n), q_i \geq 0, \sum_{i=1}^{n} q_i = 1, i = 1, 2, ..., n,$ the inequality

$$\psi\left[\sum_{i=1}^{n} q_i x_i\right] \geq \sum_{i=1}^{n} q_i \psi(x_i) \text{ holds.}$$

Define the function

$$L(x) = \begin{cases} -x \log x & \text{for } x \in (0, \infty) \\ 0 & \text{for } x = 0 \end{cases}$$

It is differentiable concave function of x on $[0, \infty)$ and continuous at 0 (from right), as

$$\frac{\delta^2}{\delta x^2}(x \log x) > 0, \qquad \underset{x \to 0+}{lt}\, x \log x = 0 \log 0 = 0$$

Putting $x_i = \frac{p_i}{q_i}, i = 1, 2, ..., n$ in Lemma (1.7.2), we get

$$\sum_{i=1}^{n} q_i L\left(\frac{p_i}{q_i}\right) \leq L\left(\sum_{i=1}^{n} q_i \frac{p_i}{q_i}\right)$$

$$= L\left(\sum_{i=1}^{n} p_i\right) = L(1) = 0$$

Thus

$$0 \geq -\sum_{i=1}^{n} q_i \frac{p_i}{q_i} \log \frac{p_i}{q_i}$$

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

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

or

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

Theorem (1.7.2):(Shannon [107]): Let $\{X\}$ be a discrete message source, without memory, and x_i be any message of this source with probability of transmission p_i. If the $\{X\}$ ensemble is encoded in a sequence of uniquely decipherable characters taken from the alphabet $\{a_1, a_2, ..., a_n\}$, then

(1.7.17) $$L = \sum_{i=1}^{n} p_i l_i \geq \frac{H(P)}{\log D}$$

Proof: The condition $L \geq \frac{H(P)}{\log D}$ is equivalent to

$$\log D \sum_{i=1}^{n} p_i l_i \geq -\sum_{i=1}^{n} p_i \log p_i$$

Since $p_i l_i \log D = p_i \log D^{l_i} = -p_i \log D^{-l_i}$, The above condition may be written as

$$-\sum_{i=1}^{n} p_i \log D^{-l_i} \geq -\sum_{i=1}^{n} p_i \log p_i$$

We define $q_i = \frac{D^{-l_i}}{\sum_{i=1}^{n} D^{-l_i}}$, then q_i's add to unity and Lemma (1.7.1) yields

(1.7.18) $$-\sum_{i=1}^{n} p_i \log p_i \leq -\sum_{i=1}^{n} p_i \log \left(\frac{D^{-l_i}}{\sum_{i=1}^{n} D^{-l_i}} \right)$$

with equality iff $p_i = \frac{D^{-l_i}}{\sum_{i=1}^{n} D^{-l_i}} \quad \forall\, i = 1, 2, ..., n$

Hence by (1.7.18)

$$H(P) \leq -\sum_{i=1}^{n} p_i \log D^{-l_i} + \sum_{i=1}^{n} p_i \log \left(\sum_{i=1}^{n} D^{-l_i} \right)$$

$$H(P) \leq L \log D + \log \left(\sum_{i=1}^{n} D^{-l_i} \right)$$

with equality iff $p_i = \frac{D^{-l_i}}{\sum_{i=1}^{n} D^{-l_i}} \quad \forall\ i = 1, 2, ..., n.$

By Theorem (1.7.1), $\sum_{i=1}^{n} D^{-l_i} \leq 1$ which gives

$$\log \left(\sum_{i=1}^{n} D^{-l_i} \right) \leq 0$$

Therefore

$$H(P) \leq L \log D$$

or

$$L \geq \frac{H(P)}{\log D}$$

Theorem (1.7.3) : (Shannon [107]): Given a random variable $X = (x_1, x_2, ..., x_n)$ having probability distribution $P = (p_1, p_2, ..., p_n)$ with entropy (uncertainty) $H(P)$, there exists a base D, instantaneous code for X, whose average codeword length $L = \sum_{i=1}^{n} l_i p_i$ satisfies

(1.7.19) $$\frac{H(P)}{\log D} \leq L < \frac{H(P)}{\log D} + 1$$

Proof: In general we can not hope to construct an absolutely optimal code for a given set of probabilities $P = (p_1, p_2, ..., p_n)$, since if we choose l_i to satisfy $p_i = D^{-l_i}$, then $l_i = \frac{-\log p_i}{\log D}$ may not be an integer. However we can do the next best thing and select the integer l_i such that

(1.7.20) $$\frac{-\log p_i}{\log D} \leq l_i < \frac{-\log p_i}{\log D} + 1\ ,\ i = 1, 2, ..., n$$

We claim that an instantaneous code can be constructed with word lengths $l_1, l_2, ..., l_n$. To prove this we must show that $\sum_{i=1}^{n} D^{-l_i} \leq 1$.

For the left hand inequality of (1.7.20) it follows that

$$\log p_i \geq -l_i \log D$$

or

$$p_i \geq D^{-l_i}$$

Thus

$$\sum_{i=1}^{n} D^{-l_i} \leq \sum_{i=1}^{n} p_i = 1$$

$$\sum_{i=1}^{n} D^{-l_i} \leq 1$$

To estimate the average codeword length, we multiply (1.7.20) by p_i and sum over $i = 1, 2, ..., n$, to obtain

$$-\sum_{i=1}^{n} p_i \frac{\log p_i}{\log D} \leq \sum_{i=1}^{n} p_i l_i < -\sum_{i=1}^{n} p_i \frac{\log p_i}{\log D} + \sum_{i=1}^{n} p_i$$

or

$\frac{H(P)}{\log D} \le L < \frac{H(P)}{\log D} + 1$

1.8: Convex funtion

A real valued function $f(x)$ defined on (a,b) is said to be convex function if for every α 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

(1.8.1) $$f(\alpha x_1 + (1-\alpha)x_2) \le \alpha f(x_1) + (1-\alpha) f(x_2)$$

If we put $\alpha = \frac{1}{2}$, then (1.8.1) reduces to

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

which is also taken as definition of convexity.

A function that is convex according to (1.8.1) is also convex according to (1.8.2). We will adopt (1.8.1) as the definition for a convex function. (see the figure 1).

Remark (1.8.1): If $f^{//}(x) \ge 0$, then $f(x)$ is convex function.

1.9. Strictly convex funtion

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

(1.9.1) $$f(\alpha x_1 + (1-\alpha)x_2) < \alpha f(x_1) + (1-\alpha) f(x_2)$$

Remark (1.9.1): If $f^{//}(x) > 0$, then $f(x)$ is strictly convex function.

1.10. Concave function

A function $f(x)$ is said to be concave if $-f(x)$ is convex. (see the figure 2)

Remark (1.10.1): If $f^{//}(x) \le 0$ then $f(x)$ is concave.

1.11. Strictly concave function

A function $f(x)$ is said to be strictly concave if $-f(x)$ is strictly convex.

Remark (1.11.1): If $f^{//}(x) < 0$ then $f(x)$ is strictly concave function.

1.12. Jensen's inequality: (Rao [100]): If X is a random variable such that $E(X) = \mu$ exists and $f(.)$ is a convex function, then

(1.12.1) $$E[f(X)] \geq f[E(X)]$$

with equality iff the random variable X has a degenerate distribution at μ.

1.13. Holder's inequality: (Shisha [116]): If $x_i, y_i > 0, i = 1, 2, ..., n$ and $\frac{1}{p} + \frac{1}{q} = 1, p > 1$, then the following inequality holds

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

with equality iff $x_i^p = c y_i^q$. The inequality is reversed for $p < 1 (\neq 0), q < 0$ or $q < 1 (\neq 0), p < 0$.

1.14. Kantorovic inequality: (Mitrinovic [89]): Kantrovic proved the following inequality for the sequences of real numbers

(1.14.1) $$\sum_{k=1}^{n} r_k u_k^2 \sum_{k=1}^{n} \frac{1}{r_k} u_k^2 \leq \frac{1}{4} \left(\sqrt{\frac{M}{m}} + \sqrt{\frac{m}{M}}\right)^2 \left(\sum_{k=1}^{n} u_k^2\right)^2$$

where $0 < m \leq r_k \leq M$ for $k = 1, 2, ..., n$

1.15. Diaz- Metcalf inequality: (Mitrinovic [89]): Let $p_k > 0 \quad (k = 1, 2, ..., n)$ with $\sum_{k=1}^{n} p_k = 1$. If $a_k (\neq 0)$ and $b_k (k = 1, 2, ..., n)$ are real numbers and

(1.15.1) $$m \leq \frac{b_k}{a_k} \leq M \quad \text{for } k = 1, 2, ..., n$$

Then

(1.15.2) $$\sum_{k=1}^{n} p_k b_k^2 + mM \sum_{k=1}^{n} p_k a_k^2 \leq (M + m) \sum_{k=1}^{n} p_k a_k b_k$$

equality holds in (1.15.2) iff for each k, $1 \leq k \leq n$ either $b_k = m a_k$ or $b_k = M a_k$

1.16. Interior point

If $I \subseteq \Re$ be a set of real numbers then any point $x \in I$ is said to be interior point if there exists at least one r (r is real number) positive such that $(x - r, x + r) \subset I$. The set of all interior points of I is denoted by $\overset{\bullet}{I}$ and we call $\overset{\bullet}{I}$ as interior of I.

It is true that the Shannon entropy is fundamental from the application point of view. But during past several years researchers have paid attention to the applications of generalized entropies in different branches and found them as good as Shannon entropy and sometimes better because of flexibility of the parameters, specially in comparisons purposes.

In this chapter, several coding theorems have been obtained by considering some parametric entropy functions involving utilities. In the literature of information theory several type of coding theorems involving entropy functions exists. The coding theorems obtained here are not only new but also generalizes some well known results available in the literature.

Also, codes of variable length that are capable of error correction are studied in this chapter. A lower bound on the generalized mean length of such codes under the criterion of "promptness" is obtained. This generalizes the result due to Bernard and Sharma [20], Baig and Zaheeruddin [14] for noisy channels. Also, this generalizes the result due to Kerridge [73] which itself is a generalization of celebrated result due to Shannon [107] for noisless channel. The bounds obtained here provides a measure of optimality for variable length error correcting codes.

2.1. Introduction

Let X be a discrete random variable taking on a finite number of possible values $X = (x_1, x_2, ..., x_n)$ with respective probabilities $P = (p_1, p_2, ..., p_n)$, $p_i \geq 0, \sum_{i=1}^{n} p_i = 1$. Denote

(2.1.1) $$\chi = \begin{bmatrix} x_1 & x_2 . & . & . x_n \\ p_1 & p_2 . & . & .p_n \end{bmatrix}$$

We call the scheme (2.1.1) as a finite information scheme. Every finite scheme describes a state of uncertainty. Shannon [107] introduced a quantity which, in a reasonable way, measures the amount of uncertainty (entropy) associated with a given finite scheme. This measure is given by

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

The measure in (2.1.2) serve as a very suitable measure of entropy of the finite information scheme (2.1.1).

Let a finite set of n source symbols $X = (x_1, x_2, ..., x_n)$ be encoded using alphabet of D symbols, then it has been shown by Feinstein [49] that there is a uniquely decipherable/instantaneous code with lengths $l_1, l_2, ..., l_n$ iff the following Kraft [80] inequality is satisfied

(2.1.3) $$\sum_{i=1}^{n} D^{-l_i} \leq 1$$

If $L = \sum_{i=1}^{n} l_i p_i$ be the average codeword length, then for a code which satisfies (2.1.3), It has been shown [49] that

(2.1.4) $$L \geq H(P)$$

with equality iff $l_i = -\log p_i \quad \forall \ i = 1, 2, ..., n$. This is Shannon's coding theorem for noiseless channel. The equation (2.1.4) for Shannon's entropy with ordinary code mean length $L = \sum_{i=1}^{n} l_i p_i$ has played an important role in ordinary communication theory (see Shannon [107]).

Belis and Guiasu [17] observed that a source is not completely specified by the probability distribution P over the source alphabet X in the absence of qualitative character. So it can be assumed by Belis and Guiasu [17] that the source alphabet letters are assigned weights according to their importance or utilities in view of the experimenter.

Let $U = (u_1, u_2, ..., u_n)$ be the set of positive real numbers, where u_i is the utility or importance of outcome x_i. The utility, in general, is independent of probability of encoding of source symbol x_i i.e. p_i. The information source is thus given by

(2.1.5) $$\chi = \begin{bmatrix} x_1 & x_2. & . & . & x_n \\ p_1 & p_2. & . & . & p_n \\ u_1 & u_2 & . & . & . u_n \end{bmatrix}, \quad u_i > 0, \quad p_i \geq 0, \sum_{i=1}^{n} p_i = 1$$

Belis and Guiasu [17] introuduced the following quantitative- qualitative measure of information

(2.1.6) $$H(P; U) = -\sum_{i=1}^{n} u_i p_i \log p_i$$

which is a measure for the average of quantity of 'valuable' or 'useful' information provided by the information source (2.1.5).

Guiasu and Picard [51] considerd the problem of encoding the letter output by the source (2.1.5) by means of a single letter prefix code whose codewords $c_1, c_2, ..., c_n$ are of lengths $l_1, l_2, ..., l_n$ respectively and satisfy the Kraft's inequality (2.1.3). They introduced the following 'useful' mean length of the code

(2.1.7) $$L\left(U\right) = \frac{\sum\limits_{i=1}^{n} u_i p_i l_i}{\sum\limits_{i=1}^{n} u_i p_i}$$

Further they derived a lower bound for (2.1.7). However, Longo [86] interpreted (2.1.7) as the average transmission cost of the letters x_i and derived the bounds for this cost function.

Longo [86], Gurdial and Pessoa [53], Khan and Autar [76] , Autar and Khan [6], Jain and Tuteja [63], Taneja et al [120], Bhatia [22], Singh, Kumar and Tuteja [118], Khan and Haseen [74], Hooda and Bhaker [59], Khan et al [78] considerd the problem of 'useful' information measures and used it studying the upper and lower bounds for sources involving utilities.

In the next section, The bounds have been derived in terms of generalized 'useful' average codeword length and 'useful' information measure of order α and type β. The main aim of studying these bounds is to generalize some well known results available in the literature of information theory.

2.2. Bounds for generalized measure of cost

In the derivation of the cost measure (2.1.7) it is assumed that the cost is a linear function of code length, but this is not always the case. There are occassions when the cost behaves like an exponential function of codeword lengths. Such types of functions occur frequently in market equilibrium and growth models in economics. Thus some times it might be more appropriate to choose a code which minimizes the monotonic function of the quantity

(2.2.1) $$C = \sum_{i=1}^{n} u_i^{\beta} p_i^{\beta} D^{\frac{1-\alpha}{\alpha} l_i}$$

where $\alpha > 0\left(\neq 1\right), \ \beta > 0$ are the parameters related to cost.

In order to make the result of the chapter more comparable with the usual noiseless coding theorem, instead of minimizing (2.2.1), we minimize

(2.2.2) $$L_{\alpha}^{\beta}\left(U\right) = \frac{1}{2^{1-\alpha}-1}\left[\left(\frac{\sum\limits_{i=1}^{n}\left(u_i p_i\right)^{\beta} D^{\frac{1-\alpha}{\alpha} l_i}}{\sum\limits_{i=1}^{n}\left(u_i p_i\right)^{\beta}}\right)^{\alpha} - 1\right], \alpha > 0\left(\neq 1\right), \beta > 0$$

which is monotonic function of C and is the 'useful' average code length of order α

and type β.

Clearly, if $\alpha \to 1, \beta = 1,$ (2.2.2) reduces to (2.1.7) which further reduces to ordinary mean length given by Shannon [107] when $u_i = 1 \;\; \forall \; i = 1, 2, .., n.$ It can be also noted that (2.2.2) is monotonic non- decreasing function of α and if all the $l_i^{\text{'}}s$ are same, say $l_i = l \;\; \forall \; i = 1, 2, ..., n$ and $\alpha \to 1,$ then $L_\alpha^\beta (U) = l.$ This is an important property for any measure of length to possess.

Now, Consider a function, which is 'useful' information of order α and type β

(2.2.3) $$H_\alpha^\beta (P; U) = \frac{1}{2^{1-\alpha}-1}\left[\frac{\sum_{i=1}^n u_i^\beta p_i^{\alpha+\beta-1}}{\sum_{i=1}^n u_i^\beta p_i^\beta} - 1\right]$$

where $\alpha > 0\,(\neq 1)\,, \beta > 0, p_i \geq 0 \;\; \forall \; i = 1, 2, ..., n, \; \sum_{i=1}^n p_i \leq 1$

Remark (2.2.1)

(1) When $\beta = 1,$ (2.2.3) reduces to the measure of 'useful' information proposed and characterized by Hooda and Ram [60].

(2) When $\alpha \to 1, \beta = 1,$ (2.2.3) reduces to the measure given by Belis and Guiasu [17].

(3) When $\alpha \to 1, \beta = 1$ and $u_i = 1 \;\; \forall \; i = 1, 2, .., n.$ (2.2.3) reduces to the well known measure given by Shannon [107].

Also, the bounds are obtained for the measure (2.2.3) under the condition

(2.2.4) $$\sum_{i=1}^n u_i^\beta p_i^{\beta-1} D^{-l_i} \leq \sum_{i=1}^n u_i^\beta p_i^\beta$$

It may be seen that in case $\beta = 1, u_i = 1 \;\; \forall \; i = 1, 2, .., n.$ (2.2.4) reduces to the Kraft [80] inequality (2.1.3). Also, D is the size of the code alphabet.

Theorem (2.2.1). For all integers $D\,(D \geq 2).$ let l_i satisfies (2.2.4), then the generalized average 'useful' codeword length satisfies

(2.2.5) $$L_\alpha^\beta (U) \geq H_\alpha^\beta (P; U)$$

equality holds iff

(2.2.6) $$l_i = -\log p_i^\alpha + \log \frac{\sum_{i=1}^n u_i^\beta p_i^{\alpha+\beta-1}}{\sum_{i=1}^n u_i^\beta p_i^\beta}$$

Proof: By Holder's inequality [116]

(2.2.7) $$\sum_{i=1}^{n} x_i y_i \geq \left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} y_i^q\right)^{\frac{1}{q}}$$

for all $x_i, y_i > 0,\ i = 1, 2, .., n$ and $\frac{1}{p} + \frac{1}{q} = 1, p < 1\,(\neq 0)\,, q < 0$ or $q < 1\,(\neq 0)\,, p < 0$. We see the equality holds iff there exists a positive constant c such that

(2.2.8) $$x_i^p = c y_i^q$$

Making the substitution

$$p = \frac{\alpha-1}{\alpha}, \quad q = 1 - \alpha$$

$$x_i = \frac{(u_i p_i)^{\frac{\beta\alpha}{\alpha-1}} D^{-l_i}}{\sum_{i=1}^{n} (u_i p_i)^{\frac{\beta\alpha}{\alpha-1}}} \qquad , y_i = \frac{u_i^{\frac{\beta}{1-\alpha}} p_i^{\frac{\alpha+\beta-1}{1-\alpha}}}{\sum_{i=1}^{n} (u_i p_i)^{\frac{\beta}{1-\alpha}}}$$

in (2.2.7), we get

$$\frac{\sum_{i=1}^{n} u_i^\beta p_i^{\beta-1} D^{-l_i}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta} \geq \left[\frac{\sum_{i=1}^{n} u_i^\beta p_i^\beta D^{\frac{1-\alpha}{\alpha} l_i}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta}\right]^{\frac{\alpha}{\alpha-1}} \left[\frac{\sum_{i=1}^{n} u_i^\beta p_i^{\alpha+\beta-1}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta}\right]^{\frac{1}{1-\alpha}}$$

using the condition (2.2.4), we get

$$\left[\frac{\sum_{i=1}^{n} u_i^\beta p_i^\beta D^{\frac{1-\alpha}{\alpha} l_i}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta}\right]^{\frac{\alpha}{1-\alpha}} \geq \left[\frac{\sum_{i=1}^{n} u_i^\beta p_i^{\alpha+\beta-1}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta}\right]^{\frac{1}{1-\alpha}}$$

Taking $0 < \alpha < 1$, and raising power both sides $(1-\alpha)$, we get

$$\left[\frac{\sum_{i=1}^{n} u_i^\beta p_i^\beta D^{\frac{1-\alpha}{\alpha} l_i}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta}\right]^{\alpha} \geq \left[\frac{\sum_{i=1}^{n} u_i^\beta p_i^{\alpha+\beta-1}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta}\right]$$

Multiplying both sides by $\frac{1}{2^{1-\alpha}-1} > 0$ for $0 < \alpha < 1$ and after simplifying, we get

$$L_\alpha^\beta(U) \geq H_\alpha^\beta(P; U)$$

For $\alpha > 1$, The proof follows along the similar lines.

Theorem (2.2.2). For every code with lengths $l_1, l_2, ..., l_n$ satisfies (2.2.4), $L_\alpha^\beta(U)$ can be made to satisfy the inequality

(2.2.9) $$L_\alpha^\beta(U) < H_\alpha^\beta(P; U)\, D^{1-\alpha} + \frac{D^{1-\alpha}-1}{2^{1-\alpha}-1}$$

Proof: Let l_i be the positive integer satisfying the inequality

(2.2.10) $$-\log p_i^\alpha + \log \frac{\sum_{i=1}^{n} u_i^\beta p_i^{\alpha+\beta-1}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta} \leq l_i < -\log p_i^\alpha + \log \frac{\sum_{i=1}^{n} u_i^\beta p_i^{\alpha+\beta-1}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta} + 1$$

Consider the interval

$$(2.2.11) \qquad \delta_i = \left[-\log p_i^\alpha + \log \frac{\sum\limits_{i=1}^n u_i^\beta p_i^{\alpha+\beta-1}}{\sum\limits_{i=1}^n u_i^\beta p_i^\beta}, -\log p_i^\alpha + \log \frac{\sum\limits_{i=1}^n u_i^\beta p_i^{\alpha+\beta-1}}{\sum\limits_{i=1}^n u_i^\beta p_i^\beta} + 1\right]$$

of length 1. In every δ_i, there lies exactly one positive integer l_i such that

$$(2.2.12) \quad 0 < -\log p_i^\alpha + \log \frac{\sum\limits_{i=1}^n u_i^\beta p_i^{\alpha+\beta-1}}{\sum\limits_{i=1}^n u_i^\beta p_i^\beta} \leq l_i < -\log p_i^\alpha + \log \frac{\sum\limits_{i=1}^n u_i^\beta p_i^{\alpha+\beta-1}}{\sum\limits_{i=1}^n u_i^\beta p_i^\beta} + 1$$

We will first show that the sequence $l_1, l_2, ..., l_n$ thus defined satisfies (2.2.4). From (2.2.12), we have

$$-\log p_i^\alpha + \log \frac{\sum\limits_{i=1}^n u_i^\beta p_i^{\alpha+\beta-1}}{\sum\limits_{i=1}^n u_i^\beta p_i^\beta} \leq l_i$$

or

$$\frac{p_i^\alpha}{\frac{\sum\limits_{i=1}^n u_i^\beta p_i^{\alpha+\beta-1}}{\sum\limits_{i=1}^n u_i^\beta p_i^\beta}} \geq D^{-l_i}$$

Multiply both sides by $u_i^\beta p_i^{\beta-1}$ and summing over $i = 1, 2, ..., n$. We get (2.2.4).

The last inequality in (2.2.12) gives

$$l_i < -\log p_i^\alpha + \log \frac{\sum\limits_{i=1}^n u_i^\beta p_i^{\alpha+\beta-1}}{\sum\limits_{i=1}^n u_i^\beta p_i^\beta} + 1$$

$$l_i < \log \left(\frac{p_i^\alpha}{\frac{\sum\limits_{i=1}^n u_i^\beta p_i^{\alpha+\beta-1}}{\sum\limits_{i=1}^n u_i^\beta p_i^\beta}}\right)^{-1} D$$

or

$$D^{l_i} < \left(\frac{p_i^\alpha}{\frac{\sum\limits_{i=1}^n u_i^\beta p_i^{\alpha+\beta-1}}{\sum\limits_{i=1}^n u_i^\beta p_i^\beta}}\right)^{-1} D$$

For $0 < \alpha < 1$, raising power both sides $\frac{1-\alpha}{\alpha}$, we get

$$D^{l_i\left(\frac{1-\alpha}{\alpha}\right)} < \left(\frac{p_i^{\alpha}}{\frac{\sum_{i=1}^{n} u_i^{\beta} p_i^{\alpha+\beta-1}}{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta}}} \right)^{\frac{\alpha-1}{\alpha}} D^{\frac{1-\alpha}{\alpha}}$$

Multiplying both sides by $\frac{u_i^{\beta} p_i^{\beta}}{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta}}$ and summing over $i = 1, 2, .., n$, we get

$$\frac{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta} D^{\frac{1-\alpha}{\alpha} l_i}}{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta}} < \left(\frac{\sum_{i=1}^{n} u_i^{\beta} p_i^{\alpha+\beta-1}}{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta}} \right)^{\frac{1}{\alpha}} D^{\frac{1-\alpha}{\alpha}}$$

or

$$\left(\frac{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta} D^{\frac{1-\alpha}{\alpha} l_i}}{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta}} \right)^{\alpha} < \left(\frac{\sum_{i=1}^{n} u_i^{\beta} p_i^{\alpha+\beta-1}}{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta}} \right) D^{1-\alpha}$$

Since, $2^{1-\alpha} - 1 > 0$ for $0 < \alpha < 1$ and after suitable operations, we get

$$\frac{1}{2^{1-\alpha}-1} \left[\left(\frac{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta} D^{\frac{1-\alpha}{\alpha} l_i}}{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta}} \right)^{\alpha} - 1 \right] < \frac{1}{2^{1-\alpha}-1} \left[\frac{\sum_{i=1}^{n} u_i^{\beta} p_i^{\alpha+\beta-1}}{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta}} - 1 \right] D^{1-\alpha} + \frac{D^{1-\alpha}-1}{2^{1-\alpha}-1}$$

or we can write

$$L_{\alpha}^{\beta}(U) < H_{\alpha}^{\beta}(P; U)\, D^{1-\alpha} + \frac{D^{1-\alpha}-1}{2^{1-\alpha}-1}$$

As $D \geq 2$, we have $\frac{D^{1-\alpha}-1}{2^{1-\alpha}-1} > 1$ from which it follows that upper bound $L_{\alpha}^{\beta}(U)$ in (2.2.9) is greater than unity.

Also, for $\alpha > 1$, the proof follows along the similar lines.

In the next section, coding theorems have been obtained by considering a new parametric entropy function involving utilities and generalized 'useful' codeword mean length. The results obtained here are not only new but also generalizes some well known results available in the literature of information theory.

2.3. Coding theorems on entropy function depending upon parameter R and ν.

Consider a function

$$H_R(P^{\nu}, U) = \frac{R}{R-1} \left[1 - \left(\frac{\sum_{i=1}^{n} u_i p_i^{R+\nu-1}}{\sum_{i=1}^{n} u_i p_i^{\nu}} \right)^{\frac{1}{R}} \right] \tag{2.3.1}$$

where $R > 0\,(\neq 1)\,, \nu > 0, \sum_{i=1}^{n} p_i = 1,\ p_i \geq 0.$

Remark (2.3.1).

(1) When $\nu = 1$, (2.3.1) reduces to the 'useful' R- norm information measure due to Singh, Kumar and Tuteja [118].

(2) When $\nu = 1, u_i = 1\ \forall\ \ i = 1, 2, ..., n$, (2.3.1) reduces to the R- norm information measure due to Boekee and Lubbee [26].

(3) When $R \to 1$, $\nu = 1$ and $u_i = 1\ \ \forall\ \ i = 1, 2, ..., n$, (2.3.1) reduces to the well known measure given by Shannon [107] .

Further, consider a generalized 'useful' codeword mean length

(2.3.2) $$L_R\left(P^{\nu}, U\right) = \frac{R}{R-1}\left[1 - \frac{\sum\limits_{i=1}^{n} u_i p_i^{\nu} D^{-l_i\left(\frac{R-1}{R}\right)}}{\sum\limits_{i=1}^{n} u_i p_i^{\nu}}\right]$$

where $R > 0\,(\neq 1)\,, \nu > 0, p_i \geq 0, \sum\limits_{i=1}^{n} p_i = 1$. D is the size of the code alphabet.

Remark (2.3.2).

(1) When $\nu = 1$, (2.3.2) reduces to the 'useful' codeword mean length given by Singh, Kumar and Tuteja [118].

(2) When $\nu = 1, u_i = 1\ \ \forall\ \ \ i = 1, 2, ..., n$, (2.3.2) reduces to the codeword mean length due to Boekee et al [26].

(3) When $R \to 1$, $\nu = 1$, and $u_i = 1\ \ \ \forall\ \ \ \ i = 1, 2, ..., n$, (2.3.2) reduces to the optimal codeword mean length defined by Shannon [107].

We now establish a result, that in a sense, gives a characterization of $H_R\left(P^{\nu}, U\right)$ under the condition

(2.3.3) $$\sum\limits_{i=1}^{n} u_i p_i^{\nu-1} D^{-l_i} \leq \sum\limits_{i=1}^{n} u_i p_i^{\nu}$$

Remark (2.3.3).

When $\nu = 1,\ \ u_i = 1\ \ \forall\ i = 1, 2, ..., n$ and $\sum\limits_{i=1}^{n} p_i = 1$. (2.3.3) is a generalization of (2.1.3) which is Kraft's [80] inequality.

Theorem (2.3.1). For every code whose lengths $l_1, l_2, ..., l_n$ satisfies (2.3.3). Then the average codeword length satisfies

(2.3.4) $$L_R\left(P^{\nu}, U\right) \geq H_R\left(P^{\nu}, U\right)$$

equality holds iff

(2.3.5) $$l_i = -\log p_i^R + \log \frac{\sum\limits_{i=1}^{n} u_i p_i^{R+\nu-1}}{\sum\limits_{i=1}^{n} u_i p_i^{\nu}}$$

Proof: By Holder's inequality [116]

(2.3.6) $$\sum_{i=1}^{n} x_i y_i \geq \left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} y_i^q\right)^{\frac{1}{q}}$$

for all $x_i, y_i > 0,\ i = 1, 2, .., n$ and $\frac{1}{p} + \frac{1}{q} = 1,\ p < 1\,(\neq 0),\ \ q < 0$ or $q < 1\,(\neq 0)\,, p < 0$. we see the equality holds iff there exists a positive constant c such that

(2.3.7) $$x_i^p = c y_i^q$$

Setting

$$x_i = u_i^{\frac{R}{R-1}} p_i^{\frac{\nu R}{R-1}} D^{-l_i}, \qquad y_i = u_i^{\frac{1}{1-R}} p_i^{\frac{R+\nu-1}{1-R}}$$

$$p = \tfrac{R-1}{R}, \qquad q = 1 - R$$

in (2.3.6) and using (2.3.3), we get

(2.3.8) $$\left[\sum_{i=1}^{n} u_i p_i^{\nu} D^{-l_i\left(\frac{R-1}{R}\right)}\right]^{\frac{R}{1-R}} \geq \frac{\left[\sum\limits_{i=1}^{n} u_i p_i^{R+\nu-1}\right]^{\frac{1}{1-R}}}{\sum\limits_{i=1}^{n} u_i p_i^{\nu}}$$

Dividing both sides of (2.3.8) by $\left(\sum\limits_{i=1}^{n} u_i p_i^{\nu}\right)^{\frac{R}{1-R}}$, we get

$$\left[\frac{\sum\limits_{i=1}^{n} u_i p_i^{\nu} D^{-l_i\left(\frac{R-1}{R}\right)}}{\sum\limits_{i=1}^{n} u_i p_i^{\nu}}\right]^{\frac{R}{1-R}} \geq \left[\frac{\sum\limits_{i=1}^{n} u_i p_i^{R+\nu-1}}{\sum\limits_{i=1}^{n} u_i p_i^{\nu}}\right]^{\frac{1}{1-R}}$$

Taking $0 < R < 1$, raising both sides to the power $\frac{1-R}{R}$, $R \neq 1$, also $\frac{R}{R-1} < 0$ for $0 < R < 1$ and after suitable operations, we obtain the result (2.3.4). For $R > 1$, the inequality (2.3.4) can be obtained in a similar fashion.

