

# A Coding Theorem on Havrda-Charvat and Tsallis's Entropy

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

Satish Kumar, Arvind Kumar  
 Department of Mathematics  
 Geeta Institute of Management & Technology  
 Kanipla-136131, Kurukshetra, Haryana (India)

## ABSTRACT

A relation between Shannon entropy and Kerridge inaccuracy, which is known as Shannon inequality, is well known in information theory. In this communication, first we generalized Shannon inequality and then given its application in coding theory.

## KEYWORDS

Shannon inequality, Codeword length, Holder's inequality, Kraft inequality and Optimal code length.

## 1. INTRODUCTION

Let

$$\Delta_n = \left\{ P = (p_1, p_2, \dots, p_n); p_k \geq 0, \sum_{k=1}^n p_k = 1 \right\}$$

$n \geq 2$  be a set of  $n$ -complete probability distributions.

For  $P \in \Delta_n$ , Shannon's measure of information [14] is defined as

$$H(P) = -\sum_{k=1}^n p_k \log_D p_k. \quad (1.1)$$

The measure (1.1) has been generalized by various authors and has found applications in various disciplines such as economics, accounting, crime, physics, etc.

Havrda and Charvat [6] generalized (1.1) in the following form:

$$H(P; \alpha) = \frac{1}{2^{1-\alpha} - 1} \left[ \sum_{k=1}^n p_k^\alpha - 1 \right], \alpha > 0, \alpha \neq 1 \quad (1.2)$$

later on it studied by Vajda [20], Daroczy [4] and Tsallis [19].

For  $P, Q \in \Delta_n$ , Kerridge [7] introduced a quantity known as inaccuracy defined as:

$$H(P, Q) = -\sum_{k=1}^n p_k \log_D q_k. \quad (1.3)$$

There is well known relation between  $H(P)$  and  $H(P, Q)$  which is given by

$$H(P) \leq H(P, Q). \quad (1.4)$$

The relation (1.4) is known as Shannon inequality and its importance is well known in coding theory.

In the literature of information theory, there are many approaches to extend the relation (1.4) for other measures. Nath and Mittal [10] extended the relation (1.4) in the case of entropy of type  $\beta$ .

Using the method of Nath and Mittal [10], Lubbe [21] generalized (1.4) in the case of Renyi's entropy. On the other hand, using the method of Campbell, Lubbe [21] generalized

(1.4) for the case of entropy of type  $\beta$ . Using these generalizations, coding theorems are proved by these authors for these measures.

The objective of this communication is to generalize (1.4) for (1.2) and give its application in coding theory.

## 2. GENERALIZATION OF SHANNON INEQUALITY

For  $P, Q \in \Delta_n$ , Sharma and Autar [16] and Nath [9] defined a measure of inaccuracy, denoted by  $H(P, Q; \alpha)$  as

$$H(P, Q; \alpha) = \frac{1}{2^{1-\alpha} - 1} \left[ \sum_{k=1}^n p_k q_k^{(\alpha-1)} - 1 \right], \alpha > 0, \alpha \neq 1$$

(2.1)

Since  $H(P, Q; \alpha) \neq H(P; \alpha)$ , we will not interpret (2.1) as a measure of inaccuracy. But  $H(P, Q; \alpha)$  is a generalization of the measure of inaccuracy defined in (1.2). In spite of the fact that  $H(P, Q; \alpha)$  is not a measure of inaccuracy in its usual sense, its study is justified because it leads to meaningful new measures of length. In the following theorem, we will determine a relation between (1.2) and (2.1) of the type (1.4).

Since (2.1) is not a measure of inaccuracy in its usual sense, we will call the generalized relation as pseudo-generalization of the Shannon inequality.

## 3. APPLICATION OF HOLDER'S INEQUALITY

**Theorem 1:** If  $P, Q \in \Delta_n$ , then it holds that

$$H(P; \alpha) \leq H(P, Q; \alpha) \quad (2.2)$$

under the condition

$$\sum_{k=1}^n q_k^\alpha \leq \sum_{k=1}^n p_k^\alpha \quad (2.3)$$

and equality holds if

$$q_k = p_k; k=1, 2, \dots, n.$$

**Proof:** (a) If  $0 < \alpha < 1$

By Holder's inequality [17]

$$\left(\sum_{k=1}^n x_k^p\right)^{\frac{1}{p}} \left(\sum_{k=1}^n y_k^q\right)^{\frac{1}{q}} \leq \sum_{k=1}^n x_k y_k \quad (2.4)$$

For all  $x_k, y_k > 0, i = 1, 2, \dots, n$  and  $\frac{1}{p} + \frac{1}{q} = 1, p < 1 (\neq 0), q < 0$ , or  $q < 1 (\neq 0), p < 0$ .

