KYBERNETIKA — VOLUME 20 (1984), NUMBER 1

PSEUDO-GENERALIZATION OF SHANNON INEQUALITY FOR MITTAL'S ENTROPY AND ITS APPLICATION IN CODING THEORY

ROSIMARY PEREIRA, GUR DIAL

A relation between Shannon entropy and Kerridge inaccuracy, which is known as Shannon inequality, is well known in information theory. In this communication, Shannon inequality is generalized for Mittal's entropy using a method of Campbell and its application in coding theory is given.

1. INTRODUCTION

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

For $P \in \Gamma_n$, Shannon's measure of information [4] is defined as

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

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

Sharma and Mittal [5] generalized (1.1) in the following form:

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

For $P, Q \in \Gamma_n$, Kerridge [2] 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

$$H(P) \leqq H(P, Q) \tag{1.4}$$

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 two approaches to extend the relation (1.4) for other measures. Nath and Mittal [3] extended the relation (1.4) in the case of entropy of type β.

Using the method of Nath and Mittal [3], Lubbe [6] generalized (1.4) in the case of Renyi's entropy. On the other hand, using the method of Campbell, Lubbe [6] 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) using the method of Campbell for (1.2) and give its application in coding theory.

2. PSEUDO-GENERALIZATION OF SHANNON INEQUALITY

For $P, Q \in \Gamma_n$, define a measure of inaccuracy, denoted by $H(P, Q; \alpha, \beta)$ as

$$(2.1)\qquad H(P, Q; \alpha, \beta) = \frac{1}{2^{1-\beta} - 1}\left[\left(\sum_{k=1}^{r} p_k q_k^{\frac{\alpha-1}{\alpha}}\right)^{\frac{\beta-1}{\alpha-1}\alpha} - 1\right], \quad \alpha, \beta > 0, \quad \alpha \neq \beta, \alpha \neq 1 \neq \beta .$$

Since $H(P, P; \alpha, \beta) \neq H(P; \alpha, \beta)$, 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.

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

$$(2.2)\qquad H(P; \alpha, \beta) \leqq H(P, Q; \alpha, \beta)$$

with equality holds if

$$q_k = \frac{p_k^\alpha}{\sum_{k=1}^{n} p_k^\alpha}, \quad k = 1, 2, \ldots, n .$$

Proof. (a) $0 < \alpha < 1 < \beta$.

Using Hölder's inequality, we get

$$(2.3)\qquad \left(\sum_{k=1}^{n} p_k q_k^{\frac{\alpha-1}{\alpha}}\right)^{\frac{\alpha}{\alpha-1}} \left(\sum_{k=1}^{n} p_k^\alpha\right)^{\frac{1}{1-\alpha}} \leqq 1, \quad \alpha > 0, \quad \alpha \neq 1$$

Since $\alpha < 1$, (2.3) becomes

$$(2.4)\qquad \sum_{k=1}^{n} p_k^\alpha \leqq \left(\sum_{k=1}^{n} p_k q_k^{\frac{\alpha-1}{\alpha}}\right)^\alpha$$

Raising both sides of (2.4) with $(\beta - 1)/(\alpha - 1)$ (<0) we get

$$\left(\sum_{k=1}^{n} p_k^\alpha\right)^{\frac{\beta-1}{\alpha-1}} \geqq \left(\sum_{k=1}^{n} p_k q_k^{\frac{\alpha-1}{\alpha}}\right)^{\alpha\frac{\beta-1}{\alpha-1}} \tag{2.5}$$

Using (2.5) and the fact that $\beta > 1$, we get (2.2).

(b) $\alpha > 1$, $\beta > 1$; $0 < \alpha < 1$, $\beta > 1$ ($\alpha < \beta$ or $\beta < \alpha$); $0 < \beta < 1 < \alpha$. The proof follows on the similar lines. □

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

Then it is known [1] that there always exist a uniquely decipherable code with lengths $N_1, N_2, \ldots, N_n$ iff

$$\sum_{k=1}^{n} D^{-N_k} \leqq 1 \tag{2.6}$$

If $L = \sum_{k=1}^{n} p_k N_k$ is the average codeword length, then for a code which satisfies (2.6), it has been shown that [1],

$$L \geqq H(P) \tag{2.7}$$

with equality iff $N_k = -\log_D p_k$, $k = 1, 2, \ldots, n$.