Theorem (2.3.2). For every code with lengths $l_1, l_2, ..., l_n$ satisfies (2.3.3). Then $L_R\,(P^{\nu}, U)$ can be made to satisfy the inequality

(2.3.9) $$L_R\,(P^{\nu}, U) < H_R\,(P^{\nu}, U)\, D^{\frac{1-R}{R}} + \tfrac{R}{R-1}\left(1 - D^{\frac{1-R}{R}}\right)$$

Proof: Let l_i be the positive integer satisfying the inequality

(2.3.10) $$-\log p_i^R + \log \frac{\sum\limits_{i=1}^{n} u_i p_i^{R+\nu-1}}{\sum\limits_{i=1}^{n} u_i p_i^{\nu}} \leq l_i < -\log p_i^R + \log \frac{\sum\limits_{i=1}^{n} u_i p_i^{R+\nu-1}}{\sum\limits_{i=1}^{n} u_i p_i^{\nu}} + 1$$

Consider the interval

$$(2.3.11) \qquad \delta_i = \left[-\log p_i^R + \log \frac{\sum_{i=1}^{n} u_i p_i^{R+\nu-1}}{\sum_{i=1}^{n} u_i p_i^{\nu}}, -\log p_i^R + \log \frac{\sum_{i=1}^{n} u_i p_i^{R+\nu-1}}{\sum_{i=1}^{n} u_i p_i^{\nu}} + 1 \right]$$

of length 1. In every δ_i, there lies exactly one positive integer l_i such that

$$(2.3.12) \qquad 0 < -\log p_i^R + \log \frac{\sum_{i=1}^{n} u_i p_i^{R+\nu-1}}{\sum_{i=1}^{n} u_i p_i^{\nu}} \leq l_i < -\log p_i^R + \log \frac{\sum_{i=1}^{n} u_i p_i^{R+\nu-1}}{\sum_{i=1}^{n} u_i p_i^{\nu}} + 1$$

We will first show that the sequences $\{l_1, l_2, ..., l_n\}$, thus defined satisfies (2.3.3). From (2.3.12) we have

$$-\log p_i^R + \log \frac{\sum_{i=1}^{n} u_i p_i^{R+\nu-1}}{\sum_{i=1}^{n} u_i p_i^{\nu}} \leq l_i$$

or

$$-\log \frac{p_i^R}{\left(\frac{\sum_{i=1}^{n} u_i p_i^{R+\nu-1}}{\sum_{i=1}^{n} u_i p_i^{\nu}}\right)} \leq -\log_D D^{-l_i}$$

$$(2.3.13) \qquad \frac{p_i^R}{\left(\frac{\sum_{i=1}^{n} u_i p_i^{R+\nu-1}}{\sum_{i=1}^{n} u_i p_i^{\nu}}\right)} \geq D^{-l_i}$$

Multiplying both sides by $\sum_{i=1}^{n} u_i p_i^{\nu-1}$ and summing over $i = 1, 2, ..., n$. We get (2.3.3). The last inequality in (2.3.12) gives

$$l_i < -\log p_i^R + \log \frac{\sum_{i=1}^{n} u_i p_i^{R+\nu-1}}{\sum_{i=1}^{n} u_i p_i^{\nu}} + 1$$

$$l_i < -\log p_i^R + \log \frac{\sum_{i=1}^{n} u_i p_i^{R+\nu-1}}{\sum_{i=1}^{n} u_i p_i^{\nu}} + \log_D D$$

or

$$l_i < -\log \frac{p_i^R}{\left(\frac{\sum_{i=1}^{n} u_i p_i^{R+\nu-1}}{\sum_{i=1}^{n} u_i p_i^{\nu}}\right)} + \log_D D$$

$$D^{-l_i} > \frac{p_i^R}{\left(\frac{\sum_{i=1}^{n} u_i p_i^{R+\nu-1}}{\sum_{i=1}^{n} u_i p_i^{\nu}}\right)} D^{-1}$$

Taking $0 < R < 1$ and raising both sides to the power $\frac{R-1}{R}$, we get

$$D^{-l_i\left(\frac{R-1}{R}\right)} < \left(\frac{\frac{p_i^R}{\sum_{i=1}^{n} u_i p_i^{R+\nu-1}}}{\sum_{i=1}^{n} u_i p_i^{\nu}}\right)^{\frac{R-1}{R}} D^{\frac{1-R}{R}}$$

Multiplying both sides by $\frac{u_i p_i^{\nu}}{\sum_{i=1}^{n} u_i p_i^{\upsilon}}$ and summing over $i = 1, 2, .., n$ and after simplifying, gives (2.3.9).

For $R > 1$, The proof follows along the similar lines.

2.4. Variable length error correcting codes and lower bounds

It is well known that a lower bound on the average length is obtained in terms of Shannon entropy [107] for instantaneous codes in noiseless channel (see Abramson [1]). Bernard and Sharma [19] studied variable length codes for noisy channels and presented some combinatorial bounds for these variable length, error correcting codes. Bernard and Sharma [20] obtained a lower bound on average length for variable length error correcting codes satisfying the criterion of promptness.

In coding theory, it is assumed that Q is a finite set of alphabets and there are D code characters. A codeword is defined as a finite sequence of code characters and a variable length code C of size n is a set of n codewords denoted by $c_1, c_2, ..., c_n$ with lengths $l_1, l_2, ..., l_n$ respectively. Without loss of generality it may be assumed that $l_1 \leq l_2 \leq ... \leq l_n$.

The channel, which is considerd here, is not noiseless. In otherwords, the codes considerd here are error correcting codes. The criterion for error correcting is defined in terms of a mapping α , which depends on the noise characteristic of the channel. This mapping α is called the error admissibility mapping. Given a codeword c and error admissibilty α, the set of codewords received over the channel when c was sent, denoted by $\alpha\left(c\right)$, is the error range of c.

Various kinds of error pattern can be described in terms of mapping α . In particular α may be defined as (Bernard and Sharma [19])

$$\alpha_e\left(c\right) = \{\underline{u} \mid w\left(c - \underline{u}\right) \leq e\}$$

where e is random substitution error and $w\left(c - \underline{u}\right)$ is a Hamming weight i.e., the number of non zero co- ordinates of $\left(c - \underline{u}\right)$. It can be easily verified (Bernard and Sharma [19]) that the number of sequences in $\alpha_e\left(c\right)$ denoted as $\left|\alpha_e\left(c\right)\right|$ is given by

$$|\alpha_e(c)| = \sum_{i=0}^{e} \binom{l}{i} (D-1)^i$$

where l is the length of code c.

We may assume that α_0 corresponds to the noiseless case. In otherwords if c is sent then c is received with respect to α_0. Moreover, it is clear that $|\alpha_e(c)|$ depends only on the length l of c when α and D are given. In noiseless coding, the class of uniquely decodable instantaneous codes are studied. It is known that these codes satisfy prefix property (Abramson [1]). In the same way Hartnett [56] studied variable length code over noisy channel, satisfying the prefix property in the range. These codes are called $\alpha-$ prompt codes. Such codes have the property that they can decode promptly.

Further, Bernard and Sharma [20] gave a combinatorial inequality that must necessarily be satisfied by codeword lengths of prompt codes. Two useful concepts, namely, segment decomposition and the effective range $r_\alpha(c_i)$ of codeword c_i of length l_i under error mapping α as the cartesian product of ranges of the segment are also given by Bernard and Sharma [19]. The number of sequences in effective range of c_i denoted by $|r_\alpha|_{l_i}$ depends only on α and l_i. It is given by

$$|r_\alpha|_{l_i} = |\alpha|_{l_1} |\alpha|_{l_2-l_1} \cdots \; |\alpha|_{l_i-l_{i-1}}$$

Moreover, Bernard and Sharma [19] obtained the following inequality.

Theorem (2.4.1). An $\alpha-$ prompt code with n codewords of length $l_1 \leq l_2 \leq \ldots \leq l_n$ satisfies the following inequality

(2.4.1) $$\sum_{i=1}^{n} |r_\alpha|_{l_i} D^{-l_i} \leq 1$$

Remark (2.4.1).

If the codes of constant length l are taken then the average inequality (2.4.1) reduces to Hamming sphere packing bound given by Hamming [54].

Remark (2.4.2).

If the channel is noiseless then the inequality (2.4.1) reduces to well known Kraft [80] inequality.

Also, Baig and Zaheeruddin [14] have obtained a lower bound on codeword length of order t of prompt codes using a quantity similar to Nath's [90] inaccuracy of order β.

Theorem (2.4.2): Let an $\alpha-$ prompt code encode the n messages $s_1, s_2, ..., s_n$ in to a code alphabet of D symbols and let the length of the corresponding to message s_i be l_i . Then the code length of order t, $L(t)$ shall satisfy the inequality

(2.4.2) $$L(t) \geq \frac{1}{1-\beta} \log \sum_{i=1}^{n} p_i q_i^{\beta-1} \left(|r_\alpha|_{l_i}\right)^{1-\beta}$$

with equality iff

(2.4.3) $$l_i = -\log \left(|r_\alpha|_{l_i}\right)^{-\beta} q_i^{\beta} + \log \sum_{i=1}^{n} p_i q_i^{\beta-1} \left(|r_\alpha|_{l_i}\right)^{1-\beta}$$

where $L(t) = \frac{1}{t} \log \sum_{i=1}^{n} p_i D^{tl_i}$

In the next section, lower bound have been obtained on generalized codeword length of order α and type β of "prompt" codes using a quantity similar to Tuteja and Bhaker [129] 'useful' inaccuracy measure of order α and type β. The result obtained here generalizes the result of Bernard and Sharma [20] and Baig and Zaheeruddin [14].

2.5. A Generalized lower bound on codeword length of order α and type β of 'prompt' codes.

Consider a generalized mean codeword length of order α and type β defined by

(2.5.1) $$L_\alpha^\beta(U) = \frac{1}{\left(2^{1-\alpha}-1\right)\sum_{i=1}^{n} p_i^\beta} \left[\left(\frac{\sum_{i=1}^{n} u_i^\beta p_i^\beta D^{l_i\left(\frac{1-\alpha}{\alpha}\right)}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta} \right)^\alpha - 1 \right], \alpha > 0\,(\neq 1), \beta > 0$$

It is easy to see that for $\alpha \to 1, \beta = 1$ and $u_i = 1 \ \forall \ i = 1, 2, ..., n$, (2.5.1) reduces to ordinary mean length $L = \sum_{i=1}^{n} l_i p_i$ given by Shannon [107].

Suppose that a person believe that the probaility of i^{th} event is q_i and the code with length l_i has been constructed accordingly. But contrary to his belief the true probability is p_i.

We will now obtain a lower bound of generalized mean codeword length $L_\alpha^\beta(U)$ under the condition

(2.5.2) $$\sum_{i=1}^{n} u_i^\beta p_i^\beta q_i^{-1} D^{-l_i} \left(|r_\alpha|_{l_i}\right) \leq \sum_{i=1}^{n} u_i^\beta p_i^\beta$$

Remark (2.5.1).

If $\beta = 1, u_i = 1 \ \forall \ i = 1, 2, ..., n$ in (2.5.2). It reduces to the inequality given by Baig and Zaheeruddin [14]

(2.5.3) $$\sum_{i=1}^{n} p_i q_i^{-1} \left(\left| r_\alpha \right|_{l_i} \right) D^{-l_i} \leq 1$$

Remark (2.5.2).

Moreover if $p_i = q_i \ \forall \ i = 1, 2, ..., n$, in (2.5.3). It reduces to the inequality given by Bernard and Sharma [19]

(2.5.4) $$\sum_{i=1}^{n} \left| r_\alpha \right|_{l_i} D^{-l_i} \leq 1$$

Remark (2.5.3).

For noiseless channel $\left| r_\alpha \right|_{l_i} = 1 \ \forall \ i = 1, 2, ..., n$. Moreover if

$\beta = 1, \ u_i = 1 \ \forall \ i = 1, 2, ..., n$. Then (2.5.2) reduces to the inequality given by Autar and Soni [7]

(2.5.5) $$\sum_{i=1}^{n} p_i q_i^{-1} D^{-l_i} \leq 1$$

Remark (2.5.4).

Moreover if $p_i = q_i \ \forall \ i = 1, 2, ..., n$. Then (2.5.5) reduces to Kraft [80] inequality

(2.5.6) $$\sum_{i=1}^{n} D^{-l_i} \leq 1$$

Theorem (2.5.1). Let an $\alpha-$ prompt code encode the n messages $s_1, s_2, ..., s_n$ into a code alphabet of D symbols and let the length of corresponding to message s_i be l_i Then the generalized codeword length of order α and type β, $L_\alpha^\beta (U)$ shall satisfy the inequality

(2.5.7) $$L_\alpha^\beta (U) \geq \frac{\sum_{i=1}^{n} u_i^\beta p_i^\beta \left(q_i^{\alpha-1} \left(\left| r_\alpha \right|_{l_i} \right)^{1-\alpha} - 1 \right)}{\left(2^{1-\alpha} - 1 \right) \sum_{i=1}^{n} p_i^\beta \overline{U}} \ , \alpha > 0 \left(\neq 1 \right), \beta > 0$$

where $\overline{U} = \sum_{i=1}^{n} u_i^\beta p_i^\beta$

equality holds in (2.5.7) iff

(2.5.8) $$l_i = -\log \left(\left| r_\alpha \right|_{l_i} \right)^{-\alpha} q_i^\alpha + \log \frac{\sum_{i=1}^{n} u_i^\beta p_i^\beta q_i^{\alpha-1} \left(\left| r_\alpha \right|_{l_i} \right)^{1-\alpha}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta}$$

Proof: By Holders inequality [116]

(2.5.9) $$\sum_{i=1}^{n} x_i y_i \geq \left(\sum_{i=1}^{n} x_i^p \right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} y_i^q \right)^{\frac{1}{q}}$$

for all $x_i, y_i > 0, \ i = 1, 2, .., n$ and $\frac{1}{p} + \frac{1}{q} = 1, p < 1 \left(\neq 0 \right), q < 0$ or $q <$

$1\,(\neq 0)\,, p < 0$. We see the equality holds iff there exists a positive constant c such that

(2.5.10) $\quad x_i^p = c y_i^q$

Setting

$$x_i = \left[\frac{u_i^\beta p_i^\beta}{\sum_{i=1}^n u_i^\beta p_i^\beta}\right]^{\frac{\alpha}{\alpha-1}} D^{-l_i}, \qquad y_i = \left[\frac{u_i^\beta p_i^\beta}{\sum_{i=1}^n u_i^\beta p_i^\beta}\right]^{\frac{1}{1-\alpha}} q_i^{-1}\left(|r_\alpha|_{l_i}\right)$$

$p = \frac{\alpha-1}{\alpha}$, $\;q = 1-\alpha\;$ in (2.5.9), we get

$$\frac{\sum_{i=1}^n u_i^\beta p_i^\beta D^{-l_i} q_i^{-1}\left(|r_\alpha|_{l_i}\right)}{\sum_{i=1}^n u_i^\beta p_i^\beta} \geq \left[\frac{\sum_{i=1}^n u_i^\beta p_i^\beta D^{l_i\left(\frac{1-\alpha}{\alpha}\right)}}{\sum_{i=1}^n u_i^\beta p_i^\beta}\right]^{\frac{\alpha}{\alpha-1}} \left[\frac{\sum_{i=1}^n u_i^\beta p_i^\beta q_i^{\alpha-1}\left(|r_\alpha|_{l_i}\right)^{1-\alpha}}{\sum_{i=1}^n u_i^\beta p_i^\beta}\right]^{\frac{1}{1-\alpha}}$$

using the inequality (2.5.2), we get

$$\left[\frac{\sum_{i=1}^n u_i^\beta p_i^\beta D^{l_i\left(\frac{1-\alpha}{\alpha}\right)}}{\sum_{i=1}^n u_i^\beta p_i^\beta}\right]^{\frac{\alpha}{1-\alpha}} \geq \left[\frac{\sum_{i=1}^n u_i^\beta p_i^\beta q_i^{\alpha-1}\left(|r_\alpha|_{l_i}\right)^{1-\alpha}}{\sum_{i=1}^n u_i^\beta p_i^\beta}\right]^{\frac{1}{1-\alpha}}$$

Let $0 < \alpha < 1$. Raising both sides to the power $(1-\alpha)$ and after suitable operations, we get

$$L_\alpha^\beta(U) \geq \frac{\sum_{i=1}^n u_i^\beta p_i^\beta \left(q_i^{\alpha-1}\left(|r_\alpha|_{l_i}\right)^{1-\alpha} - 1\right)}{(2^{1-\alpha}-1)\sum_{i=1}^n p_i^\beta \overline{U}} \quad ,\alpha > 0\,(\neq 1)\,, \beta > 0$$

2.6. Particular cases

(1) For noiseless channel $|r_\alpha|_{l_i} = 1$, (2.5.7) reduces to inequality

(2.6.1) $$L_\alpha^\beta(U) \geq \frac{\sum_{i=1}^n u_i^\beta p_i^\beta \left(q_i^{\alpha-1} - 1\right)}{(2^{1-\alpha}-1)\sum_{i=1}^n p_i^\beta \overline{U}}$$

where $\frac{\sum_{i=1}^n u_i^\beta p_i^\beta \left(q_i^{\alpha-1} - 1\right)}{(2^{1-\alpha}-1)\sum_{i=1}^n p_i^\beta}$ is a information measure given by Tuteja and Bhaker [129].

(2) For $\alpha \to 1,\ \beta = 1$, the inequality (2.5.7) reduces to

(2.6.2) $$L(U) \geq \frac{\sum_{i=1}^n u_i p_i \log\left(\frac{|r_\alpha|_{l_i}}{q_i}\right)}{\sum_{i=1}^n u_i p_i}$$

(3) For $u_i = 1 \quad \forall\ i = 1, 2, ..., n$, inequality (2.6.2) reduces to the inequality given by Baig and Zaherruddin [14]

(2.6.3) $$L \geq \sum_{i=1}^{n} p_i \log\left(\frac{|r_\alpha|_{l_i}}{q_i}\right)$$

(4) Moreover if $p_i = q_i \ \forall \ i = 1, 2, ..., n.,$ inequality (2.6.3) reduces to inequality given by Bernard and Sharma [20]

(2.6.4) $$L \geq \sum_{i=1}^{n} p_i \log\left(\frac{|r_\alpha|_{l_i}}{p_i}\right)$$

(5) For noiseless channel $|r_\alpha|_{l_i} = 1$, (2.6.3) reduces to inequality given by Kerridge [73]

(2.6.5) $$L \geq -\sum_{i=1}^{n} p_i \log q_i$$

Shannon's measure of information plays a very important role for measuring uncertainty in probability distributions and also for measuring diversity in plants and animals in Biology. But this measure does not deal with growth models other than exponential. Since there are families of distributions other than exponential and there are laws of population growth other than exponential, we can not confine ourselves to exponential families only and consequently, Shannon's measure may not be much applicable. Thus we need a parametric models which are suitable for all types of distribution.

Shannon [107] introduced the ordinary mean codeword length and established bounds in terms of entropy. In this chapter, bounds on generalized mean codeword length are obtained by considering parametric measures of information with utility distribution which are suitable for all types of of distributions. The bounds obtained in this chapter are not only new but also generalizes some well established results available in the literature of information theory.

3.1 Introduction

Let $P = (p_1, p_2, ..., p_n), 0 \le p_i \le 1, \sum_{i=n}^{n} p_i = 1$ be the probability distribution associated with a finite system of events $X = (x_1, x_2, ..., x_n)$ representing the realization of some experiment. The different events x_i depend upon the experimenters goal or upon some qualitative chracteristics of the physical system taken in to account; ascribe to each event x_i a non negative number $u_i\,(> 0)$ directly proportional to its importance and call u_i the utilitiy of the event x_i . Then the weighted entropy [17] of the experiment X is defined as

(3.1.1) $$H\,(P;U) = -\sum_{i=1}^{n} u_i p_i \log p_i$$

Now let us suppose that the experimenter asserts that the probability of the i^{th} outcome x_i is q_i , whereas the true probability is p_i, with $\sum_{i=1}^{n} p_i = 1 = \sum_{i=1}^{n} q_i$, Thus, we have two utility information schemes

(3.1.2) $$S = \begin{bmatrix} x_1 & x_2 & ... & x_n \\ p_1 & p_2 & ... & p_n \\ u_1 & u_2 & ... & u_n \end{bmatrix}, 0 \le p_i \le 1,\ u_i > 0, \sum_{i=1}^{n} p_i = 1$$

of a set of n events after an experiment, and

(3.1.3) $$S^{\star}=\begin{bmatrix} x_1 & x_2 & \ldots & x_n \\ q_1 & q_2 & \ldots & q_n \\ u_1 & u_2 & \ldots & u_n \end{bmatrix}, 0\leq q_i\leq 1,\ u_i>0, \sum_{i=1}^{n} q_i=1$$

of the same set of n events before the experiment.

In both the schemes (3.1.2) and (3.1.3) the utility distribution is the same because we assume that the utility u_i of an outcome x_i is independent of its probability of occurrence p_i, or predicted probability $q_i,\ u_i$ is only a 'utility' or value of the outcome x_i for an observer relative to some specified goal (refer to [87]).

The quantitaive- qualitative measure of inaccuracy [119] associated with the above schemes

(3.1.4) $$I\left(P,Q;U\right)=-\sum_{i=1}^{n} u_i p_i \log q_i$$

Guiasu and Picard [51] considerd the problem of encoding the letter output by the source (3.1.2) by means of a single letter prefix code with codewords $c_1, c_2, ..., c_n$ having length $l_1, l_2, ..., l_n$ satisfying Kraft [80] inequality

(3.1.5) $$\sum_{i=1}^{n} D^{-l_i}\leq 1$$

D being the size of the code alphabet. They defined the useful mean length $L\left(U\right)$ of the code as

(3.1.6) $$L\left(U\right)=\frac{\sum_{i=1}^{n} u_i p_i l_i}{\sum_{i=1}^{n} u_i p_i}$$

and obtained bounds for it.

Taneja and Tuteja [119] considerd the codeword mean length given in (3.1.6) and obtained bounds in terms of (3.1.4), under the condition

(3.1.7) $$\sum_{i=1}^{n} p_i q_i^{-1} D^{-l_i}\leq 1$$

D is the size of the code alphabet. It is easy to see that for $p_i=q_i\ \forall\ i=1,2,...,n$ (3.1.7) reduces to Kraft [80] inequality.

Longo [87], Gurdial and Pessoa [53], Autar and Khan [6], Jain and Tuteja [63], Taneja et al [120], Bhatia [22], Hooda and Bhaker [59], Singh, Kumar and Tuteja [118] and Khan et al [78] considerd the problem of information measures and used it studying the bounds.

In the next section, generalized 'useful' codeword mean length are considerd and bounds have been obtained in terms of generalized 'useful' inaccuracy measure of order α and type β. The beauty of these results is that it generalizes the results which exists in the literature of information theory and the measures considerd here are suitable for the distributions other than exponential. This work is published in "International journal of pure and applied Mathematics", Vol 32 (4), PP 467-474 (2006)(Baig and Rayees [8]). All the logarithms used in this chapter are with base D, where D is the size of the code alphabet.

3.2. Generalized measures of information and their bounds

Consider a function

(3.2.1) $$I_\alpha^\beta\left(P,Q;U\right)=\frac{1}{D^{\frac{\alpha-1}{\alpha}}-1}\left[1-\left(\frac{\sum\limits_{i=1}^{n}u_ip_i^\beta q_i^{\alpha-1}}{\sum\limits_{i=1}^{n}u_ip_i^\beta}\right)^{\frac{1}{\alpha}}\right]$$

where $\alpha>0\left(\neq 1\right),\beta>0,p_i\geq 0,\sum\limits_{i=1}^{n}p_i\leq 1,i=1,2,...,n.$ D is the size of the code alphabet.

Remark (3.2.1).

(1) When $\alpha\rightarrow 1,\beta=1$ and distribution is complete, the measure (3.2.1) reduces to measure of 'useful' inaccuracy given by Taneja and Tuteja [119].

(2) When $\beta=1,p_i=q_i\ \forall\ i=1,2,...,n,$ the measure (3.2.1) reduces to the measure given by Autar and Khan [6] as 'useful' information measure.

(3) When $\alpha\rightarrow 1,\beta=1$ and $p_i=q_i\ \forall\ i=1,2,...,n,$ the measure (3.2.1) reduces to the measure of 'useful' information for incomplete probability distribution given by Belis and Guiasu [17]. Further, when utility aspect of the scheme is ignored, the measure reduces to Shannon's [107] entropy.

(4) When the probability distribution is complete and the utility aspect of the scheme is ignored as well as $\alpha\rightarrow 1,\beta=1$. The measure (3.2.1) becomes the Kerridge's [73] measure of inaccuracy. We call (3.2.1) as generalized 'useful' inaccuracy measure of order α and type β for incomplete probability distribution.

Further, consider a generalized 'useful' codeword mean length credited with utilities and probabilities as

(3.2.2) $$L_\alpha^\beta(U) = \frac{1}{D^{\frac{\alpha-1}{\alpha}}-1}\left[1-\sum_{i=1}^{n} p_i^\beta \left(\frac{u_i}{\sum_{i=1}^{n} u_i p_i^\beta}\right)^{\frac{1}{\alpha}} D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)}\right]$$

where $\alpha > 0\,(\neq 1), \beta > 0, p_i \geq 0, \sum_{i=1}^{n} p_i \leq 1, i = 1, 2, ..., n$. D is the size of the code alphabet.

Remark (3.2.2)

(1) When $\alpha \to 1, \beta = 1$ the measure (3.2.2) reduces to 'useful' mean length $L(U)$ of the code, given by Guiasu and Picard [51].

(2) When the utility aspect of the scheme is ignored by taking $u_i = 1 \ \forall \ i = 1, 2, ..., n$ also $\sum_{i=1}^{n} p_i = 1$ and $\alpha \to 1, \beta = 1$, the mean length of the code (3.2.2) becomes optimal code length identical to Shannon [107].

Now, we find the bounds for $L_\alpha^\beta(U)$ in terms of $I_\alpha^\beta(P, Q; U)$ under the condition

(3.2.3) $$\sum_{i=1}^{n} p_i^\beta q_i^{-1} D^{-l_i} \leq 1$$

where D is the size of the code alphabet. It is easy to see that for $\beta = 1$ and $p_i = q_i \ \forall\, i = 1, 2, ..., n$. Inequality (3.2.3) reduces to Kraft [80] inequality.

Theorem (3.2.1). For all integers $D\,(D > 1)$. Let l_i satisfies the condition (3.2.3), then the generalized 'useful' codeword mean length satisfies

(3.2.4) $$L_\alpha^\beta(U) \geq I_\alpha^\beta(P, Q; U)$$

equality holds iff

(3.2.5) $$l_i = -\log\left(\frac{u_i q_i^\alpha}{\sum_{i=1}^{n} u_i p_i^\beta q_i^{\alpha-1}}\right)$$

Proof: By Holder's inequality [116]

(3.2.6) $$\sum_{i=1}^{n} x_i y_i \geq \left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} y_i^q\right)^{\frac{1}{q}}$$

for all $x_i, y_i > 0,\ i = 1, 2, .., n$ and $\frac{1}{p} + \frac{1}{q} = 1, p < 1\,(\neq 0), q < 0$ or $q < 1\,(\neq 0), p < 0$. We see the equality holds iff there exists a positive constant c such that

(3.2.7) $$x_i^p = c y_i^q$$

Making the substitution

$$x_i = p_i^{\frac{\alpha\beta}{\alpha-1}} \left(\frac{u_i}{\sum\limits_{i=1}^{n} u_i p_i^{\beta}} \right)^{\frac{1}{\alpha-1}} D^{-l_i}, \quad y_i = p_i^{\frac{\beta}{1-\alpha}} \left(\frac{u_i}{\sum\limits_{i=1}^{n} u_i p_i^{\beta}} \right)^{\frac{1}{1-\alpha}} q_i^{-1}$$

$p = \frac{\alpha-1}{\alpha}, \ q = 1-\alpha$

in (3.2.6), we get

$$\sum_{i=1}^{n} p_i^{\beta} q_i^{-1} D^{-l_i} \geq \left[\sum_{i=1}^{n} p_i^{\beta} \left(\frac{u_i}{\sum\limits_{i=1}^{n} u_i p_i^{\beta}} \right)^{\frac{1}{\alpha}} D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)} \right]^{\frac{\alpha}{\alpha-1}} \left[\sum_{i=1}^{n} p_i^{\beta} \left(\frac{u_i}{\sum\limits_{i=1}^{n} u_i p_i^{\beta}} \right) q_i^{\alpha-1} \right]^{\frac{1}{1-\alpha}}$$

using the inequality (3.2.3), we get

$$\left[\sum_{i=1}^{n} p_i^{\beta} \left(\frac{u_i}{\sum\limits_{i=1}^{n} u_i p_i^{\beta}} \right)^{\frac{1}{\alpha}} D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)} \right]^{\frac{\alpha}{1-\alpha}} \geq \left[\sum_{i=1}^{n} p_i^{\beta} \left(\frac{u_i}{\sum\limits_{i=1}^{n} u_i p_i^{\beta}} \right) q_i^{\alpha-1} \right]^{\frac{1}{1-\alpha}} \tag{3.2.8}$$

Let $0 < \alpha < 1$, raising both sides of (3.2.8) to the power $\frac{1-\alpha}{\alpha}$, we get

$$\left[\sum_{i=1}^{n} p_i^{\beta} \left(\frac{u_i}{\sum\limits_{i=1}^{n} u_i p_i^{\beta}} \right)^{\frac{1}{\alpha}} D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)} \right] \geq \left[\sum_{i=1}^{n} p_i^{\beta} \left(\frac{u_i}{\sum\limits_{i=1}^{n} u_i p_i^{\beta}} \right) q_i^{\alpha-1} \right]^{\frac{1}{\alpha}}$$

After making suitable operations we get (3.2.4) for $\left(D^{\frac{\alpha-1}{\alpha}} - 1\right) \neq 0$ when $\alpha \neq 1$. For $\alpha > 1$,The proof follows along the similar lines.

