Kybernetika

Prasanna K. Sahoo; Andrew K. C. Wong

Generalized Jensen difference based on entropy functions

Kybernetika, Vol. 24 (1988), No. 4, 241--250

Persistent URL: http://dml.cz/dmlcz/124415

Terms of use:

© Institute of Information Theory and Automation AS CR, 1988

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.

This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* http://project.dml.cz

KYBERNETIKA — VOLUME 24 (1988), NUMBER 4

GENERALIZED JENSEN DIFFERENCE BASED ON ENTROPY FUNCTIONS[1]

PRASANNA K. SAHOO, A. K. C. WONG

This paper concerns an upper bound for a diversity measure between subpopulations (also called the generalized Jensen difference) based on entropy functions. We show that the diversity measure between the subpopulations with given a priori probabilities and induced by either the Shannon entropy or the entropy of degree α can never exceed the corresponding entropies of the a priori probabilities. Through this bound we prove a conjecture of Wong and You [25] in affirmative and suggest a new definition for the index of diversity based on entropy functions. An upper bound for the second order generalized Jensen difference based on entropy functions is also obtained in this paper.

1. INTRODUCTION

Consider a set of populations $\{\pi_k\}$, where the individuals of each population π_k are characterized by a set of measurements A in a measurable space $(\Omega, \mathscr{B})$. The set Ω is the sample space and $\mathscr{B}$ is a σ-algebra of subsets of Ω. The discrete probability distribution function of A in π_k is denoted by P_k which belongs to an n-dimensional simplex $\Delta_n := \{P = (p_1, p_2, \ldots, p_n) \mid 0 \leqq p_k \leqq 1, \sum_{k=1}^{n} p_k = 1\}$. Let Δ_n^0 be the interior of the n-dimensional simplex Δ_n, that is, $\Delta_n^0 = \{P = (p_1, p_2, \ldots, p_n) \mid 0 < p_k < 1, \sum_{k=1}^{n} p_k = 1\}$. Notice that Δ_n and Δ_n^0 are convex subsets of $\mathbb{R}^n$. A real valued function $D\colon \Delta_n \to \mathbb{R}$ is called a diversity measure on Δ_n if it satisfies the following conditions:

(a) $D(P) \geqq 0$ for all $P \in \Delta_n$,

(b) $D(P) = 0$ if, and only if, $p_i = 1$ for some i and $p_j = 0$ for all $j \neq i$,

(c) D is a concave function on Δ_n, that is, if $\lambda_1, \lambda_2, \ldots, \lambda_m$ are nonnegative real numbers such that $\sum_{i=1}^{m} \lambda_i = 1$ and $P_1, P_2, \ldots, P_m \in \Delta_n$, then

$$D\Big(\sum_{k=1}^{m} \lambda_k P_k\Big) \geqq \sum_{k=1}^{m} \lambda_k \, D(P_k) . \tag{1.1}$$

[1]) This work is partially supported by a NSERC of Canada grant.

While the conditions (a) and (b) are natural, the condition (c) was motivated by the consideration that the diversity in a mixture of populations should not be smaller than the average of the diversities within the component populations. Some examples of the diversity function defined on the set Δ_n are the followings (Nayak [14] and [15]):

$$H(P) = -\sum_{k=1}^{n} p_k \log_2 p_k \tag{1.2}$$

$$H_\alpha(P) = \frac{1}{2^{1-\alpha} - 1} \left(\sum_{k=1}^{n} p_k^\alpha - 1\right), \quad \alpha > 0, \quad \alpha \neq 1, \tag{1.3}$$

$$H^\alpha(P) = \frac{1}{1-\alpha} \log_2 \left(\sum_{k=1}^{n} p_k^\alpha\right), \quad \alpha \in]0, 1[, \tag{1.4}$$

$$S(P) = 1 - \sum_{k=1}^{n} p_k^2, \tag{1.5}$$

$$B(P) = \frac{1}{1 - 2^{\gamma-1}} \left[1 - \left(\sum_{k=1}^{n} p_k^{1/\gamma}\right)^\gamma\right], \quad \gamma > 0, \quad \gamma \neq 1, \tag{1.6}$$

