International Journal Of Engineering Research & Management Technology
ISSN: 2348-4039
May- 2014 Volume 1, Issue 3
Email: editor@ijermt.org Website: ijermt.org

Characterization of a Generalized Information Measure by Optimization Technique

D.S. Hooda, Keerti Upadhyay
Jaypee University of Engineering and Technology
A.B. Road, Raghogarh
Dist. Guna – M.P. (India)

Sonali Saxena
Doctoral Fellow of NBHM
Department of Atomic Energy (DAE),
(India)

Abstract
In the present paper the generalized mean codeword length is studied and characterized a new generalized information measure by obtaining bounds in terms of a new generalized information measure using Lagrange's Multiplier method. The Shannon's Noiseless coding theorem is verified by considering Huffman coding scheme and Shannon Fano coding scheme on taking empirical data. We study the monotone behaviour of the new generalized information measure with respect to parameters α and β. The important properties of the new generalized measure of information have also been studied.

Mathematical subject classification: 94A15 and 94A17.

Keywords: Kraft's inequality; Lagrange's Multiplier method; Huffman Codes; Shannon Fano Codes; Shannon's Noiseless coding theorem;

1. Introduction

Let X be a discrete random variable taking a finite number of possible values $x_1, x_2, \dots, x_n$ with probabilities $p_1, p_2, \dots, p_n$ respectively such that $p_i > 0, \forall i = 1, 2, \dots, n$ and $\sum_{i=1}^{n} p_i = 1$. The function $H(p_1, p_2, \dots, p_n)$ is to be interpreted as the average uncertainty associated with the events $\{x_1, x_2, \dots x_n\}$, $i = 1, 2, \dots, n$ given by

$$H(p_1, p_2, \dots, p_n) = -\sum_{i=1}^{n} p_i \log p_i . \qquad (1.1)$$

(1.1) plays a leading role in coding theory and provides a lower and upper bounds on the average codeword length Let a finite set of n input symbols $X = \{x_1, x_2, \dots, x_n\}$ be encoded using D size alphabets with probability distribution $P = \left\{(p_1, p_2, \dots, p_n), p_i \ge 0, \sum_{i=1}^{n} p_i = 1\right\}$. It was shown by Kraft (1949) that there is a unique decipherable code with code word lengths l_i $(i = 1, 2, \dots, n)$ satisfying the following inequality:

$$\sum_{i=1}^{n} D^{-l_i} \le 1, \qquad (1.2)$$

which is known as Kraft's inequality.

Let $L=\sum_{i=1}^{n} p_i l_i \log D$ (1.3) be the mean codeword length associated with input symbols $\{x_1, x_2, \ldots, x_n\}$, then under the Kraft's inequality Shannon (1948) proved the following result for a noiseless channel:

$$H(P) \le L < H(P) + \log D, D \ge 2, \tag{1.4}$$

with equality if and only if $l_i = -\log_D p_i$.

Shannon-Fano coding is less efficient than Huffman coding, but we have the advantage that we can go directly from the probability p_i to the codeword length l_i. Let S be set of the source symbols $s_1, s_2, \ldots, s_n$ with their corresponding probabilities $p_1, p_2, \ldots p_n$, then for each p_i there is an integer l_i such that

$$\log_D\left(\frac{1}{p_i}\right) \le l_i \le \log_D \frac{1}{p_i} + 1 \ , \tag{1.5}$$

where D is number of code's alphabets.
Now (1.5) implies

$$\frac{1}{p_i} \le D^{l_i} \le \frac{D}{p_i} \tag{1.6}$$

or

$$1 \ge \sum_{i=}^{n} D^{-l_i} > \frac{1}{D} \ , \tag{1.7}$$

which is Kraft's inequality, and that is necessary and sufficient condition for decodable code having these lengths l_i.

