International Journal of Statistics and Applied Mathematics 2017; 2(3): 28-36

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2017; 2(3): 28-36

www.mathsjournal.com
Received: 22-03-2017
Accepted: 24-04-2017

Dhanesh Garg
Maharishi Markendeshwar University, Mullana, Ambala, Haryana, India

Correspondence
Dhanesh Garg
Maharishi Markendeshwar University, Mullana, Ambala, Haryana, India

Generalized information measure and some source coding theorems

Dhanesh Garg

Abstract
A new measure $H_\alpha^\beta(P)$, called generalized information measure is defined and its relationship with a new mean codeword length is discussed. Consequently two noiseless coding theorems subject to Kraft's inequality have been proved. Further, we have shown that the mean codeword length $L_{\alpha,1:1}^\beta$ for the best one-to-one code (not necessarily uniquely decodable) are shorter than the mean codeword length L_α^β for the best uniquely decodable code by no more than $\log_D\left[\sum_{i=1}^{N}\left(\frac{1}{i+2}\right)\right]+2$ for $D = 2$.

Keywords: Entropy, Mean codeword length, Holder's inequality, Kraft inequality, Uniquely decipherable code, Best 1:1 code.
***MS Classifications*:** 94A15, 94A24, 26D15

Introduction
Throughout the paper $\mathbb{N}$ denotes the set of the natural numbers and for $N \in \mathbb{N}$ we set

$$\Delta_N = \left\{(p_1,\dots,p_N) \mid p_i \geq 0,\ i=1,\dots,N,\ \sum_{i=1}^{N} p_i = 1\right\}.$$

In case there is no rise to misunderstanding we write $P \in \Delta_N$ instead of $(p_1,\dots,p_N) \in \Delta_N$.
In case $P \in \Delta_N$ the well-known Shannon entropy is defined by

$$H(P) = H(p_1,\dots,p_N) = -\sum_{i=1}^{N} p_i \log(p_i), \tag{1.1}$$

where the convention $0\log(0) = 0$ is adapted, (see Shannon) [23].

Throughout this paper, $\sum$ will stand for $\sum_{i=1}^{N}$ unless otherwise stated and logarithms are taken to the base $D(D>1)$.

Let a finite set of N input symbols

$$X = \{x_1, x_2, \dots, x_N\}$$

be encoded using alphabet of D symbols, then it is shown in Feinstein [5] that there is a uniquely decipherable code with lengths $n_1, n_2, \dots, n_N$ if and only if the Kraft inequality holds, that is,

$$\sum_{i=1}^{N} D^{-n_i} \leq 1, \tag{1.2}$$

where D is the size of code alphabet. Kraft's [13] inequality which is uniquely determined by the condition for unique decipherability and plays an important role in proving a noiseless coding theorem without dependent on the probabilities or utilities. This inequality cannot be modified just to prove noiseless coding theorems as it will not give uniquely decipherable codes but will provide codes with a different structure. To prove some noiseless coding theorems we have made use of Kraft's inequality in original.

Furthermore, if

$$L = \sum_{i=1}^{N} n_i p_i \tag{1.3}$$

is the average codeword length, then for a code satisfying (1.2), the inequality

$$H(P) \le L < H(P) + 1 \tag{1.4}$$

is also fulfilled and equality, $L = H(P)$, holds if and only if

$$n_i = -\log_D(p_i) \ (i = 1,..., N), and \sum_{i=1}^{N} D^{-n_i} = 1. \tag{1.5}$$

If $L < H(P)$, then by suitable encoding of long input sequences, the average number of code letters per input symbol can be made arbitrarily close to $H(P)$ (see Feinstein) [5]. This is Shannon's noiseless coding theorem.

