Source Coding with Renyi's Entropy

Arun Choudhary
Department of Mathematics,
Geeta Institute of Management & Technology, Kanipla-136131, Kurukshetra, Haryana (India).

Satish Kumar, Arvind Kumar
Department of Mathematics, College of Natural Sciences, Arba-Minch University, Arba-Minch (Ethiopia).
Department of Mathematics, BRCM College of Engineering & Technology, Bahal, Bhiwani, Haryana (India).

ABSTRACT
A new measure L_α, called average code word length of order α is defined and its relationship with Renyi's entropy of order α is discussed. Using L_α, some coding theorems are proved under the condition $\sum_{i=1}^{N} D^{-n_i\alpha} \le \sum_{i=1}^{N} p_i^{\alpha}$.

AMS Subject classification: 94A15, 94A17, 94A24, 26D15.
Keywords
Codeword length, Optimal code length, Holder's inequality and Kraft inequality.

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

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

In case there is no rise to misunderstanding we write $P \in \Delta_N$ instead of $(p_1, ..., p_N) \in \Delta_N$.

In case $P \in \Delta_N$ the well-known Shannon entropy is defined by

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

where the convention $0\log(0) = 0$ is adapted, (see Shannon [13]).

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

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

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

$$\sum_{i=1}^{N} D^{-n_i} \le 1, \tag{2}$$

where *D* is the size of code alphabet.

Furthermore, if

$$L = \sum_{i=1}^{N} n_i p_i \tag{3}$$

is the average codeword length, then for a code satisfying (2), the inequality

$$H(P) \le L < H(P) + 1 \tag{4}$$

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

$$n_i = -\log_D(p_i) \ (i = 1, ..., N), \tag{5}$$
$$\text{and } \sum_{i=1}^{N} D^{-n_i} = 1.$$

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 [3], in terms of Renyi's entropy [12]:

$$H_\alpha(P) = \frac{1}{1-\alpha}\log_D \sum p_i^\alpha,\ \alpha > 0(\neq 1). \quad (6)$$

Kieffer [10] 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 [7] 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 [3].

Hooda and Bhaker [6] considered the following generalization of Campbell's mean length:

$$L^\beta(t) = \frac{1}{t}\log_D\left\{\frac{\sum p_i^\beta D^{-tn_i}}{\sum p_i^\beta}\right\}, \beta \geq 1$$

and proved

$$H_\alpha^\beta(P) \leq L^\beta(t) < H_\alpha^\beta(P) + 1,$$

$$\alpha > 0,\ \alpha \neq 1,\ \beta \geq 1$$

under the condition

$$\sum p_i^{\beta-1} D^{-n_i} \leq \sum p_i^\beta,$$

where $H_\alpha^\beta(P)$ is generalized entropy of order $\alpha = \frac{1}{1+t}$ and type β studied by Aczel and Daroczy [1] and Kapur [8]. It may be seen that the mean codeword length (3) had been generalized parametrically and their bounds had been studied in terms of generalized measures of entropies. Here we give another generalization of (3) and study its bounds in terms of generalized entropy of order α.

Generalized coding theorems by considering different information measure under the condition of unique decipherability were investigated by several authors, see for instance the papers [Aczel and Daroczy [2], Ebanks et al. [4], Hooda and Bhaker [6], Khan et al. [9], Longo [11], Singh et al. [14]].

In this paper, we study a new measure L_α, called average code word length of order α is defined and its relationship with Renyi's entropy is discussed. Using L_α, some coding theorem for discrete noiseless channel are proved.

2. CODING THEOREMS

Definition. Let $\alpha > 0(\neq 1)$ be arbitrarily fixed, then the mean length L_α corresponding to the generalized information measure $\mathrm{H}_\alpha(P)$ is given by the formula

$$L_\alpha = \frac{1}{1-\alpha}\log\left(\sum_{i=1}^{N} p_i D^{n_i(1-\alpha)}\right), \quad (7)$$

where $P = (p_1, p_2, ..., p_N)$ and $D, n_1, n_2, ..., n_N$ are positive integers so that

$$\sum_{i=1}^{N} D^{-n_i\alpha} \leq \sum_{i=1}^{N} p_i^\alpha. \quad (8)$$

Since (8) reduces to Kraft inequality (3) when $\alpha = 1$, therefore it is called generalized Kraft inequality and codes obtained under this generalized inequality are called personal codes.