Theorem (3.2.2): For every code with lengths $l_1, l_2, ..., l_n$ satisfies the condition (3.2.3), $L_{\alpha}^{\beta}(U)$ can be made to satisfy the inequality

$$L_{\alpha}^{\beta}(U) < I_{\alpha}^{\beta}(P, Q; U) D^{\frac{1-\alpha}{\alpha}} + \frac{1 - D^{\frac{1-\alpha}{\alpha}}}{D^{\frac{\alpha-1}{\alpha}} - 1} \tag{3.2.9}$$

Proof: Let l_i be the positive integer satisfying the inequality

$$-\log \left(\frac{u_i q_i^{\alpha}}{\sum\limits_{i=1}^{n} u_i p_i^{\beta} q_i^{\alpha-1}} \right) \leq l_i < -\log \left(\frac{u_i q_i^{\alpha}}{\sum\limits_{i=1}^{n} u_i p_i^{\beta} q_i^{\alpha-1}} \right) + 1 \tag{3.2.10}$$

Consider the interval

$$\delta_i = \left[-\log \left(\frac{u_i q_i^{\alpha}}{\sum\limits_{i=1}^{n} u_i p_i^{\beta} q_i^{\alpha-1}} \right), \ -\log \left(\frac{u_i q_i^{\alpha}}{\sum\limits_{i=1}^{n} u_i p_i^{\beta} q_i^{\alpha-1}} \right) + 1 \right] \tag{3.2.11}$$

of length 1. In every δ_i, there lies exactly one positive integer l_i such that

$$0 < -\log\left(\frac{u_i q_i^{\alpha}}{\sum_{i=1}^{n} u_i p_i^{\beta} q_i^{\alpha-1}}\right) \leq l_i < -\log\left(\frac{u_i q_i^{\alpha}}{\sum_{i=1}^{n} u_i p_i^{\beta} q_i^{\alpha-1}}\right) + 1 \tag{3.2.12}$$

We will first show that the sequence $l_1, l_2, ..., l_n$ thus defined satisfies (3.2.3). From (3.2.12), we have

$$-\log\left(\frac{u_i q_i^{\alpha}}{\sum_{i=1}^{n} u_i p_i^{\beta} q_i^{\alpha-1}}\right) \leq l_i$$

or

$$\left(\frac{u_i q_i^{\alpha}}{\sum_{i=1}^{n} u_i p_i^{\beta} q_i^{\alpha-1}}\right) \geq D^{-l_i}$$

Multiply both sides by $p_i^{\beta} q_i^{-1}$ and summing over $i = 1, 2, ..., n$ we get (3.2.3). The last inequality of (3.2.12) gives

$$l_i < -\log\left(\frac{u_i q_i^{\alpha}}{\sum_{i=1}^{n} u_i p_i^{\beta} q_i^{\alpha-1}}\right) + 1$$

or

$$D^{l_i} < \left(\frac{u_i q_i^{\alpha}}{\sum_{i=1}^{n} u_i p_i^{\beta} q_i^{\alpha-1}}\right)^{-1} D$$

Let $0 < \alpha < 1$, raising both sides to the power $\left(\frac{1-\alpha}{\alpha}\right)$,we get

$$D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)} < \left(\frac{u_i q_i^{\alpha}}{\sum_{i=1}^{n} u_i p_i^{\beta} q_i^{\alpha-1}}\right)^{\frac{\alpha-1}{\alpha}} D^{\frac{1-\alpha}{\alpha}}$$

Multiply both sides by $p_i^{\beta}\left(\frac{u_i}{\sum_{i=1}^{n} u_i p_i^{\beta}}\right)^{\frac{1}{\alpha}}$ and summing over $i = 1, 2, ..., n$ and after suitable operations, we get

$$L_{\alpha}^{\beta}(U) < I_{\alpha}^{\beta}(P, Q; U)\, D^{\frac{1-\alpha}{\alpha}} + \frac{1 - D^{\frac{1-\alpha}{\alpha}}}{D^{\frac{\alpha-1}{\alpha}} - 1} \tag{3.2.13}$$

Again, bounds have been obtained by considering a more generalized inaccuracy measure of order α and type β. The main aim of studying this new function is that it generalizes some information measures already existing in the literature. This new function can be used for more complex distributions other than exponential.

Consider a function

(3.2.14) $$I_{\alpha}^{\beta}\left(P,Q;U\right)=\frac{1}{1-\alpha}\log\frac{\sum\limits_{i=1}^{n}u_i^{\beta}p_i^{\beta}q_i^{\alpha-1}}{\sum\limits_{i=1}^{n}u_i^{\beta}p_i^{\beta}},\quad \alpha>0\,(\neq 1)\,,\beta>0$$

Remark (3.2.3)

(1) When $\beta=1,$ (3.2.14) reduces to the measure given by Bhatia [23].

(2) When $\beta=1,\ p_i=q_i\ \forall\ i=1,2,...,n$. (3.2.14) reduces to measure given by Taneja , Hooda and Bhaker [120].

(3) When $p_i=q_i\ \ \forall\ i=1,2,...,n$. (3.2.14) reduces to the measure given by Hooda and Bhaker [59], further it reduces to Renyi's [104] entropy when $\beta=1,\ u_i=1\ \forall\ i=1,2,...,n$.

Further, we define a parametric codeword mean length credited with utilities and probabilities as

(3.2.15) $$L_u^{\beta}\left(t\right)=\frac{1}{t}\log\frac{\sum\limits_{i=1}^{n}u_i^{\beta}p_i^{\beta}D^{tl_i}}{\sum\limits_{i=1}^{n}u_i^{\beta}p_i^{\beta}},\quad -1<t<\infty,t\neq 0,\beta>0$$

Remark (3.2.4)

(1) When $\beta=1,t\to 0,$ (3.2.15) reduces to $L\left(U\right)$ given in (3.1.6).

Now, we establish a result that in a way, gives a characterization of $I_{\alpha}^{\beta}\left(P,Q;U\right)$, under the condition

(3.2.16) $$\sum_{i=1}^{n}u_i^{\beta}p_i^{\beta}q_i^{-1}D^{-l_i}\leq\sum_{i=1}^{n}u_i^{\beta}p_i^{\beta}$$

which is generalization of Kraft [80] inequality. Also D is the size of the code alphabet.

Theorem (3.2.3): For all integers $D\left(D>1\right)$. Let l_i satisfies the condition (3.2.16), then the generalized 'useful' codeword mean length satisfies

(3.2.17) $$L_u^{\beta}\left(t\right)\geq\ I_{\alpha}^{\beta}\left(P,Q;U\right)$$

where $\alpha=\frac{1}{1+t}$, equality holds iff

(3.2.18) $$l_i=-\log q_i^{\alpha}+\log\frac{\sum\limits_{i=1}^{n}u_i^{\beta}p_i^{\beta}q_i^{\alpha-1}}{\sum\limits_{i=1}^{n}u_i^{\beta}p_i^{\beta}}$$

Proof: By Holder's inequality [116]

(3.2.19) $$\sum_{i=1}^{n} x_i y_i \geq \left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} y_i^q\right)^{\frac{1}{q}}$$

for all $x_i, y_i > 0,\ i = 1, 2, .., n$ and $\frac{1}{p} + \frac{1}{q} = 1, p < 1\,(\neq 0)\,, q < 0$ or $q < 1\,(\neq 0)\,, p < 0$. We see the equality holds iff there exists a positive constant c such that

(3.2.20) $$x_i^p = c y_i^q$$

Making the substitution

$$x_i = \left[\frac{u_i^\beta p_i^\beta}{\sum_{i=1}^{n} u_i^\beta p_i^\beta}\right]^{\frac{-1}{t}} D^{-l_i},\ y_i = \left[\frac{u_i^\beta p_i^\beta q_i^{\alpha-1}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta}\right]^{\frac{1}{1-\alpha}}$$

$$p = -t,\ \ q = 1 - \alpha$$

in (3.2.19), we get

$$\frac{\sum_{i=1}^{n} u_i^\beta p_i^\beta q_i^{-1} D^{-l_i}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta} \geq \left[\frac{\sum_{i=1}^{n} u_i^\beta p_i^\beta D^{l_i t}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta}\right]^{\frac{-1}{t}} \left[\frac{\sum_{i=1}^{n} u_i^\beta p_i^\beta q_i^{\alpha-1}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta}\right]^{\frac{1}{1-\alpha}}$$

using the inequality (3.2.16), we get

$$\left[\frac{\sum_{i=1}^{n} u_i^\beta p_i^\beta D^{l_i t}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta}\right]^{\frac{1}{t}} \geq \left[\frac{\sum_{i=1}^{n} u_i^\beta p_i^\beta q_i^{\alpha-1}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta}\right]^{\frac{1}{1-\alpha}}$$

Taking logarithms to both sides with base D, we obtain (3.2.17).

Theorem (3.2.4). For every code with lengths $l_1, l_2, ..., l_n$ satisfies the condition (3.2.16), $L_u^\beta(t)$ can be made to satisfy the inequality

(3.2.21) $$L_u^\beta(t) < I_\alpha^\beta(P, Q; U) + 1$$

Proof: Let l_i be the positive integer satisfying the inequality

(3.2.22) $$-\log q_i^\alpha + \log \frac{\sum_{i=1}^{n} u_i^\beta p_i^\beta q_i^{\alpha-1}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta} \leq l_i < -\log q_i^\alpha + \log \frac{\sum_{i=1}^{n} u_i^\beta p_i^\beta q_i^{\alpha-1}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta} + 1$$

Consider the interval

(3.2.23) $$\delta_i = \left[-\log q_i^\alpha + \log \frac{\sum_{i=1}^{n} u_i^\beta p_i^\beta q_i^{\alpha-1}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta},\ \ -\log q_i^\alpha + \log \frac{\sum_{i=1}^{n} u_i^\beta p_i^\beta q_i^{\alpha-1}}{\sum_{i=1}^{n} u_i^\beta p_i^\beta} + 1\right]$$

of length 1. In every δ_i , there lies exactly one positive integer l_i such that

$$(3.2.24) \quad 0 < -\log q_i^{\alpha} + \log \frac{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta} q_i^{\alpha-1}}{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta}} \leq l_i < -\log q_i^{\alpha} + \log \frac{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta} q_i^{\alpha-1}}{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta}} + 1$$

We will first show that the sequence $l_1, l_2, ..., l_n$ thus defined satisfy (3.2.16). From (3.2.24) we have

$$-\log q_i^{\alpha} + \log \frac{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta} q_i^{\alpha-1}}{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta}} \leq l_i$$

or

$$q_i^{-\alpha} \frac{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta} q_i^{\alpha-1}}{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta}} \leq D^{l_i}$$

$$q_i^{\alpha} \frac{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta}}{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta} q_i^{\alpha-1}} \geq D^{-l_i}$$

Multiply both sides by $u_i^{\beta} p_i^{\beta} q_i^{-1}$ and summing over $i = 1, 2, ..., n$ we get (3.2.16). The last inequality of (3.2.24) gives

$$l_i < -\log q_i^{\alpha} + \log \frac{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta} q_i^{\alpha-1}}{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta}} + 1$$

or

$$D^{l_i} < q_i^{-\alpha} \frac{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta} q_i^{\alpha-1}}{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta}} D$$

Raising both sides to the power t and multiplying both sides by $\frac{u_i^{\beta} p_i^{\beta}}{\sum_{i=1}^{n} u_i^{\beta} p_i^{\beta}}$ and also summing over $i = 1, 2, ...n$, and simplifying, we get (3.2.21).

In the next section, bounds have been obtained by considering another type of inaccuracy measure of order α and type β. This work has been published in "Sarajevo Journal of Mathematics", Vol 3(1), PP 137-143 (2007)(Baig and Rayees [12]). The function considerd here is also suitable for the distributions other than exponential.

3.3. Noiseless coding theorems of inaccuracy measure of order α and type β.

Consider a function

$$(3.3.1) \qquad I_{\alpha}^{\beta}(P, Q; U) = \frac{1}{1-\alpha} \log \left[\frac{\sum_{i=1}^{n} u_i p_i^{\beta} q_i^{\beta(\alpha-1)}}{\sum_{i=1}^{n} u_i p_i^{\beta}} \right], \alpha > 0 \, (\neq 1), \beta > 0$$

Remark (3.3.1)

(1) When $\beta = 1$, (3.3.1) reduces to 'useful' information measure of order α due to Bhatia [23].

(2) When $\beta = 1, u_i = 1 \ \forall \ i = 1, 2, ..., n$. (3.3.1) reduces to the inaccuracy measure given by Nath [90], further it reduces to Renyi's [104] entropy by taking $p_i = q_i \quad \forall \ i = 1, 2, ..., n$.

(3) When $\beta = 1, u_i = 1 \ \forall \ i = 1, 2, ..., n$ and $\alpha \to 1$. Then (3.3.1) reduces to the measure due to Kerridge [73].

(4) When $u_i = 1 \ \forall \ i = 1, 2, ..., n$ and $p_i = q_i \quad \forall \ i = 1, 2, ..., n$. The measure (3.3.1) becomes the entropy for the β power distribution derived from P studied by Roy [105]. We call $I_\alpha^\beta (P, Q; U)$ in (3.3.1) the generalized 'useful' inaccuracy measure of order α and type β.

Further, we define a paramteric codeword mean length credited with utilities and probabilities as

(3.3.2) $$L_\beta^t (U) = \frac{1}{t} \log \left[\frac{\sum_{i=1}^{n} u_i^{t+1} p_i^\beta D^{tl_i}}{\left(\sum_{i=1}^{n} u_i p_i^\beta \right)^{t+1}} \right], \quad t > -1, t \neq 0, \beta > 0$$

Remark (3.3.2).

(1) When $\beta = 1, L_\beta^t (U)$ in (3.3.2) reduces to 'useful' mean length $L^t (U)$ of the code given by Bhatia [23].

(2) When $\beta = 1, u_i = 1 \ \forall \ i = 1, 2, ..., n$, $L_\beta^t (U)$ in (3.3.2) reduces to the code mean length given by Campbell [29].

(3) When $\beta = 1, u_i = 1 \ \forall \ i = 1, 2, ..., n$ and $\alpha \to 1$. $L_\beta^t (U)$ in (3.3.2) reduces to the optimal code length identical to Shannon [107] .

(4) When $u_i = 1 \ \forall \ i = 1, 2, ..., n$, $L_\beta^t (U)$ in (3.3.2) reduces to the codeword mean length given by Khan and Haseen [74].

Now, we find the bounds for $L_\beta^t (U)$ in terms of $I_\alpha^\beta (P, Q; U)$ under the condition

(3.3.3) $$\sum_{i=1}^{n} p_i^\beta q_i^{-\beta} D^{-l_i} \leq 1$$

where D is the size of the code alphabet. Also (3.3.3) is a generalization of Kraft [80] inequality.

Theorem (3.3.1). For every code whose lengths $l_1, l_2, ..., l_n$ satisfies the condition

(3.3.3). Then the code mean length satisfies

(3.3.4) $$L_{\beta}^{t}(U) \geq \; I_{\alpha}^{\beta}(P,Q;U)$$

where $\alpha = \frac{1}{1+t}$, equality holds iff

(3.3.5) $$l_i = -\log \frac{u_i q_i^{\alpha\beta}}{\sum\limits_{i=1}^{n} u_i p_i^{\beta} q_i^{\beta(\alpha-1)}}$$

Proof: By Holder's inequality [116]

(3.3.6) $$\sum_{i=1}^{n} x_i y_i \geq \left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} y_i^q\right)^{\frac{1}{q}}$$

for all $x_i, y_i > 0,\ i = 1, 2, .., n$ and $\frac{1}{p} + \frac{1}{q} = 1, p < 1\,(\neq 0)\,, q < 0$ or $q < 1\,(\neq 0)\,, p < 0$. We see the equality holds iff there exists a positive constant c such that

(3.3.7) $$x_i^p = c y_i^q$$

Making the substitution

$$x_i = u_i^{-\left(\frac{t+1}{t}\right)} p_i^{-\frac{\beta}{t}} \left(\frac{1}{\sum\limits_{i=1}^{n} u_i p_i^{\beta}}\right)^{-\frac{1+t}{t}} D^{-l_i}$$

$$y_i = u_i^{\frac{t+1}{t}} p_i^{\beta\left(\frac{1+t}{t}\right)} \left(\frac{1}{\sum\limits_{i=1}^{n} u_i p_i^{\beta}}\right)^{\frac{1+t}{t}} q_i^{-\beta}$$

$$p = -t, \quad q = \frac{t}{1+t}$$

in (3.3.6) and using (3.3.3), we get

$$\left[\frac{\sum\limits_{i=1}^{n} u_i^{t+1} p_i^{\beta} D^{l_i t}}{\left(\sum\limits_{i=1}^{n} u_i p_i^{\beta}\right)^{1+t}}\right]^{\frac{1}{t}} \geq \left[\frac{\sum\limits_{i=1}^{n} u_i p_i^{\beta} q_i^{\beta(\alpha-1)}}{\sum\limits_{i=1}^{n} u_i p_i^{\beta}}\right]^{\frac{1+t}{t}}$$

Taking logarithms to both sides with base D, we obtain (3.3.4).

Theorem (3.3.2) . For every code with lengths $l_1, l_2, ..., l_n$ satisfies the condition (3.3.3), $L_{\beta}^{t}(t)$ can be made to satisfy the inequality

(3.3.8) $$L_{\beta}^{t}(U) < I_{\alpha}^{\beta}(P,Q;U) + 1$$

Proof: Let l_i be the positive integer satisfying

(3.3.9) $$-\log \frac{u_i q_i^{\alpha\beta}}{\sum\limits_{i=1}^{n} u_i p_i^{\beta} q_i^{\beta(\alpha-1)}} \leq l_i < -\log \frac{u_i q_i^{\alpha\beta}}{\sum\limits_{i=1}^{n} u_i p_i^{\beta} q_i^{\beta(\alpha-1)}} + 1$$

Consider the interval

$$(3.3.10) \qquad \delta_i = \left[-\log \frac{u_i q_i^{\alpha\beta}}{\sum_{i=1}^{n} u_i p_i^{\beta} q_i^{\beta(\alpha-1)}}, \; -\log \frac{u_i q_i^{\alpha\beta}}{\sum_{i=1}^{n} u_i p_i^{\beta} q_i^{\beta(\alpha-1)}} + 1 \right]$$

of length 1. In every δ_i, there lies exactly one positive integer l_i such that

$$(3.3.11) \qquad 0 < -\log \frac{u_i q_i^{\alpha\beta}}{\sum_{i=1}^{n} u_i p_i^{\beta} q_i^{\beta(\alpha-1)}} \le l_i < -\log \frac{u_i q_i^{\alpha\beta}}{\sum_{i=1}^{n} u_i p_i^{\beta} q_i^{\beta(\alpha-1)}} + 1$$

We will first show that the sequence $l_1, l_2, ..., l_n$ thus defined satisfies (3.3.3). From (3.3.11) we have

$$-\log \frac{u_i q_i^{\alpha\beta}}{\sum_{i=1}^{n} u_i p_i^{\beta} q_i^{\beta(\alpha-1)}} \le l_i$$

or

$$\frac{u_i q_i^{\alpha\beta}}{\sum_{i=1}^{n} u_i p_i^{\beta} q_i^{\beta(\alpha-1)}} \ge D^{-l_i}$$

Multiply both sides by $p_i^{\beta} q_i^{-\beta}$ and summing over $i = 1, 2, ..., n$, we get (3.3.3). The last inequality in (3.3.11) gives

$$l_i < -\log \frac{u_i q_i^{\alpha\beta}}{\sum_{i=1}^{n} u_i p_i^{\beta} q_i^{\beta(\alpha-1)}} + 1$$

or

$$D^{l_i t} < \left(\frac{u_i q_i^{\alpha\beta}}{\sum_{i=1}^{n} u_i p_i^{\beta} q_i^{\beta(\alpha-1)}} \right)^{-t} D^t$$

Multiplying both sides by $\frac{u_i^{t+1} p_i^{\beta}}{\left(\sum_{i=1}^{n} u_i p_i^{\beta}\right)^{t+1}}$ and summing over $i = 1, 2, ..., n.$ we get

$$\frac{\sum_{i=1}^{n} u_i^{t+1} p_i^{\beta} D^{t l_i}}{\left(\sum_{i=1}^{n} u_i p_i^{\beta}\right)^{t+1}} < \left[\frac{\sum_{i=1}^{n} u_i p_i^{\beta} q_i^{\beta(\alpha-1)}}{\sum_{i=1}^{n} u_i p_i^{\beta}} \right]^{t+1} D^t$$

Taking logarithms to both sides with base D and then dividing both sides by t, we obtain (3.3.8).

3.4. Some results on weighted parametric information measures.

In this section, two generalized measures of information are considerd and their bounds are obtained. The results obtained by considering first measure has been presented in the 2nd J&K science congress held in University of Kashmir, Srinagar (2006) (Baig and Rayees [9]).

Consider a 'useful' inaccuracy measure of order β given by Om Parkash [93]

(3.4.1) $$I_n^\beta(P,Q;U) = \frac{1}{2^{-\beta}-1}\left[\sum_{i=1}^{n} u_i p_i \left(q_i^\beta - 1\right)\right], \beta \neq 0$$

Remark (3.4.1).

(1) When $\beta \to 0$, $I_n^\beta(P,Q;U)$ reduces to 'useful' inaccuracy measure given by Taneja and Tuteja [119].

(2) When $\beta \to 0$ and $u_i = 1 \ \forall \ i = 1,2,...,n$, $I_n^\beta(P,Q;U)$ reduces to measure of inaccuarcy given by Kerridge [73].

Further, consider a parametric codeword mean length

(3.4.2) $$L^\beta(U) = \frac{1}{2^{-\beta}-1}\left[\left(\frac{\sum_{i=1}^{n} u_i p_i D^{-l_i\left(\frac{\beta}{\beta+1}\right)}}{\sum_{i=1}^{n} u_i p_i}\right)^{\beta+1} - 1\right], \quad \beta \neq 0$$

Remark (3.4.2):

(1) When $\beta \to 0$, $L^\beta(U)$ reduces to the 'useful' codeword mean length $L(U)$ given by Guiasu and Picard [51].

(2) When $\beta \to 0$ and $u_i = 1 \ \forall \ i = 1,2,...,n$, $L^\beta(U)$ becomes the optimal code length defined by Shannon [107] .

In the following theorem, we obtain lower bound for $L^\beta(U)$ in terms of $I_n^\beta(P,Q;U)$ under the condition

(3.4.3) $$\sum_{i=1}^{n} u_i p_i q_i^{-1} D^{-l_i} \leq \sum_{i=1}^{n} u_i p_i$$

Theorem (3.4.1). If $l_1, l_2, ..., l_n$ be the lengths of the code satisfying the inequality (3.4.3). Then the mean codeword length satisfies

(3.4.4) $$L^\beta(U) \geq \frac{I_n^\beta(P,Q;U)}{\overline{U}}, \quad \beta \neq 0$$

where $\overline{U} = \sum_{i=1}^{n} u_i p_i$, equality holds iff

(3.4.5) $$l_i = -\log q_i^{\beta+1} + \log \frac{\sum_{i=1}^{n} u_i p_i q_i^\beta}{\sum_{i=1}^{n} u_i p_i}$$

Proof: By Holder's inequality [116]

(3.4.6) $$\sum_{i=1}^{n} x_i y_i \geq \left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} y_i^q\right)^{\frac{1}{q}}$$

for all $x_i, y_i > 0,\ i = 1, 2, .., n$ and $\frac{1}{p} + \frac{1}{q} = 1, p < 1\,(\neq 0)\,, q < 0$ or $q < 1\,(\neq 0)\,, p < 0$. We see the equality holds iff there exists a positive constant c such that

(3.4.7) $\qquad x_i^p = c y_i^q$

Making the substitution

$$x_i = \left[\frac{u_i p_i}{\sum\limits_{i=1}^{n} u_i p_i}\right]^{\frac{\beta+1}{\beta}} D^{-l_i},\ y_i = \left[\frac{u_i p_i}{\sum\limits_{i=1}^{n} u_i p_i}\right]^{-\frac{1}{\beta}} q_i^{-1}$$

$$p = \tfrac{\beta}{\beta+1},\quad q = -\beta$$

in (3.4.6), we get

$$\frac{\sum\limits_{i=1}^{n} u_i p_i q_i^{-1} D^{-l_i}}{\sum\limits_{i=1}^{n} u_i p_i} \geq \left[\frac{\sum\limits_{i=1}^{n} u_i p_i D^{-l_i\left(\frac{\beta}{\beta+1}\right)}}{\sum\limits_{i=1}^{n} u_i p_i}\right]^{\frac{\beta+1}{\beta}} \left[\frac{\sum\limits_{i=1}^{n} u_i p_i q_i^{\beta}}{\sum\limits_{i=1}^{n} u_i p_i}\right]^{-\frac{1}{\beta}}$$

using the inequality (3.4.3), we get

(3.4.8) $$\left[\frac{\sum\limits_{i=1}^{n} u_i p_i D^{-l_i\left(\frac{\beta}{\beta+1}\right)}}{\sum\limits_{i=1}^{n} u_i p_i}\right]^{-\frac{\beta+1}{\beta}} \geq \left[\frac{\sum\limits_{i=1}^{n} u_i p_i q_i^{\beta}}{\sum\limits_{i=1}^{n} u_i p_i}\right]^{-\frac{1}{\beta}}$$

For $\beta > 0$, raising both sides of (3.4.8) to the power $(-\beta)$, we get

(3.4.9) $$\left[\frac{\sum\limits_{i=1}^{n} u_i p_i D^{-l_i\left(\frac{\beta}{\beta+1}\right)}}{\sum\limits_{i=1}^{n} u_i p_i}\right]^{\beta+1} \leq \left[\frac{\sum\limits_{i=1}^{n} u_i p_i q_i^{\beta}}{\sum\limits_{i=1}^{n} u_i p_i}\right]$$

Since $2^{-\beta} - 1 < 0$ for $\beta > 0$, a simple manipulation proves (3.4.4) for $\beta > 0$. The proof for $\beta < 0$ follows on the same lines.

Theorem (3.4.2). For every code with lengths $l_1, l_2, ..., l_n$ satisfies the condition (3.4.3), $L^{\beta}(U)$ can be made to satisfy the inequality

(3.4.10) $\qquad L^{\beta}(U) < \frac{I_n^{\beta}(P,Q;U) D^{-\beta}}{\overline{U}} + \frac{D^{-\beta}-1}{2^{-\beta}-1},\quad \beta \neq 0$

Proof: Let l_i be the positive integer satisfying the inequality

(3.4.11) $$-\log q_i^{\beta+1} + \log \frac{\sum\limits_{i=1}^{n} u_i p_i q_i^{\beta}}{\sum\limits_{i=1}^{n} u_i p_i} \leq l_i < -\log q_i^{\beta+1} + \log \frac{\sum\limits_{i=1}^{n} u_i p_i q_i^{\beta}}{\sum\limits_{i=1}^{n} u_i p_i} + 1$$

Consider the interval

(3.4.12) $$\delta_i = \left[-\log q_i^{\beta+1} + \log \frac{\sum_{i=1}^{n} u_i p_i q_i^{\beta}}{\sum_{i=1}^{n} u_i p_i},\ \ -\log q_i^{\beta+1} + \log \frac{\sum_{i=1}^{n} u_i p_i q_i^{\beta}}{\sum_{i=1}^{n} u_i p_i} + 1\right]$$

of length 1. In every δ_i, there lies exactly one positive integer l_i such that

(3.4.13) $$0 < -\log q_i^{\beta+1} + \log \frac{\sum_{i=1}^{n} u_i p_i q_i^{\beta}}{\sum_{i=1}^{n} u_i p_i} \leq l_i < -\log q_i^{\beta+1} + \log \frac{\sum_{i=1}^{n} u_i p_i q_i^{\beta}}{\sum_{i=1}^{n} u_i p_i} + 1$$

We will first show that sequence $\{l_1, l_2, ..., l_n\}$, thus defined satisfies (3.4.3), we have

$$-\log q_i^{\beta+1} + \log \frac{\sum_{i=1}^{n} u_i p_i q_i^{\beta}}{\sum_{i=1}^{n} u_i p_i} \leq l_i$$

or

$$-\log q_i^{\beta+1} + \log \frac{\sum_{i=1}^{n} u_i p_i q_i^{\beta}}{\sum_{i=1}^{n} u_i p_i} \leq -\log_D D^{-l_i}$$

(3.4.14) $$\frac{q_i^{\beta+1}}{\left(\frac{\sum_{i=1}^{n} u_i p_i q_i^{\beta}}{\sum_{i=1}^{n} u_i p_i}\right)} \geq D^{-l_i}$$

Multiplying both sides of (3.4.14) by $u_i p_i q_i^{-1}$ and summing over $i = 1, 2, ..., n$, we get (3.4.3). The last inequality of (3.4.13) gives

$$l_i < -\log q_i^{\beta+1} + \log \frac{\sum_{i=1}^{n} u_i p_i q_i^{\beta}}{\sum_{i=1}^{n} u_i p_i} + 1$$

(3.4.15) $$D^{l_i} < q_i^{-(\beta+1)} \frac{\sum_{i=1}^{n} u_i p_i q_i^{\beta}}{\sum_{i=1}^{n} u_i p_i} D$$

For $\beta > 0$, raising both sides of (3.4.1.5) to the power $-\left(\frac{\beta}{\beta+1}\right)$, we get

(3.4.16) $$D^{-l_i\left(\frac{\beta}{\beta+1}\right)} > q_i^{\beta} \left(\frac{\sum_{i=1}^{n} u_i p_i q_i^{\beta}}{\sum_{i=1}^{n} u_i p_i}\right)^{-\frac{\beta}{\beta+1}} D^{-\frac{\beta}{\beta+1}}$$

Multiplying both sides of (3.4.16) by $\frac{u_i p_i}{\sum_{i=1}^{n} u_i p_i}$ and summing over $i = 1, 2, ..., n$, then raising both sides to the power $(\beta + 1)$, we get

(3.4.17) $$\left[\frac{\sum_{i=1}^{n} u_i p_i D^{-l_i\left(\frac{\beta}{\beta+1}\right)}}{\sum_{i=1}^{n} u_i p_i}\right]^{\beta+1} > \frac{\sum_{i=1}^{n} u_i p_i q_i^{\beta}}{\sum_{i=1}^{n} u_i p_i} D^{-\beta}$$

Since $2^{-\beta}-1<0$ for $\beta>0$, a simple manipulation in (3.4.17) gives

$$L^{\beta}(U)<\frac{I_n^{\beta}(P,Q;U)D^{-\beta}}{\overline{U}}+\frac{D^{-\beta}-1}{2^{-\beta}-1},\quad \beta\neq 0$$

For $\beta<0$, The proof follows on the same lines.

Now, we consider weighted paramteric measure involving utilities and the bounds have been obtained.

Consider a function

(3.4.18) $$I_{\alpha}^{\beta}(P,Q;U)=\frac{1}{1-D^{(1-\alpha)\beta}}\left[1-\left(\frac{\sum_{i=1}^{n}u_ip_iq_i^{\alpha-1}}{\sum_{i=1}^{n}u_ip_i}\right)^{\beta}\right],\ \alpha>0\,(\neq 1)\,,\beta\neq 0$$

Remark (3.4.3).

(1) When $\beta\to 0$, (3.4.18) reduces to the measure given by Bhatia [23].

(2) When $\beta=\frac{1}{\alpha}$, (3.4.18) reduces to the measure given by Bhatia [22] .

(3) When $\beta=\frac{1}{\alpha},\ p_i=q_i\ \forall\ i=1,2,...,n$, (3.4.18) reduces to the measure given by Autar and Khan [6], which can be further reduced to the entropy given by Shannon [107] by taking $\alpha\to 1$ and $u_i=1\ \ \forall\ i=1,2,...,n.$

Further, consider a parametric 'useful' codeword mean length

(3.4.19) $$L_{\alpha}^{\beta}(U)=\frac{1}{1-D^{(1-\alpha)\beta}}\left[1-\left\{\sum_{i=1}^{n}p_i\left(\frac{u_i}{\sum_{i=1}^{n}u_ip_i}\right)^{\frac{1}{\alpha}}D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)}\right\}^{\alpha\beta}\right]$$

where $\alpha>0\,(\neq 1)\,,\beta\neq 0.$

Remark (3.4.4).

(1) When $\beta\to 0$, (3.4.19) reduces to the codeword mean length given by Gurdial and Pessoa [53].

(2) When $\beta=\frac{1}{\alpha}$, (3.4.19) reduces to the codeword mean length given by Autar and Khan [6], which can be further reduced to ordinary codeword mean length given by Shannon [107] by taking $\alpha\to 1,\ u_i=1\ \ \forall\ i=1,2,...,n.$

Now we find the bounds of $L_{\alpha}^{\beta}(U)$ in terms of $I_{\alpha}^{\beta}(P,Q;U)$ under the condition

(3.4.20) $$\sum_{i=1}^{n}p_iq_i^{-1}D^{-l_i}\leq 1$$

where D is the size of the code alphabet.