A coding theorem analogous to Shannon's noiseless coding theorem has been established by Campbell [2], in terms of Renyi's entropy [21]:

$$H_\alpha(P) = \frac{1}{1-\alpha} \log_D \sum p_i^\alpha, \alpha > 0 (\ne 1). \tag{1.6}$$

Kieffer [11] defined class rules and showed $H_\alpha(P)$ is the best decision rule for deciding which of the two sources can be coded with least expected cost of sequences of length n when $n \to \infty$, where the cost of encoding a sequence is assumed to be a function of length only. Further, in Jelinek [9] it is shown that coding with respect to Campbell's mean length is useful in minimizing the problem of buffer overflow which occurs when the source symbol is produced at a fixed rate and the code words are stored temporarily in a finite buffer. Concerning Campbell's mean length the reader can consult Campbell [2]. Longo [16] developed lower bound for useful mean codeword length in terms of weighted entropy introduced by Belis and Guiasu [2]. Guiasu and Picard [6] proved a noiseless coding theorem by obtaining lower bounds for another useful mean codeword length. Gurdial and Pessoa [7] extended the theorem by finding lower bounds for useful mean codeword length of order α.

Chapeau-Blondeau *et al*. [3] have presented an extension to source coding theorem traditionally based upon Shannon's entropy and later generalized to Renyi's entropy. Chapeau-Blondeau *et al.* [4] have described a practical problem of source coding and investigated an important relation stressing that Renyi's entropy emerges at an order a differing from the traditional Shannon's entropy. Some new lower and upper bounds for compression rate of binary prefix codes optimized over memoryless sources have been provided by Baer [1]. Singh *et al.* [26] have provided the application of weighted measures of entropy to the field of coding theory.

Ramamoorthy [20] considered the problem of transmitting multiple compressible sources over a network at minimum cost with the objective to find the optimal rates at which the sources should be compressed. Tu *et al.* [28] have presented a new scheme based on variable length coding, capable of providing reliable resolutions for flow media data transmission in spatial communication. Some interesting work for the construction of information theoretic source network coding in the presence of eavesdroppers has been presented by Luo *et al*. [17]. Wu *et al.* [29] have constructed a space trellis and design a low-complexity joint decoding algorithm with a variable length symbol-a posteriori probability algorithm in resource constrained deep space communication networks. The applications of coding theory to the field of marketing have been provided by Neill [19]. Some other related work concerned with the coding theory have been provided by Sharma and Raina [24] and Koski and Persson [12], etc.

The mean length of a noiseless uniquely decodable code for a discrete random variable X satisfies

$$H(X) \le L_{UD} < H(X) + 1 \tag{1.7}$$

Where $H(X) = -\sum_{i=1}^{N} p_i \log_D(p_i)$ is the Shannon's entropy [23] of the random variable X. Shannon's restriction of coding of X to prefix codes is highly justified by the implicit assumption that the description will be concatenated and thus must be uniquely decordable and instantaneous codes, cf. [1, 2], the expected codeword length is the same for both the set of codes.

There are some communication situations in which a random variable X is being transmitted rather than a sequence of random variables. For this context Leung-Yang-Cheong and Cover [15] considered one to one codes i.e., codes which assign a distinct binary code to each outcome of the random variable X without regard to the condition that concatenations of the descriptions must be uniquely decipherable. The authors showed by Lagrange multiplier arguments that the minimum length of the best 1:1 binary code satisfy the inequality

$$L_{1:1} \ge H(P) - \log_2 \sum_{i=1}^{N} \left(\frac{2}{i+2} \right) \tag{1.8}$$

where $L_{1:1} = \sum_{i=1}^{N} p_i \left[\log_D \left(\frac{i}{2} + 1 \right) \right]$, $[x]$ denotes the smallest integer greater than or equal to $x, H(P)$ is the first order approximation to minimal expected length of 1:1 codes and logarithms are taken to the base $D = 2$.
It has also been proved in [15] that

$$L_{1:1} \geq H(P) - log_D log_D N - 3, \tag{1.9}$$

and

$$L_{1:1} \geq H(P) - log_D(H(P)+1) - 0(log_D log_D(H(P)+1)), \tag{1.10}$$

where N in (1.9) is the number of values that the random variable X can take on, while in (1.10), X is allowed to have countably many values. Rissanen [22] improved the lower bound (1.8) by showing that

$$L_{1:1} \geq H(P) - log_D K(N), \tag{1.11}$$

Where

$$K(N) = n_N - 1 + r_N 2^{-n_N} \leq log_D N, \tag{1.12}$$

And n_N is the maximum integer such that for a positive remainder r_N,

$$N = 2^{n_N} + r_N - 2. \tag{1.13}$$

The inequality (1.11) is strictly better than (1.8) because it was shown by Rissanen [22] that

$$K(N) < \sum_{i=1}^{N} \left(\frac{2}{i+2} \right), \tag{1.14}$$

For every N.
In Section 2, we have introduced the new generalized information measure and consequently proved two new coding theorems. Some fascinating relations between entropy and the best 1:1 codeword lengths have been developed, the findings of which are presented in Section 3.