$$R(P) = PMP^{\mathrm{T}}, \tag{1.7}$$

where M is an $n \times n$ matrix such that PMP^{T} satisfies conditions (a), (b) and (c). The first three diversity measures are the Shannon entropy [22], the entropy of degree α [10], and the entropy of order α [19], respectively. The fourth diversity measure is a special case of the entropy of degree α and is known as the Gini-Simpson index in statistics. The fifth diversity measure in the example is the γ-entropy introduced by Arimoto [1] and axiomatically characterized by Behara and Chawla [2]. The diversity measures (1.3), (1.4) and (1.6) are entropies which depend on a parameter. In applications, one expects that these should be monotonically increasing functions of the parameter. As for the entropy of degree α and the entropy of order α, it is well known that they are monotonically increasing functions of α. Regarding the monotonicity of (1.6), it was conjectured that the γ-entropy is also an increasing function of the parameter γ (see Nayak [14]). Recently, Sahoo [20] has shown that this conjecture is true. The last diversity measure was suggested by Rao [18] and it is often referred to as the Rao's quadratic entropy. The Shannon entropy, the entropy of degree α, the entropy of order α, the Gini-Simpson index, and the γ-entropy are all invariant under permutation of their arguments. As a result, they all have certain flaws. In spite of these flaws, entropic diversity measures are widely used in genetics, sociology, statistics [14, 15, 24], anthropology, and recently in pattern recognition [3, 12, 25].

Given a set of probabilities measures $P_1, P_2, \ldots, P_m \in \Delta_n$ and $X = (x_1, x_2, \ldots, x_m) \in \Delta_m$, and a concave function D defined on Δ_n, the following decomposition was suggested by Rao ([17] and [18]).

$$D\left(\sum_{k=1}^{m} x_k P_k\right) = \sum_{k=1}^{m} x_k D(P_k) + J_D(\{P_i\}, \{x_i\}). \tag{1.8}$$

The second term, $J_D(\{P_i\}, \{x_i\})$, in (1.8) is called the generalized Jensen difference in Burbea and Rao [4] of the function D. If $P_1, P_2, \ldots, P_m$ are probability distributions in m subpopulations with a priori probabilities (or a priori weights) $x_1, x_2, \ldots \ldots, x_m$, then the term $D(\sum_{k=1}^{m} x_k P_k)$ is known as the total diversity whereas the quantity $\sum_{k=1}^{m} x_k D(P_k)$ is the average diversity within subpopulations. Thus, $J_D(\{P_i\}, \{x_i\})$ may be interpreted as the diversity between the subpopulations. In biological works, J_D when D is the Shannon entropy, is defined to be the information radius on probability distributions $P_1, P_2, \ldots, P_m$ (see Sibson [23]) and some applications of this concept to cluster analysis are discussed in Sibson [23] and Jardine and Sibson [11]. The ratio

$$G_D = \frac{J_D(\{P_i\}, \{x_i\})}{J_D(\{P_i\}, \{x_i\}) + \sum_{k=1}^{m} x_k D(P_k)}$$

is the index of diversity induced by the functional D between subpopulations compared to the total. The index G_D has been used widely in genetics by Lewontin [13], Nei [16] and Chakraborty [6] when D is the Shannon entropy. For convenience, we will refer $J_D(\{P_i\}, \{x_i\})$ to as the diversity (between the subpopulations) induced by the function D.

This paper is organized as follows: In Section 2, we prove a upper bound for J_D when D is the Shannon entropy. Section 3 presents an upper bound for J_D assuming D to be entropy of degree α. In Section 4, some inequalities regarding the generalized Jensen difference are proven. In Section 5, we propose a new index of diversity.

2. UPPER BOUND FOR J_D INDUCED BY THE SHANNON ENTROPY

In this section, we prove the following theorem.