We see that equality holds if and only if there exists a positive constant  $c$  such that

$$x_k^p = c y_k^q.$$

Making the substitutions

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

$$x_k = p_k^{\frac{\alpha}{\alpha-1}} q_k^\alpha, \quad y_k = p_k^{\frac{\alpha}{1-\alpha}}$$

in (2.4), we get

$$\left(\sum_{k=1}^n p_k q_k^{\alpha-1}\right)^{\frac{\alpha}{\alpha-1}} \left(\sum_{k=1}^n p_k^\alpha\right)^{\frac{1}{1-\alpha}} \leq \sum_{k=1}^n q_k^\alpha; \alpha > 0, \alpha \neq 1.$$

Using the condition (2.3), we get

$$\left(\sum_{k=1}^n p_k q_k^{\alpha-1}\right)^{\frac{\alpha}{\alpha-1}} \left(\sum_{k=1}^n p_k^\alpha\right)^{\frac{1}{1-\alpha}} \leq \sum_{k=1}^n p_k^\alpha; \alpha > 0, \alpha \neq 1. \quad (2.5)$$

Since  $0 < \alpha < 1$ , (2.5) becomes

$$\sum_{k=1}^n p_k q_k^{\alpha-1} \geq \sum_{k=1}^n p_k^\alpha \quad (2.6)$$

using (2.6) and the fact that  $0 < \alpha < 1$ , we get (2.2).

(b) If  $\alpha > 1$  the proof follows on the similar lines.

#### 4. APPLICATION IN CODING THEORY

We will now give an application of Theorem 1 in coding theory. Let a finite set of  $n$ -input symbols with probabilities  $p_1, p_2, \dots, p_n$  be encoded in terms of symbols taken from the alphabet  $\{a_1, a_2, \dots, a_n\}$ .

Then it is known Feinstein [5] that there always exist a uniquely decipherable code with lengths  $N_1, N_2, \dots, N_n$  iff

$$\sum_{k=1}^n D^{-N_k} \leq 1. \quad (2.7)$$

If  $L = \sum_{k=1}^n p_k N_k$  is the average codeword length, then for a code which satisfies (2.7), it has been shown that Feinstein [5],

$$L \geq H(P) \quad (2.8)$$

with equality iff  $N_k = -\log_D p_k; k = 1, 2, \dots, n$

and that by suitable encoded into words of long sequences, the average length can be made arbitrary close to  $H(P)$ . This is Shannon's noiseless coding theorem. By considering Renyi's [12] entropy, a coding theorem and analogous to the above

noiseless coding theorem has been established by Campbell [3] and the authors obtained bounds for it in terms of

$$H_\alpha(P) = \frac{1}{1-\alpha} \log_D \sum_{k=1}^n p_k^\alpha; \alpha \neq 1, \alpha > 0.$$

It may be seen that the mean codeword length  $L = \sum_{k=1}^n p_k N_k$  had been generalized parametrically and their bounds had been studied in terms of generalized measures of entropies.

We define the measure of length  $L(\alpha)$  by

$$L(\alpha) = \frac{1}{2^{1-\alpha} - 1} \left[ \sum_{k=1}^n p_k D^{N_k(1-\alpha)} - 1 \right]; \alpha > 0, \alpha \neq 1. \quad (2.9)$$

Also, we have used the condition

$$\sum_{k=1}^n D^{-N_k \alpha} \leq \sum_{k=1}^n p_k^\alpha \quad (2.10)$$

to find the bounds. It may be seen that in the case when  $\alpha = 1$ , then (2.10) reduces to Kraft Inequality (2.7).

**Theorem2:** If  $N_k, k = 1, 2, \dots, n$  are the lengths of codewords satisfying (2.10), then

$$H(P; \alpha) \leq L(\alpha) < D^{1-\alpha} H(P; \alpha) + \frac{1 - D^{1-\alpha}}{1 - 2^{1-\alpha}}. \quad (2.11)$$

**Proof :** In (2.2) choose  $Q = (q_1, q_2, \dots, q_n)$  where

$$q_k = D^{-N_k} \quad (2.12)$$

with choice of  $Q$ , (2.2) becomes

$$H(P; \alpha) \leq \frac{1}{2^{1-\alpha} - 1} \left[ \sum_{k=1}^n p_k D^{N_k(1-\alpha)} - 1 \right]$$

i.e.,  $H(P; \alpha) \leq L(\alpha)$  which proves the first part of (2.11).