APPLICATIONS OF HOLDER'S INEQUALITY IN CODING THEORY

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

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

$$L_\alpha \geq \mathrm{H}_\alpha(P) \quad (9)$$

is fulfilled. Furthermore, equality holds if and only if

$$n_i = -\log_D(p_i). \quad (10)$$

Proof: We have two possibilities:

Case 1. Let $\alpha > 1$. The Hölder inequality, that is,

$$\left(\sum_{i=1}^{N} x_i^p\right)^{\frac{1}{p}} \cdot \left(\sum_{i=1}^{N} y_i^q\right)^{\frac{1}{q}} \leq \sum_{i=1}^{N} x_i y_i. \quad (11)$$

If $\frac{1}{p} + \frac{1}{q} = 1,\ 0 < p < 1,\ q < 0$

$or\ 0 < q < 1,\ p < 0.$

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

$$x_i = p_i^{\frac{\alpha}{\alpha-1}} D^{-n_i\alpha},\ y_i = p_i^{\frac{\alpha}{1-\alpha}}\ (i = 1, ..., N).$$

Putting these values into (11), we get

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

$$\Rightarrow\left(\sum_{i=1}^{N}p_iD^{n_i(1-\alpha)}\right)^{\frac{\alpha}{\alpha-1}}\left(\sum_{i=1}^{N}p_i^{\alpha}\right)^{\frac{1}{1-\alpha}}\le\sum_{i=1}^{N}p_i^{\alpha}$$

where we used (8), too. This implies however that

$$\left(\sum_{i=1}^{N}p_iD^{n_i(1-\alpha)}\right)\le\left(\sum_{i=1}^{N}p_i^{\alpha}\right) \quad (12)$$

Taking logarithm on both sides and we obtain the result (9) after simplification for $\frac{1}{1-\alpha}<0.$

i.e., $L_\alpha \ge H_\alpha^\beta(P).$

It is clear that the equality in (9) is true if and only if

$$D^{-n_i} = p_i. \quad (13)$$

Which is equivalent to (10).
From (13) and after simplification, we get

$$p_iD^{-n_i(\alpha-1)} = p_i^\alpha$$

it implies

$$\left(\sum_{i=1}^{N}p_iD^{n_i(1-\alpha)}\right)=\left(\sum_{i=1}^{N}p_i^{\alpha}\right). \quad (14)$$

Which gives $L_\alpha = \mathrm{H}_\alpha(P).$

Case 2. If $0<\alpha<1$, the proof follows on the same lines as for $\alpha>1.$

In the following theorem, we give an upper bound for L_α in terms of $\mathrm{H}_\alpha(P).$

Theorem 2. For α as in Theorem 1, there exist positive integers $\{n_i\}$ satisfying (8) such that

$$L_\alpha<\mathrm{H}_\alpha(P)+\log D. \quad (15)$$

Proof: Let n_i be the positive integer satisfying the inequalities

$$-\log_D(p_i)\le n_i<-\log_D(p_i)+1 \quad (16)$$

Consider the intervals

$$\delta_i=\left[-\log_D(p_i),-\log_D(p_i)+1\right] \quad (17)$$

of length 1. In every δ_i, there lies exactly one positive number n_i such that

$$0<-\log_D(p_i)\le n_i<-\log_D(p_i)+1. \quad (18)$$

It is easy to see that the sequence $\{n_i\}, i=1,2,...,N$ thus defined, satisfies (8).
From the right inequality of (18), we have

$$n_i<-\log_D(p_i)+1$$

$$\Rightarrow D^{n_i}<Dp_i^{-1}. \quad (19)$$

Now consider two cases:

(i) When $0<\alpha<1$, then raising power $(1-\alpha)$ to both sides of (19), we have

$$D^{n_i(1-\alpha)}<D^{(1-\alpha)}p_i^{\alpha-1}. \quad (20)$$

Multiplying (20) throughout by p_i and then summing up from $i=1$ to $i=N,$ we have

$$\sum_{i=1}^{N}p_iD^{n_i(1-\alpha)}<D^{(1-\alpha)}\sum_{i=1}^{N}p_i^{\alpha}. \quad (21)$$

Taking logarithm on both sides and we obtain the result (15) after simplification for $\frac{1}{1-\alpha}>0.$

i.e., $L_\alpha<\mathrm{H}_\alpha(P)+\log D.$

(ii) When $\alpha>1,$ then raising power $(1-\alpha)$ to both sides of (19), we have

$$D^{n_i(1-\alpha)}>D^{(1-\alpha)}p_i^{\alpha-1}. \quad (22)$$

Multiplying (22) throughout by p_i and then summing up from $i=1$ to $i=N,$ we have

$$\sum_{i=1}^{N}p_iD^{n_i(1-\alpha)}>D^{(1-\alpha)}\sum_{i=1}^{N}p_i^{\alpha}. \quad (23)$$