2. Coding Theorems

Definition: Let $N \in \mathbb{N}$ be arbitrarily fixed, $\alpha, \beta > 0, \alpha \neq 1$ be given real numbers. Then the information measure $H_\alpha^\beta : \Delta_N \to \mathbb{R}$ is defined by

$$H_\alpha^\beta(p_1, ..., p_N) = \frac{\alpha}{1-\alpha} \log_D \frac{\sum_{i=1}^{N} p_i^\beta}{\left(\sum_{i=1}^{N} p_i^{\alpha\beta} \right)^{1/\alpha}} \quad ((p_1, ..., p_N) \in \Delta_N). \tag{2.1}$$

Remarks

(i) If $\beta = 1$ or $\alpha = \frac{1}{\beta}$, then the information measure H_α^β reduces to Renyi's [21] entropy.

(ii) When $\beta \to 1$ and $\alpha \to 1$, then the information measure H_α^β reduces to Shannon [23] entropy.

Definition: Let $N \in \mathbb{N}$, $\alpha, \beta > 0, \alpha \neq 1$ be arbitrarily fixed, then the mean length L_α^β corresponding to the generalized information measure H_α^β is given by the formula

$$L_\alpha^\beta = \frac{1}{\alpha - 1} \log_D \left(\frac{\sum_{i=1}^{N} p_i^{\alpha\beta} D^{n_i(\alpha-1)}}{\sum_{i=1}^{N} p_i^{\alpha\beta}} \right). \tag{2.2}$$

It is also studied by Kumar and Choudhary [14].
Where $D, n_1, n_2, ..., n_N$ are positive integers so that

$$\sum_{i=1}^{N} D^{-n_i} \leq 1. \tag{2.3}$$

(i) If $\beta = 1$, then L_α^β reduces to Nath [18] average code word length,

$$i.e., \quad L_\alpha = \frac{1}{\alpha-1}\log_D\left(\frac{\sum_{i=1}^{N} p_i^\alpha D^{n_i(\alpha-1)}}{\sum_{i=1}^{N} p_i^\alpha}\right). \tag{2.4}$$

(ii) If $\alpha = \frac{1}{\beta}$, then it reduces to Cambell's Codeword Length,

$$i.e., \quad L^\beta = \frac{\beta}{1-\beta}\log_D\left(\sum_{i=1}^{N} p_i D^{n_i\left(\frac{1-\beta}{\beta}\right)}\right). \tag{2.5}$$

(iii)When $\beta \to 1$ and $\alpha \to 1$, then $L_\alpha^\beta \to L$, where $L = \sum_{i=1}^{N} p_i n_i$.

(iv) When $n_1 = n_2 = n_3 = \cdots = n_N = n$, then $L_\alpha^\beta = 1$.

(v) L_α^β lies between minimum and maximum values of $n_1, n_2, ..., n_N$.

(vi)When $\alpha \to 1$ and $\beta \neq 1$, then (2.2) reduces to $L^\beta = \frac{\sum n_i p_i^\beta}{\sum p_i^\beta}$.

Applications of Holder's Inequality in Coding Theory

In the following theorem, we find lower bound for L_α^β.

