

Optimum Probability Distribution for Minimum Redundancy of Source Coding

Om Parkash¹, Priyanka Kakkar²

Department of Mathematics, Guru Nanak Dev University, Amritsar, India
Email: omparkash777@yahoo.co.in, priyanka_kakkar85@yahoo.com

Received October 8, 2013; revised November 8, 2013; accepted November 15, 2013

Copyright © 2014 Om Parkash, Priyanka Kakkar. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. In accordance of the Creative Commons Attribution License all Copyrights © 2014 are reserved for SCIRP and the owner of the intellectual property Om Parkash, Priyanka Kakkar. All Copyright © 2014 are guarded by law and by SCIRP as a guardian.

ABSTRACT

In the present communication, we have obtained the optimum probability distribution with which the messages should be delivered so that the average redundancy of the source is minimized. Here, we have taken the case of various generalized mean codeword lengths. Moreover, the upper bound to these codeword lengths has been found for the case of Huffman encoding.

KEYWORDS

Mean Codeword Length; Uniquely Decipherable Code; Kraft's Inequality; Entropy; Optimum Probability Distribution; Escort Distribution; Source Coding

1. Introduction

Any message that brings a specification in a problem which involves a certain degree of uncertainty is called information and it was Shannon [1] who named this measure of information as entropy. In coding theory, the operational role of entropy comes from the source coding theorem which states that if H is the entropy of the source letters for a discrete memoryless source, then the sequence of source outputs cannot be represented by a binary sequence using fewer than H binary digits per source digit on the average, but it can be represented by a binary sequence using as close to H binary digits per source digit on the average as desired. To be clearer, let us consider the discrete source S that emits symbols x_1, x_2, \dots, x_n with probability distribution

$p = (p_1, p_2, \dots, p_n)$ where $\sum_{i=1}^n p_i = 1$. The aim of source coding is to encode the source using an alphabet of size D , that is, to map each symbol x_i to a codeword c_i of length l_i expressed using the D letters of the alphabet. It is known that if the set of lengths l_i satisfies Kraft's [2] inequality

$$\sum_{i=1}^n D^{-l_i} \leq 1 \quad (1.1)$$

then there exists a uniquely decodable code with these lengths, which means that any sequence $c_{i_1}c_{i_2}\dots c_{i_n}$ can be decoded unambiguously into a sequence of symbols $x_{i_1}x_{i_2}\dots x_{i_n}$. In this respect, Shannon [1] proved the first noiseless coding theorem for the uniquely decipherable code in the form of following inequality

$$H(p) \leq L < H(p) + 1 \quad (1.2)$$

where $H(p) = -\sum_{i=1}^n p_i \log_D p_i$ is a Shannon's entropy and $L = \sum_{i=1}^n p_i l_i$ is the mean codeword length.

Later, Campbell [3] and Kapur [4] proved the source coding theorems for their own exponentiated mean co-

deword length in the form of following inequalities

$$R_\alpha(p) \leq L_\alpha < R_\alpha(p) + 1 \quad (1.3)$$

and

$$R_\alpha(p) \leq L^\alpha < R_\alpha(p) + 1 \quad (1.4)$$

respectively, where

$$L_\alpha = \frac{\alpha}{1-\alpha} \log_D \left(\sum_{i=1}^n p_i D^{\frac{l_i(1-\alpha)}{\alpha}} \right), \alpha \neq 1, \alpha > 0$$

is Campbell's [3] mean codeword length,

$$L^\alpha = \frac{1}{\alpha-1} \log_D \left(\frac{\sum_{i=1}^n p_i^\alpha D^{l_i(\alpha-1)}}{\sum_{i=1}^n p_i^\alpha} \right), \alpha \neq 1, \alpha > 0$$

is Kapur's [4] mean codeword length and

$$R_\alpha(p) = \frac{1}{1-\alpha} \log_D \sum_{i=1}^n p_i^\alpha, \alpha \neq 1, \alpha > 0$$

is Renyi's [5] measure of entropy.