Theorem (3.4.3). For every code with lengths $l_1, l_2, ..., l_n$ satisfies the condition (3.4.20). Then generalized codeword mean length satisfies

(3.4.21) $$L_\alpha^\beta(U) \geq I_\alpha^\beta(P, Q; U)$$

equality holds iff

(3.4.22) $$l_i = -\log \frac{u_i q_i^\alpha}{\sum_{i=1}^n u_i p_i q_i^{\alpha-1}}$$

Proof: By Holder's inequality [116]

(3.4.23) $$\sum_{i=1}^n x_i y_i \geq \left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^n y_i^q\right)^{\frac{1}{q}}$$

for all $x_i, y_i > 0,\ i = 1, 2, .., n$ and $\frac{1}{p} + \frac{1}{q} = 1, p < 1\,(\neq 0), q < 0$ or $q < 1\,(\neq 0), p < 0$. We see the equality holds iff there exists a positive constant c such that

(3.4.24) $$x_i^p = c y_i^q$$

Making the substitution

$$p = \frac{\alpha-1}{\alpha}, \quad q = 1-\alpha$$

$$x_i = p_i^{\frac{\alpha}{\alpha-1}} \left(\frac{u_i}{\sum_{i=1}^n u_i p_i}\right)^{\frac{1}{\alpha-1}} D^{-l_i}, \quad y_i = p_i^{\frac{1}{1-\alpha}} \left(\frac{u_i}{\sum_{i=1}^n u_i p_i}\right)^{\frac{1}{1-\alpha}} q_i^{-1}$$

in (3.4.23) and using (3.4.20), we get

(3.4.25) $$\left[\sum_{i=1}^n p_i \left(\frac{u_i}{\sum_{i=1}^n u_i p_i}\right)^{\frac{1}{\alpha}} D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)}\right]^{\frac{\alpha}{1-\alpha}} \geq \left[\frac{\sum_{i=1}^n u_i p_i q_i^{\alpha-1}}{\sum_{i=1}^n u_i p_i}\right]^{\frac{1}{1-\alpha}}$$

Case 1: For $\alpha > 1, \beta > 0$, equation (3.4.25) becomes

$$\left[\frac{\sum_{i=1}^n u_i p_i q_i^{\alpha-1}}{\sum_{i=1}^n u_i p_i}\right]^\beta \geq \left[\sum_{i=1}^n p_i \left(\frac{u_i}{\sum_{i=1}^n u_i p_i}\right)^{\frac{1}{\alpha}} D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)}\right]^{\alpha\beta}$$

Since $1 - D^{(1-\alpha)\beta} > 0$, we have

$$L_\alpha^\beta(U) \geq I_\alpha^\beta(P, Q; U)$$

Case 2: For $\alpha > 1, \beta < 0$, equation (3.4.25) becomes

$$\left[\frac{\sum_{i=1}^{n} u_i p_i q_i^{\alpha-1}}{\sum_{i=1}^{n} u_i p_i}\right]^{\beta} \leq \left[\sum_{i=1}^{n} p_i \left(\frac{u_i}{\sum_{i=1}^{n} u_i p_i}\right)^{\frac{1}{\alpha}} D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)}\right]^{\alpha\beta}$$

Since $1 - D^{(1-\alpha)\beta} < 0$, we have

$$L_{\alpha}^{\beta}(U) \geq I_{\alpha}^{\beta}(P, Q; U)$$

Case 3: For $\alpha < 1, \beta > 0$, equation (3.4.25) becomes

$$\left[\frac{\sum_{i=1}^{n} u_i p_i q_i^{\alpha-1}}{\sum_{i=1}^{n} u_i p_i}\right]^{\beta} \leq \left[\sum_{i=1}^{n} p_i \left(\frac{u_i}{\sum_{i=1}^{n} u_i p_i}\right)^{\frac{1}{\alpha}} D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)}\right]^{\alpha\beta}$$

Since $1 - D^{(1-\alpha)\beta} < 0$, we have

$$L_{\alpha}^{\beta}(U) \geq I_{\alpha}^{\beta}(P, Q; U)$$

Case 4: For $\alpha < 1, \beta < 0$, equation (3.4.25) becomes

$$\left[\frac{\sum_{i=1}^{n} u_i p_i q_i^{\alpha-1}}{\sum_{i=1}^{n} u_i p_i}\right]^{\beta} \geq \left[\sum_{i=1}^{n} p_i \left(\frac{u_i}{\sum_{i=1}^{n} u_i p_i}\right)^{\frac{1}{\alpha}} D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)}\right]^{\alpha\beta}$$

Since $1 - D^{(1-\alpha)\beta} > 0$, we have

$$L_{\alpha}^{\beta}(U) \geq I_{\alpha}^{\beta}(P, Q; U)$$

Theorem (3.4.4): For every code with lengths $l_1, l_2, ..., l_n$ satisfies the condition (3.4.20), $L_{\alpha}^{\beta}(U)$ can be made to satisfy

(3.4.26) $$I_{\alpha}^{\beta}(P, Q; U) \leq L_{\alpha}^{\beta}(U) < D^{(1-\alpha)\beta}\, I_{\alpha}^{\beta}(P, Q; U) + 1$$

Proof: In general we can not hope to construct an absolute optimal code for a given set of probabilities $p_1, p_2, ..., p_n$. Since if we choose l_i satisfy (3.4.22) then l_i may not be an integer. However, can do the next thing and select the integer $\overline{l_i}$ such that

(3.4.27) $$l_i \leq \overline{l_i} < l_i + 1$$

(3.4.28) $$-\log \frac{u_i q_i^{\alpha}}{\sum_{i=1}^{n} u_i p_i q_i^{\alpha-1}} \leq \overline{l_i} < -\log \frac{u_i q_i^{\alpha}}{\sum_{i=1}^{n} u_i p_i q_i^{\alpha-1}} + 1$$

We claim that prefix code can be constructed with word lengths $\overline{l_1}, \overline{l_2}, ..., \overline{l_n}$. To prove this we must show that sequences $\overline{l_1}, \overline{l_2}, ..., \overline{l_n}$ satisfies (3.4.20). From left hand inequality of (3.4.28), it follows that

$$\frac{u_i q_i^{\alpha}}{\sum_{i=1}^{n} u_i p_i q_i^{\alpha-1}} \geq D^{-\overline{l_i}}$$

Multiplying both sides by $p_i q_i^{-1}$ and summing over $i = 1, 2, ..., n$, we get (3.4.20). Considering $L_\alpha^\beta(U)$ as a function of $l_1, l_2, ..., l_n$ only and using differentiable technique it can be easily proved that $L_\alpha^\beta(U)$ is an increasing function of $l_1, l_2, ..., l_n$.

From (3.4.27) , we have

$$\frac{1}{1-D^{(1-\alpha)\beta}}\left[1-\left(\sum_{i=1}^{n} p_i\left(\frac{u_i}{\sum\limits_{i=1}^{n} u_i p_i}\right)^{\frac{1}{\alpha}} D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)}\right)^{\alpha\beta}\right]$$

$$\leq \frac{1}{1-D^{(1-\alpha)\beta}}\left[1-\left(\sum_{i=1}^{n} p_i\left(\frac{u_i}{\sum\limits_{i=1}^{n} u_i p_i}\right)^{\frac{1}{\alpha}} D^{-\overline{l_i}\left(\frac{\alpha-1}{\alpha}\right)}\right)^{\alpha\beta}\right]$$

$$< \frac{1}{1-D^{(1-\alpha)\beta}}\left[1-D^{(1-\alpha)\beta}\left(\sum_{i=1}^{n} p_i\left(\frac{u_i}{\sum\limits_{i=1}^{n} u_i p_i}\right)^{\frac{1}{\alpha}} D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)}\right)^{\alpha\beta}\right]$$

Clearly,

$$\left[\sum_{i=1}^{n} p_i\left(\frac{u_i}{\sum\limits_{i=1}^{n} u_i p_i}\right)^{\frac{1}{\alpha}} D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)}\right]^{\alpha\beta} = \left[\frac{\sum\limits_{i=1}^{n} u_i p_i q_i^{\alpha-1}}{\sum\limits_{i=1}^{n} u_i p_i}\right]^{\beta}$$

We get

$$\frac{1}{1-D^{(1-\alpha)\beta}}\left[1-\left(\frac{\sum\limits_{i=1}^{n} u_i p_i q_i^{\alpha-1}}{\sum\limits_{i=1}^{n} u_i p_i}\right)^{\beta}\right]$$

$$\leq L_\alpha^\beta(U) < \frac{1}{1-D^{(1-\alpha)\beta}}\left[1-D^{(1-\alpha)\beta}\left(\frac{\sum\limits_{i=1}^{n} u_i p_i q_i^{\alpha-1}}{\sum\limits_{i=1}^{n} u_i p_i}\right)^{\beta}\right]$$

Implies that

$$I_\alpha^\beta(P,Q;U) \leq L_\alpha^\beta(U) < D^{(1-\alpha)\beta} I_\alpha^\beta(P,Q;U) + 1$$

3.5. Bounds on generalized inaccuracy measures with two and three parameters

In this section, bounds have been obtained on generalized inaccuracy measures with two and three parameters given by Tuteja and Bhaker [129] .

Consider a function of inaccuracy measure given by Tuteja and Bhaker [129]

(3.5.1) $$I_{\alpha}^{\beta}(P,Q;U)=\frac{\sum_{i=1}^{n}(u_i p_i)^{\beta}\left(q_i^{\alpha-1}-1\right)}{(2^{1-\alpha}-1)\sum_{i=1}^{n}p_i^{\beta}}$$

where $\alpha>0\,(\neq 1)\,,\beta>0,p_i\geq 0,\ \ \sum_{i=1}^{n}p_i\leq 1$

Remark (3.5.1)

(1) When $\beta=1$, (3.5.1) reduces to the measure given by Sharma, Mittal and Mohan [113].

(2) When $\alpha\rightarrow 1,\beta=1$, (3.5.1) reduces to the non- additive 'useful' inaccuracy measure for generalized probability distribution characterized by Taneja and Tuteja [119].

(3) When $\beta=1,p_i=q_i\ \ \forall\ i=1,2,...,n$, (3.5.1) reduces to the measure given by Jain and Tuteja [63].

(4) When $\alpha\rightarrow 1,\beta=1,\ u_i=1\ \ \forall\ i=1,2,...,n$, and probabiltiy distribution is complete then (3.5.1) reduces to the measure given by Kerridge [73].

Consider a Parametric 'useful' generalized codeword mean length

(3.5.2) $$L_{\alpha}^{\beta}(U)=\frac{1}{(2^{1-\alpha}-1)\sum_{i=1}^{n}p_i^{\beta}}\left[\left(\frac{\sum_{i=1}^{n}(u_i p_i)^{\beta}D^{l_i\left(\frac{1-\alpha}{\alpha}\right)}}{\sum_{i=1}^{n}(u_i p_i)^{\beta}}\right)^{\alpha}-1\right]$$

where $\alpha>0\,(\neq 1)\,,\beta>0,p_i\geq 0,\sum_{i=1}^{n}p_i\leq 1$.

Remark (3.5.2)

(1) When $\beta=1$ and distribution is complete, then (3.5.2) reduces to codeword mean length given by Jain and Tuteja [63].

(2) When $\alpha\rightarrow 1,\beta=1$ and distribution is complete, then (3.5.2) reduces to the codeword mean length given by Guiasu and Picard [51] and further reduces to ordinary mean length given by Shannon [107] by taking $u_i=1\ \ \forall\ i=1,2,...,n.$

The bounds are obtained here, under the condition

(3.5.3) $$\sum_{i=1}^{n}(u_i p_i)^{\beta}\,q_i^{-1}D^{-l_i}\leq\sum_{i=1}^{n}(u_i p_i)^{\beta}$$

which is a generalization of Kraft [80] inequality.

Theorem (3.5.1): If $l_1, l_2, ..., l_n$ denote the code lengths satisfying the condition (3.5.3). Then

(3.5.4) $$L_\alpha^\beta(U) \geq \frac{I_\alpha^\beta(P,Q;U)}{\overline{U}},\ \alpha > 0\,(\neq 1)\,, \beta > 0$$

where $\overline{U} = \sum_{i=1}^{n}(u_i p_i)^\beta$, equality holds iff

(3.5.5) $$l_i = -\log q_i^\alpha + \log \frac{\sum_{i=1}^{n}(u_i p_i)^\beta q_i^{\alpha-1}}{\sum_{i=1}^{n}(u_i p_i)^\beta}$$

Proof: By Holder's inequality [116]

(3.5.6) $$\sum_{i=1}^{n} x_i y_i \geq \left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} y_i^q\right)^{\frac{1}{q}}$$

for all $x_i, y_i > 0,\ i = 1, 2, .., n$ and $\frac{1}{p} + \frac{1}{q} = 1, p < 1\,(\neq 0)\,, q < 0$ or $q < 1\,(\neq 0)\,, p < 0$. We see the equality holds iff there exists a positive constant c such that

(3.5.7) $$x_i^p = c y_i^q$$

Making the substitution

$$x_i = \left[\frac{(u_i p_i)^\beta}{\sum_{i=1}^{n}(u_i p_i)^\beta}\right]^{\frac{\alpha}{\alpha-1}} D^{-l_i},\ y_i = \left[\frac{(u_i p_i)^\beta}{\sum_{i=1}^{n}(u_i p_i)^\beta}\right]^{\frac{1}{1-\alpha}} q_i^{-1}$$

$$p = \frac{\alpha-1}{\alpha},\quad q = 1-\alpha$$

in (3.5.6), we get

$$\frac{\sum_{i=1}^{n}(u_i p_i)^\beta D^{-l_i} q_i^{-1}}{\sum_{i=1}^{n}(u_i p_i)^\beta} \geq \left[\frac{\sum_{i=1}^{n}(u_i p_i)^\beta D^{l_i\left(\frac{1-\alpha}{\alpha}\right)}}{\sum_{i=1}^{n}(u_i p_i)^\beta}\right]^{\frac{\alpha}{\alpha-1}} \left[\frac{\sum_{i=1}^{n}(u_i p_i)^\beta q_i^{\alpha-1}}{\sum_{i=1}^{n}(u_i p_i)^\beta}\right]^{\frac{1}{1-\alpha}}$$

using the inequality (3.5.3), we get

(3.5.8) $$\left[\frac{\sum_{i=1}^{n}(u_i p_i)^\beta D^{l_i\left(\frac{1-\alpha}{\alpha}\right)}}{\sum_{i=1}^{n}(u_i p_i)^\beta}\right]^{\frac{\alpha}{1-\alpha}} \geq \left[\frac{\sum_{i=1}^{n}(u_i p_i)^\beta q_i^{\alpha-1}}{\sum_{i=1}^{n}(u_i p_i)^\beta}\right]^{\frac{1}{1-\alpha}}$$

Let $0 < \alpha < 1$, raising both sides of (3.5.8) to the power $(1-\alpha)$, we get

(3.5.9) $$\left[\frac{\sum_{i=1}^{n}(u_i p_i)^\beta D^{l_i\left(\frac{1-\alpha}{\alpha}\right)}}{\sum_{i=1}^{n}(u_i p_i)^\beta}\right]^{\alpha} \geq \left[\frac{\sum_{i=1}^{n}(u_i p_i)^\beta q_i^{\alpha-1}}{\sum_{i=1}^{n}(u_i p_i)^\beta}\right]$$

Since $2^{1-\alpha} - 1 > 0$, for $0 < \alpha < 1$, a simple manipulation proves (3.5.4) for $0 < \alpha < 1$. The proof for $1 < \alpha < \infty$ follows on the same lines.

Theorem (3.5.2): For every code with lengths $l_1, l_2, ..., l_n$ satisfying the condition (3.5.3), $L_\alpha^\beta(U)$ can be made to satisfy the inequality

$$L_\alpha^\beta(U) < \frac{I_\alpha^\beta(P,Q;U)D^{1-\alpha}}{\overline{U}} + \frac{D^{1-\alpha}-1}{(2^{1-\alpha}-1)\sum_{i=1}^n p_i^\beta}, \alpha > 0\,(\neq 1), \beta > 0 \tag{3.5.10}$$

Proof: Let l_i be the positive integer satisfying the inequality

$$-\log q_i^\alpha + \log \frac{\sum_{i=1}^n (u_i p_i)^\beta q_i^{\alpha-1}}{\sum_{i=1}^n (u_i p_i)^\beta} \le l_i < -\log q_i^\alpha + \log \frac{\sum_{i=1}^n (u_i p_i)^\beta q_i^{\alpha-1}}{\sum_{i=1}^n (u_i p_i)^\beta} + 1 \tag{3.5.11}$$

Consider the interval

$$\delta_i = \left[-\log q_i^\alpha + \log \frac{\sum_{i=1}^n (u_i p_i)^\beta q_i^{\alpha-1}}{\sum_{i=1}^n (u_i p_i)^\beta},\ -\log q_i^\alpha + \log \frac{\sum_{i=1}^n (u_i p_i)^\beta q_i^{\alpha-1}}{\sum_{i=1}^n (u_i p_i)^\beta} + 1 \right] \tag{3.5.12}$$

of length 1. In every δ_i, there lies exactly one positive integer l_i such that

$$0 < -\log q_i^\alpha + \log \frac{\sum_{i=1}^n (u_i p_i)^\beta q_i^{\alpha-1}}{\sum_{i=1}^n (u_i p_i)^\beta} \le l_i < -\log q_i^\alpha + \log \frac{\sum_{i=1}^n (u_i p_i)^\beta q_i^{\alpha-1}}{\sum_{i=1}^n (u_i p_i)^\beta} + 1 \tag{3.5.13}$$

We will first show that the sequence $\{l_1, l_2, ..., l_n\}$, thus defined satisfies (3.5.3). From (3.5.13) we have

$$-\log q_i^\alpha + \log \frac{\sum_{i=1}^n (u_i p_i)^\beta q_i^{\alpha-1}}{\sum_{i=1}^n (u_i p_i)^\beta} \le l_i$$

$$-\log q_i^\alpha + \log \frac{\sum_{i=1}^n (u_i p_i)^\beta q_i^{\alpha-1}}{\sum_{i=1}^n (u_i p_i)^\beta} \le -\log_D D^{-l_i}$$

$$\frac{q_i^\alpha}{\left(\frac{\sum_{i=1}^n (u_i p_i)^\beta q_i^{\alpha-1}}{\sum_{i=1}^n (u_i p_i)^\beta} \right)} \ge D^{-l_i} \tag{3.5.14}$$

Multiplying both sides of (3.5.14) by $(u_i p_i)^\beta q_i^{-1}$ and summing over $i = 1, 2, ..., n.$ we get (3.5.3).

The last inequality of (3.5.13) gives

$$l_i < -\log q_i^\alpha + \log \frac{\sum_{i=1}^n (u_i p_i)^\beta q_i^{\alpha-1}}{\sum_{i=1}^n (u_i p_i)^\beta} + 1$$

$$D^{l_i} < D q_i^{-\alpha} \frac{\sum_{i=1}^n (u_i p_i)^\beta q_i^{\alpha-1}}{\sum_{i=1}^n (u_i p_i)^\beta} \tag{3.5.15}$$

Let $0 < \alpha < 1$, raising both sides of (3.5.15) to the power $\frac{1-\alpha}{\alpha}$, we get

(3.5.16) $$D^{l_i\left(\frac{1-\alpha}{\alpha}\right)} < D^{\frac{1-\alpha}{\alpha}} q_i^{\alpha-1}\left[\frac{\sum_{i=1}^{n}(u_ip_i)^{\beta}q_i^{\alpha-1}}{\sum_{i=1}^{n}(u_ip_i)^{\beta}}\right]^{\frac{1-\alpha}{\alpha}}$$

Multiply both sides of (3.5.16) by $\frac{(u_ip_i)^{\beta}}{\sum_{i=1}^{n}(u_ip_i)^{\beta}}$, summing over $i = 1, 2, ..., n$ and after then raising both sides to the power α, we get

(3.5.17) $$\left[\frac{\sum_{i=1}^{n}(u_ip_i)^{\beta}D^{l_i\left(\frac{1-\alpha}{\alpha}\right)}}{\sum_{i=1}^{n}(u_ip_i)^{\beta}}\right]^{\alpha} < D^{1-\alpha}\left[\frac{\sum_{i=1}^{n}(u_ip_i)^{\beta}q_i^{\alpha-1}}{\sum_{i=1}^{n}(u_ip_i)^{\beta}}\right]$$

Since $2^{1-\alpha}-1 > 0$, for $0 < \alpha < 1$, a simple manipulation in (3.5.17) proves (3.5.10) for $0 < \alpha < 1$. Also for $\alpha > 1$, the proof follows along the similar lines.

Now, bounds have been obtained here by considering a function studied by Tuteja and Bhaker [129] which involves three parameters. The results obtained here are more generalized than previous results. The function used here is applicable to more complex distributions.

Consider a function of inaccuracy measure given by Tuteja and Bhaker [129]

(3.5.18) $$I_{\alpha}^{\beta,\gamma}(P,Q;U) = \frac{\sum_{i=1}^{n}(u_ip_i)^{\beta}\left(p_i^{\gamma-1}q_i^{\alpha-\gamma}-1\right)}{(2^{1-\alpha}-1)\sum_{i=1}^{n}p_i^{\beta}} \quad , \quad \alpha,\beta,\gamma > 0, \alpha \neq 1$$

Remark (3.5.3)

(1) When $\gamma = 1$ in (3.5.18), it reduces to measure (3.5.1) given by Tuteja and Bhaker [129].

(2) When $\beta = 1, \gamma = 1$ and $\alpha \to 1$, then (3.5.18) reduces to measure given by Taneja and Tuteja [119].

(3) When $\beta = 1, \gamma = 1$, $\alpha \to 1$ and $u_i = 1\ \forall\ i = 1, 2, ..., .n$. Then (3.5.18) reduces to the result given by Kerridge [73] . Further if $p_i = q_i\ \ \forall\ \ i = 1, 2, ..., n$, it reduces to Shannon's [107] entropy.

(4) When $\beta = 1, \gamma = 1$ and $p_i = q_i\ \ \forall\ i = 1, 2, ..., n$. (3.5.18) reduces to measure given by Jain and Tuteja [63].

Let us consider the three parametric codeword mean length

(3.5.19) $$L_{\alpha}^{\beta,\gamma}(U) = \frac{1}{(2^{1-\alpha}-1)\sum_{i=1}^{n}p_i^{\beta}}\left[\left(\frac{\sum_{i=1}^{n}(u_ip_i)^{\beta}p_i^{\gamma-1}D^{-l_i\left(\frac{\gamma-\alpha}{\gamma-\alpha-1}\right)}}{\sum_{i=1}^{n}(u_ip_i)^{\beta}}\right)^{\alpha+1-\gamma}-1\right]$$

where $\alpha, \beta, \gamma > 0, \alpha \neq 1$.

Remark (3.5.4)

(1) When $\beta=1,\gamma=1$ and $\alpha\rightarrow 1$, then (3.5.19) reduces to codeword mean length given by Guiasu and Picard [51].

(2) When $\beta=1,\gamma=1$,$\alpha\rightarrow 1$ and $u_i=1\ \forall\ i=1,2,...,.n$. Then (3.5.19) reduces to codeword mean length given by Shannon [107].

(3) When $\beta=1,\gamma=1$, Then (3.5.19) reduces to the codeword mean length given by Jain and Tuteja [63].

Now, we obtain the bounds under the condition

(3.5.20) $$\sum_{i=1}^{n}(u_ip_i)^{\beta}p_i^{\gamma-1}q_i^{-1}D^{-l_i}\leq\sum_{i=1}^{n}(u_ip_i)^{\beta}$$

which is a generalization of Kraft [80] inequality.

Theorem (3.5.3): If $l_1,l_2,...,l_n$ denote the codeword lengths and satisfying the condition (3.5.20). Then

(3.5.21) $$L_{\alpha}^{\beta,\gamma}(U)\geq\frac{I_{\alpha}^{\beta,\gamma}(P,Q;U)}{\overline{U}},\ \ \alpha,\beta,\gamma>0,\alpha\neq 1$$

where $\overline{U}=\sum_{i=1}^{n}(u_ip_i)^{\beta}$, equality holds iff

(3.5.22) $$l_i=-\log q_i^{\alpha+1-\gamma}+\log\frac{\sum_{i=1}^{n}(u_ip_i)^{\beta}p_i^{\gamma-1}q_i^{\alpha-\gamma}}{\sum_{i=1}^{n}(u_ip_i)^{\beta}}$$

Proof: By Holder's inequality [116]

(3.5.23) $$\sum_{i=1}^{n}x_iy_i\geq\left(\sum_{i=1}^{n}x_i^p\right)^{\frac{1}{p}}\left(\sum_{i=1}^{n}y_i^q\right)^{\frac{1}{q}}$$

for all $x_i,y_i>0,\ i=1,2,..,n$ and $\frac{1}{p}+\frac{1}{q}=1,p<1\,(\neq 0)\,,q<0$ or $q<1\,(\neq 0)\,,p<0$. We see the equality holds iff there exists a positive constant c such that

(3.5.24) $$x_i^p=cy_i^q$$

For $\alpha>\gamma$,

Making the substitution

$$x_i=\left[\frac{(u_ip_i)^{\beta}}{\sum_{i=1}^{n}(u_ip_i)^{\beta}}\right]^{\frac{\gamma-\alpha-1}{\gamma-\alpha}}p_i^{\frac{(\gamma-1)(\gamma-1-\alpha)}{\gamma-\alpha}}D^{-l_i}$$

$$y_i = \left[\frac{(u_i p_i)^\beta}{\sum_{i=1}^{n}(u_i p_i)^\beta}\right]^{\frac{1}{\gamma-\alpha}} p_i^{\frac{\gamma-1}{\gamma-\alpha}} q_i^{-1}$$

$$p = \frac{\gamma-\alpha}{\gamma-\alpha-1}, q = \gamma - \alpha$$

in (3.5.23), we get

$$\frac{\sum_{i=1}^{n}(u_i p_i)^\beta p_i^{\gamma-1} q_i^{-1} D^{-l_i}}{\sum_{i=1}^{n}(u_i p_i)^\beta} \geq \left[\frac{\sum_{i=1}^{n}(u_i p_i)^\beta p_i^{\gamma-1} D^{-l_i\left(\frac{\gamma-\alpha}{\gamma-\alpha-1}\right)}}{\sum_{i=1}^{n}(u_i p_i)^\beta}\right]^{\frac{\gamma-\alpha-1}{\gamma-\alpha}} \left[\frac{\sum_{i=1}^{n}(u_i p_i)^\beta p_i^{\gamma-1} q_i^{\alpha-\gamma}}{\sum_{i=1}^{n}(u_i p_i)^\beta}\right]^{\frac{1}{\gamma-\alpha}}$$

using the inequality (3.5.20), we get

$$\left[\frac{\sum_{i=1}^{n}(u_i p_i)^\beta p_i^{\gamma-1} D^{-l_i\left(\frac{\gamma-\alpha}{\gamma-\alpha-1}\right)}}{\sum_{i=1}^{n}(u_i p_i)^\beta}\right]^{-\left(\frac{\gamma-\alpha-1}{\gamma-\alpha}\right)} \geq \left[\frac{\sum_{i=1}^{n}(u_i p_i)^\beta p_i^{\gamma-1} q_i^{\alpha-\gamma}}{\sum_{i=1}^{n}(u_i p_i)^\beta}\right]^{\frac{1}{\gamma-\alpha}} \tag{3.5.25}$$

Raising both sides of (3.5.25) to the power $(\gamma - \alpha)$, we get

$$\left[\frac{\sum_{i=1}^{n}(u_i p_i)^\beta p_i^{\gamma-1} D^{-l_i\left(\frac{\gamma-\alpha}{\gamma-\alpha-1}\right)}}{\sum_{i=1}^{n}(u_i p_i)^\beta}\right]^{1+\alpha-\gamma} \leq \left[\frac{\sum_{i=1}^{n}(u_i p_i)^\beta p_i^{\gamma-1} q_i^{\alpha-\gamma}}{\sum_{i=1}^{n}(u_i p_i)^\beta}\right] \tag{3.5.26}$$

Since $2^{1-\alpha} - 1 < 0$ for $\alpha > 1$, a simple manipulation in (3.5.26) proves (3.5.21) for $\alpha > 1$.For other cases, proof follows on the same lines.

Theorem (3.5.4): For every code with lengths $l_1, l_2, ..., l_n$ satisfying the condition (3.5.20). $L_\alpha^{\beta,\gamma}(U)$ can be made to satisfy the inequality

$$L_\alpha^{\beta,\gamma}(U) < \frac{I_\alpha^{\beta,\gamma}(P,Q;U) D^{\gamma-\alpha}}{\overline{U}} + \frac{D^{\gamma-\alpha}-1}{(2^{1-\alpha}-1)\sum_{i=1}^{n} p_i^\beta}, \alpha, \beta, \gamma > 0, \alpha \neq 1. \tag{3.5.27}$$

where $\overline{U} = \sum_{i=1}^{n}(u_i p_i)^\beta$

Proof: Let l_i be the positive integer satisfying the inequality

$$-\log q_i^{\alpha+1-\gamma} + \log \frac{\sum_{i=1}^{n}(u_i p_i)^\beta p_i^{\gamma-1} q_i^{\alpha-\gamma}}{\sum_{i=1}^{n}(u_i p_i)^\beta} \leq l_i < -\log q_i^{\alpha+1-\gamma} + \log \frac{\sum_{i=1}^{n}(u_i p_i)^\beta p_i^{\gamma-1} q_i^{\alpha-\gamma}}{\sum_{i=1}^{n}(u_i p_i)^\beta} + 1 \tag{3.5.28}$$

Consider the interval

$$\delta_i = \left[-\log q_i^{\alpha+1-\gamma} + \log \frac{\sum_{i=1}^{n}(u_i p_i)^\beta p_i^{\gamma-1} q_i^{\alpha-\gamma}}{\sum_{i=1}^{n}(u_i p_i)^\beta}, -\log q_i^{\alpha+1-\gamma} + \log \frac{\sum_{i=1}^{n}(u_i p_i)^\beta p_i^{\gamma-1} q_i^{\alpha-\gamma}}{\sum_{i=1}^{n}(u_i p_i)^\beta} + 1\right] \tag{3.5.29}$$

of lenght 1. In every δ_i, there lies exactly one positive l_i such that

$$0 < -\log q_i^{\alpha+1-\gamma} + \log \frac{\sum_{i=1}^{n}(u_ip_i)^\beta p_i^{\gamma-1}q_i^{\alpha-\gamma}}{\sum_{i=1}^{n}(u_ip_i)^\beta} \tag{3.5.30}$$

$$\leq l_i < -\log q_i^{\alpha+1-\gamma} + \log \frac{\sum_{i=1}^{n}(u_ip_i)^\beta p_i^{\gamma-1}q_i^{\alpha-\gamma}}{\sum_{i=1}^{n}(u_ip_i)^\beta} + 1$$

We will first show that the sequence $\{l_1, l_2, ..., l_n\}$, thus defined satisfies (3.5.20). From (3.5.30) we have

$$-\log q_i^{\alpha+1-\gamma} + \log \frac{\sum_{i=1}^{n}(u_ip_i)^\beta p_i^{\gamma-1}q_i^{\alpha-\gamma}}{\sum_{i=1}^{n}(u_ip_i)^\beta} \leq l_i$$

$$-\log q_i^{\alpha+1-\gamma} + \log \frac{\sum_{i=1}^{n}(u_ip_i)^\beta p_i^{\gamma-1}q_i^{\alpha-\gamma}}{\sum_{i=1}^{n}(u_ip_i)^\beta} \leq -\log_D D^{-l_i}$$

or

$$\frac{q_i^{\alpha+1-\gamma}}{\left(\frac{\sum_{i=1}^{n}(u_ip_i)^\beta p_i^{\gamma-1}q_i^{\alpha-\gamma}}{\sum_{i=1}^{n}(u_ip_i)^\beta}\right)} \geq D^{-l_i}$$

Multiplying both sides by $\frac{(u_ip_i)^\beta p_i^{\gamma-1}q_i^{-1}}{\sum_{i=1}^{n}(u_ip_i)^\beta}$ and summing over $i = 1, 2, ..., n.$ we get (3.5.20).