Theorem 2.1. Let $\alpha, \beta > 0, \alpha \neq 1$ be arbitrarily fixed real numbers, then for all integers $D > 1$ inequality

$$L_\alpha^\beta \geq H_\alpha^\beta(P) \tag{2.6}$$

is fulfilled. Furthermore, equality holds if and only if

$$n_i = -\log_D \frac{p_i^\beta}{\sum_{i=1}^{N} p_i^\beta}. \tag{2.7}$$

Proof

By Shisha [25] Holder's inequality, we have

$$\left(\sum_{i=1}^{N} x_i^p\right)^{\frac{1}{p}}\left(\sum_{i=1}^{N} y_i^q\right)^{\frac{1}{q}} \leq \sum_{i=1}^{N} x_i y_i, \tag{2.8}$$

where $p^{-1} + q^{-1} = 1;\ p(\neq 0) < 1, q < 0\ \ or\ \ q(\neq 0) < 1,\ p < 0;\ \ x_i, y_i > 0$ for each i.

Let $p = \dfrac{\alpha-1}{\alpha}, q = 1-\alpha, \quad x_i = p_i^{\frac{\alpha\beta}{\alpha-1}}, y_i = p_i^{\frac{\alpha\beta}{1-\alpha}} D^{-n_i}, \quad (i = 1,..., N).$

Putting these values into (2.8), we get

$$\left(\sum_{i=1}^{N} p_i^\beta\right)^{\frac{\alpha}{\alpha-1}}\left(\sum_{i=1}^{N} p_i^{\alpha\beta} D^{-n_i(1-\alpha)}\right)^{\frac{1}{1-\alpha}} \leq \sum_{i=1}^{N} D^{-n_i} \leq 1,$$

where we used (2.3), too. This implies however that

$$\left(\sum_{i=1}^{N} p_i^{\alpha\beta} D^{-n_i(1-\alpha)}\right)^{\frac{1}{1-\alpha}} \leq \left(\sum_{i=1}^{N} p_i^\beta\right)^{\frac{\alpha}{1-\alpha}} \tag{2.9}$$

Now consider two cases:

(i) For $0 < \alpha < 1$, when we devide (2.9) by $\left(\sum_{i=1}^{N} p_i^{\alpha\beta}\right)^{\frac{1}{1-\alpha}}$ and take logarithms,

We obtain the result (2.6) after simplification, i.e., $L_\alpha^\beta \geq H_\alpha^\beta(P)$.

(ii) The proof for $\alpha > 1$, follows on similar case (i).

It is clear that the equality in (2.6) is true if $n_i = -\log_D \dfrac{p_i^\beta}{\sum_{i=1}^N p_i^\beta}$. The necessity of this condition for equality in (2.6) follows from the condition for equality in Holder's inequality: In the case of the Holder's inequality given above, equality holds if and only if for some c,

$$x_i^p = c y_i^q, i = 1,2,\ldots,N. \tag{2.10}$$

Plugging this condition into our situation, with the x_i, y_i, p and q as specified, and using the fact that the $\sum_{i=1}^N p_i = 1$, the necessity is proven.

In the following theorem, we give an upper bound for L_α^β in terms of $H_\alpha^\beta(P)$.

Theorem 2.2. For α and β as in Theorem 2.1, there exist positive integers n_i satisfying (2.3) such that

$$L_\alpha^\beta < \mathrm{H}_\alpha^\beta(P) + 1. \tag{2.11}$$

Proof: Let n_i be the positive integer satisfying the inequalities

$$-\log_D \frac{p_i^\beta}{\sum_{i=1}^N p_i^\beta} \leq n_i < -\log_D \frac{p_i^\beta}{\sum_{i=1}^N p_i^\beta} + 1. \tag{2.12}$$

It is easy to see that the sequence $\{n_i\}$, $i = 1,2,\ldots,N$ thus defined, satisfies (2.5).

Now, from the right inequality of (2.12), we have

$$n_i < -\log_D \frac{p_i^\beta}{\sum_{i=1}^N p_i^\beta} + 1$$

$$\Rightarrow D^{-n_i} > D^{-1} \frac{p_i^\beta}{\sum_{i=1}^N p_i^\beta}. \tag{2.13}$$

Now consider two cases:

(i) Let $0 < \alpha < 1$, then raising power $1-\alpha$, to both sides of (2.13), we have

$$D^{-n_i(1-\alpha)} > D^{(\alpha-1)} \left(\frac{p_i^\beta}{\sum_{i=1}^N p_i^\beta} \right)^{(1-\alpha)} \tag{2.14}$$

(i) When we multiply (2.14) by $p_i^{\alpha\beta}$, and then summing up from $i = 1$ to $i = N$, devide (2.14) by $\sum_{i=1}^N p_i^{\alpha\beta}$ and take logarithms, We obtain the result (2.11) after simplification.