Recently, Parkash and Kakkar [6] introduced two mean codeword lengths given by

$$L(\alpha, \beta) = \frac{\alpha}{1-\alpha} \log_D \left(\frac{\sum_{i=1}^n p_i^\alpha D^{\frac{l_i(1-\alpha)}{\alpha}}}{\sum_{i=1}^n p_i^\alpha} \right), \alpha > 0, \beta > 0, \alpha \neq 1 \quad (1.5)$$

and

$$L(\beta) = \frac{\sum_{i=1}^n p_i^\beta l_i}{\sum_{i=1}^n p_i^\beta}, \beta > 0 \quad (1.6)$$

Further, the authors provided two source coding theorems which show that for all uniquely decipherable codes, the mean codeword lengths $L(\alpha, \beta)$ and $L(\beta)$ satisfy the relation:

$$E_\beta^\alpha(p) \leq L(\alpha, \beta) < E_\beta^\alpha(p) + 1 \quad (1.7)$$

and

$$K^\beta(p) \leq L(\beta) < K^\beta(p) + 1 \quad (1.8)$$

respectively where

$$E_\beta^\alpha(p) = \frac{1}{\alpha-1} \log_D \left[\frac{\left(\sum_{i=1}^n p_i^\alpha \right)^\alpha}{\sum_{i=1}^n p_i^\beta} \right]$$

is a Kapur's [4] two parameter additive measure of entropy and

$$K^\beta(p) = -\frac{\sum_{i=1}^n p_i^\beta \log_D \left(\frac{p_i^\beta}{\sum_{i=1}^n p_i^\beta} \right)}{\sum_{i=1}^n p_i^\beta}, \beta > 0$$

is measure of entropy developed by Parkash and Kakkar [6].

This is to emphasize that in the entire literature of source coding theorems, one can observe that the mean codeword length is lower bounded by the entropy of the source and it can never be less than the entropy of the source but can be made closer to it. This phenomenon provides the idea of absolute redundancy which is the number of bits used to transmit a message minus the number of bits of actual information in the message, that is, the mean codeword length minus the entropy of the source. The objective of the present communication is to minimize this redundancy in order to increase the efficiency of the source encoding. For this purpose we have made use of the concept of escort distribution as follows:

If $p = (p_1, p_2, \dots, p_n)$ is the original distribution, then its escort distribution is given by $P = (P_1, P_2, \dots, P_n)$ where $P_i = \frac{p_i^\beta}{\sum_{i=1}^n p_i^\beta}$ for some parameter $\beta \neq 0$. Many researchers including Harte [7], Bercher [8,9], Beck and

Schloegl [10] etc. used this distribution in their respective findings.

The aim of the present paper is to obtain the optimum probability distribution with which the source should deliver messages in order to minimize the absolute redundancy. To obtain our goal, we have taken into consideration the above mentioned generalized mean codeword lengths. Moreover, the upper bound to these codeword lengths has been found for Huffman [4] encoding.

2. Optimum Probability Distribution to Minimize Absolute Redundancy

Let us assume that for discrete source S that emits symbols x_1, x_2, \dots, x_n with probability distribution $p = (p_1, p_2, \dots, p_n)$, the codewords c_i having lengths $l_i, i = 1, 2, \dots, n$, have been obtained using some encoding procedure on noiseless channel. Further, we assume that entropy of the source is $E_\beta^\alpha(p)$ and average codeword length is $L(\alpha, \beta)$. Since from (1.7), we have $E_\beta^\alpha(p) \leq L(\alpha, \beta)$, therefore, the average redundancy of the source code is given by

$$\begin{aligned} f(p_1, p_2, \dots, p_n) &= L(\alpha, \beta) - E_\beta^\alpha(p) = \frac{\alpha}{1-\alpha} \log_D \left(\frac{\sum_{i=1}^n (p_i^\beta)^{\frac{1}{\alpha}} D^{\frac{l_i(1-\alpha)}{\alpha}}}{\left(\sum_{i=1}^n p_i^\beta\right)^{\frac{1}{\alpha}}} \right) \\ &= \frac{\alpha}{1-\alpha} \log_D f_1(P_1, P_2, \dots, P_n) \end{aligned} \tag{2.1}$$

where $P_i = \frac{p_i^\beta}{\sum_{i=1}^n p_i^\beta}$ and $f_1(P_1, P_2, \dots, P_n) = \sum_{i=1}^n P_i^{\frac{1}{\alpha}} D^{\frac{l_i(1-\alpha)}{\alpha}}$.