Theorem 1. Let $P_i = (p_{i1}, p_{i2}, \ldots, p_{in}) \in \Delta_n^0$ for $i = 1, 2, \ldots, m$ be the m complete probability distributions. Let $X = (x_1, x_2, \ldots, x_m) \in \Delta_m^0$, then inequality

$$J_H(\{P_i\}, \{x_i\}) < H(X) \tag{2.1}$$

holds, where H is the Shannon entropy.

Proof. Using (1.8) and the form of the Shannon entropy in (1.2), one obtains

$$J_H(\{P_i\}, \{x_i\}) - H(X) = \sum_{j=1}^{m} \sum_{k=1}^{n} p_{jk} \log_2 \frac{p_{jk}}{r_k} + \sum_{i=1}^{m} x_i \log_2 x_i\,, \tag{2.2}$$

where $r_k = \sum_{i=1}^{m} x_i p_{ik}$ $(k = 1, 2, \ldots, n)$. Since $\sum_{k=1}^{n} p_{ik} = 1$ for all $i = 1, 2, \ldots, m$, therefore (2.2) can be rewritten as

$$J_H(\{P_i\}, \{x_i\}) - H(X) = \sum_{i=1}^{m} \sum_{k=1}^{n} x_i p_{ik} \log_2 \frac{x_i p_{ik}}{r_k}. \tag{2.3}$$

The use of the inequality

$$\log_2 x \leqq (x - 1) \log_2 e$$

(with equality if, and only if, $x = 1$) in (2.3) results in the following:

$$J_H(\{P_i\}, \{x_i\}) - H(X) < \sum_{i=1}^{m} \sum_{k=1}^{n} x_i p_{ik} \left(\frac{x_i p_{ik}}{r_k} - 1 \right) \log_2 e. \tag{2.4}$$

This strict inequality is due to the fact that $P_i \in \Delta_n^0$ and $X \in \Delta_m^0$. Some algebraic simplifications of (2.4) yield

$$J_H(\{P_i\}, \{x_i\}) - H(X) < \log_2 e \sum_{i=1}^{m} \sum_{k=1}^{n} \frac{(x_i^2 p_{ik}^2 - x_i p_{ik} r_k)}{r_k}.$$

Some further simplifications of the above inequality lead to

$$J_H(\{P_i\}, \{x_i\}) - H(X) < \log_2 (e^{-1}) \sum_{k=1}^{n} \frac{1}{r_k} \sum_{s=1}^{n} \sum_{t \neq s} x_s x_t p_{sk} p_{tk}, \tag{2.5}$$

where $\sum_{t \neq s}$ denotes the summation over all t except $t = s$. Since x_i and p_{ik} are strictly greater than zero, we have

$$\sum_{k=1}^{n} \frac{1}{r_k} \sum_{s=1}^{n} \sum_{t \neq s} x_s x_t p_{sk} p_{tk} > 0. \tag{2.6}$$

Hence, from (2.6) and (2.5) one obtains

$$J_H(\{P_i\}, \{x_i\}) < H(X).$$

This completes the proof of the theorem. □

Remark 1. The proof of Theorem 1 can also be obtained using the analytic and algebraic properties of mutual information, since $J_H(\{P_i\}, \{x_i\})$ (see Gallager [9] p. 90) is the mutual information associated with the m-input symbol distribution and the so called channel matrix $(p_{ij})_{m \times n}$. However the proof given in this paper is straightforward and does not use the properties of mutual information.

The proof of the following corollary is similar to that of Theorem 1 and is therefore omitted.

Corollary 1. Let $P_i \in \Delta_n$ $(i = 1, 2, \ldots, m)$ and $X \in \Delta_m$. Then the following inequality holds

$$J_H(\{P_i\}, \{x_i\}) \leqq H(X)$$

with equality if, and only if, $x_j = 1$ for some j and $x_i = 0$ for all $i \neq j$.

3. UPPER BOUND FOR J_D INDUCED BY THE ENTROPY OF DEGREE α

In this section, we shall derive an upper bound for the diversity (between the sub-population) induced by the entropy of degree α. We denote the entropy of degree α by H_α. Notice that the definition of H_α is valid only for all $\alpha\,(>0)$ and $\alpha \neq 1$. When $\alpha = 1$, we define H_α through the limit $\alpha \to 1$. That is,

$$H_1 := \lim_{\alpha \to 1} H_\alpha\,. \tag{3.1}$$

It is not difficult to prove, using l'Hopital rule, that the right hand side of (3.1) is the Shannon entropy H.

