Tamsui Oxford Journal of Information and Mathematical Sciences **29(1)** (2013) 19-27
Aletheia University

Sharma-Mittal Entropy and Coding Theorem *

Satish Kumar and Arun Choudhary†

Department of Mathematics, Geeta Institute of Management & Technology, Kanipla-136131, Kurukshetra, Haryana, India

Received September 9, 2010, Accepted January 8, 2013.

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 and discuss some particular cases.

Keywords and Phrases: *Shannon inequality, Codeword length, Holder's inequality, Kraft inequality, Optimal code length.*

1. Introduction

Let $\Delta_n = \{P = (p_1, p_2, ..., p_n)\,;\ p_k \geq 0,\ \sum_{k=1}^{n} p_k = 1\}\,,\ n \geq 2$ be a set of n-complete probability distributions.

For P$\in \Delta_n$, Shannon's measure of information [9] is defined as

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

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

*2000 *Mathematics Subject Classification.* Primary 94A15, 94A17, 94A24, 26D15.

†Corresponding author. E-mail: arunchoudhary07@gmail.com

Sharma and Mittal [10] generalized (1.1) in the following form:

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

where $\alpha,\beta>0,\ \alpha\neq\beta,\ \alpha\neq1\neq\beta$.
For $P,Q\in\Delta_n$, Kerridge [6] introduced a quantity known as inaccuracy defined as:

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

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

$$\mathrm{H(P)}\ \leq\ \mathrm{H(P,Q)}. \tag{1.4}$$

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 [7] extended the relation (1.4) in the case of entropy of type β.

Using the method of Nath and Mittal [7], Lubbe [14] generalized (1.4) in the case of Renyi's entropy. On the other hand, using the method of Campbell, Lubbe [14] generalized (1.4) for the case of entropy of type β. 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 Gupta [4] defined a measure of inaccuracy, denoted by $H(P,Q;\alpha,\beta)$ as

$$H(P,Q;\alpha,\beta)=\frac{1}{2^{1-\beta}-1}\left[\left(\sum_{k=1}^{n}p_kq_k^{(\alpha-1)}\right)^{\left(\frac{\beta-1}{\alpha-1}\right)}-1\right], \tag{2.1}$$

where $\alpha, \beta > 0,\ \alpha \neq \beta,\ \alpha \neq 1 \neq \beta$.
Since $H(P, Q; \alpha, \beta) \neq H(P; \alpha, \beta)$, we will not interpret (2.1) as a measure of inaccuracy. But $H(P, Q; \alpha, \beta)$ is a generalization of the measure of inaccuracy defined in (1.2). In spite of the fact that $H(P, Q; \alpha, \beta)$ 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.

Application of Holder's Inequality

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

$$H(P; \alpha, \beta) \leq H(P, Q; \alpha, \beta) \tag{2.2}$$

under the condition

$$\sum_{k=1}^{n} q_k^{\alpha} \leq \sum_{k=1}^{n} p_k^{\alpha} \tag{2.3}$$

and equality holds if

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

Proof: (a) If $0 < \alpha < 1 < \beta$.
By Holder's inequality [11]

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

for all $x_k, y_k > 0,\ i = 1, 2, ..., n$ 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

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

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

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

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

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

$$\left(\sum_{k=1}^{n} p_k q_k^{\alpha-1}\right)^{\left(\frac{\beta-1}{\alpha-1}\right)} \leq \left(\sum_{k=1}^{n} p_k^{\alpha}\right)^{\left(\frac{\beta-1}{\alpha-1}\right)} \tag{2.6}$$

using (2.6) and the fact that $\beta > 1$, we get (2.2).
(b) If $\alpha > 1, \ \beta > 1; \ \ 0 < \alpha < 1; \ \ \beta > 1 \ (\alpha < \beta \textit{ or } \beta < \alpha); \ \ 0 < \beta < 1 < \alpha.$ $\square$

The proof follows on the similar lines.

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 $p_1, p_2, ..., p_n$ be encoded in terms of symbols taken from the alphabet $\{a_1, a_2, ..., a_n\}$.

Then it is known Feinstein [3] that there always exist a uniquely decipherable code with lengths $N_1, N_2, ..., N_n$ iff

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

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 [3],

$$L \geq H(P) \tag{2.8}$$

with equality iff $N_k = -\log_D p_k; \quad k = 1, 2, ..., n$
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 [8] entropy, a coding theorem and analogous to the above noiseless coding theorem has been established by Campbell [1] 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 $L(\alpha, \beta)$ by

$$L(\alpha, \beta) = \frac{1}{2^{1-\beta} - 1}\left[\left(\sum_{k=1}^n p_k D^{N_k(1-\alpha)}\right)^{\left(\frac{\beta-1}{\alpha-1}\right)} - 1\right], \tag{2.9}$$

where $\alpha, \beta > 0,\ \alpha \neq \beta,\ \alpha \neq 1 \neq \beta$.
Also, we have used the condition

$$\sum_{k=1}^n D^{-N_k\alpha} \leq \sum_{k=1}^n p_k^\alpha, \quad \alpha > 0 \tag{2.10}$$