The last inequality in (3.5.30) gives

$$l_i < \log q_i^{-(\alpha+1-\gamma)} \frac{\sum_{i=1}^{n}(u_ip_i)^\beta p_i^{\gamma-1}q_i^{\alpha-\gamma}}{\sum_{i=1}^{n}(u_ip_i)^\beta} D$$

$$D^{l_i} < q_i^{-(\alpha+1-\gamma)} \frac{\sum_{i=1}^{n}(u_ip_i)^\beta p_i^{\gamma-1}q_i^{\alpha-\gamma}}{\sum_{i=1}^{n}(u_ip_i)^\beta} D \tag{3.5.31}$$

Let $\alpha > \gamma$, raising both sides of (3.5.31) to the power $-\left(\frac{\gamma-\alpha}{\gamma-\alpha-1}\right)$, we get

$$D^{-l_i\left(\frac{\gamma-\alpha}{\gamma-\alpha-1}\right)} > q_i^{\alpha-\gamma}\left(\frac{\sum_{i=1}^{n}(u_ip_i)^\beta p_i^{\gamma-1}q_i^{\alpha-\gamma}}{\sum_{i=1}^{n}(u_ip_i)^\beta}\right)^{\left(\frac{\alpha-\gamma}{\gamma-\alpha-1}\right)} D^{\left(\frac{\alpha-\gamma}{\gamma-\alpha-1}\right)} \tag{3.5.32}$$

Multiply both sides of (3.5.32) by $\frac{\sum_{i=1}^{n}(u_ip_i)^\beta p_i^{\gamma-1}}{\sum_{i=1}^{n}(u_ip_i)^\beta}$ and summing over $i = 1, 2, .., n$ and after raising both sides to the power $(\alpha+1-\gamma)$, we get

$$\left[\frac{\sum_{i=1}^{n}(u_ip_i)^\beta p_i^{\gamma-1}D^{-l_i\left(\frac{\gamma-\alpha}{\gamma-\alpha-1}\right)}}{\sum_{i=1}^{n}(u_ip_i)^\beta}\right]^{\alpha+1-\gamma} > \frac{\sum_{i=1}^{n}(u_ip_i)^\beta p_i^{\gamma-1}q_i^{\alpha-\gamma}}{\sum_{i=1}^{n}(u_ip_i)^\beta}D^{\gamma-\alpha} \tag{3.5.33}$$

Since for $\alpha > 1, 2^{1-\alpha}-1 < 0$, after simple manipulation in (3.5.33) we get for $\alpha > 1$ and $\alpha > \gamma$

$$L_\alpha^{\beta,\gamma}(U) < \frac{I_\alpha^{\beta,\gamma}(P,Q;U)D^{\gamma-\alpha}}{\overline{U}} + \frac{D^{\gamma-\alpha}-1}{(2^{1-\alpha}-1)\sum_{i=1}^{n}p_i^\beta}.$$

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 measures for this purpose have been proposed and extensively studied by Jeffreys [65], Kullback and Leibler [81], Renyi [104], Havrada and Charvat [57], Kapur [68], Sharma and Mittal [108], Burbea and Rao [28], Rao [99], Lin [85], Csiszar [36], Ali and Silvey [4], Vajda [131], Shioya and Da-te [115] and others .

These measures have been applied in a variety of fields such as: anthropology [99], Genetics [88], Finance [106], Economics [126], Poltical science [127], Biology [96], The analysis of contingency tables [50], Approximation of probability distributions [32,72], Signal processing [66,67] and Pattern recognition [21,31].

In this chapter, generalizations of relative entropy and of other different divergence measures are considerd and several information inequalities related to these divergence measures have been obtained.

4.1 Introduction

Let $\Delta_n = \left\{ P = (p_1, p_2, ..., p_n) / p_i > 0, \sum\limits_{i=1}^{n} p_i = 1 \right\}, n \geq 2$ be the set of complete finite discrete probability distributions.

The relative entropy is a measure of the distance between two distributions. In Statistics, it arises as the expectation of the logarithm of the likelihood ratio. The relative entropy $K(P//Q)$ is a measure of the inefficiency of assuming that the distribution is Q when the true distribution is P, e.g., if we know the true distribution of the random variable, then we could construct a code with average description length $H(P)$. If, instead, we used the code for a distribution Q, we would need $H(P)$+$K(P//Q)$ bits on the average to describe the random variable.

The relative entropy or Kullback- Leibler [81] distance, between two distributions is defined by

(4.1.1) $$K(P//Q) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i}$$

for all $P, Q \in \Delta_n$. In Δ_n, we have taken all $p_i > 0$. If we take $p_i \geq 0 \quad \forall \; i = 1, 2, ..., n$, then in this case we have to suppose that $0 \ln 0 = 0 \ln \left(\frac{0}{0}\right) = 0$. From the information theoretic point of view, we generally take all the logarithms with base 2, but here we have taken only natural logarithms.

In information theory, various divergence measures are applied in addition to the Kullback- Leibler divergence measure. These are given by Pearson [94], Sibson [117] and Lin [85], Taneja [122,123,124,125], Jeffrey [65], Hellinger [58], Topsoe [128], Ben-tal et al [18], Renyi [104], Sharma and Autar [109] and Vajda [131].

Given a convex function $f : [0, \infty) \to \Re$, the f-divergence functional

(4.1.2) $$C_f\left(P//Q\right) = \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right)$$

was introduced by Csiszar [35]- [36] as a generalized measure of information, a 'distance function' on the set of probability distribution P^n. The restriction here to discrete distributions is only for convenience, similar results hold for general distributions. As in Csiszar [35]- [36], we interpret undefined expression by

$$f(0) = \lim_{t\to 0+} f(t), \qquad 0f\left(\frac{0}{0}\right) = 0$$

$$0f\left(\frac{a}{0}\right) = \lim_{\varepsilon\to 0+} \varepsilon f\left(\frac{a}{\varepsilon}\right) = a \lim_{t\to\infty} \frac{f(t)}{t}, a > 0$$

The following results (Theorem (4.1.1), (4.1.2) and Corollary (4.1.1)) were given by Csiszar and Korner [34].

Theorem (4.1.1). (Joint convexity). If $f : [0, \infty) \to \Re$ is convex, then $C_f\left(P//Q\right)$ is jointly convex in P and Q.

Theorem (4.1.2). (Jensen's inequality). Let $f : [0, \infty) \to \Re$ be convex. Then for any $P, Q \in \Re_+^n$ with $P_n = \sum_{i=1}^{n} p_i > 0,\ Q_n = \sum_{i=1}^{n} q_i > 0$, we have the inequality

(4.1.3) $$C_f\left(P//Q\right) \geq Q_n f\left(\frac{P_n}{Q_n}\right)$$

If f is strictly convex, equality holds in (4.1.3) iff

(4.1.4) $$\frac{p_1}{q_1} = \frac{p_2}{q_2} = \ldots = \frac{p_n}{q_n}$$

It is natural to consider the following corollary.

Corollary (4.1.1). (Non negativity). Let $f : [0, \infty) \to \Re$ be convex and normalized, i.e $f(1) = 0$.

Then for any $P, Q \in \Re_+^n$ with $P_n = Q_n$, we have the inequality

(4.1.5) $$C_f\left(P//Q\right) \geq 0$$

If f is strictly convex, equality holds in (4.1.5) iff

(4.1.6) $\qquad p_i = q_i \quad \forall\ i = 1, 2, ..., n$

In particular, if P,Q are probability vectors, then the corollary (4.1.1) shows that, for strictly convex and normalized $f : [0, \infty) \to \Re$ that

(4.1.7) $\qquad C_f\left(P//Q\right) \geq 0$ and $C_f\left(P//Q\right) = 0$ iff P=Q.

We now give some examples of divergence measures in information theory which are particular cases of Csiszar f-divergence.

(1) Symmetric J-divergence [65]

The Symmetric J-divergence is defined by

(4.1.8) $\qquad K\left(P//Q\right) + K\left(Q//P\right) = J\left(P//Q\right) = \sum_{i=1}^{n} \left(p_i - q_i\right) \ln \frac{p_i}{q_i}$

where $K\left(P//Q\right) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i}$ studied by Kullback- Leibler [81].

If we choose $f\left(t\right) = \left(t-1\right) \ln t,\ \ t > 0$. Then

(4.1.9) $\qquad C_f\left(P//Q\right) = J\left(P//Q\right)$

(2) χ^2-distance [94]

The χ^2-distance (chi square distance) is defined by

(4.1.10) $\qquad \chi^2\left(P//Q\right) = \sum_{i=1}^{n} \frac{\left(p_i - q_i\right)^2}{q_i}$

It is clear that if $f\left(t\right) = \left(t-1\right)^2,\ \ t \in [0, \infty)$. Then

(4.1.11) $\qquad C_f\left(P//Q\right) = \chi^2\left(P//Q\right)$

(3) Hellinger discrimination [58]

The Hellinger discrimination $h^2\left(P//Q\right)$ is defined by

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

It is obvious that if $f\left(t\right) = \frac{1}{2}\left(\sqrt{t} - 1\right)^2$. Then

(4.1.13) $\qquad C_f\left(P//Q\right) = h^2\left(P//Q\right)$

(4) Bhattacharya distance [24]

The Bhattacharya distance $B\left(P//Q\right)$ is defined by

(4.1.14) $$B(P//Q) = \sum_{i=1}^{n} \sqrt{p_i q_i}$$

If we choose $f(t) = \sqrt{t}, \;\; t \in (0, \infty)$. Then

(4.1.15) $$C_f(P//Q) = B(P//Q)$$

(5) Triangular discrimination [128]

The triangular discrimination between P and Q is defined by

(4.1.16) $$\Delta(P//Q) = \sum_{i=1}^{n} \frac{|p_i - q_i|^2}{p_i + q_i}$$

It is obvious that if $f(t) = \frac{(t-1)^2}{t+1}, t \in (0, \infty)$. Then

(4.1.17) $$C_f(P//Q) = \Delta(P//Q)$$

(6) Renyi's $\alpha-$ order entropy [104]

The Renyi's $\alpha-$ order entropy $(\alpha > 1)$ is defined by

(4.1.18) $$R_\alpha(P//Q) = \sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha}$$

It is obvious that $f(t) = t^\alpha, t \in (0, \infty)$. Then

(4.1.19) $$C_f(P//Q) = R_\alpha(P//Q)$$

Dragomir [45,48] proved the following inequalities for Csiszar f-divergence.

Theorem (4.1.3). Let $\phi : \Re_+ \to \Re$ be differentiable convex. Then for all $P, Q \in \Re_+^n$, we have the inequality

(4.1.20) $$\phi'(1)(P_n - Q_n) \leq C_\phi(P//Q) - Q_n \phi(1) \leq C_{\phi'}\left(\frac{P^2}{Q}//P\right) - C_{\phi'}(P//Q)$$

where $P_n = \sum_{i=1}^{n} p_i > 0, \; Q_n = \sum_{i=1}^{n} q_i > 0$ and $\phi' : (0, \infty) \to \Re$ is the derivative of ϕ. If ϕ is strictly convex and $p_i, q_i > 0 \;\; (i = 1, 2, ..., n)$, then the equality holds in (4.1.20) iff $P = Q$. If we assume that $P_n = Q_n$ and ϕ is normalized, then we obtain the simpler inequality

(4.1.21) $$0 \leq C_\phi(P//Q) \leq C_{\phi'}\left(\frac{P^2}{Q}//P\right) - C_{\phi'}(P//Q)$$

Applications for particular divergences which are instances of Csiszar f-divergence were also given.

Theorem (4.1.4). Let $\phi : \Re_+ \to \Re$ be differentiable convex. Then for all $P, Q \in \Re^n_+$. Then we have the inequality

(4.1.22) $$0 \le C_\phi (P//Q) - Q_n \phi\left(\frac{P_n}{Q_n}\right) \le C_{\phi^/}\left(\frac{P^2}{Q}//P\right) - \frac{P_n}{Q_n} C_{\phi^/}(P//Q)$$

If ϕ is strictly convex and $p_i, q_i > 0, \ \ (i = 1, 2, ..., n)$, then the equality holds in (4.1.22) iff

$$\frac{p_1}{q_1} = \frac{p_2}{q_2} = ... = \frac{p_n}{q_n}.$$

Obviously, If $P_n = Q_n$ and ϕ is normalized, then (4.1.22) becomes (4.1.21).

Dragomir [42,43,44] has obtained many results related to Kullback- Leibler distance, Hellinger discrimination and variational distance for the Csiszar f-divergence. In the next section, upper and lower bounds for the Csiszar f-divergence in terms of symmetric J-divergence measure have been obtained. Also some particular cases are obtained in terms of symmetric J-divergence measure by comparing it with a number of other divergence measures arising in information theory. This work has been accepted for publication in "Indian Journal of Mathematics" (Baig and Rayees [13]).

4.2 Some inequalities between Csiszar f-divergence and symmetric J-divergence measure.

Theorem (4.2.1). Assume that the generating mapping $f : [0, \infty) \to \Re$ is normalized, i.e $f(1) = 0$ and satisfies the assumption:

(1) f is twice differentiable on (r, R), where $0 \le r \le 1 \le R < \infty$;

(2) there exists the real constants m,M such that

(4.2.1) $$m \le \frac{t^2}{t+1} f^{//}(t) \le M \quad \forall \ t \in (r, R)$$

If P,Q are discrete probability distributions satisfying the assumption $r \le r_i = \frac{p_i}{q_i} \le R \ \forall i \in \{1, 2, ..., n\}$. Then we have the inequality

(4.2.2) $$mJ(P//Q) \le C_f (P//Q) \le MJ(P//Q)$$

Proof: Define the mapping $F_m : (0, \infty) \to \Re, \ F_m(t) = f(t) - m(t-1)\ln t$

is normalized, twice differentiable and since

$$F_m^{//}(t) = f^{//}(t) - \frac{m(t+1)}{t^2} = \frac{t+1}{t^2}\left[\frac{t^2}{t+1} f^{//}(t) - m\right] \ge 0 \ \ \forall t \in (r, R)$$

It follows that $F_m(t)$ is convex on (r, R). Applying the non negative property of the Csiszar f-divergence functional $F_m(t)$ and the linearity property, we may state that

$$0 \leq C_{F_m}(P//Q) = C_f(P//Q) - mC_{(t-1)\ln t}(P//Q)$$
$$= C_f(P//Q) - mJ(P//Q)$$

or

$$mJ(P//Q) \leq C_f(P//Q)$$

from where we get the first inequality in (4.2.2).

Define $F_M : (0, \infty) \to \Re,\ F_M(t) = M(t-1)\ln t - f(t)$, which is obviously normalized, twice differentiable and by (4.2.1) convex on (r, R). Applying the non negativity property of Csiszar f-divergence for C_{F_M}, we obtain the second part of inequality (4.2.2).

Remark (4.2.1). If in (4.2.1), we have the strict inequality "$<$" for any $t \in (r, R)$, then the mapping F_m and F_M are strictly convex and the case of equality holds in (4.2.2) iff P=Q.

Using the inequality (4.1.21) which holds for ϕ differentiable convex and normalized functions for P,Q probability distributions, we can state the following theorem as well.

Theorem (4.2.2). Let $f : [0, \infty) \to \Re$ be a normalized mapping, i.e., $f(1) = 0$ and satisfies the assumption:

(1) f is twice differentiable on (r, R), where $0 \leq r \leq 1 \leq R < \infty$;

(2) there exists the real constants m, M such that

(4.2.3) $$m \leq \frac{t^2}{t+1} f^{//}(t) \leq M \quad \forall\ t \in (r, R)$$

If P,Q are discrete probability distributions satisfying the assumption $r \leq r_i = \frac{p_i}{q_i} \leq R\ \forall i \in \{1, 2, ..., n\}$. Then we have the inequality

(4.2.4) $$C_{f^/(t)}\left(\frac{P^2}{Q}//P\right) - C_{f^/(t)}(P//Q) + MC(P//Q)$$
$$\leq C_f(P//Q)$$
$$\leq C_{f^/(t)}\left(\frac{P^2}{Q}//P\right) - C_{f^/(t)}(P//Q) + mC(P//Q)$$

where $C(P//Q) = 1 - \sum\limits_{i=1}^{n} \frac{q_i^2}{p_i}$

Proof: We know that (see the proof of theorem (4.2.1)) that the mapping

$F_m(t) = f(t) - m(t-1)\ln t$ is normalized, twice differentiable and convex on (r, R).

If we apply the second inequality from (4.1.21) for F_m, we may write

(4.2.5) $$C_{F_m}(P//Q) \leq C_{F_m^{/}}\left(\frac{P^2}{Q}//P\right) - C_{F_m^{/}}(P//Q)$$

However,

$$C_{F_m}(P//Q) = C_f(P//Q) - mJ(P//Q)$$

$$\begin{aligned} C_{F_m^{/}}\left(\frac{P^2}{Q}//P\right) &= C_{f^{/}(t)-m(1-t^{-1}+\ln t)}\left(\frac{P^2}{Q}//P\right) \\ &= C_{f^{/}_{(t)}}\left(\frac{P^2}{Q}//P\right) - m + mC_{t^{-1}}\left(\frac{P^2}{Q}//P\right) - mC_{\ln t}\left(\frac{P^2}{Q}//P\right) \\ &= C_{f^{/}_{(t)}}\left(\frac{P^2}{Q}//P\right) - mK(P//Q) \end{aligned}$$

Also,

$$\begin{aligned} C_{F_m^{/}}(P//Q) &= C_{f^{/}(t)-m(1-t^{-1}+\ln t)}(P//Q) \\ &= C_{f^{/}(t)}(P//Q) - m + mC_{t^{-1}}(P//Q) - mC_{\ln t}(P//Q) \\ &= C_{f^{/}(t)}(P//Q) - m + m\sum_{i=1}^{n}\frac{q_i^2}{p_i} + mK(Q//P) \end{aligned}$$

And then by (4.2.5), we may write

$$\begin{aligned} &C_f(P//Q) - mJ(P//Q) \\ &\quad \leq C_{f^{/}_{(t)}}\left(\frac{P^2}{Q}//P\right) - mK(P//Q) \\ &\qquad - C_{f^{/}(t)}(P//Q) + m - m\sum_{i=1}^{n}\frac{q_i^2}{p_i} - mK(Q//P) \\ &\quad = C_{f^{/}_{(t)}}\left(\frac{P^2}{Q}//P\right) - C_{f^{/}(t)}(P//Q) - mJ(P//Q) + mC(P//Q) \end{aligned}$$

or

$$C_f(P//Q) \leq C_{f^{/}(t)}\left(\frac{P^2}{Q}//P\right) - C_{f^{/}(t)}(P//Q) + mC(P//Q)$$

which is equivalent to the second inequality in (4.2.4).

Consider $F_M(t) = M(t-1)\ln t - f(t)$, which is obviously normalized, twice differentiable and convex on (r, R).

If we apply the second inequality from (4.1.21) for F_M, we may write

(4.2.6) $$C_{F_M}(P//Q) \leq C_{F_M^{/}}\left(\frac{P^2}{Q}//P\right) - C_{F_M^{/}}(P//Q)$$

However,

$$C_{F_M}(P//Q) = MJ(P//Q) - C_f(P//Q)$$

$$C_{F_M^{/}}\left(\frac{P^2}{Q}//P\right) = C_{M(1-t^{-1}+\ln t)-f^{/}(t)}\left(\frac{P^2}{Q}//P\right)$$
$$= M - MC_{t^{-1}}\left(\frac{P^2}{Q}//P\right) + MC_{\ln t}\left(\frac{P^2}{Q}//P\right) - C_{f^{/}_{(t)}}\left(\frac{P^2}{Q}//P\right)$$
$$= MK\left(P//Q\right) - C_{f^{/}_{(t)}}\left(\frac{P^2}{Q}//P\right)$$

Also,

$$C_{F_M^{/}}\left(P//Q\right) = C_{M(1-t^{-1}+\ln t)-f^{/}(t)}\left(P//Q\right)$$
$$= M - M\sum_{i=1}^{n}\frac{q_i^2}{p_i} - MK\left(Q//P\right) - C_{f^{/}(t)}\left(P//Q\right)$$
$$= MC\left(P//Q\right) - MK\left(Q//P\right) - C_{f^{/}(t)}\left(P//Q\right)$$

and then by (4.2.6) , we may obtain

$$MJ\left(P//Q\right) - C\left(P//Q\right)$$
$$\leq MK\left(P//Q\right) - C_{f^{/}_{(t)}}\left(\frac{P^2}{Q}//P\right) - M + M\sum_{i=1}^{n}\frac{q_i^2}{p_i}$$
$$+MK\left(Q//P\right) + C_{f^{/}(t)}\left(P//Q\right)$$
$$= -C_{f^{/}_{(t)}}\left(\frac{P^2}{Q}//P\right) + C_{f^{/}(t)}\left(P//Q\right) + MJ\left(P//Q\right) - MC\left(P//Q\right)$$

or

$$C_f\left(P//Q\right) \geq C_{f^{/}_{(t)}}\left(\frac{P^2}{Q}//P\right) - C_{f^{/}(t)}\left(P//Q\right) + MC\left(P//Q\right)$$

which is equivalent to the first inequality in (4.2.4).

4.3 Some particular cases

The results in section (4.2) have natural applications when the symmetric J-divergence measure is compared with a number of other divergence measures arising in information theory.

Proposition (4.3.1). Let P,Q be two probability distributions with property that

(4.3.1) $$0 < r \leq \frac{p_i}{q_i} \leq R < \infty \quad \forall\, i \in \{1, 2, ..., n\}$$

Then we have the inequality

(4.3.2) $$\frac{2r^2}{r+1}J\left(P//Q\right) \leq \chi^2\left(P//Q\right) \leq \frac{2R^2}{R+1}J\left(P//Q\right)$$

Proof: Consider the mapping $f : (0, \infty) \to \Re,\ f(t) = (t-1)^2$. Then $f^{//}(t) = 2$

Define the mapping $g : [r, R] \to \Re,\ g(t) = \frac{t^2}{t+1}f^{//}(t)$

$g(t) = \frac{2t^2}{t+1}$

Then obviously

$\inf_{t\in[r,R]} g(t) = \frac{2r^2}{r+1}$

$\sup_{t\in[r,R]} g(t) = \frac{2R^2}{R+1}$

Since

$$C_f(P//Q) = \chi^2(P//Q)$$

Now, using the inequality (4.2.2) with $m = \frac{2r^2}{r+1},\ M = \frac{2R^2}{R+1}$, we get the inequality (4.3.2).

Proposition (4.3.2). Let P,Q be two probability distributions satisfying the condition (4.3.1). Then we have the inequality

(4.3.3) $$\frac{\sqrt{r}}{4(r+1)} J(P//Q) \le h^2(P//Q) \le \frac{\sqrt{R}}{4(R+1)} J(P//Q)$$

Proof: Consider the mapping $f(t) = \frac{1}{2}\left(\sqrt{t}-1\right)^2$. Then $f^{//}(t) = \frac{1}{4\sqrt{t^3}}$

Define the mapping $g : [r,R] \to \Re,\ g(t) = \frac{t^2}{t+1} f^{//}(t)$

$g(t) = \frac{\sqrt{t}}{4(t+1)}$

Then obviously

$\inf_{t\in[r,R]} g(t) = \frac{\sqrt{r}}{4(r+1)}$

$\sup_{t\in[r,R]} g(t) = \frac{\sqrt{R}}{4(R+1)}$

Since

$$C_f(P//Q) = h^2(P//Q)$$

Now, using the inequality (4.2.2) with $m = \frac{\sqrt{r}}{4(r+1)},\ \ M = \frac{\sqrt{R}}{4(R+1)}$, we get the desired inequality (4.3.3).

Proposition (4.3.3). Let P,Q be two probability distributions satisfying the condition (4.3.1). Then we have the inequality

(4.3.4) $$\frac{\alpha(\alpha-1)r^\alpha}{r+1} J(P//Q) + 1 \le R_\alpha(P//Q) \le \frac{\alpha(\alpha-1)R^\alpha}{R+1} J(P//Q) + 1$$

Proof: Consider the mapping $f : (0,\infty) \to \Re,\ f(t) = t^\alpha - 1,\ \alpha > 1$

Then $f^{//}(t) = \alpha(\alpha-1)t^{\alpha-2}$

Define the mapping $g:[r,R]\to\Re,\ g(t)=\frac{t^2}{t+1}f^{//}(t)$

$$g(t)=\frac{\alpha(\alpha-1)t^\alpha}{t+1}$$

Then obviously

$$\inf_{t\in[r,R]}g(t)=\frac{\alpha(\alpha-1)r^\alpha}{r+1}$$

$$\sup_{t\in[r,R]}g(t)=\frac{\alpha(\alpha-1)R^\alpha}{R+1}$$

Also,

$$C_f(P//Q)=R_\alpha(P//Q)-1$$

Now, using the inequality (4.2.2) with $m=\frac{\alpha(\alpha-1)r^\alpha}{r+1},\quad M=\frac{\alpha(\alpha-1)R^\alpha}{R+1}$, we get the desired inequality (4.3.4).

Proposition (4.3.4). Let P,Q be two probability distributions. Then we have the inequality

(4.3.5) $$\Delta(P//Q)\le\frac{1}{2}J(P//Q)$$

Proof: Consider the mapping $f(t)=\frac{(t-1)^2}{t+1}$. Then obviously $f^{//}(t)=\frac{8}{(t+1)^3}$

Define the mapping $g:[r,R]\to\Re,\ g(t)=\frac{t^2}{t+1}f^{//}(t)$

$$g(t)=\frac{8t^2}{(t+1)^4}$$

A simple calculation shows that

$$g^{/}(t)=\frac{16t(t+1)^3(1-t)}{(t+1)^8}$$

Consequently, the mapping g is increasing on the interval (0,1) and decreasing on (1,∞).

Moreover,

$$\sup_{t\in(0,\infty)}g(t)=g(1)=\frac{1}{2}$$

and

$$C_f(P//Q)=\Delta(P//Q)$$

Applying the inequality (4.2.2) for $M=\frac{1}{2}$, we deduce (4.3.5).

If we know more about $r_i=\frac{p_i}{q_i}$, i.e., the condition (4.3.1) holds, then we can improve the inequality (4.3.5) as follows.

Proposition (4.3.5). Assume that the probability distributions P,Q satisfies (4.3.1). Then we have the inequality

(4.3.6) $$8\min\left\{\frac{r^2}{(r+1)^4},\frac{R^2}{(R+1)^4}\right\}J\left(P//Q\right)\le\Delta\left(P//Q\right)$$

Proof: Taking in to account that the mapping $g\left(t\right)=\frac{8t^2}{(t+1)^4}$ is monotonic increasing on (0,1) and decreasing on $(1,\infty)$, we may assert that

$$\inf_{t\in[r,R]}g\left(t\right)=\min\left\{g\left(r\right),g\left(R\right)\right\}$$

$$=8\min\left\{\frac{r^2}{(r+1)^4},\frac{R^2}{(R+1)^4}\right\}$$

using the inequality (4.2.2), we deduce (4.3.6).

Proposition (4.3.6). Let P,Q be two probability distributions. Then we have the inequality

(4.3.7) $$1-\frac{1}{8}J\left(P//Q\right)\le B\left(P//Q\right)$$

Proof: Consider the mapping $f:(0,\infty)\to\Re,\ \ f\left(t\right)=1-\sqrt{t}$

Then

$$f^{//}\left(t\right)=\frac{1}{4}t^{\frac{-3}{2}}$$

Define $g:[r,R]\to\Re,\ \ g\left(t\right)=\frac{t^2}{t+1}f^{//}\left(t\right)$

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

which shows with simple calculations that $g^{/}\left(t\right)=\frac{t^{\frac{-1}{2}}(1-t)}{8(t+1)^2}$

Consequently, the mapping g is increasing on the interval (0,1) and decreasing on (1,∞).

Moreover,

$$\sup_{t\in(0,\infty)}g\left(t\right)=g\left(1\right)=\frac{1}{8}$$

Since

$$C_f\left(P//Q\right)=1-B\left(P//Q\right)$$

Then, applying the inequality (4.2.2) for $M=\frac{1}{8}$, we deduce (4.3.7).

If we know more about $r_i=\frac{p_i}{q_i}$, i.e., the condition (4.3.1) holds. Then we can point out an upper bound for $B\left(P//Q\right)$ as follows.

Proposition (4.3.7). If $0 < r \le \frac{p_i}{q_i} \le R < \infty \quad \forall\, i \in \{1, 2, ..., n\}$. Then we have the inequality

(4.3.8) $$\frac{1}{4}\min\left\{\frac{r^{\frac{1}{2}}}{(r+1)}, \frac{R^{\frac{1}{2}}}{(R+1)}\right\} J\left(P//Q\right) \le 1 - B\left(P//Q\right)$$

Proof: Taking in to account that the mapping $g\left(t\right) = \frac{1}{4}\frac{t^{\frac{1}{2}}}{(t+1)}$ is monotonic increasing on the interval (0,1) and decreasing on $(1,\infty)$.

Moreover

$$\inf_{t\in[r,R]} g\left(t\right) = \min\left\{g\left(r\right), g\left(R\right)\right\}$$

$$= \min\left\{\frac{1}{4}\frac{r^{\frac{1}{2}}}{(r+1)}, \frac{1}{4}\frac{R^{\frac{1}{2}}}{(R+1)}\right\}$$

$$= \frac{1}{4}\min\left\{\frac{r^{\frac{1}{2}}}{(r+1)}, \frac{R^{\frac{1}{2}}}{(R+1)}\right\}$$

using the inequality (4.2.2), we deduce the desired upper bound (4.3.8).

Let us consider the Harmonic distance by

$$M\left(P//Q\right) = \sum_{i=1}^{n}\frac{2p_iq_i}{p_i+q_i}$$

The following proposition holds.

Proposition (4.3.8). Let P,Q be two probability distributions. Then we have the inequality

(4.3.9) $$1 - \frac{1}{4}J\left(P//Q\right) \le M\left(P//Q\right)$$

Proof: Consider the mapping $f : (0,\infty) \to \Re,\ \ f\left(t\right) = 1 - \frac{2t}{t+1},\ f^{//}\left(t\right) = \frac{4}{(t+1)^3}$

Define $g : [r, R] \to \Re,\ \ g\left(t\right) = \frac{t^2}{t+1}f^{//}\left(t\right)$

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

A simple calculation shows that

$$g^{/}\left(t\right) = \frac{32t(1-t)}{(t+1)^5}$$

Consequently, the mapping g is increasing on the interval (0,1) and decreasing on $(1,\infty)$.

Moreover,

$$\sup_{t\in(0,\infty)} g\left(t\right) = g\left(1\right) = \frac{1}{4}$$

Then, applying the inequality (4.2.2) for $M = \frac{1}{4}$, we have (4.3.9).

If we know more about $r_i = \frac{p_i}{q_i}$, i.e., the condition (4.3.1) holds. Then we can point out an upper bound for $M(P//Q)$ as follows.

Proposition (4.3.9). If $0 < r \leq \frac{p_i}{q_i} \leq R < \infty \quad \forall\, i \in \{1, 2, ..., n\}$. Then we have the inequality

(4.3.10) $$M(P//Q) \leq 1 - 4\min\left\{\frac{r^2}{(r+1)^4}, \frac{R^2}{(R+1)^4}\right\} J(P//Q)$$

Proof: Taking in to account that the mapping $g(t) = \frac{4t^2}{(t+1)^4}$ is monotonic increasing on the interval (0,1) and decreasing on $(1, \infty)$.