Theorem 2. Let $P_1, P_2, \ldots, P_m \in \Delta_n^0$ and $X \in \Delta_m^0$, then $J_{H_\alpha}(\{P_i\}, \{x_i\})$ satisfies the inequality

$$J_{H_\alpha}(\{P_i\}, \{x_i\}) < H_\alpha(X)\,, \quad \alpha \in \mathbb{R}_+\,, \tag{3.2}$$

where H_α is the entropy of degree α.

Proof. Let $\alpha > 0$ and consider the following

$$\sum_{j=1}^{n} \Big[\Big(\sum_{i=1}^{m} x_i p_{ij}\Big)^\alpha - \sum_{i=1}^{m} (x_i p_{ij})^\alpha\Big] - \sum_{i=1}^{m} x_i^\alpha + 1\,.$$

Since $\sum_{i=1}^{m} x_i = 1 = \sum_{j=1}^{n} p_{ij}$, the above expression can be written as

$$\sum_{j=1}^{n} \Big[\Big(\sum_{i=1}^{m} x_i p_{ij}\Big)^\alpha - \sum_{i=1}^{m} (x_i p_{ij}^\alpha + x_i^\alpha p_{ij} - x_i p_{ij})\Big]\,.$$

Let

$$\Phi_j(\alpha) := \Big(\sum_{i=1}^{m} x_i p_{ij}\Big)^\alpha - \sum_{i=1}^{m} (x_i p_{ij}^\alpha + x_i^\alpha p_{ij} - x_i p_{ij})\,. \tag{3.3}$$

At this point we would like to prove the following inequalities

$$\sum_{j=1}^{n} \Phi_j(\alpha) \begin{cases} <0 & \text{for} \quad \alpha \in \,]0, 1[\\ >0 & \text{for} \quad \alpha \in \,]1, \infty[\,. \end{cases} \tag{3.4}$$

First consider the case, when $\alpha \in \,]1, \infty[$. Since $x_i, p_{ij} \in \,]0, 1[$, therefore $\big(\sum_{i=1}^{m} x_i p_{ij}\big)^\alpha > \sum_{i=1}^{m} x_i^\alpha p_{ij}^\alpha$. Thus, in view of this, (3.3) becomes

$$\Phi_j(\alpha) > \sum_{i=1}^{m} \big[x_i^\alpha p_{ij}^\alpha - x_i p_{ij}^\alpha + x_i p_{ij} - x_i^\alpha p_{ij}\big]\,. \tag{3.5}$$

Rewriting (3.5), we have

$$\Phi_j(\alpha) > \sum_{i=1}^{m} (p_{ij} - p_{ij}^\alpha)(x_i - x_i^\alpha)\,. \tag{3.6}$$

Since p_{ij} and x_i are in $]0, 1[$, therefore $(p_{ij} - p_{ij}^\alpha) > 0$ if $\alpha > 1$. Similarly we get $(x_i - x_i^\alpha) > 0$. Hence from (3.6), we obtained the second half of the inequality (3.4).

Now consider the case when $0 < \alpha < 1$. Here, the inequality (3.6) will reverse its sign, that is

$$\Phi_j(\alpha) < \sum_{i=1}^{m} (p_{ij} - p_{ij}^{\alpha})(x_i - x_i^{\alpha}), \tag{3.7}$$

since $(\sum_{i=1}^{m} x_i p_{ij})^{\alpha} < \sum_{i=1}^{m} x_i^{\alpha} p_{ij}^{\alpha}$ for $\alpha \in]0, 1[$. In this case $(p_{ij} - p_{ij}^{\alpha})$ and $(x_i - x_i^{\alpha})$ are strictly less than zero. Thus, by (3.7) the first half of the inequality (3.4) is proven.