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

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

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

Proof: In (2.2) choose $Q = (q_1, q_2, ..., q_n)$ where

$$q_k = D^{-N_k} \tag{2.12}$$

with choice of Q, (2.2) becomes

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

i.e., $H(P;\alpha,\beta) \leq L(\alpha,\beta)$ which proves the first part of (2.11).
The equality holds iff $D^{-N_k} = p_k,\ k = 1, 2, ..., n$ which is equivalent to

$$N_k = -\log_D p_k; \quad k = 1, 2, ..., n. \tag{2.13}$$

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

$$D^{-N_k} > p_k D^{-1}. \tag{2.14}$$

We now have two possibilities:

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

$$\left(\sum_{k=1}^{n} p_k D^{N_k(1-\alpha)}\right) > \sum_{k=1}^{n} p_k^{\alpha} D^{1-\alpha}. \tag{2.15}$$

Now consider two cases:

i) let $0 < \beta < 1$. Raising both sides of (2.15) with $(\beta-1)/(\alpha-1)$, we get

$$\left(\sum_{k=1}^{n} p_k D^{N_k(1-\alpha)}\right)^{\left(\frac{\beta-1}{\alpha-1}\right)} < \left(\sum_{k=1}^{n} p_k^{\alpha}\right)^{\left(\frac{\beta-1}{\alpha-1}\right)} D^{1-\beta}. \tag{2.16}$$

Since $2^{1-\beta} - 1 > 0$ for $\beta < 1$, we get from (2.16) the right hand side in (2.11).
ii) Let $\beta > 1$. The proof follows similarly.

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

Particular's cases:
(1) Since $D \geq 2$, we have

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

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

(2) If $\beta = \alpha$, then (2.11) becomes

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

where

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

be the Havrda-Charvat [5] Entropy and later on it studied by Vajda [13], Daroczy [2] and Tsallis [12].

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

be the new mean codeword length.
(3) If $\beta \to 1$ then (2.11) becomes

$$H(P;\alpha) \leq L(\alpha) < H(P;\alpha) + \log D.$$

Where $H_\alpha(P) = \frac{1}{1-\alpha}\log_D \sum_{k=1}^{n} p_k^{\alpha},\ \alpha > 0,\ \alpha \neq 1$ be the Renyi's [8] Entropy and $L(\alpha) = \frac{1}{1-\alpha}\log_D \sum_{k=1}^{n} p_k D^{-N_k(\alpha-1)},\ \alpha > 0,\ \alpha \neq 1$ be the new mean codeword length.
(4) If $\beta = \alpha$ and $\alpha \to 1$ then (2.11) becomes

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

Which is the Shannon [9] classical noiseless coding theorem.

Conclusion:

We know that optimal code is that code for which the value $L(\alpha,\beta)$ 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 two parameters α and β, 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 parameters.

References

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

[2] Z. Daroczy, Generalized Information Functions, *Information and Control*, **16** (1970), 36-51.

[3] A. Feinstein, *Foundations of Information Theory*, McGraw-Hill, New York, 1958.

[4] H. C. Gupta and Bhu Dev Sharma, On non-additive measures of inaccuracy, *Czechoslovak Mathematical Journal*, **26** no. 4 (1976), 584-595.

[5] J. F. Havrda and F. Charvat, Quantification Methods of Classification Process, The Concept of structural α-entropy, *Kybernetika*, **3** (1967), 30-35.

[6] D. F. Kerridge, Inaccuracy and inference, *J. Roy. Statist Soc. Sec. B*, **23** (1961), 184-194.

[7] P. Nath and D. P. Mittal, A generalization of Shannon's inequality and its application in coding theory, *Inform. and Control*, **23** (1973), 439-445.

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

[9] C. E. Shannon, A mathematical theory of information, *Bell System Techn. J.*, **27** (1948), 378-423, 623-656.

[10] B. D. Sharma and D. P. Mittal, New non-additive measures of entropy for discrete Probability distributions, *J. Math. Sci.*, **10** (1975), 28-40.

[11] O. Shisha, *Inequalities*, Academic Press, New York, 1967.

[12] C. Tsallis, Possible Generalization of Boltzmann Gibbs Statistics, *J. Stat. Phy.*, **52** (1988), 479.

[13] I. Vajda, On measure of entropy and information, *Proc. Fourth Berk. Symp. in Math, Stat. and Prob.*, **1** (1961), 547-561.

[14] J. C. A. van der Lubbe, On certain coding theorems for the information of order α and of type β, *In: Trans. Eighth Prague Conf. Inform. Theory, Statist. Dec. Functions, Random Processes, Academia, Prague*, **C** (1978), 253-266.