Moreover

$$\begin{aligned} \inf_{t\in[r,R]} g(t) &= \min\{g(r), g(R)\} \\ &= \min\left\{\frac{4r^2}{(r+1)^4}, \frac{4R^2}{(R+1)^4}\right\} \\ &= 4\min\left\{\frac{r^2}{(r+1)^4}, \frac{R^2}{(R+1)^4}\right\}, t \in [r, R] \end{aligned}$$

using the inequality (4.2.2), we deduce the desired upper bound (4.3.10).

Numerical illustration (4.3.1). We consider two examples of symmetrical and asymmetrical probability distributions. We calculate $\chi^2(P//Q)$ and $J(P//Q)$ and compare bounds.

Example (4.3.1). Let P be the binomial probability distribution for the random variable X with parameters $(n = 8, p = 0.5)$ and Q its approximated normal probability distribution. Then

Table 1. Binomial probability distribution $(n = 8, p = 0.5)$

x	0	1	2	3	4	5	6	7	8
$p(x)$	0.004	0.031	0.109	0.219	0.274	0.219	0.109	0.031	0.004
$q(x)$	0.005	0.030	0.104	0.220	0.282	0.220	0.104	0.030	0.005
$\frac{p(x)}{q(x)}$	0.774	1.042	1.0503	0.997	0.968	0.997	1.0503	1.042	0.774

The measure $\chi^2(P//Q)$ and $J(P//Q)$ are:

$\chi^2(P//Q) =$0.00145837, $J(P//Q) = 0.00151848$

It is noted that

$$r\,(=0.77417993) \le \frac{p_i}{q_i} \le R\,(=1.050330018)$$

The lower and upper bounds for $\chi^2\,(P//Q)$ from (4.3.2).

lower bound=$\frac{2r^2}{r+1}J\,(P//Q) = 0.0010258824$

upper bound= $\frac{2R^2}{R+1}J\,(P//Q) = 0.0016339321$

Thus,

$$0.0010258824 < \chi^2\,(P//Q) = 0.00145837 < 0.0016339321$$

Example (4.3.2). Let P be the binomial distribution for the random variable X with parameters $(n = 8, p = 0.4)$ and Q its approximated normal probability distribution. Then

Table 2. Binomial probability distribution $(n = 8, p = 0.4)$

x	0	1	2	3	4	5	6	7	8
$p\,(x)$	0.017	0.090	0.209	0.279	0.232	0.124	0.041	0.008	0.001
$q\,(x)$	0.020	0.082	0.198	0.285	0.244	0.124	0.037	0.007	0.0007
$\frac{p(x)}{q(x)}$	0.850	1.102	1.056	0.979	0.952	1.001	1.097	1.194	1.401

From the above data, measures $\chi^2\,(P//Q)$ and $J\,(P//Q)$ are calculated:

$$\chi^2\,(P//Q) = 0.00333883,\ J\,(P//Q) = 0.00327778$$

Note that

$$r\,(=0.849782156) \le \frac{p_i}{q_i} \le R\,(=1.401219652)$$

Lower and upper bounds for $\chi^2\,(P//Q)$ from (4.3.2).

lower bound=$\frac{2r^2}{r+1}J\,(P//Q) = 0.0025591122$

upper bound= $\frac{2R^2}{R+1}J\,(P//Q) = 0.0053601511$

Thus,

$$0.0025591122 < \chi^2\,(P//Q) = 0.00333883 < 0.0053601511$$

Dragomir [47] has obtained many interesting information inequalities by using the results in terms of χ^2 distance with the help of Csiszar's f divergence. In the next section, some generalized information inequalities have been obtained, which generalizes not only the result of Dragimir [47] for chi -square distance but also gives the result in terms of relative J- divergence measure. The work has been published in " Journal of Mathematics and System Sciences", Vol 2(2), PP 57-75 (2006) (Baig and Rayees [10]).

4.4 Generalized divergence measure and information inequalities.

Let $\Delta_n = \{P = (p_1, p_2, ..., p_n) / p_i > 0\}, \sum_{i=1}^{n} p_i = 1, n \geq 2$ be the set of complete finite discrete probability distribution.

The relative J-divergene measure given by Dragomir [41] is

(4.4.1) $$D(P//Q) = \sum_{i=1}^{n} (p_i - q_i) \ln \left(\frac{p_i + q_i}{2q_i} \right) \quad \forall \ P, Q \in \Delta_n$$

In Δ_n, we have taken all $p_i > 0$ but if we take $p_i \geq 0 \ \forall \ i = 1, 2, .., n$, then in this case we have to suppose that $0 \ln 0 = 0 \ln \left(\frac{0}{0} \right) = 0$. From the information theoretic point of view we generally take all the logarithms with base 2, but here we have taken only natural logarithms.

We observe that the measure (4.4.1) is not symmetric in P,Q. Its symmetric version, famous as J- divergence measure and is given by Jeffreys [65] as

(4.4.2) $$J(P//Q) = D(P//Q) + D(Q//P) = \sum_{i=1}^{n} (p_i - q_i) \ln \left(\frac{p_i}{q_i} \right)$$

The one parametric generalization of the measure (4.4.1) given by Taneja [121] and is called relative J-divergence of type s as

(4.4.3) $$D_s(P//Q) = (s-1)^{-1} \sum_{i=1}^{n} (p_i - q_i) \left(\frac{p_i + q_i}{2q_i} \right)^{s-1}, \ s \neq 1$$

In this case we have the following limiting case.

$$\lim_{s \to 1} D_s(P//Q) = \sum_{i=1}^{n} (p_i - q_i) \ln \left(\frac{p_i + q_i}{2q_i} \right) = D(P//Q)$$

We have a interesting particular case of the measure (4.4.3).

When s=2, we have

$$D_2(P//Q) = \frac{1}{2} \chi^2(P//Q)$$

where

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

is the χ^2- distance given by Pearson [94].

For simplicity, let us write the measure (4.4.3) in the unified way:

(4.4.5) $$\xi_s(P//Q) = \begin{cases} D_s(P//Q) & ;s \neq 1 \\ D(P//Q) & ;s = 1 \end{cases}$$

Summarizing, we have the following particular cases of the measure (4.4.5).

(1) $\xi_1(P//Q) = D(P//Q)$

(2) $\xi_2(P//Q) = \frac{1}{2}\chi^2(P//Q)$

In view of theorem (4.1.1) and (4.1.2), we have the following results.

Result (4.4.1). For all $P, Q \in \Delta_n$, we have

(1) $\xi_s(P//Q) \geq 0$ for $0 \leq s \leq 4$ with equality iff P=Q.

(2) $\xi_s(P//Q)$ is convex function of the pair of distribution $(P, Q) \in \Delta_n \times \Delta_n$, for $0 \leq s \leq 4$.

Proof: Take

(4.4.6) $$\phi_s(u) = \begin{cases} (s-1)^{-1}(u-1)\left[\left(\frac{u+1}{2}\right)^{s-1} - 1\right]; & s \neq 1 \\ (u-1)\ln\left(\frac{u+1}{2}\right); & s = 1 \end{cases}$$

for all $u > 0$ in (4.1.2), we have

$$C_f(P//Q) = \xi_s(P//Q) = \begin{cases} D_s(P//Q) & ;s \neq 1 \\ D(P//Q) & ;s = 1 \end{cases}$$

The above result holds for all $0 \leq s \leq 4$.

Moreover,

(4.4.7) $$\phi_s^{/}(u) = \begin{cases} \frac{1}{2}(u-1)\left(\frac{u+1}{2}\right)^{s-2} + (s-1)^{-1}\left[\left(\frac{u+1}{2}\right)^{s-1} - 1\right]; & s \neq 1 \\ \frac{u-1}{u+1} + \ln\left(\frac{u+1}{2}\right); & s = 1 \end{cases}$$

and

(4.4.8) $$\phi_s^{//}(u) = \left(\frac{u+1}{2}\right)^{s-3}\left(\frac{su+4-s}{4}\right)$$

Thus we have $\phi_s^{//}(u) > 0\ \forall\ u > 0$ and $0 \leq s \leq 4$. Hence $\phi_s(u)$ is convex for all $u > 0$ and $0 \leq s \leq 4$. Also, we have $\phi_s(1) = 0$. In view of Theorem (4.1.1) and (4.1.2), we have the proof of the result (4.4.1).

The following theorem summarizes some of the result studied by Dragomir [45,48]. For simplicity we have taken $f(1) = 0$ and $P, Q \in \Delta_n$.

Theorem (4.4.1). Let $f : \Re_+ \to \Re$ be differentiable convex and normalized i.e., $f(1) = 0$. If $P, Q \in \Delta_n$ are such that $0 < r \leq \frac{p_i}{q_i} \leq R < \infty\ \ \forall\ i \in \{1, 2, ..., n\}$ for some r and R with $0 < r \le 1 \le R < \infty$, then we have the following inequalities

(4.4.9) $$0 \le C_f(P//Q) \le \frac{1}{4}(R-r)\left(f^{/}(R) - f^{/}(r)\right)$$

(4.4.10) $$0 \le C_f(P//Q) \le \beta_f(r,R)$$

and

(4.4.11) $$\begin{aligned} 0 &\le \beta_f(r,R) - C_f(P//Q) \\ &\le \frac{f^{/}(R)-f^{/}(r)}{R-r}\left[(R-1)(1-r) - \chi^2(P//Q)\right] \\ &\le \frac{1}{4}(R-r)\left(f^{/}(R) - f^{/}(r)\right) \end{aligned}$$

where

(4.4.12) $$\beta_f(r,R) = \frac{(R-1)f(r)+(1-r)f(R)}{R-r}$$

$\chi^2(P//Q)$ and $C_f(P//Q)$ are as given by (4.4.4) and (4.1.2) respectively. In view of theorem (4.4.1), we have the following inequality

Result (4.4.2). Let $P, Q \in \Delta_n$ and $0 \le s \le 4$. If there exists r,R such that $0 < r \le \frac{p_i}{q_i} \le R < \infty \ \ \forall \ i \in \{1, 2, ..., n\}$ with $0 < r \le 1 \le R < \infty$, then we have

(4.4.13) $$0 \le \xi_s(P//Q) \le \mu_s(r,R)$$

(4.4.14) $$0 \le \xi_s(P//Q) \le \phi_s(r,R)$$

and

(4.4.15) $$\begin{aligned} 0 &\le \phi_s(r,R) - \xi_s(P//Q) \\ &\le K_s(r,R)\left[(R-1)(1-r) - \chi^2(P//Q)\right] \le \mu_s(r,R) \end{aligned}$$

where

(4.4.16) $$\mu_s(r,R) = \begin{cases} \left\{ \frac{1}{4}(R-r) \left\{ \begin{array}{l} \frac{1}{2}\left[(R-1)\left(\frac{R+1}{2}\right)^{s-2} - (r-1)\left(\frac{r+1}{2}\right)^{s-2}\right] \\ +(s-1)^{-1}\left[\left(\frac{R+1}{2}\right)^{s-1} - \left(\frac{r+1}{2}\right)^{s-1} - 2\right] \end{array} \right. ; s \ne 1 \right\} \\ \frac{1}{4}(R-r)\left[\frac{R-1}{R+1} - \frac{r-1}{r+1} + \ln\left(\frac{R+1}{r+1}\right)\right]; \qquad s = 1 \end{cases}$$

(4.4.17) $$\begin{aligned} K_s(r,R) &= \frac{\phi_s(R)-\phi_s(r)}{R-r} \\ &= \begin{cases} \begin{array}{l} \frac{1}{2}\left[(R-1)\left(\frac{R+1}{2}\right)^{s-2} - (r-1)\left(\frac{r+1}{2}\right)^{s-2}\right] \\ +(s-1)^{-1}\left[\left(\frac{R+1}{2}\right)^{s-1} - \left(\frac{r+1}{2}\right)^{s-1} - 2\right]; \ s \ne 1 \end{array} \\ \frac{R-1}{R+1} - \frac{r-1}{r+1} + \ln\left(\frac{R+1}{r+1}\right); \qquad s = 1 \end{cases} \end{aligned}$$

Also,

$$\mu_s(r,R) = \frac{1}{4}(R-r)^2 K_s(r,R)$$

(4.4.18)
$$\phi_s(r,R) = \frac{(R-1)\phi_s(r)+(1-r)\phi_s(R)}{R-r}$$
$$= \begin{cases} \frac{(R-1)(1-r)}{(R-r)(s-1)}\left[\left(\frac{R+1}{2}\right)^{s-1} - \left(\frac{r+1}{2}\right)^{s-1}\right]; & s \neq 1 \\ \frac{(1-r)(R-1)}{R-r}\ln\left(\frac{R+1}{r+1}\right); & s = 1 \end{cases}$$

Proof: The above results follows immediately from theorem (4.4.1), by taking $f(u) = \phi_s(u)$ where $\phi_s(u)$ is given by (4.4.6), then in this case we have $C_f(P//Q) = \xi_s(P//Q)$.

We have the following corollaries as particular cases of result (4.4.2).

Corollary (4.4.1). Under the condition of result (4.4.2), we have

(4.4.19)
$$0 \leq D(P//Q) \leq \frac{1}{4}(R-r)\left\{\frac{R-1}{R+1} - \frac{r-1}{r+1} + \ln\left(\frac{R+1}{r+1}\right)\right\}$$

(4.4.20)
$$0 \leq \frac{1}{2}\chi^2(P//Q) \leq \frac{1}{4}(R-r)(R-r-2)$$

Proof: (4.4.19) follows by taking s=1, (4.4.20) follows by taking s=2 in (4.4.13).

Corollary (4.4.2) .Under the conditions of result (4.4.2), we have

(4.4.21)
$$0 \leq D(P//Q) \leq \frac{(1-r)(R-1)}{R-r}\ln\left(\frac{R+1}{r+1}\right)$$

(4.4.22)
$$0 \leq \frac{1}{2}\chi^2(P//Q) \leq \frac{(R-1)(1-r)}{2}$$

Proof: (4.4.21) follows by taking s=1, (4.4.22) follows by taking s=2 in (4.4.14).

4.5 Main results

In this section, we present a theorem which generalizes the one obtained by Dragomir [47]. The result due to Dragomir [47] are limited only to chi-square divergence, while the theorem established here is given in terms of relative J- divergence of type s, that in particular lead us to bounds in terms of chi-square and relative J-divergence.

Theorem (4.5.1). Let $f: I \subset \Re_+ \to \Re$ the generating mapping be normalized i.e., $f(1) = 0$ and satisfies the assumption:

(1) f is twice differentiable on (r,R), where $0 \leq r \leq 1 \leq R < \infty$;

(2) there exists the real constants m,M with m<M such that

(4.5.1)
$$m \leq \frac{4}{sx+4-s}\left(\frac{x+1}{2}\right)^{3-s} f^{//}(x) \leq M \qquad \forall\ x \in (r,R), 0 \leq s \leq 4$$

If $P,Q\in\Delta_n$ are discrete probability distribution satisfying the assumption $0<r\leq\frac{p_i}{q_i}\leq R<\infty$. Then we have the inequality

(4.5.2) $$m\left[\phi_s\left(r,R\right)-\xi_s\left(P//Q\right)\right]\leq\beta_f\left(r,R\right)-C_f\left(P//Q\right)\leq M\left[\phi_s\left(r,R\right)-\xi_s\left(P//Q\right)\right]$$

where $C_f\left(P//Q\right),\xi_s\left(P//Q\right),\beta_f\left(r,R\right)$ and $\phi_s\left(r,R\right)$ are as given by (4.1.2), (4.4.5), (4.4.12) and (4.4.18) respectively.

Proof: Let us consider the function $F_{m,s}\left(u\right)$ and $F_{M,s}\left(u\right)$ given by

(4.5.3) $$F_{m,s}\left(u\right)=f\left(u\right)-m\phi_s\left(u\right)$$

and

(4.5.4) $$F_{M,s}\left(u\right)=M\phi_s\left(u\right)-f\left(u\right)$$

respectively, where m and M are as given by (4.5.1) and function $\phi_s\left(u\right)$ is as given by (4.4.6).

Since $f\left(u\right)$ and $\phi_s\left(u\right)$ are normalized, then $F_{m,s}\left(u\right)$ and $F_{M,s}\left(u\right)$ are also normalized i.e., $F_{m,s}\left(1\right)=0$ and $F_{M,s}\left(1\right)=0$. However the functions $f\left(u\right)$ and $\phi_s\left(u\right)$ are twice differentiable. Then in view of (4.4.8) and (4.5.1), we have

$$\begin{aligned}F^{//}_{m,s}\left(u\right)&=f^{//}\left(u\right)-m\phi^{//}_s\left(u\right)\\&=f^{//}\left(u\right)-m\left(\tfrac{u+1}{2}\right)^{s-3}\left(\tfrac{su+4-s}{4}\right)\\&=\left(\tfrac{u+1}{2}\right)^{s-3}\left(\tfrac{su+4-s}{4}\right)\left\{\tfrac{4}{su+4-s}\left(\tfrac{u+1}{2}\right)^{3-s}f^{//}\left(u\right)-m\right\}\geq0\end{aligned}$$

and

$$\begin{aligned}F^{//}_{M,s}\left(u\right)&=M\phi^{//}_s\left(u\right)-f^{//}\left(u\right)\\&=M\left(\tfrac{u+1}{2}\right)^{s-3}\left(\tfrac{su+4-s}{4}\right)-f^{//}\left(u\right)\\&=\left(\tfrac{u+1}{2}\right)^{s-3}\left(\tfrac{su+4-s}{4}\right)\left\{M-\tfrac{4}{su+4-s}\left(\tfrac{u+1}{2}\right)^{3-s}f^{//}\left(u\right)\right\}\geq0\end{aligned}$$

for all $u\in\left(r,R\right)$ and $0\leq s\leq4$. Then the functions $F_{m,s}\left(u\right)$ and $F_{M,s}\left(u\right)$ are convex on $\left(r,R\right)$.

We have seen that the real mappings $F_{m,s}\left(u\right)$ and $F_{M,s}\left(u\right)$ given by (4.5.3) and (4.5.4) respectively are normalized, twice differentiable and convex on (r, R). Applying the r.h.s of the inequality (4.4.10), we have

(4.5.5) $$C_{F_{m,s}}\left(P//Q\right)\leq\beta_{F_{m,s}}\left(r,R\right)$$

and

(4.5.6) $$C_{F_{M,s}}(P//Q) \leq \beta_{F_{M,s}}(r,R)$$

respectively.

Moreover,

(4.5.7) $$C_{F_{m,s}}(P//Q) = C_f(P//Q) - m\xi_s(P//Q)$$

and

(4.5.8) $$C_{F_{M,s}}(P//Q) = M\xi_s(P//Q) - C_f(P//Q)$$

In view of (4.5.5) and (4.5.7), we have

$$C_f(P//Q) - m\xi_s(P//Q) \leq \beta_{F_{m,s}}(r,R)$$

i.e.,

$$C_f(P//Q) - m\xi_s(P//Q) \leq \beta_f(r,R) - m\phi_s(r,R)$$

or

$$m\left[\phi_s(r,R) - \xi_s(P//Q)\right] \leq \beta_f(r,R) - C_f(P//Q)$$

Thus we have l.h.s of inequality (4.5.2).

Again in view of (4.5.6) and (4.5.8), we have

$$M\xi_s(P//Q) - C_f(P//Q) \leq \beta_{F_{M,s}}(r,R)$$

i.e.,

$$M\xi_s(P//Q) - C_f(P//Q) \leq M\phi_s(r,R) - \beta_f(r,R)$$

or

$$\beta_f(r,R) - C_f(P//Q) \leq M\left[\phi_s(r,R) - \xi_s(P//Q)\right]$$

Thus we have r.h.s of inequality (4.5.2).

4.6 Information bounds in terms of χ^2- divergence

In particular for s=2 in theorem (4.5.1), we have the following proposition.

Proposition (4.6.1). Let $f : I \subset \Re_+ \to \Re$ the generating mapping be normalized i.e., $f(1) = 0$ and satisfies the assumption:

(1) f is twice differentiable on (r,R), where $0 \leq r \leq 1 \leq R < \infty$;

(2) there exists the real constants m,M with m<M such that

(4.6.1) $$m \leq f^{//}(x) \leq M \qquad \forall\ x \in (r,R),$$

If $P, Q \in \Delta_n$ are discrete probability distribution satisfying the assumption $0 < r \leq \frac{p_i}{q_i} \leq R < \infty$, then we have the inequalities

(4.6.2) $$\frac{m}{2}\left[(R-1)(1-r) - \chi^2(P//Q)\right]$$
$$\leq \quad \beta_f(r,R) - C_f(P//Q)$$
$$\leq \frac{M}{2}\left[(R-1)(1-r) - \chi^2(P//Q)\right]$$

where $C_f(P//Q)$, $\beta_f(r,R)$ and $\chi^2(P//Q)$ are as given by (4.1.2), (4.4.12) and (4.4.4) respectively.

Result (4.6.1). Let $P, Q \in \Delta_n$ and $0 \leq s \leq 4$. If there exists r,R $(0 < r \leq 1 \leq R < \infty)$ such that $0 < r \leq \frac{p_i}{q_i} \leq R < \infty \quad \forall\ i \in \{1, 2, ..., n\}$. Then in view of proposition (4.6.1), we have

(4.6.3) $$\frac{1}{2}\left(\frac{R+1}{2}\right)^{s-3}\left(\frac{sR+4-s}{4}\right)\left[(R-1)(1-r) - \chi^2(P//Q)\right]$$
$$\leq \phi_s(r,R) - \xi_s(P//Q)$$
$$\leq \frac{1}{2}\left(\frac{r+1}{2}\right)^{s-3}\left(\frac{sr+4-s}{4}\right)\left[(R-1)(1-r) - \chi^2(P//Q)\right] ;0 \leq s < 2$$

(4.6.4) $$\frac{1}{2}\left(\frac{r+1}{2}\right)^{s-3}\left(\frac{sr+4-s}{4}\right)\left[(R-1)(1-r) - \chi^2(P//Q)\right]$$
$$\leq \phi_s(r,R) - \xi_s(P//Q)$$
$$\leq \frac{1}{2}\left(\frac{R+1}{2}\right)^{s-3}\left(\frac{sR+4-s}{4}\right)\left[(R-1)(1-r) - \chi^2(P//Q)\right] ;2 < s \leq 4$$

We observe that the expression (4.6.4) is valid for $2 < s \leq 6$, but we have taken the range $0 \leq s \leq 4$, because of the convexity and non-negativity of the function $\xi_s(P//Q)$ in the range.

Proof: Let us consider $f(u) = \phi_s(u)$, where $\phi_s(u)$ is as given by (4.4.6), then according the expression (4.4.8), we have

$$\phi_s^{//}(u) = \left(\frac{u+1}{2}\right)^{s-3}\left(\frac{su+4-s}{4}\right)$$

Let us define the function $g : [r,R] \to \Re$ such that

$$g(u) = \left(\frac{u+1}{2}\right)^{s-3}\left(\frac{su+4-s}{4}\right)$$

Then

(4.6.5) $$\sup_{u\in[r,R]} g(u) = \begin{cases} \left(\frac{r+1}{2}\right)^{s-3}\left(\frac{sr+4-s}{4}\right) & ;0 \leq s < 2 \\ \left(\frac{R+1}{2}\right)^{s-3}\left(\frac{sR+4-s}{4}\right) & ;2 < s \leq 4 \end{cases}$$

(4.6.6) $$\inf_{u\in[r,R]} g(u) = \begin{cases} \left(\frac{R+1}{2}\right)^{s-3}\left(\frac{sR+4-s}{4}\right) & ;0 \leq s < 2 \\ \left(\frac{r+1}{2}\right)^{s-3}\left(\frac{sr+4-s}{4}\right) & ;2 < s \leq 4 \end{cases}$$

In view of (4.6.5) and (4.6.6) and proposition (4.6.1), we have proof of the result.

Corollary (4.6.1). In view of result (4.6.1), we have following corollaries.

(4.6.7)
$$\frac{R+3}{2(R+1)^2}\left[(R-1)(1-r)-\chi^2(P//Q)\right] \leq \frac{(1-r)(R-1)}{R-r}\ln\left(\frac{R+1}{r+1}\right)-D(P//Q) \leq \frac{r+3}{2(r+1)^2}\left[(R-1)(1-r)-\chi^2(P//Q)\right]$$

Proof: (4.6.7) follows by taking s=1 in (4.6.3), while for s=2, we have equality sign.

4.7. Information bounds in terms of relative J- divergence

In particular for s=1 in theorem (4.5.1), we have the following proposition.

Proposition (4.7.1). Let $f : I \subset \Re_+ \to \Re$ the generating mapping be normalized i.e., $f(1) = 0$ and satisfies the assumption:

(1) f is twice differentiable on (r, R), where $0 \leq r \leq 1 \leq R < \infty$;

(2) there exists the real constants m,M with m<M such that

(4.7.1)
$$m \leq \frac{(x+1)^2}{x+3} f^{//}(x) \leq M \qquad \forall\ x \in (r, R),$$

If $P, Q \in \Delta_n$ are discrete probability distribution satisfying the assumption $0 < r \leq \frac{p_i}{q_i} \leq R < \infty$, then we have the inequalities

(4.7.2)
$$m\left[\frac{(1-r)(R-1)}{R-r}\ln\left(\frac{R+1}{r+1}\right)-D(P//Q)\right] \leq \beta_f(r,R)-\xi_s(P//Q) \leq M\left[\frac{(1-r)(R-1)}{R-r}\ln\left(\frac{R+1}{r+1}\right)-D(P//Q)\right]$$

where $C_f(P//Q)$, $\beta_f(r,R)$ and $D(P//Q)$ are as given by (4.1.2), (4.4.12) and (4.4.1) respectively.

Result (4.7.1). Let $P, Q \in \Delta_n$ and $0 \leq s \leq 4$. If there exists r,R $(0 < r \leq 1 \leq R < \infty)$ such that $0 < r \leq \frac{p_i}{q_i} \leq R < \infty \ \ \forall\ i \in \{1, 2, ..., n\}$. Then in view of proposition (4.7.1), we have

(4.7.3)
$$\left(\frac{R+1}{2}\right)^{s-1}\left(\frac{sR+4-s}{R+3}\right)\left[\frac{(1-r)(R-1)}{R-r}\ln\left(\frac{R+1}{r+1}\right)-D(P//Q)\right] \leq \phi_s(r,R)-\xi_s(P//Q) \leq \left(\frac{r+1}{2}\right)^{s-1}\left(\frac{sr+4-s}{r+3}\right)\left[\frac{(1-r)(R-1)}{R-r}\ln\left(\frac{R+1}{r+1}\right)-D(P//Q)\right];0 \leq s < 1$$

$$\text{(4.7.4)}\quad \left(\frac{r+1}{2}\right)^{s-1}\left(\frac{sr+4-s}{r+3}\right)\left[\frac{(1-r)(R-1)}{R-r}\ln\left(\frac{R+1}{r+1}\right)-D\left(P//Q\right)\right]$$
$$\leq \phi_s\left(r,R\right)-\xi_s\left(P//Q\right)$$
$$\leq \left(\frac{R+1}{2}\right)^{s-1}\left(\frac{sR+4-s}{R+3}\right)\left[\frac{(1-r)(R-1)}{R-r}\ln\left(\frac{R+1}{r+1}\right)-D\left(P//Q\right)\right];1<s\leq 4$$

Proof: Let us consider $f\left(u\right)=\phi_s\left(u\right)$, where $\phi_s\left(u\right)$ is as given by (4.4.6). Then according to the expression (4.4.8), we have

$$\phi_s^{//}\left(u\right)=\left(\frac{u+1}{2}\right)^{s-3}\left(\frac{su+4-s}{4}\right)$$

Let us define the function $g:[r,R]\to\Re$ such that

$$g\left(u\right)=\frac{(u+1)^2}{u+3}\phi_s^{//}\left(u\right)$$

Then

$$g\left(u\right)=\left(\frac{u+1}{2}\right)^{s-1}\left(\frac{su+4-s}{u+3}\right)$$

$$\text{(4.7.5)}\qquad \sup_{u\in[r,R]} g\left(u\right)=\begin{cases}\left(\frac{r+1}{2}\right)^{s-1}\left(\frac{sr+4-s}{r+3}\right);0\leq s<1\\ \left(\frac{R+1}{2}\right)^{s-1}\left(\frac{sR+4-s}{R+3}\right);1<s\leq 4\end{cases}$$

$$\text{(4.7.6)}\qquad \inf_{u\in[r,R]} g\left(u\right)=\begin{cases}\left(\frac{R+1}{2}\right)^{s-1}\left(\frac{sR+4-s}{R+3}\right);0\leq s<1\\ \left(\frac{r+1}{2}\right)^{s-1}\left(\frac{sr+4-s}{r+3}\right);1<s\leq 4\end{cases}$$

In above cases we have taken the range $0\leq s\leq 4$, because of the convexity and non negativity of the function $\xi_s\left(P//Q\right)$ in the range. In view of (4.7.5) and (4.7.6) and proposition (4.7.1), we have proof of the result (4.7.1).

Corollary (4.7.1) . In view of result (4.7.1), we have the following corollary.

$$\text{(4.7.7)}\qquad \frac{(r+1)^2}{(r+3)}\left[\frac{(1-r)(R-1)}{R-r}\ln\left(\frac{R+1}{r+1}\right)-D\left(P//Q\right)\right]$$
$$\leq\frac{(R-1)(1-r)}{2}-\frac{1}{2}\chi^2\left(P//Q\right)$$
$$\leq\frac{(R+1)^2}{(R+3)}\left[\frac{(1-r)(R-1)}{R-r}\ln\left(\frac{R+1}{r+1}\right)-D\left(P//Q\right)\right]$$

Proof: (4.7.7) follows by taking s=2 in (4.7.4) and for s=1, we have equality sign.

Shannon's inequalities are well known in the field of information theory. Researchers have found many inequalities by using well known Holders's inequality, Jensen's inequality etc and have found many applications for Shannon's entropy , Renyi's entropy etc. Dragomir [46] have found many information inequalities for the logarithmic mapping and convex mappings by using Jensen's inequality. In this chapter, several generalized information inequalities have been obtained by introducing a independent variable 's' and applied for Shannon's entropy, Renyi's entropy and mutual information . The information inequalities obtained here are not only new but also generalizes some established inequalities in information theory given by Dragomir [46].

Also some upper bounds for the relative arithmetic geometric divergence measure have been obtained by using some classical inequalities like Kantorovic inequality, Diaz-Metcalf inequality and other inequality for logarithmic function.

5.1. Introduction

The following converse of Jensen's discrete inequality for convex mappings of a real variable was proved by Dragomir and Ionescu [40].

Theorem (5.1.1). Let $f : I \subseteq R \to R$ be a differentiable convex function on the interval $I,\ x_i \in \overset{\bullet}{I}\left(\overset{\bullet}{I} \text{ is the interior of } I\right),\ p_i \geq 0\,(i = 1, 2, ..., n)$ and $\sum\limits_{i=1}^{n} p_i = 1$. Then we have the inequality

$$(5.1.1)\qquad 0 \leq \sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i x_i f^{/}(x_i) - \sum_{i=1}^{n} p_i x_i \sum_{i=1}^{n} p_i f^{/}(x_i)$$

They also pointed out some natural applications of (5.1.1) in connection to the arithmetic geometric mean inequality, the generalized triangular inequality etc. A generalization of (5.1.1) for differentiable convex mappings of several variables was obtained by Dragomir and Goh [39]. They also considerd the following analytical inequality for the logarithmic mapping.

Theorem (5.1.2) . Let $x_i, p_i > 0, (i = 1, 2, ..., n)$ with $\sum\limits_{i=1}^{n} p_i = 1$ and $b > 1$. Then

$$(5.1.2)\qquad 0 \leq \log_b\left(\sum_{i=1}^{n} p_i x_i\right) - \sum_{i=1}^{n} p_i \log_b x_i$$

$$\leq \frac{1}{\ln b}\left[\sum_{i=1}^{n}\frac{p_i}{x_i}\sum_{i=1}^{n}p_i x_i - 1\right]$$
$$= \frac{1}{\ln b}\sum_{1\leq i\leq j\leq n} p_i p_j \frac{(x_i - x_j)^2}{x_i x_j}$$

equality holds in (5.1.2) iff $x_1 = x_2 = ... = x_n$.