By using the definition of H_α and the definition of $J_{H_\alpha}(\{P_i\}, \{x_i\})$, we obtain

$$J_{H_\alpha}(\{P_i\}, \{x_i\}) - H_\alpha(X) = (2^{1-\alpha} - 1)^{-1} \sum_{j=1}^{n} \Phi_j(\alpha). \tag{3.8}$$

Note that if $\alpha > 1$ then $(2^{1-\alpha} - 1)^{-1} < 0$. Thus, using (3.4) in (3.8) we obtain (3.2). Again, if $0 < \alpha < 1$, then $(2^{1-\alpha} - 1)^{-1} > 0$ and (3.2) follows. The case, when $\alpha = 1$ can be obtained from Theorem 1 since $\lim_{\alpha \to 1} J_{H_\alpha}(\{P_i\}, \{x_i\}) = J_H(\{P_i\}, \{x_i\})$. This completes the proof of the theorem. □

Remark 2. Daróczy [7] defined channel capacity of degree α using the quantity $I_\alpha(X, Y) := H_\alpha(Y) - H_\alpha(Y \mid X)$, where Y is the distribution associated with the mixture, that is, $\sum_{k=1}^{m} x_k P_k$ and $H_\alpha(Y \mid X) = \sum_{k=1}^{m} x_k^{\alpha} H_\alpha(P_k)$. Thus, $J_{H\alpha}(\{P_i\}, \{x_i\})$ is not the mutual information of Daróczy and hence one cannot use the properties of mutual information of degree α to obtain another proof of Corollary 1 unlike the case in Theorem 1. For definition and some applications of $H_\alpha(Y \mid X)$, the conditional entropy of degree α, refer to El-Sayed [8] and Sahoo [21].

The following corollary can be proven by mimicking the proof of Theorem 2.

Corollary 2. Let $P_i \in \Delta_n$ $(i = 1, 2, \ldots, m)$ and $X \in \Delta_m$, then the inequality

$$J_{H_\alpha}(\{P_i\}, \{x_i\}) \leqq H_\alpha(X)$$

holds, with equality if, and only if, $x_j = 1$ for some j and $x_i = 0$ for all $i \neq j$.

4. UPPER BOUND FOR SECOND ORDER GENERALIZED JENSEN DIFFERENCE

The concept of the generalized Jensen difference was extended to higher order in [5] by Burbea and Rao. In view of this extension, the Jensen difference J_D defined through (1.8), is the Jensen difference of order one. In this section, we derive an upper bound for the second order Jensen difference based on the Shannon entropy or the entropy of degree α.

Consider the set of n-ary complete probability distributions $\{P_{i_1 i_2} \in \Delta_n \mid i_i = 1, 2, \ldots, m_1; i_2 = 1, 2, \ldots, m_2\}$ index by combinations of the levels i_1 and i_2 of two factors with independent a priori distributions $X^{(j)}$ in Δ_{m_j}, where $X^{(j)} = (x_1^{(j)}, x_2^{(j)}, \ldots, x_{m_j}^{(j)})$, $j = 1, 2$. The conditional distributions subject to individual levels of i_1 and i_2 and the unconditional distributions are

$$P_{i_1} = \sum_{i_2=1}^{m_2} x_{i_2}^{(2)} P_{i_1 i_2}, \tag{4.1}$$

$$P_{i_2} = \sum_{i_1=1}^{m_1} x_{i_1}^{(1)} P_{i_1 i_2}, \tag{4.2}$$

and

$$P = \sum_{i_1=1}^{m_1} \sum_{i_2=1}^{m_2} x_{i_1}^{(1)} x_{i_2}^{(2)} P_{i_1 i_2}. \tag{4.3}$$

The conditional Jensen difference between the levels of $X^{(1)}$ for a given level i_2 of $X^{(2)}$ is, given by (see Burbea and Rao [5])

$$J_D^{(1,i_2)} = D\Big(\sum_{i_1=1}^{m_1} x_{i_1}^{(1)} P_{i_1 i_2}\Big) - \sum_{i_1=1}^{m_1} x_{i_1}^{(1)} D(P_{i_1 i_2}). \tag{4.4}$$

