International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 5 Issue: 6 253 – 257

Generalized Coding Theorem with Different Source Coding Schemes

D. K. Sharma
Jaypee University of Engineering and Technology,
Raghogarh, Dist. Guna(M.P.)-473226, India
dilipsharmajiet@gmail.com

Sonali Saxena
Jaypee University of Engineering and Technology,
Raghogarh, Dist. Guna(M.P.)-473226, India

Abstract— In the present paper we have consider a parametric generalization of mean codeword length and for this we have proved generalized source coding theorems. To transmit text written in the source alphabet in the form of code alphabetic character we have to associate a code word to represent each source alphabet word that we might wish to send, hence we have also verified generalized source coding theorem by using source coding schemes. Kraft's theorem states the condition which the lengths of codewords must meet to be a prefix codes. It may seem restrictive to limit ourselves to prefix codes, as uniquely decipherable codes are not always prefix codes. The Shannon-Fano encoding scheme is based on the principle that each code bit, which can be described by a random variable, must have a maximum entropy, so we have also discussed different type of source coding schemes by taking some suitable examples.

Keywords- *Mean codeword length, source coding theorem, Code Alphabets, Huffman and Shannon Fano coding schemes, Kraft's inequality.*

*****

I. INTRODUCTION

Shanonn [9] discussed the notion of source coding in his seminal paper which laid the framwork of the Information Theory. The primary objective of the source coding is to compress the data by efficient representation of the symbols. Let we have a Discrete Memoryless Source(DMS) generated a symbol every seconds and let X be the discrete random variable taking a finite number of symbols $x_1, x_2, x_3, \dots \dots \dots x_n$ with probabilities $p_1, p_2, p_3, \dots \dots \dots p_n$ such that $p_i \geq 0, i = 1,2,3, \dots \dots \dots, n$ and $\sum_{i=1}^{n} p_i = 1$. Then entropy of DMS in bits per source symbols is given as:

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

Entropy given by (1) plays a vital role in theory of coding and provided lower and upper bounds for average code word length. For a finite set of n symbols$(x_1, x_2, x_3, \dots \dots \dots x_n)$,encoded by the D size alphabets with probabilities $(p_1, p_2, p_3, \dots \dots \dots p_n)$ and codeword lengths$(l_1, l_2, l_3, \dots \dots \dots l_n)$,then average mean codeword length is defined as follows:

$$L = \sum_{i=1}^{n} p_i l_i log D, \tag{2}$$

Further, it was shown by Kraft[6] that uniquely decipherable codes with code word length satisfy the following inequality which is known as Kraft's inequality :

$$\sum_{i=1}^{n} D^{-l_i} \leq 1. \tag{3}$$

Therefore under the Kraft inequality source coding theorem is stated as follows:

$$H(P) \leq L \leq H(P) + log D, D \geq 2 \tag{4}$$

In other words source coding theorem says that for any prefix code used to represent the symbol from a source the minimum number of bits required to represent the source symbols on an average on any event equal to the entropy of the source.

In the present work we consider the parametric generalization of mean code word length defined by Hooda and Bhakar [4] and proved source coding theorem for the same generalized code word length in section 2. In section 3 by using different source coding techniques source coding theorem is verified by taking some illustrations. Also some other coding techniques with their applications to communication theory are also discussed. Sections 4 conclude the work presented.

II. GENERALIZED MEAN CODE WORD LENGTH AND SHANNON'S NOISELESS CODING THEOREM

In this section we consider generalized mean codeword length as follows:

$$L_\beta^\alpha(P) = \frac{\alpha}{1-\alpha} \log_D \left(\frac{\sum p_i^{\beta} D^{\left(\frac{1-\alpha}{\alpha}\right) l_i}}{\sum p_i^{\beta}} \right), 0 < \alpha < 1, \beta > 1, \alpha \neq 1. \tag{5}$$

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 (5) 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. $L_\beta^\alpha(P)$ lies between minimum and maximum values of $l_1, l_2, \ldots, l_n$.