(ii) The proof for $\alpha > 1$, follows on similar case (i).

Remark

1. Huffman [8] introduced a procedure for designing a variable length source code which achieves performance close to Shannon's entropy bound. For individual codeword lengths n_i, the average length $L = \sum_{i=1}^N \mathrm{p_i n_i}$ of Huffman code is always within one unit of Shannon's measure of entropy, that is,

$$H(X) \leq L_{UD} < H(X) + 1$$

Where $H(X) = -\sum_{i=1}^{N} p_i \log_D(p_i)$ is the Shannon's measure of entropy. Huffman coding scheme can also be applied to codeword length L_α^β i.e., for individual codeword lengths n_i, the average length L_α^β of Huffman code satisfies

$$\mathrm{H}_\alpha^\beta(P) \le L_\alpha^\beta < \mathrm{H}_\alpha^\beta(P) + 1,$$

as numerically shown in following Table 1.

Table 1: Relation between L_α^β and $\mathrm{H}_\alpha^\beta(P)$

p_i	**Huffman codewords**	**Length of Huffman codewords (n_i)**	α	β	L_α^β	$\mathrm{H}_\alpha^\beta(P)$
0.3	00	2	3	2	0.183	0.148
0.25	10	2				
0.2	11	2				
0.1	011	3				
0.1	0100	4				
0.05	0101	4				

Now we prove the coding theorem for noiseless channels with independent input symbols. Let S^M denote the Mth order extension of a discrete memoryless source S which emits the symbols $(x_1, x_2, \dots, x_N)$ with positive probabilities $(p_1, p_2, \dots, p_N), \sum_{i=1}^{N} p_i = 1.$ Then S^M is also a discrete memoryless source which emits $M-$tuples of the form $s = (a_1, a_2, \dots, a_N),$ where

$$a_1 = x_{i_1}, \dots, a_M = x_{i_M}.$$

Since the symbols are emitted independently,

$$P(s) = P(a_1)P(a_2)\dots P(a_N) = p_{i_1} p_{i_2} \cdots p_{i_M}.$$

If $N(s)$ denotes the length of the code word assigned to $M-$tuple s and $L_{M,\alpha}^\beta$ denotes the average code length, then

$$L_{M,\alpha}^\beta = \frac{1}{\alpha-1} \log_D \left(\frac{\sum_s P^{\alpha\beta}(s) D^{N(s)(\alpha-1)}}{\sum_s P^{\alpha\beta}(s)} \right), \quad \alpha, \beta > 0, \alpha \ne 1$$

$$= \sum_s P(s) N(s), \beta = 1, \alpha = 1.$$

Let $\alpha, \beta > 0, \alpha \ne 1.$ Since Renyi's entropy is finitely additive, by using (2.7) and **(2.12)** we get

$$\mathrm{H}_\alpha^\beta(P) \le L_{M,\alpha}^\beta / M < \mathrm{H}_\alpha^\beta(P) + 1/M, \alpha, \beta > 0, \alpha \ne 1.$$

Obviously,

$$\lim_{M \to \infty} L_{M,\alpha}^\beta / M = \mathrm{H}_\alpha^\beta(P).$$

Thus, we have proved the following theorem.

Theorem 2.3. Let there be a discrete memoryless source emitting independent input symbols $x_1, x_2, \dots, x_N$ governed by the distribution $P = (p_1, \dots, p_N) \in \Delta_N.$ By suitable encoding in a uniquely decipherable way sufficiently long M-sequences of input symbols, it is possible to make $L_{M,\alpha}^\beta / M,$ the average code length of order α and type β per input symbol, as closely to $\mathrm{H}_\alpha^\beta(P)$ as desired. It is not possible to find a uniquely decipherable code whose average code length L_α^β of order α and type β is less than Renyi's entropy order α.