Multiplying (1.7) by p_i and summing over i, we have

$$\sum_{i=1}^{n} p_i \log_D \frac{1}{p_i} \le \sum_{i=1}^{n} p_i l_i \le \sum_{i=1}^{n} p_i \log_D \frac{1}{p_i} + 1 \tag{1.8}$$

$$H_D(P) \le L \le H_D(P) + 1, \tag{1.9}$$

where L is average codeword length given by

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

The above defined average codeword length has been generalized by so many authors. Hooda and Bhaker (1992) gave the following generalization of (1.9):

$$L_\alpha^\beta(P) = \frac{\alpha}{1-\alpha} \log\left(\frac{\sum_{i=1}^{n} p_i^\beta D^{\left(\frac{1-\alpha}{\alpha}\right) l_i}}{\sum_{i=1}^{n} p_i^\beta} \right), 0 < \alpha < 1, \ \beta \ge 1, \alpha \ne 1. \tag{1.10}$$

and studied its lower and upper bounds by applying Holder's inequality. They proved the following result:

$$H_\alpha^\beta(P) \le L_\alpha^\beta(P) < H_\alpha^\beta(P) + 1 \tag{1.11}$$

where $H_\alpha^\beta(P)=\frac{1}{1-\alpha}\log\left(\frac{\sum_{i=1}^{n}p_i^{\alpha+\beta-1}}{\sum_{i=1}^{n}p_i^{\beta}}\right),\ 0<\alpha<1, 0<\beta\le 1, \alpha\ne 1$. (1.12)

This generalized entropy (1.12) was characterized by Aczel and Daroczy (1963) and was also characterized by Kapur (1967) by following a different method.

In the present paper we study the generalized mean codeword length and characterize a new generalized information measure by obtaining bounds in terms of a new generalized measure of information as a biproduct in section 2. In section 3 we verify the Shannon's Noiseless Coding theorem in cases of Shannon Fano coding scheme and Huffman Coding scheme. We study the monotone behaviour of the new generalized information measure with respect to parameters α and β in section 4. In section 5 we study the properties of a new generalized measure of information.

2. A Generalized Mean Codeword Length and its Bounds

Hooda and Bhaker (1992) gave the following generalization of mean codeword length:

$$L_\alpha^\beta(P)=\frac{\alpha}{1-\alpha}\log_D\left(\frac{\sum p_i^{\beta}D^{\left(\frac{1-\alpha}{\alpha}\right)l_i}}{\sum p_i^{\beta}}\right), 0<\alpha<1, \beta\ge 1, \alpha\ne 1. \quad (2.1)$$

where l_i is the length of the codeword x_i and p_i is the probability of occurrence of codeword x_i

The codeword length defined in (2.1) satisfies the following essential properties of being a mean codeword length:

1. When $l_1=l_2=\cdots=l_n=l$, then $L_\beta^\alpha(P)=l$
2. $L_\alpha^\beta(P)$ lies between minimum and maximum values of $l_1,l_2,\ldots,l_n$.
3. When $\beta=1$ and $\alpha\to 1$, then $L_\alpha^\beta\to L$, where $L=\sum_{i=1}^{n}p_il_i$

Next we obtain the lower and upper bounds of (2.1) in the following theorem.

Theorem 2.1. For all uniquely decipherable codes the exponentiated mean codeword length $L_\alpha^\beta(P)$ defined in (2.1) satisfies the following relation

$$H_\alpha^\beta(P)\le L_\alpha^\beta(P)<H_\alpha^\beta(P)+1, \quad (2.2)$$

where $H_\alpha^\beta(P)=\frac{1}{1-\alpha}\log_D\left[\frac{\left(\sum p_i^{\left(\frac{\beta+2\alpha-2}{\alpha}\right)}\right)^{\alpha}}{\sum p_i^{\beta}}\right], 0<\alpha<1, \beta\ge 1, \alpha\ne 1.$ (2.3)

under the generalized Kraft inequality given by

$$\sum p_i^{\beta-1} D^{-l_i} \le \sum p_i^{\beta}. \tag{2.4}$$

Proof: Let us choose $\frac{p_i^{\beta-1} D^{-l_i}}{\sum p_i^{\beta}} = x_i$, for each $i = 1,2,\ldots,n.$ (2.5)

Substituting (2.5) in (2.1) we have

$$L_\alpha^\beta(P) = \frac{\alpha}{1-\alpha}\log_D\left[\frac{\sum p_i^{\frac{\alpha+\beta-1}{\alpha}} x_i^{\left(\frac{\alpha-1}{\alpha}\right)}}{\left(\sum p_i^{\beta}\right)^{1/\alpha}}\right] \tag{2.6}$$

Thus we are to minimize (2.6) subject to the following constraints:

$$\sum_{i=1}^{n} x_i = \frac{\sum p_i^{\beta-1} D^{-l_i}}{\sum p_i^{\beta}} \le 1 \tag{2.7}$$

Since $L_\alpha^\beta(P)$ is pseudo convex function for each $i = 1,2,\ldots,n$, therefore, we can obtain the minimum value of $L_\alpha^\beta(P)$ by applying the Lagrange's multiplier method.