They applied inequality (5.1.2) in information theory for the entropy mapping, conditional entropy, mutual information etc. An integral version of (5.1.2) was employed by Dragomir and Goh [39] to obtain different bounds for the entropy, conditional entropy and mutual information for continuous random variables.

In the next section, some generalized inequalities have been obtained and applications for Shannon's entropy, Renyi's entropy and mutual information are also given. The results obtained in this section generalizes some results of Dragomir [46]. This work is published in International Journal of pure and applied Mathematics, Vol 31 (2) PP 253-263 (2006) (Baig and Rayees [11]).

5.2 Some generalized inequalities for convex functions

Theorem (5.2.1). Let $f : [a, b] \to \Re$ be twice differentiable on (a, b) , continuous in $[a, b]$ and $m \leq x^{2-s} f^{//}(x) \leq M \;\; \forall\, x \in (a, b)$ and $s \in \Re$. If $x_i \in [a, b]$, $i = 1, 2, ..., n$ and $p = p_i$, $i = 1, 2, ..., n$ is a probability distribution, then

(5.2.1) $$m\left[\sum_{i=1}^{n}p_i\phi_s(x_i) - \phi_s\left(\sum_{i=1}^{n}p_i x_i\right)\right]$$
$$\leq \sum_{i=1}^{n}p_i f(x_i) - f\left(\sum_{i=1}^{n}p_i x_i\right)$$
$$\leq M\left[\sum_{i=1}^{n}p_i\phi_s(x_i) - \phi_s\left(\sum_{i=1}^{n}p_i x_i\right)\right]$$

Proof: Let us consider the function

(5.2.2) $$\phi_s(x) = \left\{\begin{array}{ll} [s(s-1)]^{-1}[x^s - 1 - s(x-1)] \;; & s \neq 0, 1 \\ x - 1 - \ln x \;; & s = 0 \\ 1 - x + \ln x \;; & s = 1 \end{array}\right\}$$

Then

(5.2.3) $$\phi_s^{/}(x) = \left\{\begin{array}{ll} (s-1)^{-1}[x^{s-1} - 1] \;; & s \neq 0, 1 \\ 1 - x^{-1} \;; & s = 0 \\ \ln x \;; & s = 1 \end{array}\right\}$$

and

(5.2.4) $$\phi_s^{//}(x) = \left\{ \begin{array}{c} x^{s-2}\ ;\ s \neq 0,1 \\ x^{-2}\ ;\quad s=0 \\ x^{-1};\quad s=1 \end{array} \right\}$$

Here $\phi_s^{//}(x) > 0, \quad \forall\ x > 0$, here $\phi_s(x)$ is strictly convex for all $x > 0$ and $s \in \Re$.

Let us consider the function $g : [a,b] \to \Re$

$$g(x) = f(x) - m\phi_s(x) \quad , \quad x \in (a,b)\ , \quad s \in \Re$$

$$g^{//}(x) = f^{//}(x) - m\phi_s^{//}(x) = x^{s-2}\left(x^{2-s} f^{//}(x) - m\right) \geq 0$$

which shows that the mapping $g(x)$ is convex on $[a,b]$.

Applying Jensen's discrete inequality for the convex mapping $g(x)$, i.e.

$$g\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i g(x_i)$$

Therefore,

$$f\left(\sum_{i=1}^{n} p_i x_i\right) - m\phi_s\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i\left[f(x_i) - m\phi_s(x_i)\right]$$

$$\sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i\right) \geq m\left[\sum_{i=1}^{n} p_i\phi_s(x_i) - \phi_s\left(\sum_{i=1}^{n} p_i x_i\right)\right]$$

The first inequality in (5.2.1) is proved.

The proof of the second inequality goes likewise for the mapping $h : [a,b] \to \Re,\ h(x) = M\phi_s(x) - f(x)$ which is convex on $[a,b]$.

Corollary (5.2.1) . Let $x_i, w_i > 0 \quad (i = 1,2,...,n)$ and put $W_n = \sum_{i=1}^{n} w_i$. Also consider Arithmetic mean $A_n(w,a) = \frac{1}{W_n}\sum_{i=1}^{n} w_i a_i$. If $x_i \in [m,M] \subset (0,\infty)\ ,\ i = 1,2,...,n$ and $s \in \Re$, then we have the inequalities

(5.2.5) $$\exp\left[\frac{1}{MW_n}\left\{\sum_{i=1}^{n} w_i\phi_s(x_i) - \phi_s\left(\sum_{i=1}^{n} w_i x_i\right)\right\}\right]$$

$$\leq \frac{\left[\prod_{i=1}^{n} x_i^{w_i x_i}\right]^{\frac{1}{W_n}}}{\left[A_n(w,a)\right]^{A_n(w,a)}}$$

$$\leq \exp\left[\frac{1}{mW_n}\left\{\sum_{i=1}^{n} w_i\phi_s(x_i) - \phi_s\left(\sum_{i=1}^{n} w_i x_i\right)\right\}\right]$$

Proof: Consider the mapping $f(x) = x\ln x,\ x > 0$.Then

$$f^{/}(x) = \ln x + 1,\ x \in (0,\infty)$$

$$f^{//}(x) = \frac{1}{x}, \quad x \in (0,\infty)$$

which shows that f is strictly convex on the interval $(0,\infty)$.

$$\underset{x\in[m,M]}{Inf}\ f^{//}(x)=\frac{1}{M},\ \ \underset{x\in[m,M]}{Sup}\ f^{//}(x)=\frac{1}{m}$$

Applying theorem (5.2.1) for this mapping and $p_i=\frac{w_i}{W_n},\ i=1,2,...,n.$ We deduce (5.2.5).

The case of equality follows by the strict convexity of the mapping

$g(x)=x\ln x-\frac{1}{M}\phi_s(x)\,,\ h(x)=\frac{1}{m}\phi_s(x)-x\ln x$ on (m,M).

Theorem (5.2.2). Let $f:[a,b]\to\Re_+$ be twice differetiable on (a,b), continuous in $[a,b]$ and $m\le x^{2-s}f^{//}(x)\le M\quad\forall\,x\in[a,b]$ and $s\in\Re$. If $x_i\in[a,b]\,,i=1,2,...,n$ and $p=p_i\ (i=1,2,...,n)$ is a probability distribution. Then we have the inequalities

(5.2.6)
$$\frac{1}{2}\sum_{i,j}p_ip_j(x_i-x_j)\left(f^{/}(x_i)-f^{/}(x_j)\right)$$
$$+M\left[\sum_{i=1}^{n}p_i\phi_s(x_i)-\phi_s\left(\sum_{i=1}^{n}p_ix_i\right)-\frac{1}{2}\sum_{i,j}p_ip_j(x_i-x_j)\left(\phi_s^{/}(x_i)-\phi^{/}(x_j)\right)\right]$$
$$\le\sum_{i=1}^{n}p_if(x_i)-f\left(\sum_{i=1}^{n}p_ix_i\right)$$
$$\le\ \frac{1}{2}\sum_{i,j}p_ip_j(x_i-x_j)\left(f^{/}(x_i)-f^{/}(x_j)\right)$$
$$+m\left[\sum_{i=1}^{n}p_i\phi_s(x_i)-\phi_s\left(\sum_{i=1}^{n}p_ix_i\right)-\frac{1}{2}\sum_{i,j}p_ip_j(x_i-x_j)\left(\phi_s^{/}(x_i)-\phi^{/}(x_j)\right)\right]$$

Proof: Consider the mapping

$$g:[a,b]\to\Re,\quad g(x)=f(x)-m\phi_s(x)\,,\quad s\in\Re,\quad x\in(a,b)$$

where $\phi_s(x)$ is given in (5.2.2).

Then g is twice differentiable on (a,b) and

$$g^{//}(x)=f^{//}(x)-m\phi_s^{//}(x)=f^{//}(x)-mx^{s-2}$$
$$g^{//}(x)=x^{s-2}[x^{2-s}f^{//}(x)-m]\ge0\ \ \forall\ x\in(a,b)\ \ ,s\in\Re.$$

which shows that the mapping is convex on $x\in[a,b]\,,s\in\Re$. Also $\phi_s^{//}(x)$ is given by (5.2.4). We apply inequality (5.1.1) for the convex mapping g, i.e

$$0\le\sum_{i=1}^{n}p_ig(x_i)-g\left(\sum_{i=1}^{n}p_ix_i\right)\le\frac{1}{2}\sum_{i,j}p_ip_j(x_i-x_j)\left(g^{/}(x_i)-g^{/}(x_j)\right)$$

to obtain

$$0 \leq \sum_{i=1}^{n} p_i \left[f\left(x_i\right) - m\phi_s\left(x_i\right)\right] - \left[f\left(\sum_{i=1}^{n} p_i x_i\right) - m\phi_s\left(\sum_{i=1}^{n} p_i x_i\right)\right]$$

$$\leq \frac{1}{2}\sum_{ij} p_i p_j \left(x_i - x_j\right)\left[f^{/}\left(x_i\right) - f^{/}\left(x_j\right) - m\phi_s^{/}\left(x_i\right) + m\phi_s^{/}\left(x_j\right)\right]$$

$$= \frac{1}{2}\sum_{ij} p_i p_j \left(x_i - x_j\right)\left(f^{/}\left(x_i\right) - f^{/}\left(x_j\right)\right) - \frac{m}{2}\sum_{ij} p_i p_j \left(x_i - x_j\right)\left(\phi_s^{/}\left(x_i\right) - \phi_s^{/}\left(x_j\right)\right)$$

Thus

$$\sum_{i=1}^{n} p_i f\left(x_i\right) - f\left(\sum_{i=1}^{n} p_i x_i\right)$$

$$\leq \frac{1}{2}\sum_{ij} p_i p_j \left(x_i - x_j\right)\left(f^{/}\left(x_i\right) - f^{/}\left(x_j\right)\right)$$

$$+m\left[\sum_{i=1}^{n} p_i \phi_s\left(x_i\right) - \phi_s\left(\sum_{i=1}^{n} p_i x_i\right) - \frac{1}{2}\sum_{ij} p_i p_j \left(x_i - x_j\right)\left(\phi_s^{/}\left(x_i\right) - \phi_s^{/}\left(x_j\right)\right)\right]$$

and the second inequality is proved.

The proof of the first inequality goes likewise for the mapping $h : [a, b] \to \Re, h\left(x\right) = M\phi_s\left(x\right) - f\left(x\right).$

Corollary (5.2.2). Let $x_i \in [m, M] \subset (0, \infty)$ and $p_i > 0 \quad (i = 1, 2, ..., n)$ with $\sum_{i=1}^{n} p_i = 1$. Then we have the inequality

(5.2.7) $$\frac{1}{2}\sum_{ij} p_i p_j \frac{\left(x_i - x_j\right)^2}{x_i x_j}$$

$$+\frac{1}{m^2}\left[\sum_{i=1}^{n} p_i \phi_s\left(x_i\right) - \phi_s\left(\sum_{i=1}^{n} p_i x_i\right) - \frac{1}{2}\sum_{ij} p_i p_j \left(x_i - x_j\right)\left(\phi_s^{/}\left(x_i\right) - \phi_s^{/}\left(x_j\right)\right)\right]$$

$$\leq \ln\left(\sum_{i=1}^{n} p_i x_i\right) - \sum_{i=1}^{n} p_i \ln x_i$$

$$\leq \frac{1}{2}\sum_{ij} p_i p_j \frac{\left(x_i - x_j\right)^2}{x_i x_j}$$

$$+\frac{1}{M^2}\left[\sum_{i=1}^{n} p_i \phi_s\left(x_i\right) - \phi_s\left(\sum_{i=1}^{n} p_i x_i\right) - \frac{1}{2}\sum_{ij} p_i p_j \left(x_i - x_j\right)\left(\phi_s^{/}\left(x_i\right) - \phi_s^{/}\left(x_j\right)\right)\right]$$

Proof: Consider the mapping $f : [m, M] \subset (0, \infty)$ given by $f\left(x\right) = -\ln x$

Then

$$f^{/}\left(x\right) = -\frac{1}{x}, \qquad f^{//}\left(x\right) = \frac{1}{x^2}$$

Also

$$\underset{x\in[m,M]}{Inf} f^{//}(x)=\frac{1}{M^2},\quad \underset{x\in[m,M]}{Sup} f^{//}(x)=\frac{1}{m^2}$$

Applying inequality (5.2.6) for this mapping, we can write

$$\frac{1}{2}\sum_{ij}p_ip_j\frac{(x_i-x_j)^2}{x_ix_j}$$
$$+\frac{1}{m^2}\left[\sum_{i=1}^{n}p_i\phi_s(x_i)-\phi_s\left(\sum_{i=1}^{n}p_ix_i\right)-\frac{1}{2}\sum_{ij}p_ip_j(x_i-x_j)\left(\phi_s^{/}(x_i)-\phi_s^{/}(x_j)\right)\right]$$
$$\leq \ln\left(\sum_{i=1}^{n}p_ix_i\right)-\sum_{i=1}^{n}p_i\ln x_i$$
$$\leq \frac{1}{2}\sum_{ij}p_ip_j\frac{(x_i-x_j)^2}{x_ix_j}$$
$$+\frac{1}{M^2}\left[\sum_{i=1}^{n}p_i\phi_s(x_i)-\phi_s\left(\sum_{i=1}^{n}p_ix_i\right)-\frac{1}{2}\sum_{ij}p_ip_j(x_i-x_j)\left(\phi_s^{/}(x_i)-\phi_s^{/}(x_j)\right)\right]$$

which is equivalent to (5.2.7). The case of equality follows by the strict convexity of the mapping $g(x)=-\ln x-\frac{1}{M^2}\phi_s(x)$, $h(x)=\frac{1}{m^2}\phi_s(x)+\ln x$ on the interval $[m,M]$.

Corollary (5.2.3). Let $x_i\in[m,M]\subset(0,\infty)$, also $s\in\Re,\ p_i>0\ \ (i=1,2,...,n)$ with $\sum_{i=1}^{n}p_i=1$. Then we have the inequlity

$$\text{(5.2.8)}\qquad \exp\left[\begin{array}{c}\frac{1}{2}\sum_{ij}p_ip_j\frac{(x_i-x_j)^2}{x_ix_j}\\ +\frac{1}{m^2}\left[\sum_{i=1}^{n}p_i\phi_s(x_i)-\phi_s\left(\sum_{i=1}^{n}p_ix_i\right)-\frac{1}{2}\sum_{ij}p_ip_j(x_i-x_j)\left(\phi_s^{/}(x_i)-\phi_s^{/}(x_j)\right)\right]\end{array}\right]$$
$$\leq\frac{A_n(p,x)}{G_n(p,x)}$$
$$\leq \exp\left[\begin{array}{c}\frac{1}{2}\sum_{ij}p_ip_j\frac{(x_i-x_j)^2}{x_ix_j}\\ +\frac{1}{M^2}\left[\sum_{i=1}^{n}p_i\phi_s(x_i)-\phi_s\left(\sum_{i=1}^{n}p_ix_i\right)-\frac{1}{2}\sum_{ij}p_ip_j(x_i-x_j)\left(\phi_s^{/}(x_i)-\phi_s^{/}(x_j)\right)\right]\end{array}\right]$$

The equality holds in (5.2.8) iff $x_1=x_2=...=x_n$.

The proof is obvious by (5.2.7). Also, $A_n(p,x)=\sum_{i=1}^{n}p_ix_i\ (A.M)$, $G_n(p,x)=\prod_{i=1}^{n}x_i^{p_i}\ (G.M)$.

If in (5.2.8), we put x instead of $\frac{1}{x}$, we obtain the following corollary.

Corollary (5.2.4). Let $x_i, p_i,\ i = 1, 2, ..., n$ be as in corollary (5.2.3). Then we have the inequality

$$\text{(5.2.9)}\quad \exp\left[\begin{array}{c} \frac{1}{2}\sum_{ij} p_i p_j \frac{(x_i - x_j)^2}{x_i x_j} \\ +\frac{1}{m^2}\left[\begin{array}{c} \sum_{i=1}^{n} p_i \phi_s(x_i) - \phi_s\left(\sum_{i=1}^{n} p_i x_i\right) \\ -\frac{1}{2}\sum_{ij} p_i p_j (x_i - x_j)\left(\phi_s^{/}(x_i) - \phi_s^{/}(x_j)\right) \end{array}\right] \end{array}\right]$$

$$\leq \frac{G_n(p,x)}{H_n(p,x)}$$

$$\leq \exp\left[\begin{array}{c} \frac{1}{2}\sum_{ij} p_i p_j \frac{(x_i - x_j)^2}{x_i x_j} \\ +\frac{1}{M^2}\left[\begin{array}{c} \sum_{i=1}^{n} p_i \phi_s(x_i) - \phi_s\left(\sum_{i=1}^{n} p_i x_i\right) \\ -\frac{1}{2}\sum_{ij} p_i p_j (x_i - x_j)\left(\phi_s^{/}(x_i) - \phi_s^{/}(x_j)\right) \end{array}\right] \end{array}\right]$$

The equality holds in (5.2.9) iff $x_1 = x_2 = ... = x_n$.

Also, $H_n(p,x) = \frac{1}{\sum_{i=1}^{n} \frac{p_i}{x_i}}$ (Harmonic mean) .

5.3 Application in Shannon's entropy

The following inequalities for the logarithmic mapping holds.

Lemma (5.3.1). Let $\xi_i \in [m, M] \subset (0, \infty), p_i > 0,\ \ i = 1, 2, ..., n$ with $\sum_{i=1}^{n} p_i = 1$ and $s \in \Re$. Then we have the inequality

$$\text{(5.3.1)}\qquad \frac{1}{M}\left[\sum_{i=1}^{n} p_i \phi_s(\xi_i) - \phi_s\left(\sum_{i=1}^{n} p_i \xi_i\right)\right] \leq \sum_{i=1}^{n} p_i \xi_i \ln \xi_i - \sum_{i=1}^{n} p_i \xi_i \ln\left(\sum_{i=1}^{n} p_i \xi_i\right)$$

$$\leq \frac{1}{m}\left[\sum_{i=1}^{n} p_i \phi_s(\xi_i) - \phi_s\left(\sum_{i=1}^{n} p_i \xi_i\right)\right]$$

The case of equality holds iff $\xi_1 = \xi_2 = ... = \xi_n$.

The proof is obvious by theorem (5.2.1) for the convex mapping

$$f : [0, \infty) \to \Re, \qquad f(x) = x \ln x$$

Corollary (5.3.1). Under the assumption for $\xi_i\ (i = 1, 2, ..., n)$, we have

$$(5.3.2)\quad 0\leq\frac{1}{M}\left[\sum_{i=1}^{n}\phi_s(\xi_i)-\phi_s\left(\sum_{i=1}^{n}\xi_i\right)\right]\leq\sum_{i=1}^{n}\xi_i\ln\xi_i-\sum_{i=1}^{n}\xi_i\ln\left(\sum_{i=1}^{n}\frac{1}{n}\xi_i\right)$$
$$\leq\frac{1}{m}\left[\sum_{i=1}^{n}\phi_s(\xi_i)-\phi_s\left(\sum_{i=1}^{n}\xi_i\right)\right]$$

equality holds iff $\xi_1=\xi_2=....=\xi_n$.

The proof is obvious by lemma (5.3.1), choosing $p_i=\frac{1}{n}\,(i=1,2,...,n)$.

Theorem (5.3.1). Let X be a random variable with probability distribution $p_i\,(i=1,2,...,n)$. Assume that $p=\min\{p_i/i=1,2,...,n\}>0$ and $P=\max\{p_i/i=1,2,...,n\}<1$.Then

$$(5.3.3)\quad \frac{1}{2}\sum_{ij}p_ip_j(p_j-p_i)^2$$
$$+P^2\left[\sum_{i=1}^{n}p_i\phi_s\left(\frac{1}{p_i}\right)-\phi_s(n)-\frac{1}{2}\sum_{ij}(p_j-p_i)\left(\phi_s^{/}\left(\frac{1}{p_i}\right)-\phi_s^{/}\left(\frac{1}{p_j}\right)\right)\right]$$
$$\leq\ln(x)-H(X)$$
$$\leq\frac{1}{2}\sum_{ij}p_ip_j(p_j-p_i)^2$$
$$+p^2\left[\sum_{i=1}^{n}p_i\phi_s\left(\frac{1}{p_i}\right)-\phi_s(n)-\frac{1}{2}\sum_{ij}(p_j-p_i)\left(\phi_s^{/}\left(\frac{1}{p_i}\right)-\phi_s^{/}\left(\frac{1}{p_j}\right)\right)\right]$$

Proof: If we choose $x_i=\frac{1}{p_i}\in\left[\frac{1}{P},\frac{1}{p}\right]$ in (5.2.7). We can deduce that (5.3.3) with $\left(m=\frac{1}{P},M=\frac{1}{p}\right)$.

5.4 Application in Renyi's entropy

(1) If $x_i=p_i^{\alpha-1}\,(i=1,2,...,n);\alpha\in(0,1)$, then $P^{\alpha-1}\leq x_i\leq p^{\alpha-1}$, then by (5.2.7), we deduce that

$$\frac{1}{2}\sum_{ij}p_ip_j\frac{\left(p_i^{\alpha-1}-p_j^{\alpha-1}\right)}{p_i^{\alpha-1}p_j^{\alpha-1}}$$
$$+\frac{1}{P^{2(\alpha-1)}}\left[\sum_{i=1}^{n}p_i\phi_s\left(p_i^{\alpha-1}\right)-\phi_s\left(\sum_{i=1}^{n}p_i^{\alpha}\right)-\frac{1}{2}\sum_{ij}p_ip_j\left(p_i^{\alpha-1}-p_j^{\alpha-1}\right)\left(\phi_s^{/}\left(p_i^{\alpha-1}\right)-\phi_s^{/}\left(p_j^{\alpha-1}\right)\right)\right]$$
$$\leq\ln\left(\sum_{i=1}^{n}p_i^{\alpha}\right)-\sum_{i=1}^{n}p_i\ln p_i^{\alpha-1}$$
$$\leq\frac{1}{2}\sum_{ij}p_ip_j\frac{\left(p_i^{\alpha-1}-p_j^{\alpha-1}\right)}{p_i^{\alpha-1}p_j^{\alpha-1}}$$

$$+\frac{1}{p^{2(\alpha-1)}}\left[\sum_{i=1}^{n}p_i\phi_s\left(p_i^{\alpha-1}\right)-\phi_s\left(\sum p_i^{\alpha}\right)-\frac{1}{2}\sum_{ij}p_ip_j\left(p_i^{\alpha-1}-p_j^{\alpha-1}\right)\left(\phi_s^{/}\left(p_i^{\alpha-1}\right)-\phi_s^{/}\left(p_j^{\alpha-1}\right)\right)\right]$$

$$\frac{1}{2}\sum_{ij}p_ip_j\frac{\left(p_i^{\alpha-1}-p_j^{\alpha-1}\right)}{p_i^{\alpha-1}p_j^{\alpha-1}}$$

$$+\frac{1}{P^{2(\alpha-1)}}\left[\sum_{i=1}^{n}p_i\phi_s\left(p_i^{\alpha-1}\right)-\phi_s\left(\sum_{i=1}^{n}p_i^{\alpha}\right)-\frac{1}{2}\sum_{ij}p_ip_j\left(p_i^{\alpha-1}-p_j^{\alpha-1}\right)\left(\phi_s^{/}\left(p_i^{\alpha-1}\right)-\phi_s^{/}\left(p_j^{\alpha-1}\right)\right)\right]$$

$$\leq(1-\alpha)\left[H_\alpha(X)-H(X)\right]$$

$$\leq\frac{1}{2}\sum_{ij}p_ip_j\frac{\left(p_i^{\alpha-1}-p_j^{\alpha-1}\right)}{p_i^{\alpha-1}p_j^{\alpha-1}}$$

$$+\frac{1}{p^{2(\alpha-1)}}\left[\sum_{i=1}^{n}p_i\phi_s\left(p_i^{\alpha-1}\right)-\phi_s\left(\sum_{i=1}^{n}p_i^{\alpha}\right)-\frac{1}{2}\sum_{ij}p_ip_j\left(p_i^{\alpha-1}-p_j^{\alpha-1}\right)\left(\phi_s^{/}\left(p_i^{\alpha-1}\right)-\phi_s^{/}\left(p_j^{\alpha-1}\right)\right)\right]$$

(2) If If $x_i=p_i^{\alpha-1}\,(i=1,2,...,n)\,;\alpha\in(1,\infty)$, Then $p^{\alpha-1}\leq x_i\leq P^{\alpha-1}$, then by (5.2.7), we deduce that

$$\frac{1}{2}\sum_{ij}p_ip_j\frac{\left(p_i^{\alpha-1}-p_j^{\alpha-1}\right)}{p_i^{\alpha-1}p_j^{\alpha-1}}$$

$$+\frac{1}{P^{2(\alpha-1)}}\left[\sum_{i=1}^{n}p_i\phi_s\left(p_i^{\alpha-1}\right)-\phi_s\left(\sum_{i=1}^{n}p_i^{\alpha}\right)-\frac{1}{2}\sum_{ij}p_ip_j\left(p_i^{\alpha-1}-p_j^{\alpha-1}\right)\left(\phi_s^{/}\left(p_i^{\alpha-1}\right)-\phi_s^{/}\left(p_j^{\alpha-1}\right)\right)\right]$$

$$\leq(\alpha-1)\left[H(X)-H_\alpha(X)\right]$$

$$\leq\frac{1}{2}\sum_{ij}p_ip_j\frac{\left(p_i^{\alpha-1}-p_j^{\alpha-1}\right)}{p_i^{\alpha-1}p_j^{\alpha-1}}$$

$$+\frac{1}{p^{2(\alpha-1)}}\left[\sum_{i=1}^{n}p_i\phi_s\left(p_i^{\alpha-1}\right)-\phi_s\left(\sum_{i=1}^{n}p_i^{\alpha}\right)-\frac{1}{2}\sum_{ij}p_ip_j\left(p_i^{\alpha-1}-p_j^{\alpha-1}\right)\left(\phi_s^{/}\left(p_i^{\alpha-1}\right)-\phi_s^{/}\left(p_j^{\alpha-1}\right)\right)\right]$$

5.5 Application in mutual information

Theorem (5.5.1). Let X and Y be two random variables with a joint probability mass function $r(x,y)$ and marginal probability mass function $p(x)$ and $q(y)$ respectively, also $0<m\leq\frac{r(x,y)}{p(x)q(y)}\leq M<\infty\ ,\forall\ \ (x,y)\in X\times Y$. Then we have the inequalities

(5.5.1) $$\frac{1}{M}\left[\sum_{(x,y)\in X\times Y}p(x)\,q(y)\,\phi_s\left(\frac{r(x,y)}{p(x)q(y)}\right)\right]$$

$$\leq I(X;Y)$$

$$\leq \frac{1}{m}\left[\sum_{(x,y)\in X\times Y} p(x)\,q(y)\,\phi_s\left(\frac{r(x,y)}{p(x)q(y)}\right)\right]$$

Proof: Choosing $p_i = p(x)\,q(y)$, $\xi_i = \frac{r(x,y)}{p(x)q(y)}$ and $((x,y)\in X\times Y)$ in lemma (5.3.1) and taking

$$\sum_{(x,y)\in X\times Y} p(x)\,q(y)\,\frac{r(x,y)}{p(x)q(y)}\ln\frac{r(x,y)}{p(x)q(y)} = I(X;Y),$$

Also, $\phi_s\left(\sum_{(x,y)\in X\times Y} p(x)\,q(y)\,\frac{r(x,y)}{p(x)q(y)}\right) = \phi_s(1) = 0$. Then we have the desired inequality.

Dragomir et al [38] has obtained many interesting bounds for relative entropy and he has shown applications of these bounds in the field of information theory. In the next section, some upper bounds have been obtained for the relative arithmetic geometric divergence measure with the help of some well known inequalities.

5.6 Upper bounds for the relative arithmetic geometric divergence measure.

Let $p(x), q(x), x\in\chi, card(\chi)<\infty$, be two probability mass functions. Taneja [125] defined the relative arithmetic geometric divergence measure as

(5.6.1) $$G(p//q) = \sum_{x\in\chi}\frac{p(x)+q(x)}{2}\log\frac{p(x)+q(x)}{2p(x)}$$

Also, the χ^2- distance measure is given by Pearson [94]

(5.6.2) $$D_{\chi^2}(q//p) = \sum_{x\in\chi}\frac{q^2(x)}{p(x)} - 1$$

Theorem (5.6.1). Let $p(x), q(x), x\in\chi$ be two probability mass functions. Then

(5.6.3) $$G(p//q)\geq 0$$

with equality iff $p(x) = q(x) \quad \forall\, x\in\chi$.

Proof: Let $A = \{x : p(x) > 0\}$ be the support of $p(x)$. Then

$$-G(p//q) = -\sum_{x\in A}\frac{p(x)+q(x)}{2}\log\frac{p(x)+q(x)}{2p(x)}$$

$$= \sum_{x\in A}\frac{p(x)+q(x)}{2}\log\frac{2p(x)}{p(x)+q(x)}$$

$$\leq \log\left(\sum_{x\in A}\frac{p(x)+q(x)}{2}\frac{2p(x)}{p(x)+q(x)}\right)$$

$$= \log\left(\sum_{x\in A} p(x)\right)$$

$$\leq \log\left(\sum_{x\in\chi} p(x)\right) = \log(1) = 0$$

Thus,

$$G(p//q) \geq 0$$

Where the first inequality follows from Jensen's inequality. Since log is strictly concave, we have equality above iff $\frac{q(x)}{p(x)} = 1$ everywhere. Hence we have $G(p//q) = 0$ iff $p(x) = q(x) \quad \forall\, x \in \chi$.

Theorem (5.6.2). Let $p(x), q(x), x \in \chi$ be two probability mass functions. Then

(5.6.4) $\qquad G(p//q) \leq \frac{1}{4}\left(\sum_{x\in\chi} \frac{q^2(x)}{p(x)} - 1\right)$

with equality iff $p(x) = q(x) \quad \forall\, x \in \chi$.

Proof: We know that for every differentiable real valued strictly convex function f defined on an interval I of the real line, we have the inequality

(5.6.5) $\qquad f^{/}(b)(b-a) \geq f(b) - f(a) \quad \forall\ a, b \in I$

The equality holds iff $a = b$.

Now, apply (5.6.5) to $f(x) = -\log(x)$ and $I = (0, \infty)$ to get

(5.6.6) $\qquad \frac{1}{b}(a-b) \geq \log a - \log b \quad \forall\ a, b > 0$

choose $\quad a = p(x) + q(x),\, b = 2p(x),\ x \in \chi$

Then by (5.6.6), we get

$$\frac{1}{2p(x)}(q(x) - p(x)) \geq \log\frac{p(x)+q(x)}{2p(x)},\ x \in \chi$$

Multiplying by $\frac{p(x)+q(x)}{2} > 0$, we get

$$\frac{(p(x)+q(x))(q(x)-p(x))}{4p(x)} \geq \frac{p(x)+q(x)}{2}\log\frac{p(x)+q(x)}{2p(x)},\ \forall\ x \in \chi.$$

Summing over $x \in \chi$, we get

$$G(p//q) = \sum_{x\in\chi} \frac{p(x)+q(x)}{2}\log\frac{p(x)+q(x)}{2p(x)} \leq \sum_{x\in\chi} \frac{q^2(x)-p^2(x)}{4p(x)}$$

$$G(p//q) \leq \frac{1}{4}\left[\sum_{x\in\chi} \frac{q^2(x)}{p(x)} - 1\right]$$

The case of equality follows by the strict convexity of -log(.).