The equality holds iff  $D^{-N_k} = p_k, k=1, 2, \dots, n$  which is equivalent to

$$N_k = -\log_D p_k; k = 1, 2, \dots, n. \quad (2.13)$$

Choose all  $N_k$  such that

$$-\log_D p_k \leq N_k < -\log_D p_k + 1.$$

Using the above relation, it follows that

$$D^{-N_k} > p_k D^{-1}. \quad (2.14)$$

We now have two possibilities:

1) If  $\alpha > 1$ ; (2.14) gives us

$$\left( \sum_{k=1}^n p_k D^{N_k(1-\alpha)} \right) > \sum_{k=1}^n p_k^\alpha D^{1-\alpha}$$

using (2.15) and the fact  $\alpha > 1$ , we get right hand side in (2.11).

2) If  $0 < \alpha < 1$ , The proof follows on the same lines.

#### Particular's cases:

(1) Since  $D \geq 2$ , we have

$$\frac{1 - D^{1-\alpha}}{1 - 2^{1-\alpha}} \geq 1.$$

It follows then the upper bound of  $L(\alpha)$  in (2.11) is greater than unity.

(2) If  $\alpha \rightarrow 1$ , then (2.11) becomes

$$\frac{H(P)}{\log D} \leq \frac{L}{\log D} < H(P) + 1.$$

Which is the Shannon [14] classical noiseless coding theorem.

## 5. CONCLUSION

We know that optimal code is that code for which the value  $L(\alpha)$  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 parameter  $\alpha$ , 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.

## 6. REFERENCES

- [1] C. Arndt, Information Measure-Information and its description in Science and Engineering, Springer, Berlin, 2001.
- [2] M.A.K. Baig and Rayees Ahmad Dar, Coding theorems on a generalized information measures, J. KSIAM, vol. 11 (2), 2007, 3-8.
- [3] L.L. Campbell, A coding theorem and Renyi's entropy, Information and Control, vol. 8, 1965, 423-429.
- [4] Z. Daroczy, Generalized Information Functions, Information and Control, vol. 16, 1970, 36-51.
- [5] A. Feinstein, Foundations of Information Theory, McGraw-Hill, New York, 1958.
- [6] J.F. Havrda and F. Charvat, Quantification Methods of Classification Process, The Concept of structural  $\alpha$ -entropy, Kybernetika, vol. 3, 1967, 30-35.
- [7] D. F. Kerridge, Inaccuracy and inference, J. Roy. Statist Soc. Sec., vol. B23, 1961, 184-194.
- [8] A.B. Khan, B.A. Bhat and S.Pirzada, Some Results on a Generalized Useful Information Measure, Journal of Inequalities in Pure and Applied Mathematics, Vol. 6, Issue 4, Article 117, 2005.
- [9] P. Nath, An axiomatic characterization of inaccuracy for finite generalized discrete probability Distributions, Opsearch, vol. 7, 1970, 115-133.
- [10] P. Nath and D. P. Mittal, A generalization of Shannon's inequality and its application in coding theory, Inform. and Control, vol. 23, 1973, 439-445.
- [11] S. Pirzada and B.A. Bhat, Some more results in coding theory, J. KSIAM, vol. 10(2), 2006, 123-131.
- [12] A. Renyi, On Measure of entropy and information, Proc. 4th Berkeley Symp. Maths. Stat. Prob., Vol. 1, 1961, 547-561.
- [13] Satish Kumar, Some More Results on R-Norm Information Measure, Tamkang Journals of Mathematics, Vol. 40(1), 2009, 41-58.
- [14] C.E. Shannon, A mathematical theory of information, Bell System Techn. J., vol. 27, 1948, 378-423, 623-656.
- [15] B. D. Sharma and D. P. Mittal, New non-additive measures of entropy for discrete Probability distributions, J. Math. Sci., vol. 10, 1975, 28-40.
- [16] B. D. Sharma and Ram Autar, On characterization of a generalized inaccuracy measure in information theory, Journal of Applied Probability, vol. 10, 1973, 464-468.
- [17] O. Shisha, Inequalities, Academic Press, New York, 1967.
- [18] R.P. Singh, R. Kumar and R.K. Tuteja, Application of Holder's Inequality in Information Theory, Information Sciences, vol. 152, 2003, 145-154.
- [19] C. Tsallis, Possible Generalization of Boltzmann Gibbs Statistics, J. Stat. Phy., vol. 52, 1988, 479.
- [20] I. Vajda, On measure of entropy and information, Proc. Fourth Berk. Symp. in Math, Stat. and Prob., vol. 1, 1961, 547-561.
- [21] J. C. A. van der Lubbe, On certain coding theorems for the information of order  $\alpha$  and of type  $\beta$ . In: Trans. Eighth Prague Conf. Inform. Theory, Statist. Dec. Functions, Random Processes, vol. C, Academia, Prague, 1978, pp. 253-266.