Let us consider the corresponding Lagrangian as given below:

$$L = \frac{\alpha}{1-\alpha}\log\left[\frac{\sum p_i^{\frac{\alpha+\beta-1}{\alpha}} x_i^{\left(\frac{\alpha-1}{\alpha}\right)}}{\left(\sum p_i^{\beta}\right)^{1/\alpha}}\right] + \lambda\left(\sum_{i=1}^{n} x_i - 1\right)$$

Differentiating w.r.t. x_i and equating to zero, we get

$$\left(\frac{dL}{dx_i}\right)_{\alpha=\beta=1} = -p_i x_i^{-1} + \lambda = 0$$

It implies

$$\therefore x_i = cp_i \text{ , where } c = \frac{1}{\lambda} \ne 0 \tag{2.8}$$

(2.8) together with (2.5) gives

$$\frac{p_i^{\beta-1} D^{-l_i}}{\left(\sum p_i^{\beta}\right) p_i} \le 1$$

It implies

$$D^{-l_i} \le \frac{\sum p_i^{\beta}}{p_i^{\beta-2}}.$$

Taking log of both sides, we have

$$-l_i \le \log_D \frac{\sum p_i^{\beta}}{p_i^{\beta-2}}$$

or

$$l_i \ge -\log_D \frac{\sum p_i^{\beta}}{p_i^{\beta-2}} \tag{2.9}$$

Multiplying both sides of (2.9) by $\left(\frac{1-\alpha}{\alpha}\right) \geq 0$ as $0 < \alpha < 1$, we get

$$\left(\frac{1-\alpha}{\alpha}\right) l_i \geq -\left(\frac{1-\alpha}{\alpha}\right) \log_D \left(\frac{\sum p_i^\beta}{p_i^{\beta-2}}\right)$$

or

$$\frac{1-\alpha}{\alpha} l_i \geq -\log_D \left(\frac{\sum p_i^\beta}{p_i^{\beta-2}}\right)^{\frac{1-\alpha}{\alpha}} \tag{2.10}$$

From (2.1) and (2.10), we get the minimum value of $L_\beta^\alpha(P)$ as follows:

$$L_\alpha^\beta(P)_{\min} = \frac{1}{1-\alpha} \log \left[\frac{\left(\sum p_i^{\frac{\beta+2\alpha-2}{\alpha}}\right)^\alpha}{\sum p_i^\beta}\right] = H_\alpha^\beta(P) \tag{2.11}$$

l_i is always integral value in (2.9), so it must be equal to

$$l_i = a_i + \varepsilon_i, \tag{2.12}$$

where $a_i = \log \frac{p_i^{\beta-2}}{\sum p_i^\beta}$ and $0 \leq \varepsilon_i < 1$

Putting (2.12) in (2.1), we have

$$L_\alpha^\beta(P) = \frac{\alpha}{1-\alpha} \frac{\left[\log \sum p_i^\beta \left(\frac{p_i^{\beta-2}}{\sum p_i^\beta}\right)^{\frac{1-\alpha}{\alpha}} D^{\varepsilon_i\left(\frac{1-\alpha}{\alpha}\right)}\right]}{\sum p_i^\beta}$$

$$= \frac{1}{1-\alpha} \log \frac{\left(\sum p_i^{\frac{\beta+2\alpha-2}{\alpha}}\right)^\alpha}{\sum p_i^\beta} + \varepsilon_i \; . \tag{2.13}$$

Since $0 \leq \varepsilon_i < 1$, therefore, (2.13) reduce to

$$L_\alpha^\beta(P) < \frac{1}{1-\alpha} \log \left[\frac{\left(\sum p_i^{\frac{\beta+2\alpha-2}{\alpha}}\right)^\alpha}{\sum p_i^\beta}\right] + 1 = H_\alpha^\beta(P) + 1 \tag{2.14}$$

Hence from (2.11) and (2.14), we get

$H_\alpha^\beta(P) \leq L_\alpha^\beta(P) < H_\alpha^\beta(P) + 1$, which is (2.2).

Thus by applying optimization technique in studying bounds of mean code word length $L_\alpha^\beta(P)$ and we obtain a new generalized measure of information $H_\alpha^\beta(P)$ given by (2.3).

3. Application of Shannon-Fano Coding and Huffman Coding schemes

In this section we illustrate the veracity of the theorem 2.1 by taking empirical data as given in table (3.1) and (3.2) on the lines of Prakash and Priyanka (2012).