In order to minimize the average redundancy, we resort to the following theorem:

Theorem 1: The optimum probability distribution that minimizes the absolute redundancy $f(p_1, p_2, \dots, p_n)$ of the source with entropy $E_\beta^\alpha(p)$ and the mean codeword length $L(\alpha, \beta)$ is the escort distribution, given by

$$P_i = \frac{D^{-l_i}}{\sum_{i=1}^n D^{-l_i}}, i = 1, 2, \dots, n \tag{2.2}$$

Proof: To minimize the redundancy, we need to minimize

$$f(p_1, p_2, \dots, p_n) = \frac{\alpha}{1-\alpha} \log_D f_1(P_1, P_2, \dots, P_n) \tag{2.3}$$

subject to the constraint

$$\sum_{i=1}^n P_i = 1 \tag{2.4}$$

To prove this, we first of all, find the extremum of $\log_D f_1(P_1, P_2, \dots, P_n)$ which is equivalent to extremizing $f_1(P_1, P_2, \dots, P_n)$ and then use the fact that $f(p_1, p_2, \dots, p_n)$ is minimum or maximum will depend upon the value of parameter α .

So, in order to extremize $f_1(P_1, P_2, \dots, P_n)$, we consider the Lagrangian given by

$$\phi = \sum_{i=1}^n P_i^\alpha D^{\frac{1}{\alpha} \frac{l_i(1-\alpha)}{\alpha}} + \lambda \left(1 - \sum_{i=1}^n P_i \right)$$

where $\lambda \geq 0$ is Lagrange's multiplier.

Now

$$\frac{\partial \phi}{\partial P_i} = \frac{1}{\alpha} P_i^{\alpha-1} D^{\frac{1}{\alpha} \frac{l_i(1-\alpha)}{\alpha}} - \lambda \tag{2.5}$$

Letting $\frac{\partial \phi}{\partial P_i} = 0, i = 1, 2, \dots, n$, we get

$$P_i = \frac{D^{-l_i}}{\alpha^{\alpha-1} \lambda^{\alpha-1}} \tag{2.6}$$

Substituting (2.6) in (2.4), we get

$$\lambda = \frac{\left(\sum_{i=1}^n D^{-l_i} \right)^{\frac{\alpha-1}{\alpha}}}{\alpha} \tag{2.7}$$

Substituting (2.7) in (2.6), we get the result (2.2).

Now,

$$\left[\frac{\partial^2 \phi}{\partial P_i^2} \right]_{P_i = \frac{D^{-l_i}}{\sum_{i=1}^n D^{-l_i}}} = \frac{(1-\alpha)}{\alpha^2} \left(\frac{D^{-l_i}}{\sum_{i=1}^n D^{-l_i}} \right)^{\frac{1}{\alpha}-2} D^{\frac{l_i(1-\alpha)}{\alpha}} \tag{2.8}$$

We see that $\frac{\partial^2 \phi}{\partial P_i^2} > 0$ for $0 < \alpha < 1$ and $\frac{\partial^2 \phi}{\partial P_i^2} < 0$ for $\alpha > 1$.

Also,

$$\frac{\partial^2 \phi}{\partial P_i \partial P_j} = 0, i \neq j$$

So, $f_1(P_1, P_2, \dots, P_n)$ has minimum value for $0 < \alpha < 1$ and maximum for $\alpha > 1$.

Thus, $\log_D f_1(P_1, P_2, \dots, P_n)$ has minimum value for $0 < \alpha < 1$ and maximum for $\alpha > 1$ and consequently, observing the function $f(p_1, p_2, \dots, p_n)$, we see that it has minimum value for $0 < \alpha < 1, \alpha > 1$.

Thus, the minimum value is given by

$$f(p_1, p_2, \dots, p_n)_{\min} = -\log_D \sum_{i=1}^n D^{-l_i} . \tag{2.9}$$

Again, the necessary condition for the construction of uniquely decipherable codes is given by

$$\sum_{i=1}^n D^{-l_i} \leq 1 \quad (2.10)$$

Therefore, from (2.9), we have $f(p_1, p_2, \dots, p_n)_{\min} \geq 0$.