Taking logarithm on both sides and we obtain the result (15) after simplification for $\frac{1}{1-\alpha}<0.$

i.e., $L_\alpha<\mathrm{H}_\alpha(P)+\log D.$

In the following theorem, we give a lower bound of lower bound for L_α in terms of $H_\alpha(P)$ for $\alpha>1.$

Theorem 3. For $\alpha>1$ and for every code word lengths n_i, $i=1,...,N$ of theorem 1, L_α can be made to satisfy,

$$L_\alpha\ge\mathrm{H}_\alpha(P)>H_\alpha^\beta(P)+\frac{\log D}{1-\alpha}. \quad (24)$$

Proof: Suppose

$$\overline{n}_i=-\log_D(p_i), \quad (25)$$

Clearly $\overline{n}_i$ and $\overline{n}_i$ + 1 satisfy `equality' in Holder's inequality (11). Moreover, $\overline{n}_i$ satisfies (8). Suppose n_i is the unique integer between $\overline{n}_i$ and $\overline{n}_i$ + 1, then obviously, n_i satisfies (8).
Since $\alpha>1$, we have

$$\left(\sum_{i=1}^{N} p_i D^{n_i(1-\alpha)}\right) \le \left(\sum_{i=1}^{N} p_i D^{\bar{n}_i(1-\alpha)}\right)$$

$$< D\left(\sum_{i=1}^{N} p_i D^{\bar{n}_i(1-\alpha)}\right) \quad (26)$$

Since $\left(\sum_{i=1}^{N} p_i D^{\bar{n}_i(1-\alpha)}\right) = \left(\sum_{i=1}^{N} p_i^{\alpha}\right)$.

Hence (26) becomes

$$\left(\sum_{i=1}^{N} p_i D^{n_i(1-\alpha)}\right) \le \left(\sum_{i=1}^{N} p_i^{\alpha}\right) < D\left(\sum_{i=1}^{N} p_i^{\alpha}\right)$$

which gives (24).

3. CONCLUSION

We know that optimal code is that code for which the value L_α is equal to its lower bound. From the result of the theorem 2, it can be seen that the mean codeword length of the optimal code is dependent on one parameter α, while in the case of Shannon's theorem it does not depend on any parameter. So it can be reduced significantly by taking suitable values of parameter.

4. REFERENCES

[1] Aczel, J. and Daroczy, Z. (1963). Uber Verallgemeinerte quasilineare mittelwerte, die mit Gewichtsfunktionen gebildet sind, Publ. Math. Debrecen, Vol. **10**, pp. 171-190.

[2] Aczel, J. and Daroczy, Z. (1975). On measures of information and their characterizations. Mathematics in Science and Engineering, Vol. 115, Academic Press, New York, London.

[3] Campbell, L.L. (1965). A coding theorem and Renyi's entropy, Information and Control, Vol. 8, pp. 423-429.

[4] Ebanks, B., Sahoo, P. and Sander, W. (1998). Characterizations of information measures, World Scientific Publishing Co., Inc., River Edge, NJ.

[5] Feinstein, A. (1956). Foundation of Information Theory, McGraw Hill, New York.

[6] Hooda, D.S. and Bhaker, U.S. (1997). A generalized `useful' information measure and coding theorems, Soochow J. Math., Vol. 23, pp. 53-62.

[7] Jelinek, F. (1980). Buffer overflow in variable lengths coding of fixed rate sources, IEEE, Vol. 3, pp. 490-501.

[8] Kapur, J.N. (1967). Generalized entropy of order α and type β, Maths. Seminar, Delhi, Vol. 4, pp. 78-94.

[9] Khan, A.B., Bhat, B.A. and Pirzada, S. (2005). Some Results on a Generalized Useful Information Measure, Journal of Inequalities in Pure and Applied Mathematics, Vol. 6(4), Art. 117.

[10] Kieffer, J.C. (1979). Variable lengths source coding with a cost depending only on the codeword length, Information and Control, Vol. 41, pp. 136-146.

[11] Longo, G. (1976). A Noiseless Coding Theorem for Sources Having Utilities, Siam J. Appl. Math., Vol. 30, pp. 739-748.

[12] Renyi, A. (1961). On Measure of entropy and information, Proc. 4th Berkeley Symp. Maths. Stat. Prob., Vol. 1, pp. 547-561.

[13] Shannon, C.E. (1948). A Mathematical Theory of Communication, Bell System Tech. J., Vol. 27, pp. 379-423, 623-656.

[14] Singh, R.P., Kumar, R. and Tuteja, R.K. (2003). Application of Hölder's Inequality in Information Theory, Information Sciences, Vol. 152, pp. 145-154.