3. When $\beta \to 1$ and $\alpha \to 1$, then $L_\beta^\alpha \to L$,where

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

Next we prove Shaanon's noise less coding theorem for generalized mean codeword length defined in (5) as follows:

Theorem 2.1. For all uniquely decipherable codes the exponentiated mean codeword length $L_\beta^\alpha(P)$ defined in (5) satisfies the following relation

$$H_\beta^\alpha(P) \le L_\beta^\alpha(P) < H_\beta^\alpha(P) + 1, \tag{6}$$

where

$$H_\beta^\alpha(P) = \frac{1}{1-\alpha} \log_D \frac{\left(\sum p_i^{\left(\frac{\beta+2\alpha-2}{\alpha}\right)}\right)^\alpha}{\sum p_i^\beta}, 0 < \alpha < 1, \beta \ge 1, \alpha \ne 1. \tag{7}$$

under the generalized Kraft inequality given by

$$\sum p_i^{\beta-1} D^{-l_i} \le \sum p_i^\beta. \tag{8}$$

Proof: Let us choose

$$\frac{p_i^{\beta-1} D^{-l_i}}{\sum p_i^\beta} = x_i, \text{ for each } i = 1,2,\ldots,n. \tag{9}$$

Substituting (9) in (5) we have

$$L_\beta^\alpha(P) = \frac{\alpha}{1-\alpha} \log_D \left[\frac{\sum p_i^{\frac{\alpha+\beta-1}{\alpha}} x_i^{\left(\frac{\alpha-1}{\alpha}\right)}}{\left(\sum p_i^\beta\right)^{1/\alpha}} \right] \tag{10}$$

Thus we are to minimize (10) subject to the following constraints:

$$\sum_{i=1}^{n} x_i = \frac{\sum p_i^{\beta-1} D^{-l_i}}{\sum p_i^\beta} \le 1 \tag{11}$$

Since L_β^α is pseudo convex function for each $i = 1,2,\ldots,n$,therefore, we can obtain the minimum value of $L_\beta^\alpha(P)$ 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

$$\therefore x_i = c p_i, \text{ when } 0 < c \le 1 \tag{12}$$

(12) together with (9) 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

$$l_i \ge -\log_D \frac{\sum p_i^\beta}{p_i^{\beta-2}} \tag{13}$$

Multiplying both sides of (13) by $\left(\frac{1-\alpha}{\alpha}\right) \ge 0$ as $0 < \alpha < 1$, we get

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

or

$$\frac{1-\alpha}{\alpha} l_i \ge -\log_D \left(\frac{\sum p_i^\beta}{p_i^{\beta-2}}\right)^{\frac{1-\alpha}{\alpha}} \tag{14}$$

From (5) and (10), we get the minimum value of $L_\beta^\alpha(P)$ as follows:

$$L_\beta^\alpha(P)_{\min} = \frac{1}{1-\alpha} \log \frac{\left(\sum p_i^{\frac{\beta+2\alpha-2}{\alpha}}\right)^\alpha}{\left(\sum p_i^\beta\right)} = H_\beta^\alpha(P) \tag{15}$$

l_i is always integral value in (13), so it must be equal to

$$l_i = a_i + \varepsilon_i \tag{16}$$

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

Putting (16) in (5), we have

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

$$= \frac{1}{1-\alpha} \log \frac{\left(\sum p_i^{\frac{\beta+2\alpha-2}{\alpha}}\right)^\alpha}{\sum p_i^\beta} + \varepsilon_i \tag{17}$$

Since $0 \le \varepsilon_i < 1$,therefore, (17) reduce to

 ISSN: 2321-8169
253 – 257

$$L_\beta^\alpha(P) < \frac{1}{1-\alpha}\log\frac{\left(\sum p_i^{\frac{\beta+2\alpha-2}{\alpha}}\right)^\alpha}{\sum p_i^\beta} + 1 = H_\beta^\alpha(P) + 1 \qquad (18)$$

Hence from (15) and (18), we get

$H_\beta^\alpha(P) \le L_\beta^\alpha(P) < H_\beta^\alpha(P) + 1,$ which is (6).

Thus by following optimization technique we get new generalized entropy given by (7).

III SOURCE CODING SCHEMES

In this section we review different type of source coding schemes. Basically the source coding schemes are characterized in two parts.

1). Source specific

2). Universal

In the first case (i.e. source specific) source encoder require full knowledge of the source statistics in order to perform function. Such type of coding scheme could not perform well for source other than that for which it is designed. However, on the other end universal coding scheme tend to perform well (in an asymptotic sense) for all source in some comparatively large class.

These source coding schemes can be listed as follows:

(a). Shannon Fano

(b). Huffman

(c). Lynch-Davision

(d). Elias-Willems

(e). Lampbel –Ziv

Shannon-Fano coding [2] is a technique for realizing the message encoder that explicitly aims to make the resulting sequence of codeword digits a good guess to the output of the Binary symmetric source(BSS). The Shannon-Fano algorithm is an "insatiable" algorithm in the sense that it makes each successive codeword digit as nearly as equally likely to be a 0 or a 1 as possible, at the cost of possible severe biasing of later codeword digits. Its algorithm is simple for which first makes a list of all possible messages in order of decreasing probability. Then splits this list at the point where the two resulting values are as nearly equally probable as possible, assigning the first codeword digit as a 0 for messages in the first o value and as 1 in the second value. And repeats this splitting process on all the values to assign subsequent codeword digits to messages until all values contain a single message.