We remark that the last part of the above theorem needs no explanation as it is obvious from **Theorem 2.1 and Theorem 2.2.**

In the next section, we have developed the relation between entropy and best 1:1 codeword length.

3. Codeword Length for the best 1:1 code and its Lower bound

In the above section, we have proved in Theorem 2.1 and Theorem 2.2 that new codeword length L_{α}^{β} satisfies $\mathrm{H}_{\alpha}^{\beta}(P) \leq L_{\alpha}^{\beta} < \mathrm{H}_{\alpha}^{\beta}(P) + 1$ for all uniquely decipherable codes. We consider this codeword length for the best 1:1 code and find its lower bound. Let X be a random variable taking on a finite number of values $(x_1, x_2, ..., x_N)$ which have to be encoded in terms of an alphabet of size D with probabilities $(p_1, p_2, ..., p_N)$. Without any loss of generality, we may assume that $p_1 \geq p_2 \geq p_3 \geq ... \geq p_N$. Here we take $D = 2$, that is, the set of available codewords is $(0,1)$, the set of all finite binary sequences. Let $n_i, i = 1,2,...,N$ denotes the length of the codeword for x_i. For the best 1:1 code, it is clear that $n_1 \leq n_2 \leq n_3 \leq ... \leq n_N$. Since there are 2^n 1:1 codewords of length n so $n_1 = 1, n_2 = 1, n_3 = 2, ...$ and in general $n_i = \left[\log_D\left(\frac{i}{2}+1\right)\right]$ [15], where $[x]$ denotes the smallest integer greater than or equal to x. Thus

$$L_{\alpha,1:1}^{\beta} = \frac{1}{\alpha-1}\log_D\left(\frac{\sum_{i=1}^{N} p_i^{\alpha\beta} D^{\left[\log_D\left(\frac{i}{2}+1\right)\right](\alpha-1)}}{\sum_{i=1}^{N} p_i^{\alpha\beta}}\right), \tag{3.1}$$

We will now prove the following theorem, which gives a lower bound on $L_{\alpha,1:1}^{\beta}$.

Theorem 3.1 For $H_{\alpha}^{\beta}(P)$, L_{α}^{β} and $L_{\alpha,1:1}^{\beta}$ as given in (2.1), (2.2) and (3.1) respectively, the following estimates hold:

$$L_{\alpha,1:1}^{\beta} \geq H_{\alpha}^{\beta}(P) - \log_D\left[\sum\left(\frac{2}{i+2}\right)\right], \tag{3.2}$$

and

$$L_{\alpha,1:1}^{\beta} \geq L_{\alpha}^{\beta} - \log_D\left[\sum_{i=1}^{N}\left(\frac{2}{i+2}\right)\right] - 1. \tag{3.3}$$

Proof: From (3.3), we have

$$L_{\alpha,1:1}^{\beta} \geq \frac{1}{\alpha-1}\log_D\left(\frac{\sum_{i=1}^{N} p_i^{\alpha\beta} D^{\left[\log_D\left(\frac{i}{2}+1\right)\right](\alpha-1)}}{\sum_{i=1}^{N} p_i^{\alpha\beta}}\right) \tag{3.4}$$

Now

$$H_{\alpha}^{\beta}(P) - L_{\alpha,1:1}^{\beta} \leq \frac{\alpha}{1-\alpha}\log_D \frac{\sum_{i=1}^{N} p_i^{\beta}}{\left(\sum_{i=1}^{N} p_i^{\alpha\beta}\right)^{1/\alpha}} - \frac{1}{\alpha-1}\log_D\left(\frac{\sum_{i=1}^{N} p_i^{\alpha\beta} D^{\left[\log_D\left(\frac{i}{2}+1\right)\right](\alpha-1)}}{\sum_{i=1}^{N} p_i^{\alpha\beta}}\right)$$

$$= \log_D\left(\sum_{i=1}^{N} p_i^{\beta}\right)^{\left(\frac{\alpha}{\alpha-1}\right)}\left(\sum_{i=1}^{N} p_i^{\alpha\beta}\left(\frac{i+2}{2}\right)^{\alpha-1}\right)^{\frac{1}{1-\alpha}} \tag{3.5}$$

Applying Holder's inequality to (3.5), we obtain

$$H_{\alpha}^{\beta}(P) - L_{\alpha,1:1}^{\beta} \leq \log_D\left[\sum_{i=1}^{N}\left(\frac{2}{i+2}\right)\right] \tag{3.6}$$

Which gives (3.2).