Table-3.1

Probabilities p_i	Shannon Fano code words	Length of Shannon Fano code words l_i	α	β	$L_\alpha^\beta(P)$	$H_\alpha^\beta(P)$	$\eta=\frac{H_\alpha^\beta(P)}{L_\alpha^\beta(P)}\times 100$
.3846	00	2	.5	2	2.21979	2.03595	91.78%
.1795	01	2					
.1538	10	2					
.1538	110	3					
.1282	111	3					

Table-3.2

Probabilities p_i	Huffman code words	Length of Huffman code words l_i	α	β	$L_\alpha^\beta(P)$	$H_\alpha^\beta(P)$	$\eta=\frac{H_\alpha^\beta(P)}{L_\alpha^\beta(P)}\times 100$
.3846	0	1	.5	2	2.12484	2.03595	95.81%
.1795	100	3					
.1538	101	3					
.1538	110	3					
.1282	111	3					

From table (3.1) and (3.2) we infer the following:
(i) Theorem 2.1 holds in both cases of Shannon -Fano codes and Huffman codes.
(ii) Huffman mean codeword length is less than Shannon –Fano mean codeword length.
(iii) Coefficient of efficiency of Huffman Codes is greater than Coefficient of efficiency of

Shannon -Fano Codes i.e. it is concluded that Huffman coding Scheme is more efficient than Shannon -Fano coding scheme.

4. Monotone Behaviour of the New Generalized Information Measure $H_\alpha^\beta(P)$

In this section we study the monotone behaviour of the new generalized information measure given by (2.3) with respect to parameters α and β.

Let $P = \{0.3846, 0.1795, 0.1538, 0.1538, 0.1282\}$ be a set of probabilities.

Assuming $\beta = 3$. We tabulate the values of $H_\alpha^\beta(P)$ for different values of α as given in the following table:

Table 4.1: Monotone behaviour of $H_\alpha^\beta(P)$ with respect to α

α	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$H_\alpha^\beta(P)$	2.37855	2.33423	2.29160	2.25586	2.22722	2.20438	2.18603	2.17104	2.15861

Next we draw the graph of the table (4.1) and illustrate from figure (4.1) that $H_\alpha^\beta(P)$ is monotonic decreasing with increasing values of α.

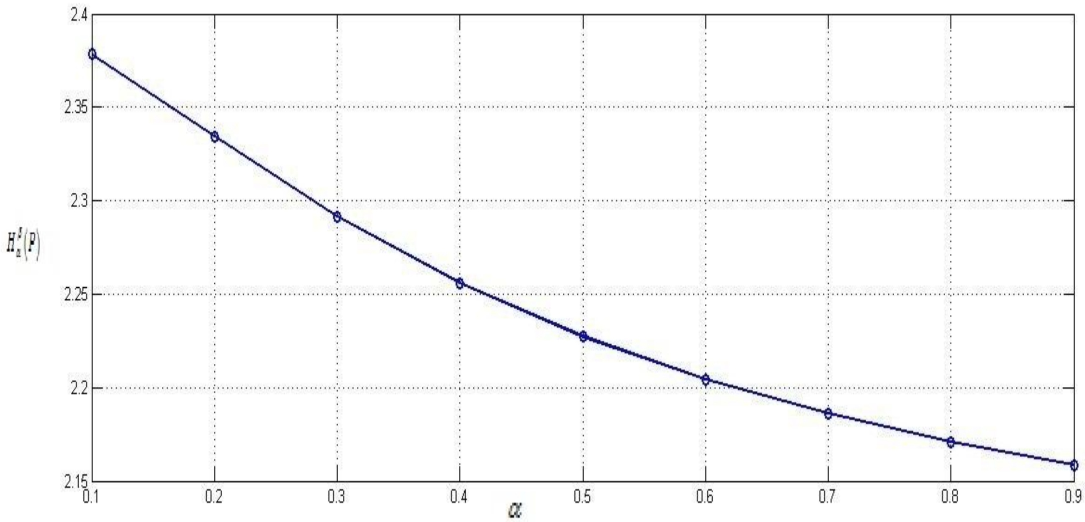

Fig: 4.1 Monotone behaviour of $H_\alpha^\beta(P)$ with respect to β

Assuming $\alpha = 0.5$. We tabulate the values of $H_\alpha^\beta(P)$ for different values of β as given in the following table:

Table: 4.2 Monotone behaviour of $H_\alpha^\beta(P)$ with respect to β

β	2	4	6	8	10
$H_\alpha^\beta(P)$	2.03594	2.48223	2.70162	2.74676	2.75534

Next we draw the graph of the table (4.2) and illustrate from figure (4.2) that $H_\alpha^\beta(P)$ is monotonic increasing with regard values of β.