The algorithm for optimum prefix-free encoding of a message set was given by Huffman [5]. The trick is to be completely "satiable " and to choose the last digits of codewords first. The algorithm is extremely simple. One assigns a last digit of 0 and 1, respectively, to the two least probable messages, then merges these two messages to a single message whose probability is the sum of those of the two merged messages. One then repeats this combination on the new message set until one has just a solitary message left.

Further we will verify generalized Shannon's coding theorem for generalized mean codeword length by using source specific coding scheme (i.e. Shannon Fano and Huffman). Here we have used some empirical data as given in the following two tables. For $\alpha = 0.5\ and\ \beta = 2$ then from (5) and (7) ,we have

Table 3.1: Shannon-Fano coding Scheme

Probabilities p_i	.3846	.1795	.1538	.1538	.1282
Shannon-Fano code words	00	01	10	110	111
Length of Shannon -Fano code words	2	2	2	3	3

We have $L_\beta^\alpha(P) = 2.21979, H_\beta^\alpha(P) = 2.03595,$ and $\eta = \frac{H_\beta^\alpha(P)}{L_\beta^\alpha(P)} \times 100 = 91.78\%$

Table 3.2: Huffman coding scheme

Probabilities p_i	.3846	.1795	.1538	.1538	.1282
Huffman codewords	0	100	101	110	111
Length of Huffman codewords	1	1	3	3	3

We have $L_\beta^\alpha(P) = 2.12484, H_\beta^\alpha(P) = 2.03595,$ and $\eta = \frac{H_\beta^\alpha(P)}{L_\beta^\alpha(P)} \times 100 = 95.81\%$

From table (3.1) and (3.2) we conclude the following:

(i) Shannon' Noiselss Coding theorem 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) Huffman Coding is more efficient then Shannon-Fano Coding scheme.

Next we discuss universal coding schemes as following:

The Lynch-Davisson coding scheme utilizes an L-block message parser. The message encoder first determines the number of 1's (i.e. hamming weight)) W_H in the message $v_1 = \{u_1, u_2, \dots \dots \dots . u_n\}$, then determines the index I of this message in an indexed list of all binary n-tuples of Hamming weight W_H. The codeword s_1 is then the $\lceil \log(L + 1) \rceil$ bit binary code for W_H followed by the $\lceil \log({}^{n}_{W_H}C) \rceil$ bit binary code for I. Here we consider base 2 for the algorithm. Because the length of the code for W_H does not rely on upon the specific message v_1, the decoder can decide W_H from this code to figure out where the codeword will end, so this encoding of the message v_1 is undoubtedly free from prefix .Hence it can be says that the Lynch-Davisson source-coding scheme is universally asymptotically optimum for the class of all binary memoryless sources. It perform good for Discrete Stationary and Ergodic Sources (DSES's) with weak memory but can be very inefficient for such sources with strong memory.

In **Elias [2]-Willems [8]** two prefix-free coding schemes for the positive integers $\mathbb{Z}+=\{1,2,3,\ldots\ldots\}$ let we consider the natural binary coding $B(n)$ for $n \in \mathbb{Z}+$, i.e., B(1) = 1, B(2) =10, etc. We note that this natural binary code is not a prefix-free code for $\mathbb{Z}+$ [infact B(1) = 1 is a prefix of every other codeword] and that the length L(n) of B(n) is log n + 1. Elias' first coding scheme for $\mathbb{Z}+$ encodes n as B1(n) where B1(n) consists of L(n) - 1 0's followed by B(n). For instance, because L(13) = log 13 one obtains (n) = 0001101. The length of B_1(n) is L_1(n) = 2 L(n) - 1 = 2 log n 1, about twice that of B(n). In any case, the encoding B_1(n) is without prefix in light of the fact that the number L(n) - 1 of driving 0's in B1(n) decides the length of the codeword i.e., where the codeword will end. Elias's second prefix-free coding scheme for $\mathbb{Z}+$ builds on the first. The codeword B2(n) is B_1(L(n)) [i.e., the first coding applied to the length of n in the natural binary code] followed by B(n) with its now "useless" leading 1 removed. The Elias-Willems source-coding scheme is universal for the class of all discrete stationary and ergodic sources.

The Lempel-Ziv((1977) coding scheme is quite different from above defined schemes. It uses variable-length message parsing; indeed this parsing is its most distinctive attribute. There are rather several versions of the Lempel-Ziv scheme,all of which are based on the ideas originally proposed in Ziv. We will consider the version described by Welch (1984), which seems to be the one most often actualized, and we will consign to this version as the LZ-W source-coding scheme.