The Jensen difference between the levels of $X^{(1)}$ averaged over all the levels of $X^{(2)}$ is

$$J_D^{(1,\bar{2})} = D(P) - \sum_{i_1=1}^{m_1} D(P_{i_1}). \tag{4.5}$$

The second order generalized Jensen difference (or interaction between factor 1 and 2) is defined in [5] as

$$J_D^{(1,2)} = \sum_{i_2=1}^{m_2} x_{i_2}^{(2)} J_D^{(1,i_2)} - J_D^{(1,\bar{2})}. \tag{4.6}$$

In [5], Burbea and Rao studied the convexity of higher order generalized Jensen difference. Regarding second order generalized Jensen difference the followings

(a) $J_D^{(1,2)} = J_D^{(2,1)}$,

(b) $J_D^{(1,2)} \geqq 0$ if and only $J_D \geqq 0$,

are true. For details readers should refer to [5]. Now we prove the following lemma.

Lemma 1. If D is a concave function, then the inequality

$$J_D^{(1,2)} \leqq \sum_{i_1=1}^{m_1} x_{i_1}^{(1)} \Big[D\Big(\sum_{i_2=1}^{m_2} x_{i_2}^{(2)} P_{i_1 i_2}\Big) - \sum_{i_2=1}^{m_2} x_{i_2}^{(2)} D(P_{i_1 i_2})\Big] + \tag{4.7}$$
$$+ \sum_{i_2=1}^{m_2} x_{i_2}^{(2)} \Big[D\Big(\sum_{i_1=1}^{m_1} x_{i_1}^{(1)} P_{i_1 i_2}\Big) - \sum_{i_1=1}^{m_1} x_{i_1}^{(1)} D(P_{i_1 i_2})\Big]$$

holds, where $P_{i_1 i_2} \in \Delta_n$ $(i_1 = 1, 2, \ldots, m_1;\ i_2 = 1, 2, \ldots, m_2)$ and $X^{(j)} \in \Delta_m$ $(j = 1, 2)$.

Proof. By (4.4) and (4.3), (4.5) becomes

$$J_D^{(1,2)} = \sum_{i_2=1}^{m_2} x_{i_2}^{(2)} \Big[D\Big(\sum_{i_1=1}^{m_1} x_{i_1}^{(1)} P_{i_1 i_2}\Big) - \sum_{i_1=1}^{m_1} x_{i_1}^{(1)} D(P_{i_1 i_2})\Big] + \tag{4.8}$$
$$+ \sum_{i_1=1}^{m_1} x_{i_1}^{(1)} D\Big(\sum_{i_2=1}^{m_2} x_{i_2}^{(2)} P_{i_1 i_2}\Big) - D\Big(\sum_{i_1=1}^{m_1} \sum_{i_2=1}^{m_2} x_{i_1}^{(1)} x_{i_2}^{(2)} P_{i_1 i_2}\Big).$$

Since D is a concave function,

$$\text{(4.9)} \qquad \mathsf{D}\Big(\sum_{i_1=1}^{m_1}\sum_{i_2=1}^{m_2} x_{i_1}^{(1)} x_{i_2}^{(2)} P_{i_1 i_2}\Big) \geqq \sum_{i_1=1}^{m_1}\sum_{i_2=1}^{m_2} x_{i_1}^{(1)} x_{i_2}^{(2)}\, \mathsf{D}(P_{i_1 i_2})\,.$$

In view of (4.9), (4.8) yields (4.7). This completes the proof of the lemma. □

Now we proceed to determine an upper bound for $J_{\mathsf{D}}^{(1,2)}$ when D is either the Shannon entropy of the entropy of degree α. We denote $J_{H_\alpha}^{(1,2)}$ as the second order generalized Jensen difference based on the entropy of degree α.

Theorem 3. Let $P_{i_1 i_2} \in \Delta_n (i_1 = 1, 2, \dots, m_1;\ i_2 = 1, 2, \dots, m_2)$ and $X^{(j)} \in \Delta_{m_j} (j = 1, 2)$. Then the inequality

$$\text{(4.10)} \qquad J_{H_\alpha}^{(1,2)} \leqq \sum_{j=1}^{2} H_\alpha(X^{(j)})$$

holds for all α in $\mathbb{R}_+$.