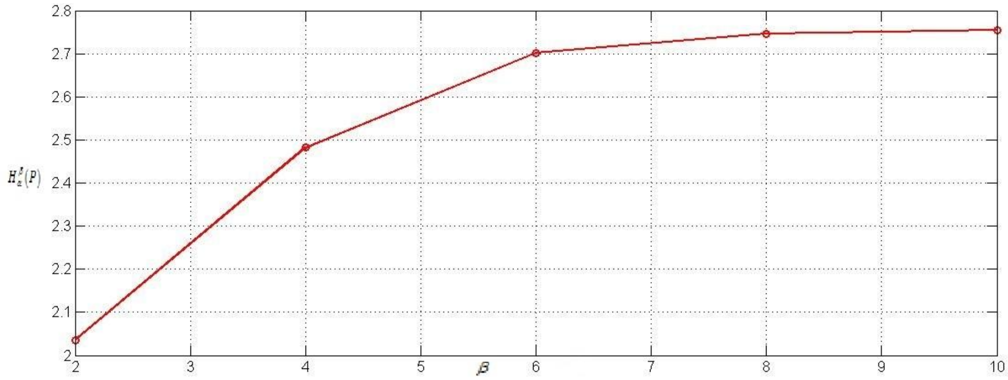

Fig4.2: Monotone behaviour of $H_\alpha^\beta(P)$ with respect to β

5. Properties of the new generalized Information Measure $H_\alpha^\beta(P)$

In this section we shall discuss properties of the new generalized measure of Information $H_\alpha^\beta(P)$ given by (2.3)

Property5.1 $H_\alpha^\beta(P)$ satisfies the additivity of the following form:

$$H_\alpha^\beta(P*Q)=H_\alpha^\beta(P)+H_\alpha^\beta(Q),$$

where $P*Q=\left(p_1q_1,\ldots,p_1q_m,p_2q_1,\ldots,p_2q_m,p_nq_1,\ldots,p_nq_m\right)$

Proof: Let $H_\alpha^\beta(P*Q)=H_\alpha^\beta(P)+H_\alpha^\beta(Q)$

R.H.S. $=H_\alpha^\beta(P)+H_\alpha^\beta(Q)$

$$=\frac{1}{1-\alpha}\log\frac{\left(\sum p_i^{\frac{\beta+2\alpha-2}{\alpha}}\right)^\alpha}{\sum p_i^\beta}+\frac{1}{1-\alpha}\log\frac{\left(\sum q_j^{\frac{\beta+2\alpha-2}{\alpha}}\right)^\alpha}{\sum q_j^\beta}$$

$$=\frac{1}{1-\alpha}\log\left[\frac{\left(\sum p_i^{\frac{\beta+2\alpha-2}{\alpha}}\right)^\alpha}{\sum p_i^\beta}\cdot\frac{\left(\sum q_j^{\frac{\beta+2\alpha-2}{\alpha}}\right)^\alpha}{\sum q_j^\beta}\right]$$

$$=\frac{1}{1-\alpha}\log\left[\frac{\sum\sum\left(p_i^{\frac{\beta+2\alpha-2}{\alpha}}\right)^\alpha\left(q_j^{\frac{\beta+2\alpha-2}{\alpha}}\right)^\alpha}{\sum\sum\left(p_iq_j\right)^\beta}\right]$$

$=H_\alpha^\beta(P*Q)$ = L.H.S.

Property5.2 $H_\alpha^\beta(P)$ is continuous if and only if $H_\alpha^\beta(P)$ is monotonic non-increasing on $q \in \left[0, \frac{1}{2}\right]$.