The key plan in every Lempel-Ziv source-coding scheme is to parse the source sequence according to the subsequence or "strings" that come out for the initially within the source sequence. In the LZ-W edition, one parses a binary source by assuming that the length-one strings 0 and1 are the only earlier encountered strings. Let $L_1 = (0,1)$ refer to this initial list. The parsing rule is then as follows. For each i, i = 1, 2, ..., mark the end of the i^{th} phrase at the point where counting the next digit would give a string not in the list L_i of previously encountered strings, then position this string with the next digit added toward at the end of the list L_i to form the list L_{i+1}. Applying this parsing rule to the sequence 001000100001000 gives

110|0|1|00|01|000|010|00,

as we now explain. The initial string 0 is in $L = (0,1)$, but the string 00 is not. Thus, we place a marker after the initial 0 and form the list $L_2 = (0,1,00)$. Looking forward from this first marker, we first see 0, which is in L_2 , then we see 01, which is not. Thus we place a marker after this second 0 and form the list $L_3 = (0,1,00,01)$ etc.The messages $v_1, v_2, v_3, \ldots, v_n$ of the LZ-W scheme are the phrases of the parsed source sequence. Note that the list L_i contains exactly i+1 strings. In the LZ-W scheme, the message v_i is encoded as the $W_{i=}\lceil\log(i+1)\rceil$ bit binary code for its index in the list L_i. [Note that, for i > 1, the last string in the list L_i is placed there only after the parsing of v_{i-1}, which requires examination of the first digit of v_i. Thus, for i > 1, the decoding of the codeword x_i to the message v_i, when x_i is the codeword pointing to the last entry in L_i , cannot be performed by table look-up as the decoder will then have formed only the list L_{i-1}. But the last entry in L_i is always a string having v_i as a prefix. Thus, when i > 1 and x_i points to this last string in Li, the first digit of v_i must be the same as the first digit of v_{i-1} and hence the decoder can "prematurely" form the list Li that it needs to decode x_i.] Because the length Wi of the i-th codeword x_i does not depend on the source sequence, the LZ-W coding is prefix-free; moreover, the lengths of the first in codewords sum to

$$\sum_{i=1}^{n} W = \sum_{i=1}^{n}\lceil\log(i+1)\rceil.$$

The corresponding sum of message lengths, however, depends strongly on the statistics of the DSES encoded. Lempel and Ziv [10] have revealed (by an argument that applies additionally to the LZ-W version) that this the Lempel-Ziv source-coding scheme is universal for the class of all Discrete Stationary and Ergodic Sources. Lempel-Ziv source coding, and in particular the LZ-W version, has ended up being an exceptionally famous data-compression scheme in practice, as much result of the simplicity.

IV CONCLUSION

In the work presented here we chew over a generalized mean codeword length suggested by Hooda and Bhakar , we find out the bounds for this generalized mean codeword length in the terms of Shannon coding theorem. By taking some particular values for the parameters α and β we have illustrate the veracity of Shannon's theorem. We have also discussed some more source coding schemes.

REFERENCES

[1]. L. D. Davisson, *Comments on 'Sequence Time Coding for Data Compression*', Proc. IEEE; **54**(2010).

[2]. P. Elias,.*Universal Codeword Sets and Representations of the Integers*. IEEE Trans. Inform, Th.,IT; **21**,194-203, **(**1975).

[3]. Gallager, R. G. (1968). Information Theory and Reliable Communications. NewYork: Wiley.

[4]. Hooda, D.S. and U.S. Bhaker, A Profile on Noiseless Coding Theorems. *International Journal of Management and Systems*, 8, 76-85, (1992)..

ISSN: 2321-8169
253 – 257

[5]. Huffman, D. A.. A Method for the Construction of Minimum Redundancy Codes. *Proc. IRE*, Vol. 40, pp. 1098-1101, (1952).

[6]. Kraft, L.G.. A Device for Quantizing, Grouping and Coding Amplitude Modulated Pulses**.** *M.S.Thesis, Electrical Engineering Department, MIT,* (1949).

[7]. Welch, T. A.. A Technique for High Performance Data Compression. *IEEE Computer*,17, 8-19, (1984).

[8]. Willems, F. M. J. Universal Data Compression and Repetition Times,. *IEEE Trans. Inform. Th.* IT-35, 54-58, (1989).

[9]. Shannon, C.E.. A Mathematical Theory of Communication. *Bell System Technology Journal* 27, 379-423, (1948).

[10]. Ziv. J. and Lempel, A. A Universal Algorithm for Sequential DataCompression. *IEEE Trans. Inform. Th.*,IT-23,337-343, (1977).