Now from (2.11)

$$L_\alpha^\beta < H_\alpha^\beta(P)+1$$

So

$$L_\alpha^\beta - L_{\alpha,1:1}^\beta < H_\alpha^\beta(P) - L_{\alpha,1:1}^\beta + 1$$

$$\le 1 + \log_D\left[\sum_{i=1}^{N}\left(\frac{2}{i+2}\right)\right]$$

Which proves (3.3).

Theorem 3.2. For $K(N)$ as given in (1.12), we have the following inequality

$$L_{\alpha,1:1}^\beta \ge H_\alpha^\beta(P) - \log K(N). \tag{3.7}$$

Proof: The Kraft's sum of the best 1:1 code

$$\sum_{i=1}^{N} 2^{-n_i} = 2\times 2^{-1} + 4\times 2^{-2} + \ldots + 2^{n_N-1}\times 2^{1-n_N} + r_N \times 2^{-n_N} = K(N), \tag{3.8}$$

Where r_N satisfies (1.13).

Let $p = 1-\alpha,\ q = \dfrac{\alpha-1}{\alpha},\ x_i = p_i^{\frac{\alpha\beta}{1-\alpha}} 2^{-n_i},\ y_i = p_i^{\frac{\alpha\beta}{\alpha-1}}\ (i = 1,\ldots,N).$

Putting these values into (2.8), we get

$$\left(\sum_{i=1}^{N} p_i^{\alpha\beta} 2^{-n_i(1-\alpha)}\right)^{\frac{1}{1-\alpha}} \left(\sum_{i=1}^{N} p_i^{\beta}\right)^{\frac{\alpha}{\alpha-1}} \le \sum_{i=1}^{N} 2^{-n_i}. \tag{3.9}$$

Taking Logarithms on both sides of (3.9) and using (3.8), we get

$$\frac{1}{1-\alpha}\log\left(\sum_{i=1}^{N} p_i^{\alpha\beta} 2^{-n_i(1-\alpha)}\right) + \frac{1}{\alpha-1}\log\left(\sum_{i=1}^{N} p_i^{\beta}\right)^{\alpha} \le \log K(N). \tag{3.10}$$

Where $n_i = \left[\log_D\left(\dfrac{i}{2}+1\right)\right]$ and after simplification, (3.10) becomes

$$\frac{1}{1-\alpha}\log_D\left(\frac{\sum_{i=1}^{N} p_i^{\alpha\beta} D^{\left[\log_D\left(\frac{i}{2}+1\right)\right](\alpha-1)}}{\sum_{i=1}^{N} p_i^{\alpha\beta}}\right) + \frac{\alpha}{1-\alpha}\log\frac{\sum_{i=1}^{N} p_i^{\beta}}{\left(\sum_{i=1}^{N} p_i^{\alpha\beta}\right)^{1/\alpha}} \le \log K(N).$$

Which proves (3.7).