Proof: From (2.3) we have

$$H_\alpha^\beta(q, 1-q) = \frac{1}{1-\alpha} \log \left[\frac{\left(q^{\frac{\beta+2\alpha-2}{\alpha}} \right)^\alpha}{q^\beta} + \frac{\left\{ (1-q)^{\frac{\beta+2\alpha-2}{\alpha}} \right\}^\alpha}{(1-q)^\beta} \right].$$

Let us define function $G(q)$ by

$$G(q) = \log \left[\frac{\left(q^{\frac{\beta+2\alpha-2}{\alpha}} \right)^\alpha}{q^\beta} + \frac{\left\{ (1-q)^{\frac{\beta+2\alpha-2}{\alpha}} \right\}^\alpha}{(1-q)^\beta} \right].$$

$$\frac{dG(q)}{dq} \le 0 \quad \text{for} \quad \alpha < 1$$

It may be noted that

$$\frac{d}{dq} H_\alpha^\beta(q, 1-q) = \frac{1}{1-\alpha} \frac{dG(q)}{dq}.$$

It implies

$$\frac{d}{dq} H_\alpha^\beta(q, 1-q) \ge 0 \qquad \alpha > 0, \beta > 0, \alpha \ne 1.$$

Thus $H_\alpha^\beta(P)$ is a non- increasing monotonic function and consequently it is continuous.

Property5.3 $H_\alpha^\beta(P)$ is a symmetric function of its arguments $p_1, p_2, \ldots, p_n$.

Proof: It is evident that $H_\alpha^\beta(P)$ is a symmetric function of argument $p_1, p_2, \ldots, p_n$.

i.e. $H_\alpha^\beta(p_1, p_2, \ldots, p_{n-1}, p_n) = H_\alpha^\beta(p_n, p_1, p_2, \ldots, p_{n-1})$.

Property5.4 $H_\alpha^\beta(P)$ is non- negative.

Proof: From (2.3) we have

$$H_\alpha^\beta(P) = \frac{1}{1-\alpha} \log \left[\frac{\left(\sum p_i^{\left(\frac{\beta+2\alpha-2}{\alpha}\right)} \right)^\alpha}{\sum p_i^\beta} \right], 0 < \alpha < 1, \beta \ge 1, \alpha \ne 1.$$

From table (3.1) and (3.2) it observes that $H_\alpha^\beta(P)$ is non-negative for given values of α and β.

Property5.5 $H_\alpha^\beta(P)$ is concave function for $p_1, p_2, \ldots, p_n$.

Proof: Since the second derivative of $H_\alpha^\beta(P)$ is negative on given interval $[0,1]$.

i.e. $\frac{d^2 H_\alpha^\beta(P)}{dp_i^2} < 0$ for $p_i \in [0,1]$ and $i = 1,2,\dots,n.$,therefore,

$H_\alpha^\beta(P)$ is concave function for $p_1, p_2, \dots, p_n$.

Conclusion

The various authors have characterized the generalized information measures by various methods, but we have introduced a new generalized measure of information on studying the bounds of generalized mean codeword length by optimization technique.

Further we have established the Shannon's Noiseless Coding theorem with the help of two different coding techniques by taking experimental data and prove that Huffman coding scheme is more efficient than Shannon-Fano coding scheme. We have studied the monotone behaviour of the new generalized information measure with respect to parameters α and β. The important properties of a new generalized measure of information have also been studied.

References

1. Aczel, J. and Z. Daroczy (1963), "Uber veallegemeineste Quasilinear Mittelwerte, Die Mitgrewincbtsfunktionen Gebilelet", Sind Public Mathematics Debrecen, 10, 171-190.
2. Campbell, L.L. (1965), "A Coding Theorem and Renyi's Entropy", Information and Control, 8, 423- 429.
3. Daer, M.B. (2011), "Redundancy – Related Bounds for Generalized Huffman Codes", IEEE Transformation Information Theory, 57, 2278-2290.
4. Feinstein, A. (1958), "Foundation of Information Theory", Mc Grew-Hill, New York.
5. Hooda, D.S. and U.S. Bhaker, (1992), "A Profile on Noiseless Coding Theorems", International Journal of Management and Systems, 8, 76-85.
6. Kapur, J. N(1967), "Generalized Entropy of Order α and type β", "Math Seminar" Delhi,4.
7. Kraft, L.G. (1949), "A Device for Quantizing, Grouping and Coding Amplitude Modulated Pulses", M.S.Thesis, Electrical Engineering Department, MIT.
8. Longo, G. (1972), "A Noiseless Coding Theorem for Sources having Utilities", SIAM Journal Applied Mathematics 30, 739-748.
9. Prakash, Om and Priyanka kakkar (2012), "Development of Two new Means Codeword Lengths", Information Science 207, 90-97.
10. Shannon, C.E. (1948), "A Mathematical Theory of Communication", Bell System Technology Journal 27, 379-423.