Theorem (5.6.3). Let $p(x), q(x) > 0$, $x \in \chi$ be two probability mass functions. Then we have the inequality

(5.6.7) $$0 \le G(p//q) \le \log\left[\tfrac{1}{4}D_{\chi^2}(q//p) + 1\right] \le \tfrac{1}{4}D_{\chi^2}(q//p)$$

equality holds iff $p(x) = q(x) \quad, \forall\, x \in \chi$.

Proof: We use Jensen's discrete inequality

(5.6.8) $$f\left(\sum_{x\in\chi} \tfrac{p(x)+q(x)}{2} t(x)\right) \le \sum_{x\in\chi} \tfrac{p(x)+q(x)}{2} f(t(x))$$

provided that f is convex on a given interval I, $t(x) \in I \;\; \forall\; x \in \chi$ and $p(x), q(x) > 0$ are probability mass function on χ.

choose $f(x) = -\log x, \;\; x > 0$, we obtain from (5.6.8)

$$-\log\left(\sum_{x\in\chi} \tfrac{p(x)+q(x)}{2} t(x)\right) \le -\sum_{x\in\chi} \tfrac{p(x)+q(x)}{2} \log t(x)$$

$$\log\left(\sum_{x\in\chi} \tfrac{p(x)+q(x)}{2} t(x)\right) \ge \sum_{x\in\chi} \tfrac{p(x)+q(x)}{2} \log t(x)$$

Put $t(x) = \frac{p(x)+q(x)}{2p(x)}$

$$\log\left(\sum_{x\in\chi} \tfrac{p(x)+q(x)}{2} \tfrac{p(x)+q(x)}{2p(x)}\right) \ge \sum_{x\in\chi} \tfrac{p(x)+q(x)}{2} \log \tfrac{p(x)+q(x)}{2p(x)}$$

$$G(p//q) \le \log\left(\sum_{x\in\chi} \tfrac{(p(x)+q(x))^2}{4p(x)}\right)$$

$$G(p//q) \le \log\left(\tfrac{1}{4}D_{\chi^2}(q//p) + 1\right)$$

We use elementary inequalty $\log(u+1) \le u, \;\; u \ge 0$ with equality iff $u = 0$.

$$\log\left(\tfrac{1}{4}D_{\chi^2}(q//p) + 1\right) \le \tfrac{1}{4}D_{\chi^2}(q//p)$$

Thus we can write

$$G(p//q) \le \log\left(\tfrac{1}{4}D_{\chi^2}(q//p) + 1\right) \le \tfrac{1}{4}D_{\chi^2}(q//p)$$

Lemma (5.6.1). Let $p(x), q(x) > 0$, $x \in \chi$ be two probability mass functions. Define $r(x) = \frac{q(x)}{p(x)}, x \in \chi$ and assume that

(5.6.9) $$0 < r \le r(x) \le R < \infty \quad , \forall\; x \in \chi.$$

Then we have the inequality

(5.6.10) $$0 \le G(p//q) \le \tfrac{(R-r)^2}{16rR}$$

equality holds in (5.6.10) iff $p(x) = q(x)\,, \ \ \forall\, x \in \chi$.

Proof: Using the Kantorovic inequality [89]

(5.6.11) $$\sum_{k=1}^{n} r_k u_k^2 \sum_{k=1}^{n} \frac{1}{r_k} u_k^2 \le \frac{1}{4}\left(\sqrt{\frac{M}{m}} + \sqrt{\frac{m}{M}}\right)^2 \left(\sum_{k=1}^{n} u_k^2\right)^2$$

where $0 < m \le r_k \le M < \infty$ for $k = 1, 2, ..., n$

Put $u_k = \sqrt{q(x)}\,, \ r_k = r(x)$ in (5.6.11), we get

$$\sum_{x\in\chi} r(x)\, q(x) \sum_{x\in\chi} \frac{1}{r(x)} q(x) \le \frac{1}{4}\left(\sqrt{\frac{R}{r}} + \sqrt{\frac{r}{R}}\right)^2 \left(\sum_{x\in\chi} q(x)\right)^2$$

which is equivalent to

$$\sum_{x\in\chi} \frac{q^2(x)}{p(x)} \le \frac{1}{4}\left(\sqrt{\frac{R}{r}} + \sqrt{\frac{r}{R}}\right)^2$$

or

$$D_{\chi^2}(q//p) \le \frac{1}{4}\left(\sqrt{\frac{R}{r}} + \sqrt{\frac{r}{R}}\right)^2 - 1$$

$$= \frac{1}{4}\left(\sqrt{\frac{R}{r}} - \sqrt{\frac{r}{R}}\right)^2$$

(5.6.12) $$D_{\chi^2}(q//p) \le \frac{(R-r)^2}{4rR}$$

Also, proved in theorem (5.6.2)

$$G(p//q) \le \frac{1}{4} D_{\chi^2}(q//p)$$

Thus,

$$4G(p//q) \le D_{\chi^2}(q//p) \le \frac{(R-r)^2}{4rR}$$

or

$$G(p//q) \le \frac{(R-r)^2}{16rR} \ .$$

Theorem (5.6.4). Let $p(x), q(x) > 0,\, x \in \chi$ be two probability mass functions. Define $r(x) = \frac{q(x)}{p(x)}, x \in \chi$ and assume that

(5.6.13) $$0 < r \le r(x) \le R < \infty, \ \ \forall \ x \in \chi.$$

Then we have the inequality

(5.6.14) $$G(p//q) \le \log\left[\frac{(R-r)^2}{16rR} + 1\right] \le \frac{(R-r)^2}{16rR}$$

equality holds iff $p(x) = q(x) \ \ \forall \ x \in \chi$.

Proof: Using the inequality (5.6.7) and (5.6.12), we have

$$G\left(p//q\right) \leq \log\left[\tfrac{1}{4}D_{\chi^2}\left(q//p\right)+1\right] \leq \log\left(\tfrac{(R-r)^2}{16rR}+1\right)$$

or

$$G\left(p//q\right) \leq \log\left(\tfrac{(R-r)^2}{16rR}+1\right) \leq \tfrac{(R-r)^2}{16rR}$$

The last inequality in (5.6.14) follows by the elementary inequality $\log\left(u+1\right) \leq u, u \geq 0$ with equality iff u=0.

Lemma (5.6.2). Let $p\left(x\right), q\left(x\right) > 0,\ x \in \chi$ be two probability mass functions satisfying the condition

$$0 < r \leq \tfrac{p(x)}{q(x)} \leq R < \infty\ , \quad \forall\ x \in \chi.$$

Then we have the inequality

(5.6.15) $$G\left(p//q\right) \leq \tfrac{1}{4}\left(1-r\right)\left(R-1\right) \leq \tfrac{1}{16}\left(R-r\right)^2$$

equality holds in (5.6.15) iff $p\left(x\right) = q\left(x\right)\ ,\ \forall\ x \in \chi.$

Proof: Using the Diaz Metcalf inequality for real numbers [89]

(5.6.16) $$\sum_{k=1}^{n} q_k b_k^2 + mM\sum_{k=1}^{n} q_k a_k^2 \leq \left(m+M\right)\sum_{k=1}^{n} q_k a_k b_k$$

provided that

(5.6.17) $$m \leq \tfrac{b_k}{a_k} \leq M \ \text{ for } k = 1,2,...,n \text{ and } q_k > 0 \text{ with } \sum_{k=1}^{n} q_k = 1$$

The equality holds in (5.6.16) if either $b_k = ma_k$ or $b_k = Ma_k$ for $k = 1,2,...,n$

Define

$$b\left(x\right) = \sqrt{\tfrac{q(x)}{p(x)}}, \quad a\left(x\right) = \sqrt{\tfrac{p(x)}{q(x)}}, \quad x \in \chi.$$

Then

$$\tfrac{b(x)}{a(x)} = \tfrac{q(x)}{p(x)} \in [r, R] \subset \left(0, \infty\right), \quad \forall\ x \in \chi.$$

From (5.6.16), we get

$$\sum_{x\in\chi} q\left(x\right)\left(\sqrt{\tfrac{q(x)}{p(x)}}\right)^2 + rR\sum_{x\in\chi} q\left(x\right)\left(\sqrt{\tfrac{p(x)}{q(x)}}\right)^2$$
$$\leq \left(R+r\right)\sum_{x\in\chi} q\left(x\right)\sqrt{\tfrac{p(x)}{q(x)}}\sqrt{\tfrac{q(x)}{p(x)}}$$

i.e

$$\sum_{x\in\chi}\tfrac{q^2(x)}{p(x)} + rR\sum_{x\in\chi} p\left(x\right) \leq \left(R+r\right)\sum_{x\in\chi} q\left(x\right)$$

In addition as $\sum_{x\in\chi} p(x) = \sum_{x\in\chi} q(x) = 1$, we obtain

$$\sum_{x\in\chi} \frac{q^2(x)}{p(x)} + rR \leq (R+r)$$

$$\sum_{x\in\chi} \frac{q^2(x)}{p(x)} \leq R + r - rR$$

or

$$\sum_{x\in\chi} \frac{q^2(x)}{p(x)} - 1 \leq (1-r)(R-1)$$

or

$$D_{\chi^2}(q//p) \leq (1-r)(R-1)$$

We use here elementary inequality

$$ab \leq \tfrac{1}{4}(a+b)^2, \quad a, b \in \Re.$$

Thus

$$(1-r)(R-1) \leq \tfrac{1}{4}(R-r)^2$$

Thus we can write

(5.6.18) $$D_{\chi^2}(q//p) \leq (1-r)(R-1) \leq \tfrac{1}{4}(R-r)^2$$

Using the inequality (5.6.4), we can write

$$G(p//q) \leq \tfrac{1}{4}(1-r)(R-1) \leq \tfrac{1}{16}(R-r)^2$$

Theorem (5.6.5). Let $p(x), q(x) > 0$, $x \in \chi$ be two probability mass functions. Define $r(x) = \frac{q(x)}{p(x)}, x \in \chi$ and assume that

(5.6.19) $$0 < r \leq r(x) \leq R < \infty \quad , \forall \ x \in \chi.$$

Then we have the inequality

(5.6.20) $$G(p//q) \leq \log\left(\frac{(1-r)(R-1)}{4} + 1\right) \leq \log\left(\frac{(R-r)^2}{16} + 1\right)$$

equality holds iff $p(x) = q(x), \ \forall \ x \in \chi$.

Proof: Using theorem (5.6.3) and inequality (5.6.8), we have

$$G(p//q) \leq \log\left(\tfrac{1}{4}D_{\chi^2}(q//p) + 1\right) \leq \log\left(\frac{(1-r)(R-1)}{4} + 1\right) \leq \log\left(\frac{(R-r)^2}{16} + 1\right)$$

or

$$G(p//q) \leq \log\left(\frac{(1-r)(R-1)}{4} + 1\right) \leq \log\left(\frac{(R-r)^2}{16} + 1\right)$$

References

[1] Abramson, N., “Information theory and coding”, McGraw Hill New York (1963).

[2] Aczel, J and Daroczy, Z., “On measure of information and their characterizations”, Academic Press, New York (1975).

[3] Aczel, J and Daroczy, Z., “Uber verallege meinerte quasilinear mittelwerte die mit Gewichtsfuncktionen gebildet sind., Pub. Math. Debrecen, Vol 10, PP 171-190 (1963).

[4] Ali, S.M and Silvey, S.D., “A general class of coefficients of divergence of one distribution from another”, Journal of Royal Statistical society, Section B, Vol 28, PP 131-142 (1966).

[5] Arimoto, S., “Information theoretic considerations in estimation theory”, Information and control, Vol 19, PP 181-190 (1971).

[6] Autar, R and Khan, A.B., “On generalized useful information for incomplete distribution”, Journal of Combinatorics, Information and System Sciences, Vol 14(4), PP 187-191 (1989).

[7] Autar, R and Soni, R.S., “Inaccuracy and Coding theorem”, Journal of Applied probability, Vol 12, PP 845-851 (1975).

[8] Baig, M.A.K and Rayees., “A Generalized ‘useful’ inaccuracy of order α and type β and coding theorems”, International Journal of Pure and Applied Mathematics, Vol 32 (4), PP 467-474 (2006).

[9] Baig, M.A.K and Rayees., ” Coding theorems on weighted parametric information measure”, Presented in 2nd J&K science congress held in University of Kashmir, Srinagar, 25-27 July (2006).

[10] Baig, M.A.K and Rayees., “Generalized divergence measures and information inequalities”, Journal of Mathematics and system sciences, Vol 2(2), PP 57-75, Dec (2006).

[11] Baig, M.A.K and Rayees., “Some generalized inequalities and their applications”, International Journal of Pure and Applied Mathematics, Vol 31(2), PP 253-263 (2006).

[12] Baig, M.A.K and Rayees., “Some Noiseless coding theorems of inaccuracy

measure of order α and type β ", Sarajevo Journal of Mathematics, Vol 3(1), PP 137-143 (2007).

[13] Baig, M.A.K and Rayees., "Upper and lower bounds for Csiszar f-divergence in terms of symmetric J-Divergence and applications", "Indian Journal of Mathematics", (2007) (To appear).

[14] Baig, M.A.K and Zaheeruddin., "Coding theorem and a measure of inaccuracy", Soochow Journal of Mathematics, Vol 19 (4), PP 1-7 (1993).

[15] Balasubrahmanyan, P and Siromoney, G., "A note on entropy of Telgu prose", Information and control, Vol 13, PP 281-285 (1968).

[16] Bar- Hillel, Y., " Language and information", Addision Wasley Pub. Co., The Jerusalum Academic press limited (1964).

[17] Belis, M and Guiasu, S., "A quantitative and qualitative measure of information in cybernetic system", IEEE Transaction on information theory, Vol IT-14, PP 593-594 (1968).

[18] Ben- Tal, A, Ben- Israel, A and Teboulle, M., "Certainty equivalents and information measures: duality and extermal", J. Math. Annal., Vol 157, PP 211-236 (1991).

[19] Bernard, M. A and Sharma, B.D., "Some combinatorial results on variable length error correcting codes", ARS Combinatorica, Vol 25(B), PP 181-194 (1988).

[20] Bernard, M.A and Sharma, B.D., "A lower bound on average codeword length of variable length error correcting codes", IEEE Transaction on information theory, Vol 36(6), PP 1474-1475 (1990).

[21] Beth Bassat, M., "f- entropies, probabilities of error and feature selection", Information and control, Vol 39, PP 227-242 (1978).

[22] Bhatia, P.K., "On a generalized 'useful' inaccuracy for incomplete probability distribution", Soochow Journal of Mathematics, Vol 25(2), PP 131-135 (1999).

[23] Bhatia, P.K., "Useful inaccuracy of order α and 1:1 coding", Soochow Journal of Mathematics, Vol 21(1), PP 81-87 (1995).

[24] Bhattacharya, A., "On a measure of divergence between two statistical populations defined by their probability distribution", Bull. Calcutta Math. Soc., Vol 35 , PP 99-109 (1943).

[25] Blake, F., "A Perspective on coding theorem", Information Sciences, Vol 57-58,

PP 111-118 (1991).

[26] Boekee, E and Van Der Lubbe, J.C.A., "The R- norm information measure", Information and Control, Vol 45, PP 136-155 (1980).

[27] Brillouin, L., "Science and information theory", New York, Academic Press.

[28] Burbea, I and Rao, C.R., "On the convexity of some divergence measures based on entropy function", IEEE Transaction on information theory, Vol 28(3), PP 489-495 (1992).

[29] Campbell, L.L., "A Coding theorem and Renyi's entropy", Information and Control, Vol 8, PP 423-429 (1965).

[30] Chaundy, T.W and Mcloed, J.B., "On a functional equation", Proc. Edinburgh Math. Soc. Notes, Vol 43, PP 7-8 (1960).

[31] Chen, C.H., "Statistical pattern recognition", Rocelle Park, New York, Hoyderc Book Co., (1973).

[32] Chow, C.K and Lin, C.N., "Approximation discrete probability distributions with dependence trees", IEEE Transaction on information theory, Vol 14(3), PP 462-467 (1968).

[33] Cover, T.M and Thomas, J.A., "Elements of information theory", John Wiley and Sons (1991).

[34] Csiszar, I and Korner, J., "Information theory: coding theorems for discrete memoryless systems", Academic Press, New York (1981).

[35] Csiszar, I., "Information measures: A critical survey", Trans. 7th Prague Conf. on Information theory, Statist. Decis. Func., Random processes and 8th European meeting of statistics, Vol B, Academia prague, PP 73-86 (1978).

[36] Csiszar, I., "Information type measures of differences of probability distributions and indirect observations", Studia Mathematika Hungarica, Vol 2, PP 299-318 (1967).

[37] Daroczy, Z., "Uber mittelwerte und entropien voll standiger wahrs-chein lichkeits verteilung", Acta Math. Acad.Sci. Hungry, Vol 15, PP 203-210 (1964).

[38] Dragomir, S.S, Scholz, M.L and Sunde, J., "Some upper bounds for relative entropy and applications", Comp. and Mathematics with applications, Vol 39, PP 91-100 (2000).

[39] Dragomir, S.S and Goh, C.J., "A counterpart of Jensen's discrete inequality for differentiable convex mappings and applications in information theory", Math. Comput. Modelling, Vol 24 (2), PP 1-11 (1996).

[40] Dragomir, S.S and Ionescu, N.M., "Some converse of Jensen's inequality and applications", Annal. Num. Thero. Approx. (Cluj-Napoca), Vol 23, PP 71-78 (1994).

[41] Dragomir, S.S, Gluscevic, V and Pearce, C.E.M., "Approximation for the Csiszar f-divergence via midpoints inequalities", In inequality theory and applications- Vol 1, Y.J Cho, J.K Kim and S.S Dragomir (Eds), Nova Science publishers, Inc. Huntington, New York, PP 139-154 (2001).

[42] Dragomir, S.S., "An upper bound for Csiszar f-divergence in terms of the variational distance and applications", Pan American Mathematical Journal, Vol 12 (4), PP 105-117 (2002).

[43] Dragomir, S.S., " Upper and lower bounds for Csiszar f-divergence in terms of the Kullback Leibler distance and applications", Bull Math. Soc. Math, Roumania, Vol 43 (91)(1), PP 25-37 (2000).

[44] Dragomir, S.S., " Upper and lower bounds for Csiszar f-divergence in terms of the Hellinger discrimination and applications", Non linear analysis Forum, Vol 7(1), PP 1-13 (2002).

[45] Dragomir, S.S., "A converse inequality for the Csiszar ϕ divergence", Tamsui Oxford Journal of Mathematical Sciences, Vol 20 (1), PP 35-53 (2004).

[46] Dragomir, S.S., "An inequality for twice differentiable convex mappings and applications for the Shannon and Renyi's entropies", Mathematical Journal of Ibaraki University, Vol 32, PP 19-28 (2000).

[47] Dragomir, S.S., "Other inequalities for Csiszar divergence and applications", Acta Math. Vietnam, Vol 27(2), PP 203-217 (2002).

[48] Dragomir, S.S., "Some inequalities for the Csiszar ϕ divergence", Journal of Korean society of industrial and applied Mathematics, Vol 7(1), PP 63-77 (2003).

[49] Feinstein, A., "Foundation of information theory", McGraw Hill, New York (1958).

[50] Gokhale, D.V and Kullback, S., "Information in contingency tables", New York, Marcell Dekker (1978).

[51] Guiasu, S and Picard, C.F., "Borne inferieure de la longueur de certian codes", C.R Academic Sciences, Paris, Vol 273, PP 248-251 (1971).

[52] Guiasu, S., "Information theory with application", Mc Graw Hill, Mass., (1976).

[53] Gurdial and Pessoa, F., "On useful information of order α", Journal of combinatorics, Information and System Sciences, Vol 2, PP 158-162 (1977).

[54] Hamming, R.W., "Error detecting and error correcting codes", Bell System Technical Journal, Vol 29, PP 147-160 (1950).

[55] Hartley, R.V.L., "Transmission of information", Bell. Sys. Tech. Journal, Vol 7, PP 535-563 (1928).

[56] Hartnett, W.E (ed)., "Foundation of coding theory", D. Reidling Publishing Co., Derdietcht, Holland (1974).

[57] Havrada, J and Charvat, F., "Quantification method of classification process: concept of structural α entropy", Kybernetika, Vol 3, PP 30-35 (1967).

[58] Hellinger, E., "Neue Begrundung der Theorie der quadratischen Formen Von unendlichen Vielen Veranderlichen", J. Rein Aug. Math. Vol 136, PP 210-271 (1909).

[59] Hooda, D.S and Bhaker, U.S., "A generalized 'useful' information measure and coding theorems", Soochow Journal of Mathematics, Vol 23 (1), PP 53-62 (1997).

[60] Hooda, D.S and Ram, A., "Characterization of Non-additive 'useful' information measure", Recent Advances in information theory, Statistics and Computer applications, CCS Haryana, Agricultural University, Hisar, PP 64-77(1998).

[61] Hooda, D.S and Singh, U., "On useful information generating functions", Statistica, XLVI.4, PP 527-536 (1986).

[62] Hooda, D.S and Tuteja, R.K., "Two generalized measures of 'useful' information", Information Sciences, Vol 23, PP 11-24 (1981).

[63] Jain, P and Tuteja, R.K., "On coding theorem connected with 'useful' entropy of order β", International Journal of Mathematics and Mathematical Sciences, Vol 12(1), PP 193-198 (1989).

[64] Jaynes, E.T., "Information theory and statistical mechanics", Physical Review, Vol 106, PP 620-630 (1957).

[65] Jeffreys, H., "An invariant form for the prior probability in estimation problems", Proc. Roy. Soci. London, Serial A, Vol 186, PP 453-461 (1946).

[66] Kadota, T.T and Shepp, L.A., “On the best finite set of linear observables for discriminating two gaussian signals”, IEEE Transaction on information theory, Vol 13, PP 288-294 (1967).

[67] Kailath, T., “The divergence and Bhattacharya distance measures in signal selection”, IEEE Transaction Comm. Technology, Vol Com-15, PP 52-60 (1967).

[68] Kapur, J.N., “A comperative assesment of various measures of directed divergence”, Advances in Management studies, Vol 3, PP 1-16 (1984).

[69] Kapur, J.N., “Generalized entropy of order α and type β”, The Mathematical Seminar (Delhi), Vol 4, PP 78- 94 (1967).

[70] Kapur, J.N., “Maximum entropy models in science and engineering”, Wiley Eastern, New Delhi (1990).

[71] Kapur, J.N., “On some applications of dynamic programming to information theory”, Proceedings 2nd Academic Science, Vol 68, I-II (1968).

[72] Kazakos, D and Cotsidas, T., “A decision theory approach to the approximation of discrete probability densities”, IEEE Transaction Perform. Annal. Machine intell., Vol 1, PP 61-67 (1980).

[73] Kerridge, D.F., “Inaccuracy and inference”, Journal of Royal Statistical society, Serial B, Vol 13(23), PP 184-194 (1961).

[74] Khan, A.B and Haseen., “Some noiseless coding theorems of entropy of order α of the power distribution P^{β}”, Metron, Vol 39(3-4), PP 87-94 (1981).

[75] Khan, A.B and Autar, R., “Generalized Non-additive coding theorems for personal probability codes”, Soochow Journal of Mathematics, Vol 5, PP 171-179 (1979).

[76] Khan, A.B and Autar, R., “On unified information of order α and type β”, Soochow Journal of Mathematics, Vol 5, PP 93-99 (1979).

[77] Khan, A.B, Autar, R and Haseen., “Noiseless coding theorems for generalized non additive entropy”, Tamkang Journal of Mathematics, Vol 12(1), PP 15-20 (1981).

[78] Khan, A.B, Bhat, B.A and Pirzada, S., “Some results on a generalized useful information measure”, Journal of inequalities in pure and applied Mathematics, vol 6, Issue 4, Article 117 (2005).

[79] Khinchin, A.Ja., “The concept of entropy in the theory of probability”, (Russian) Uspeki Math. Nauk; 8, 3-20 (English Transl: In Mathematical foundation of infor-

mation theory, 1-28 Dovar, New York (1957).

[80] Kraft, L.J., “A device for quantizing grouping and coding amplitude modulates pulses”, M.S Thesis, Department of Electrical Engineering, MIT, Cambridge (1949).

[81] Kullback, S and Leibler, R.A., “On information and Sufficiency”, Ann. Math. Statist., Vol 22, PP 79-86 (1951).

[82] Kullback, S., “Information theory and Statistics”, John Wiley and Sons, New York (1959).

[83] Kumar, P and Taneja, I.J., “On unified generalizations of relative Jensen Shannon and Arithmatic- Geometric divergence measure, and their properties”, Indian Journal of Mathematics and Mathematical Sciences, Vol 1(1), PP 77-79 (2005).

[84] Lee, P.M., “On the axioms of information theory”, Ann. Math. Stat., Vol 35, PP 414-418 (1964).

[85] Lin, J., “Divergence measures based on the Shannon entropy”, IEEE Transaction on information theory”, Vol IT-37, PP 145-151 (1991).

[86] Longo, G., “A Noiseless coding theorem for source having utilities”, SIAM Journal of Applied Mathematics, Vol 30, PP 739-748 (1976).

[87] Longo, G., “Quantitative-qualitative measure of information”, Springer- Verlag, New York (1972).

[88] Mei, M., “The theory of genetic distance and evaluation of human races”, Japan Journal of Human genetics, Vol 23, PP 341-369 (1978).

[89] Mitrinovic, D.S., “Analytical inequalities”, Springer Verlag (1970).

[90] Nath, P., “An axiomatic characterizations of inaccuracy for discrete generalized probability distributions”, Opsearch, Vol 7, PP 115-133 (1970).

[91] Nyquist, H., “Certain topics in telegraphy transmission theory”, AIEE Trans., Vol 47, PP 617 (1928).

[92] Nyquist, H., “Cetain factors affecting telegraphy speed”, Bell. System Tech. Journal, Vol 3, PP 324-346 (1924).

[93] Parkash, O., “On weighted parametric information measure”, MATEMATNPKN BECHNK, Vol 49, PP 93-97 (1997).

[94] Pearson, K., “On the criterion that a given system of deviations from the proba-

ble in the case of correlated system of variables is such that it can be reasonable supposed to have arisen from random sampling", Phil. Mag., Vol 50, PP 157-172 (1900).

[95] Picard, C.F., "Weighted probabilistic information measures", Journal of Combinatorics, Information and System Scienes, Vol 4, PP 343-356 (1979).

[96] Pielou, E.C., "Ecological diversity", Wiley, New York (1975).

[97] Quaster, H., "Information theory and Biology", University of Illinois Press, Urbana (1953).

[98] Quaster, H., "Information theory and Psychology", Free Press, Glencoe, Illinois (Proceeding of conference held in 1954).

[99] Rao, C.R., "Diversity and Dissimilarity coefficents: a unified approach", Theoretic population Biology, Vol 21, PP 24-43 (1982).

[100] Rao, C.R., "Linear Statistical inference and its application", John Wiley and Sons, Inc. New York (1965)

[101] Rathie, P.M and Kannappan, M., "An inaccuracy function of type β", Ann. inst. Stat. Math., Vol 27, PP 205-214 (1978).

[102] Rathie, P.N., "Generalized entropies in coding theory", Metrika, Vol 18, PP 216-219 (1972).

[103] Rathie, P.N., "On a generalized entropy and coding theorem", Queen's university Kingston Pre print, No. 17 (1968)

[104] Renyi, A., "On measure of entropy and information", Proceeding Fourth Berkely Symposium on Math. Stat. and probability, University of California Press, Vol 1, PP 547-561(1961).

[105] Roy, L.K., "Comparison of Renyi's entropy of power distribution", ZAMM, Vol 56, PP 217-218 (1976).

[106] Sen, A., "On economic inequality", Oxford university Press, London (1973).

[107] Shannon, C.E., "A Mathematical theory of Communication", Bell System Technical Journal, Vol 27, PP 379-423, 623-659 (1948).

[108] Sharma, B.D and Mittal, D.P., "New non-additive measures of relative information", Journal of combinatorics, Information and system sciences, Vol 2(4), PP 122-132 (1997).

[109] Sharma, B.D and Autar, R., “Relative information function and their type (α, β) generalizations”, Metrika, Vol 21, PP 41-50 (1974).

[110] Sharma, B.D and Gupta, H.C., “Subadditive measures of relative information and inaccuracy”, Mathematika, Vol 23, PP 755-765 (1976).

[111] Sharma, B.D and Mittal, D.P., “New non additive measures of entropy for discrete probability distributions”, Journal of Mathematical Science, Vol 10, PP 28-40 (1975).

[112] Sharma, B.D and Taneja, I.J., “Three generalized additive measures of entropy:, Elektronische informations verarbeitung und Kybernetik Eik, Vol 13, 7/8, PP 419 -433 (1977).

[113] Sharma, B.D, Mittel. J and Man Mohan., “On measure of ‘useful’ information”, Information and Control, Vol 39 (3), PP 323-336 (1978).

[114] Sharma, B.D.,“The mean value study quantities in information theory”, Ph.D Thesis, University of Delhi (1970).

[115] Shioya, H and Da-te, T., “A generalization of Lin divergence and the derivative of a new information divergence”, Elec. and communication in Japan, Vol 78 (7), PP 37-40 (1995).

[116] Shisha, O., “Inequalities”, Academic press, New York (1967).

[117] Sibson, R., “Information radius”, Z.Wahrs. Verw Geb., Vol 14, PP 149-160 (1969).

[118] Singh, R.P, Kumar, R and Tuteja, R.K., “Application of Holder’s inequality in information theory”, Information Sciences, Vol 152, PP 145-154 (2003).

[119] Taneja, H.C and Tuteja, R.K., “Characterization of a quatitative-qualitative measure of inaccuracy”, Kybernetika, Vol 22, PP 393-402 (1986).

[120] Taneja, H.C, Hooda, D.S and Tuteja, R.K., “Coding theorems on a generalized ‘useful’ information”, Soochow Journal of Mathematics, Vol 11, PP 123-131 (1985).

[121] Taneja, I.J and Kumar,P., “Generalized non symmetric divergence measures and inequalities”, Journal of Interdisplinary Mathematics, accepted, (2005).

[122] Taneja, I.J., “A Joint characterizations of directed divergence, Inaccuracy and their generalizations”, Journal of Statistical Physics, Vol 11(2), PP 169-176 (1974).

[123] Taneja, I.J., “New developments in generalized information measure”, Ad-

vances in Imaging and Electron Physics, Edited by P.W Hawkes, Academic Press, PP 37-136 (1995).

[124] Taneja, I.J., "On symmetric and non symmetric divergence measures and their generalizations", Advances in imaging and electron physics, Vol 138, PP 177-250 (2005).

[125] Taneja, I.J., "Relative divergence measure and information inequalities", To appear in Inequality theory and applications, Vol 4 (2004).

[126] Theil, H., "Economics and information theory", North Holland publishing company, Amsterdam (1967).

[127] Theil, H., "Statistical decomposition analysis", North Holland, Amsterdam (1972).

[128] Topsoe, F., "Some inequalities for information divergence and related measure of discrimination", IEEE Transaction on information theory, Vol IT-46, PP 1602-1609 (2000).

[129] Tuteja, R.K and Bhaker, U.S., "On characterization of some non additive measure of 'useful' information", Information Sciences, Vol 78, PP 119-128 (1994).

[130] Tverberg, H., "A new derivative of information function", Math. Scand., Vol 6, PP 297 -298 (1958).

[131] Vajda, I., "Theory of Statistical inference and information", London: Kluwer Academic Press (1989).

[132] Varma, R.S., "Generalizations of Renyi's entropy of order α", Journal of Mathematical Science, Vol 1, PP 34-48 (1966).

[133] Wiener, N., "Cybernetics", The MIT Press and John Wiley and Sons Inc., New York, London (1948).

[134] Zadeh, L.A., "Fuzzy sets", Information and Control, Vol 8, PP 94-102 (1996).

[135] Zaheeruddin, "A basic theorem in information theory", The Aligarh Journal of Statistics, Vol 7, PP 1-10 (1987).