NOTE: It is to be noted that $f(p_1, p_2, \dots, p_n)_{\min} = 0$ if the source is Huffman [11] encoded since for the Huffman encoding, we have

$$\sum_{i=1}^n D^{-l_i} = 1. \quad (2.11)$$

Therefore, for this case, (2.2) becomes

$$P_i = D^{-l_i}, i = 1, 2, \dots, n. \quad (2.12)$$

Similarly, if we consider the codeword length $L(\beta)$ which satisfies the relation $L(\beta) \geq K^\beta(p)$, then the absolute redundancy of the source code in this case is given by

$$g(p_1, p_2, \dots, p_n) = L(\beta) - K^\beta(p) = \frac{\sum_{i=1}^n p_i^\beta l_i}{\sum_{i=1}^n p_i^\beta} + \frac{\sum_{i=1}^n p_i^\beta \log_D \left(\frac{p_i^\beta}{\sum_{i=1}^n p_i^\beta} \right)}{\sum_{i=1}^n p_i^\beta} g_1(p_1, p_2, \dots, p_n)$$

where $P_i = \frac{p_i^\beta}{\sum_{i=1}^n p_i^\beta}$ and $g_1(p_1, p_2, \dots, p_n) = \sum_{i=1}^n P_i l_i + \sum_{i=1}^n P_i \log_D P_i$.

Theorem 2. The optimum probability distribution that minimizes the absolute redundancy $g(p_1, p_2, \dots, p_n)$ of the source with entropy $K^\beta(p)$ and mean codeword length $L(\beta)$ is the escort distribution, given by

$$P_i = \frac{D^{-l_i}}{\sum_{i=1}^n D^{-l_i}}, i = 1, 2, \dots, n \quad (2.13)$$

Proof: We will find the extremum of $g(p_1, p_2, \dots, p_n)$ which is equivalent to extremizing $g_1(p_1, p_2, \dots, p_n)$ subject to constraint

$$\sum_{i=1}^n P_i = 1 \quad (2.14)$$

Let us consider the Lagrangian given by

$$\psi = g_1(p_1, p_2, \dots, p_n) + \mu \left(1 - \sum_{i=1}^n P_i \right) \quad (2.15)$$

where $\mu \geq 0$ is a Lagrange's multiplier.

For an extremum, let $\frac{\partial \psi}{\partial P_i} = 0, i = 1, 2, \dots, n$, that is,

$$l_i + \log_D P_i + \frac{1}{\log_e D} - \mu, i = 1, 2, \dots, n \quad (2.16)$$

$$p_i = e^{-1} D^\mu D^{-l_i}, i = 1, 2, \dots, n$$

Using (2.14), we get

$$\sum_{i=1}^n D^{-l_i} = e D^{-\mu}, i = 1, 2, \dots, n \quad (2.17)$$

Substituting (2.17) in (2.16), we get (2.13).
Also,

$$\left[\frac{\partial^2 \psi}{\partial P_i^2} \right]_{P_i = \frac{D^{-i}}{\sum_{i=1}^n D^{-i}}} = \frac{\sum_{i=1}^n D^{-i}}{D^{-i} \log_e D} > 0$$

and

$$\frac{\partial^2 \psi}{\partial P_i \partial P_j} = 0, i \neq j$$

So, $g_1(P_1, P_2, \dots, P_n)$ reaches its minimum value when $P_i = \frac{D^{-i}}{\sum_{i=1}^n D^{-i}}, i = 1, 2, \dots, n$ and is given by

$$g_1(P_1, P_2, \dots, P_n)_{\min} = -\log_D \sum_{i=1}^n D^{-i}$$

that is,

$$g(p_1, p_2, \dots, p_n)_{\min} = -\log_D \sum_{i=1}^n D^{-i}$$

Note: Again in this case also, if the source is Huffman [11] encoded, then the probabilities are given by $P_i = D^{-i}, i = 1, 2, \dots, n$.

Next, we will find the upper bound on the codeword lengths $L(\alpha, \beta)$ and $L(\beta)$ when the source is Huffman encoded.

Theorem 3. The exponentiated codeword length $L(\alpha, \beta)$ satisfies the following inequality

$$L(\alpha, \beta) \leq \log_D n \tag{2.18}$$

if the source is encoded using Huffman procedure.