Proof. Since for α in $\mathbb{R}_+ - \{1\}$, H_α is a concave function of the probabilities, using Lemma 1 we obtain

$$\text{(4.11)} \qquad J_{H_\alpha}^{(1,2)} \leqq \sum_{i_2=1}^{m_2} x_{i_2}^{(2)}\Big[H_\alpha\Big(\sum_{i_1=1}^{m_1} x_{i_1}^{(1)} P_{i_1 i_2}\Big) - \sum_{i_1=1}^{m_1} x_{i_1}^{(1)} H_\alpha(P_{i_1 i_2})\Big] + $$
$$+ \sum_{i_1=1}^{m_1} x_{i_1}^{(1)}\Big[H_\alpha\Big(\sum_{i_2=1}^{m_2} x_{i_2}^{(2)} P_{i_1 i_2}\Big) - \sum_{i_2=1}^{m_2} x_{i_2}^{(2)} H_\alpha(P_{i_1 i_2})\Big]\,.$$

Now Corollary 2 in (4.11) yields

$$\text{(4.12)} \qquad J_{H_\alpha}^{(1,2)} \leqq \sum_{i_2=1}^{m_2} x_{i_2}^{(2)} H_\alpha(X^{(1)}) + \sum_{i_1=1}^{m_1} x_{i_1}^{(1)} H_\alpha(X^{(2)})\,.$$

Using the fact $\sum_{i_1=1}^{m_1} x_{i_1}^{(1)} = 1$ and $\sum_{i_2=1}^{m_2} x_{i_2}^{(2)} = 1$, we get (4.10) from (4.12). If $\alpha = 1$, by Corollary 1 we again obtain (4.10). This completes the proof of the theorem. □

5. COMMENTS

By maximality of the Shannon entropy, we obtain the following inequality, an obvious consequence of Corollary 1:

$$\text{(5.1)} \qquad J_H(\{P_i\}, \{x_i\}) \leqq \log_2 m\,,$$

where $P_1, P_2, \dots, P_m \in \Delta_n$ and $X \in \Delta_m$.

In the context of structural pattern recognition, Wong and You [25] observed the necessity of an upper bound for J_{D} when D is the Shannon entropy. They conjectured on the basis of computer simulation that, for all $x \in [0, 1[$ and for every $P_1, P_2 \in \Delta_n$, the quantity $H(xP_1 + (1 - x) P_2) - x H(P_1) - (1 - x) H(P_2)$, which they call 'increment of entropy', is bounded from above by 1 (see [25, p. 604]). From (5.1), it is quite obvious that the conjecture is true.

Since D is a concave function, by Jensen inequality for concave function, we see from (1.8), that $J_D(\{P_i\}, \{x_i\})$ is nonnegative. Thus, the 'increment of entropy' (which is equivalent to the mutual information associated with a 2-symbol source and channel matrix (p_{ij})) can be used as an index of diversity. This index is successfully used in the comparison of random graphs in pattern recognition (see Wong and You [25]). Similar to the index of diversity G_D one can define another index of diversity by normalizing the diversity between the subpopulations by its (best) upper bound. If the concave function used in the definition of J_D is either the Shannon entropy or the entropy of degree α, then the index of diversity can be defined as follows:

$$G_D^* = \begin{cases} \dfrac{J_D(\{P_i\}, \{x_i\})\,(2^{1-\alpha} - 1)}{n^{1-\alpha} - 1} & \text{if} \quad D = H_\alpha \\[2ex] \dfrac{J_D(\{P_i\}, \{x_i\})}{\log_2 m} & \text{if} \quad D = H\,. \end{cases} \tag{5.2}$$

In summary, we have shown that the diversity between subpopulations (or the generalized Jensen difference) induced by either the Shannon entropy or the entropy of degree α can never exceed the corresponding entropies of the prior probabilities of subpopulations. The second order generalized Jensen difference (which is also interaction between factor 1 and 2) is bounded from above by the sum of the entropies of prior distributions of the factor 1 and 2.