References

1. Baer MB. Redundancy-related bounds for generalized Huffman codes. IEEE Trans. Inf. Theory. 2011; 57:2278-2290.
2. Campbell LL. A coding theorem and Renyi's entropy. Information and Control. 1965; 8:423-429.
3. Chapeau-Blondeau F, Delahaies A, Rousseau D. Source coding with Tsallis entropy. Electron. Lett., 2011; 47:187-188.
4. Chapeau-Blondeau F, Rousseau D, Delahaies A. Renyi entropy measure of noise aided information transmission in a binary channel. Phys. Rev. E, 2010; 81:051112(1)-051112(10).
5. Feinstein A. Foundation of Information Theory. McGraw Hill, New York, 1956.
6. Guiasu S, Picard CF. Borne infericutre de la Longuerur utile de certain codes. C.R. Acad. Sci. Paris. 1971; 273(A):248-251.
7. Gurdial, Pessoa F. On useful information of order α. J. Combut. Inf. Syst. Sci., 1977; 2:158-162.
8. Huffman DA. A method for the construction of minimum redundancy codes. In: Proceedings of the IRE. 1952; 40:1098-1101.
9. Jelinek F. Buffer overflow in variable lengths coding of fixed rate sources. IEEE, 1980; 3:490-501.
10. Kapur JN. Entropy and Coding. MSTS, New Delhi, 1998.
11. Kieffer JC. Variable lengths source coding with a cost depending only on the codeword length. Information and Control. 1979; 41:136-146.
12. Koski T, Persson LE. Some properties of generalized exponential entropies with applications to data compression. Inform. Sci., 1992; 62:103-132.
13. Kraft LG. A Device for Quantizing Grouping and Coding Amplitude Modulated Pulses. M.S. Thesis, Electrical Engineering Department, MIT, 1949.

14. Kumar S, Choudhary A. A Noiseless Coding theorem connected with generalized Renyi's entropy of order α for incomplete power probability distribution P^{β}. Asian Journal of Applied Sciences. 2011; 4(6):649-656.
15. Leung-Yan-Cheong SK, Cover T. Some equivalence between Shannon entropy and Kolmogrov complexity. IEEE Trans. On Inform. Theory. 1978; 24:331-338.
16. Longo G. A noiseless coding theorem for sources having utilities. SIAM J. Appl. Math. 1976; 30:739-748.
17. Luo MX, Yang YX, Wang LC, Niu XX. Secure network coding in the presence of eavesdroppers. Sci. China Inform. Sci. 2011; 53:648-658.
18. Nath P. An axiomatic characterization of inaccuracy for discrete generalized probability distribution. Opsearch. 1970; 7:115-133.
19. Neill WDO. An application of Shannon's coding theorem to information transmission in economic markets. Inform. Sci. 1987; 41:171-185.
20. Ramamoorthy A. Minimum cost distributed source coding over a network. IEEE Trans. Inf. Theory. 2011; 57:461-475.
21. Renyi A. On Measure of entropy and information. Proc. 4th Berkeley Symp. Maths. Stat. Prob., 1961; 1:547-561.
22. Rissanen J. Tight lower bounds for optimum code length. IEEE Trans. Inf. Theory, 1982; 28:348-349.
23. Shannon CE. A Mathematical Theory of Communication. Bell System Tech. J. 1948; 27:379-423, 623-656.
24. Sharma BD, Raina N. Coding theorem for partial received information. Inform. Sci. 1980; 20:181-189.
25. Shisha O. Inequalities. Academic Press, New York, 1967.
26. Singh RP, Kumar R, Tuteja RK. Applications of holder's inequality in information theory. Inform. Sci. 2003; 152:145-154.
27. Taneja IJ. A short note on exponentiated mean codeword length for the best 1:1 code, Computational and Applied Mathematics. 1984; 3:199-204.
28. Tu GF, Liu JJ, Zhang C, Gao SS, Li SD. Studies and advances on joint source-channel encoding/decoding techniques in flow media communication. Sci. China Inform. Sci., 2011; 54:1883-1894.
29. Wu WR, Tu J, Tu GF, Gao SS, Zhang C. Joint source channel VL coding/decoding for deep space communication networks based on a space trellis. Sci. China Inform. Sci., 2010; 53:1-17.
30. Garg D, Kumar S. Parametric R-norm directed-divergence convex function, Infinite Dimensional Analysis, Quantum Probability and Related Topics, World Science Journals. 2010; 19(2), 2016, DOI: 10.1142/S0219025716500144
31. Garg D, Kumar S. An application of generalized Tsalli's-Havrda-Charvat entropy in coding theory through a generalization of Kraft inequality, International Journal of Applied Research. 2016; 1(4):01-05, ISSN: 2394-5869.
32. Garg D, Kumar S. Some Remarks On 'Useful' Renyi's Entropy of Order α in International Journal of trends in Mathematics and Statistics (IJTIMS). 2015; 4(10):178-183, ISSN: 2279-0318.