Proof: The exponentiated codeword length $L(\alpha, \beta)$ can be written in the following form

$$L(\alpha, \beta) = \frac{\alpha}{1-\alpha} \log_D \left(\frac{\sum_{i=1}^n P_i^\alpha D^{\frac{l_i(1-\alpha)}{\alpha}}}{\sum_{i=1}^n P_i^\alpha} \right) \tag{2.19}$$

where $P_i = \frac{p_i^\beta}{\sum_{i=1}^n p_i^\beta}$.

Considering (2.12), (2.19) becomes

$$L(\alpha, \beta) = \frac{\alpha}{\alpha-1} \log_D \left(\sum_{i=1}^n D^{-l_i} \sum_{i=1}^n D^{\frac{l_i}{\alpha}} \right) = \frac{\alpha}{\alpha-1} \log_D h(l_1, l_2, \dots, l_n) \tag{2.20}$$

where $h(l_1, l_2, \dots, l_n) = \sum_{i=1}^n D^{-l_i} \sum_{i=1}^n D^{\frac{l_i}{\alpha}}$.

We need to find the extremum of $L(\alpha, \beta)$ subject to constraint $\sum_{i=1}^n D^{-l_i} = 1$ (as the source is encoded using Huffman Procedure).

For this purpose, we first of all, find the extremum of $\log_D h(l_1, l_2, \dots, l_n)$ which is equivalent to extremizing $h(l_1, l_2, \dots, l_n)$ and then use the fact that $L(\alpha, \beta)$ is minimum or maximum depending upon the value of parameter α .

So, we consider the Lagrangian given by

$$A = h(l_1, l_2, \dots, l_n) + \nu \left(1 - \sum_{i=1}^n D^{-l_i} \right) \tag{2.21}$$

where $\nu \geq 0$ is a Lagrange's multiplier

Put $x_i = D^{-l_i}, i = 1, 2, \dots, n$, (2.21) becomes

$$A = \sum_{i=1}^n x_i^\alpha + \nu \left(1 - \sum_{i=1}^n x_i \right)$$

Letting $\frac{\partial A}{\partial x_i} = 0, i = 1, 2, \dots, n$, we get

$$x_i = (\nu \alpha)^{\frac{1}{1-\alpha}}, i = 1, 2, \dots, n \tag{2.22}$$

Now, $\sum_{i=1}^n x_i = 1$ gives

$$\nu = \frac{n^{\frac{\alpha-1}{\alpha}}}{\alpha} \tag{2.23}$$

Using (2.23) in (2.22), we get

$$x_i = \frac{1}{n}, i = 1, 2, \dots, n$$

that is, $l_i = \log_D n, i = 1, 2, \dots, n$

Now,

$$\left[\frac{\partial^2 A}{\partial x_i^2} \right]_{x_i = \frac{1}{n}} = \frac{(1-\alpha)}{\alpha^2} (n)^{2\alpha-1}.$$

We see that $\frac{\partial^2 A}{\partial x_i^2} > 0$ for $0 < \alpha < 1$ and $\frac{\partial^2 A}{\partial x_i^2} < 0$ for $\alpha > 1$.

Also,

$$\frac{\partial^2 A}{\partial P_i \partial P_j} = 0, i \neq j.$$

So, $h(l_1, l_2, \dots, l_n)$ has minimum value for $0 < \alpha < 1$ and maximum for $\alpha > 1$.

Therefore, $\log_D h(l_1, l_2, \dots, l_n)$ has minimum value for $0 < \alpha < 1$ and maximum for $\alpha > 1$ and consequently, observing the exponentiated mean codeword length $L(\alpha, \beta)$, we see that it has maximum value for $0 < \alpha < 1, \alpha > 1$.

Thus, the maximum value is given by

$$L(\alpha, \beta)_{\max} = \log_D n.$$

Theorem 4. The mean codeword length $L(\beta)$ is upper bounded by $\log_D n$, that is,

$$L(\beta) \leq \log_D n \tag{2.24}$$

if the source is encoded using Huffman procedure.

Proof: The exponentiated codeword length $L(\beta)$ can be written in the following form

$$L(\beta) = \sum_{i=1}^n P_i l_i = \sum_{i=1}^n D^{-l_i} l_i \text{ (as } P_i = D^{-l_i} \text{)}. \tag{2.25}$$

We need to find the extremum of $L(\beta)$ subject to constraint $\sum_{i=1}^n D^{-l_i} = 1$ (as the source is encoded using Huffman Procedure).