ACKNOWLEDGEMENT

The authors are thankful to Prof. J. Aczél, Y. C. Chen and the referee for comments and suggestion for the improvement of the paper.

(Received July 16, 1987.)

REFERENCES

[1] S. Arimoto: Information theoretic considerations of estimations problems. Inform. and Control *19* (1971), 181—194.

[2] M. Behara and J. S. Chawla: Generalized γ-entropy. Entropy and ergodic theory, pp. 15—38. Selecta Statistica Canadiana, Vol. 2, Univ. Press Canada, Toronto, Ontario 1975.

[3] M. Ben-Bassat: f-Entropies, probability of error, and feature selection. Inform. and Control *39* (1978), 227—242.

[4] J. Burbea and C. R. Rao: On the convexity of some divergence measures based on entropy functions. IEEE Trans. Inform. Theory *IT-28* (1982), 489—495.

[5] J. Burbea and C. R. Rao: On the convexity of higher order Jensen differences based on entropy functions. IEEE Trans. Inform. Theory *IT-28* (1982), 961—963.

[6] R. Chakraborty: A note on Nei's measure of gene diversity in a substructured population. Humangenetik *21* (1974), 85—88.

[7] Z. Daróczy: Generalized information functions. Inform. and Control *16* (1970), 36—51.

[8] A. El-Sayed: A generalized entropy form of the Fano inequality. Utilitas Math. *12* (1977), 289—298.

[9] R. G. Gallager: Information Theory and Reliable Communication. Wiley, New York 1968.
[10] J. Havrda and F. Charvát: Quantification method of classification process, concept of structural a-entropy. Kybernetika *3* (1967), 30—35.
[11] N. Jardine and R. Sibson: Mathematical Taxonomy. Wiley, New York 1971.
[12] L. Kanal: Patterns in pattern recognition: 1968—1974. IEEE Trans. Inform. Theory *IT-20* (1974), 697—722.
[13] R. C. Lewotin: The apportionment of human diversity. Evolutionary Biology *6* (1972), 381—398.
[14] T. K. Nayak: On diversity measures based on entropy functions. Comm. Statist. A *14* (1985), 203—215.
[15] T. K. Nayak: Sampling distributions in analysis of diversity. Sankhya Ser. B *48* (1986), 1—9.
[16] M. Nei: Analysis of gene diversity in subdivided populations. Proc. Nat. Acad. Sci. *70* (1973), 3321—3323.
[17] C. R. Rao: Diversity and dissimilarity coefficients: A unified approach. J. Theor. Pop. Biol. *21* (1982), 24—43.
[18] C. R. Rao: Convexity properties of entropy function and analysis of diversity. In: Inequalities in Statistics and Probability (Y. L. Tong, ed.). IMS Lecture Notes, 1984, Vol. 5, pp. 68—77.
[19] A. Rényi: On measures of entropy and information. In: Proc. Fourth Berkeley Symp. 1960, Vol. 1, Univ. of California Press, Berkeley—Los Angeles 1961, pp. 547—561.
[20] P. K. Sahoo: On the monotonicity of generalized gamma entropy. Submitted to Bull. Inst. Math. Acad. Sinica (1985).
[21] P. K. Sahoo: On a generalization of Shannon's random cipher result. Kybernetika *22* (1986), 385—392.
[22] C. E. Shannon: Mathematical theory of communication. Bell Systems Tech. J. *27* (1948), 379—423.
[23] R. Sibson: Information radius. Z. Wahrsch. Verw. Gebiete *14* (1969), 149—160.
[24] I. Vajda: Bounds of the minimal error probability on checking a finite or countable number of hypotheses. Problems Inform. Transmissions *4* (1968), 6—14.
[25] A. K. C. Wong, and M. L. You: Distance and entropy measure of random graph with application to structural pattern recognition. IEEE Trans. Pattern Anal. Mach. Intell. *PAMI-7* (1985), 599—609.

Dr. Prasanna K. Sahoo, Dr. A. K. C. Wong, Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1. Canada.