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

$$L(\alpha, \beta) = \frac{1}{2^{1-\beta} - 1}\left[\left(\sum_{k=1}^{n} p_k D^{N_k \frac{1-\alpha}{\alpha}}\right)^{\alpha\frac{\beta-1}{\alpha-1}} - 1\right], \tag{2.8}$$

$$\alpha, \beta > 0\,, \quad \alpha \neq \beta\,, \quad \alpha \neq 1 \neq \beta\,.$$

We prove the following theorem in respect to the relation between $L(\alpha, \beta)$ and $H(P; \alpha, \beta)$.

Theorem 2. If N_k, $k = 1, \ldots, n$ are the lengths of codewords satisfying (2.6), then

$$H(P; \alpha, \beta) \leqq L(\alpha, \beta) < D^{1-\beta} H(P; \alpha, \beta) + \frac{1 - D^{1-\beta}}{1 - 2^{1-\beta}}\,. \tag{2.9}$$

Proof. In (2.2), choose $Q = (q_1, \ldots, q_n)$ where

$$q_k = \frac{D^{-N_k}}{\sum_{k=1}^{n} D^{-N_k}} \tag{2.10}$$

with choice of Q, (2.2) becomes

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

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

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

Using the relation (2.6), we get

$H(P; \alpha, \beta) \leqq L(P; \alpha, \beta)$ which proves the first part of (2.9).

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

(2.11) $$N_k = -\log_D p_k^\alpha + \log_D \left[\sum_{k=1}^{n} p_k^\alpha \right], \quad k = 1, 2, \ldots, n$$

Choose all N_k such that

$$-\log_D \frac{p_k^\alpha}{\sum_{k=1}^{n} p_k^\alpha} \leqq N_k < -\log_D \frac{p_k^\alpha}{\sum_{k=1}^{n} p_k^\alpha} + 1$$

Using the above relation, it follows that

(2.12) $$D^{-N_k} \geqq \frac{p_k^\alpha}{\sum_{k=1}^{n} p_k^\alpha D}$$

We now have two possibilites:

1) If $\alpha > 1$, (2.12) gives us

(2.13) $$\left[\sum_{k=1}^{n} p_k D^{-N_k \frac{1-\alpha}{\alpha}} \right]^\alpha \geqq \sum_{k=1}^{n} p_k^\alpha D^{1-\alpha}$$

Now consider two cases:

(a) Let $0 < \beta < 1$

Raising both sides of (2.13) with $\beta - 1/(\alpha - 1)$ we get

(2.14) $$\left[\sum_{k=1}^{n} p_k D^{-N_k \frac{1-\alpha}{\alpha}} \right]^{\alpha \frac{\beta-1}{\alpha-1}} \leqq \left[\sum_{k=1}^{n} p_k^\alpha D^{1-\alpha} \right]^{\frac{\beta-1}{\alpha-1}}$$

Since $2^{1-\beta} - 1 > 0$ for $\beta < 1$, we get from (2.14) the right hand side in (2.9).

(b) Let $\beta > 1$. The proof follows similarly.

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

Remarks.

1) Since $D \geqq 2$, we have

$$\frac{1 - D^{1-\beta}}{1 - 2^{1-\beta}} \geqq 1$$

It follows then the upper bound of $L(\alpha, \beta)$ in (2.9) is greater than unity.

2) If $\beta = \alpha$, (2.9) reduces to the relation proved in [3] for entropy of type β.

(Received September 6, 1982.)

REFERENCES

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

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

[3] 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.

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

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

[6] 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, Vol. C. Academia, Prague 1978, pp. 253–266.

Dr. Rosimary Pereira, Prof. Dr. Gur Dial, Departamento de Matematica, U.F.S.C., 88000 Florianópolis, SC, Brasil.