So, we consider the Lagrangian given by

$$B = L(\beta) + \theta \left(1 - \sum_{i=1}^n D^{-l_i} \right) \tag{2.26}$$

where $\theta \geq 0$ is a Lagrange's multiplier .

Letting $\frac{\partial B}{\partial l_i} = 0, i = 1, 2, \dots, n$, we get

$$l_i = \log_D e + \theta. \tag{2.27}$$

Since $\sum_{i=1}^n D^{-l_i} = 1$, we have

$$\theta = \log_D \left(\frac{n}{e} \right). \tag{2.28}$$

Substitute (2.28) in (2.27), we get

$$l_i = \log_D n, i = 1, 2, \dots, n.$$

Now,

$$\left[\frac{\partial^2 B}{\partial l_i^2} \right]_{l_i = \log_D n} = -\frac{1}{n} \log_e D < 0.$$

Also,

$$\frac{\partial^2 B}{\partial P_i \partial P_j} = 0, i \neq j.$$

So, the mean codeword length $L(\beta)$ has maximum value when $l_i = \log_D n, i = 1, 2, \dots, n$, and is given by

$$L(\beta)_{\max} = \log_D n .$$

Note-I: For the case of Campbell's codeword length L_α , we have from (1.3), $L_\alpha \geq R_\alpha(p)$. So, the average redundancy of the source code in this case is given by

$$\begin{aligned} K(p_1, p_2, \dots, p_n) &= L_\alpha - R_\alpha(p) = \frac{\alpha}{1-\alpha} \log_D \left(\frac{\sum_{i=1}^n (p_i^\alpha)^{\frac{1}{\alpha}} D^{\frac{l_i(1-\alpha)}{\alpha}}}{\left(\sum_{i=1}^n p_i^\alpha \right)^{\frac{1}{\alpha}}} \right) \\ &= \frac{\alpha}{1-\alpha} \log_D K_1(p_1, p_2, \dots, p_n) \end{aligned}$$

where $P_i = \frac{p_i^\alpha}{\sum_{i=1}^n p_i^\alpha}$ and $K_1(p_1, p_2, \dots, p_n) = \sum_{i=1}^n P_i^{\frac{1}{\alpha}} D^{\frac{l_i(1-\alpha)}{\alpha}}$.

The absolute redundancy in the case of Campbell’s [3] mean codeword length is the same as in case of exponentiated mean codeword length $L(\alpha, \beta)$ developed by Parkash and Kakkar [6] as given in (2.1). Thus, we see that similar results as proved in theorem (2.1) and theorem (2.3) hold for Campbell’s case also.

Note-II: Absolute redundancy when we use Kapur’s[4] mean codeword length is given by

$$J(p_1, p_2, \dots, p_n) = L^\alpha - R_\alpha(p) = \frac{1}{\alpha - 1} \log_D J_1(p_1, p_2, \dots, p_n)$$

where $J_1(p_1, p_2, \dots, p_n) = \sum_{i=1}^n p_i^\alpha D^{i(\alpha-1)}$.

Theorem 5: The optimum probability distribution that minimizes the absolute redundancy of the source with entropy $R_\alpha(p)$ and mean codeword length L^α is given by

$$p_i = \frac{D^{-i}}{\sum_{i=1}^n D^{-i}}, i = 1, 2, \dots, n. \tag{2.29}$$

Proof: To minimize the redundancy, we need to minimize

$$J(p_1, p_2, \dots, p_n) = \frac{1}{\alpha - 1} \log_D J_1(p_1, p_2, \dots, p_n) \tag{2.30}$$

subject to the constraint

$$\sum_{i=1}^n p_i = 1. \tag{2.31}$$

To prove this, we first of all find the extremum of $\log_D J_1(p_1, p_2, \dots, p_n)$ which is equivalent to extremizing $J_1(p_1, p_2, \dots, p_n)$ and then using the fact that $J(p_1, p_2, \dots, p_n)$ is minimum or maximum depending upon the value of parameter α .

So, in order to extremize $J_1(p_1, p_2, \dots, p_n)$, we consider the Lagrangian given by

$$L = \sum_{i=1}^n p_i^\alpha D^{(\alpha-1)i} + \gamma \left(1 - \sum_{i=1}^n p_i \right)$$

where $\gamma \geq 0$ is Lagrange’s multiplier.

$$\frac{\partial L}{\partial p_i} = \alpha p_i^{\alpha-1} D^{i(1-\alpha)} - \gamma. \tag{2.32}$$

Letting $\frac{\partial L}{\partial p_i} = 0, i = 1, 2, \dots, n$, we get

$$p_i = \frac{\lambda^{\alpha-1} D^{-i}}{\alpha^{\alpha-1}}. \tag{2.33}$$

Substituting (2.33) in (2.31), we get

$$\lambda = \frac{\alpha}{\left(\sum_{i=1}^n D^{-i} \right)^{\frac{1}{\alpha-1}}}. \tag{2.34}$$

Substituting (2.34) in (2.33), we get the result (2.29).

Now, $\left[\frac{\partial^2 L}{\partial p_i^2} \right]_{p_i = \frac{D^{-i}}{\sum_{i=1}^n D^{-i}}} = \alpha(\alpha - 1) \frac{D^i}{\left(\sum_{i=1}^n D^{-i} \right)^{\alpha-2}}$.

We see that $\frac{\partial^2 L}{\partial p_i^2} < 0$ for $0 < \alpha < 1$ and $\frac{\partial^2 \phi}{\partial p_i^2} > 0$ for $\alpha > 1$.

Also, $\frac{\partial^2 L}{\partial p_i \partial p_j} = 0, i \neq j$.

So, $J_1(p_1, p_2, \dots, p_n)$ has maximum value for $0 < \alpha < 1$ and minimum value for $\alpha > 1$.

Therefore, $\log_D J_1(p_1, p_2, \dots, p_n)$ has maximum value for $0 < \alpha < 1$ and minimum value for $\alpha > 1$ and consequently observing the function $J(p_1, p_2, \dots, p_n)$, we see that it has minimum value for $0 < \alpha < 1, \alpha > 1$. The minimum value is given by

$$J(p_1, p_2, \dots, p_n)_{\min} = -\log_D \sum_{i=1}^n D^{-l_i}.$$

Theorem 6. The Kapur's [8] mean codeword length L^α satisfies the following inequality

$$L^\alpha \leq \log_D n \quad (2.35)$$

if the source is encoded using Huffman procedure.

Proof: Proceeding as in Theorem 2.3, we can prove the Theorem 6.

Acknowledgements

The authors are thankful to Council of Scientific and Industrial Research, New Delhi, for providing the financial assistance for the preparation of the manuscript.

REFERENCES

- [1] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, Vol. 27, 1948, pp. 379-423 (Part I) 623-656 (Part II).
- [2] L. G. Kraft, "A Device for Quantizing Grouping and Coding Amplitude Modulated Pulses," M.S. Thesis, MIT, Cambridge, 1949.
- [3] L. L. Campbell, "A Coding Theorem and Renyi's Entropy," *Information and Control*, Vol. 8, No. 4, 1965, pp. 423-429.
- [4] J. N. Kapur, "Entropy and Coding," Mathematical Sciences Trust Society (MSTS), New Delhi, 1998.
- [5] A. Renyi, "On Measures of Entropy and Information," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 1961, pp. 547-561.
- [6] O. Parkash and P. Kakkar, "Development of Two New Mean Codeword Lengths," *Information Sciences*, Vol. 207, 2012, pp. 90-97.
- [7] D. Harte, "Multifractals: Theory and Applications," Chapman and Hall, London, 2001.
- [8] J. F. Bercher, "Source Coding with Escort Distributions and Renyi Entropy Bounds," *Physics Letters A*, Vol. 373, No. 36, 2009, 3235-3238.
- [9] J. F. Bercher, "Tsallis Distribution as a Standard Maximum Entropy Solution with 'Tail' Constraint," *Physics Letters A*, Vol. 372, No. 35, 2008, pp. 5657-5659.
- [10] C. Beck and F. Schloegl, "Thermodynamics of Chaotic Systems," Cambridge University Press, Cambridge, 1993.
- [11] D. A. Huffman, "A Method for the Construction of Minimum Redundancy Codes," *Proceedings of the Institute of Radio Engineers*, Vol. 40, No. 10, 1952, pp. 1098-1101.