INFORMATION AND CONTROL 36, 133–156 (1978)

Nonnegative Entropy Measures of Multivariate Symmetric Correlations

TE SUN HAN

Department of Mathematical Engineering, Faculty of Engineering, Sagami Institute of Technology, Tsujido Nishikaigan 1-1-25, Fujisawa 251, Japan

A study of nonnegativity "in general" in the symmetric (correlative) entropy space as well as discussions of some related problems is presented. The main result is summarized as Theorems 4.1 and 5.3, which give the necessary and sufficient condition for an element of the symmetric (correlative) entropy space to be nonnegative. In particular, Theorem 4.1 may be regarded as establishing a mathematical foundation for information-theoretic analysis of multivariate symmetric correlation. On the basis of these results, we propose a "hierarchical structure" of probabilistic dependence relations where it is shown that any symmetric correlation associated with a nonnegative entropy is decomposed into pairwise conditional and/or nonconditional correlations. A systematic duality existing in the set of nonnegative entropies is also considerably clarified.

1. INTRODUCTION

The problem of constructing "effective" measures of correlation is of fundamental significance in information-theoretic multivariate analysis. Extensive investigations in this respect have been developed by a number of researchers (e.g., Watanabe, 1954, 1960, 1969; McGill, 1954; Garner and McGill, 1956; Garner, 1958; Kullback, 1959; Baldwin, 1966; Ku and Kullback, 1968) along diverse directions of theoretic developments and/or practical applications.

In most cases, the problem has been treated and put forward with special reference to "nonnegativity" of measures. In fact, nonnegativity plays a dominant role in the multivariate interpretation of measures such as Shannon's (conditional) entropy/mutual information, Watanabe's total correlation or Baldwin's dependency capacity, Kullback's divergence. (McGill's multiple mutual information is an eminent exception. However, even this may be connected with a certain kind of nonnegativity. See Han, to appear.)

The present paper deals with the problem of nonnegativity "in general" for a collection of various entropy measures as well as several related topics on the basis of the concept of (*correlative*) *entropy space*, which has been introduced by Han (1975) to study the linear dependence relations underlying the set of information-theoretic measures. We first define, as a fundamental tool of ana-

0019-9958/78/0362–0133$02.00/0

643/36/2-2

lysis, the (correlative) "symmetric" entropy space which is a subspace of the (correlative) entropy space, and next discuss mainly the nonnegativity of elements in this subspace.

We establish in Section 3 a fundamental basis of the (correlative) symmetric entropy space with some nonnegativity properties. The inequalities obtained here closely resemble in form those derived by Watanabe (1960). The main result concerning nonnegativity is stated in Theorems 4.1 and 5.3, which give the necessary and sufficient condition for an element of the (correlative) symmetric entropy space to be nonnegative. The multivariate implication of the condition so derived is clarified in Section 4.2. Theorem 4.1 is to be regarded as establishing the possibilities as well as the limitations of the information-theoretic methodology in analyzing multivariate symmetric correlations in terms of nonnegative measures. In this sense, the theorem offers a mathematical foundation for information-theoretic multivariate analysis insofar as we are concerned with "symmetric" correlations. In particular, such a systematic analysis of nonnegativity leads us to the concept of "hierarchical structure" of correlations, which may be regarded as an extension of the ordinary pairwise correlation structure.

In Section 5, a remarkable duality is shown in reference to the hierarchical structure obtained. Finally, as a typical case of duality, multivariate-analytic interpretations of total correlation and dual total correlation (Han, 1975) are given in Section 6 in terms of local (lower) correlation and overall (higher) correlation, respectively.

2. Notation and Concepts

In this preparatory section, we briefly review from Han (1975) the basic notations and concepts which will be used later.

2.1. *Boolean Lattice of Random Variables*

Let the set of all natural numbers be denoted by $\Phi = \{1, 2, 3, \cdots\}$ and the direct product of n Φ's by Φ^n. We consider an n-dimensional random variable vector $X^n = (X_1, ..., X_n)$ on Φ^n and its probability distribution $p(X_1 = i_1, ..., X_n = i_n)$,[1] where $i^n = (i_1, ..., i_n) \in \Phi^n$. A ($k$-dimensional) marginal random vector $(X_{a_1}, ..., X_{a_k})$ $(1 \leqslant k \leqslant n)$ of X^n is specified by the subset $\{a_1, ..., a_k\}$ of the index set $D = \{1, ..., n\}$. The number of all distinct marginal random variable vectors of X^n is 2^n, where the zero-dimensional random variable corresponds to the empty subset $\varnothing$ of D and should be regarded as a random variable taking a constant value in Φ with probability 1, which will also be denoted by $\varnothing$.

[1] These values may be called parameters of probability of X^n.

We shall denote by $\mathbf{X}$ the set of all distinct marginal random vectors of X^n. Then, $\mathbf{X}$ is given the lattice structure isomorphic to the n-dimensional Boolean lattice formed of all distinct subsets of D with the set-inclusion relation (join $\vee$ and meet $\wedge$). Therefore, we shall call $\mathbf{X}$ the Boolean lattice of random variables for X^n. The singletons $\{X_a\}$'s are the "atoms" of the lattice $\mathbf{X}$. The elements of $\mathbf{X}$ will be denoted by Greek letters $\alpha, \beta, \gamma, \cdots$, etc. The maximum element $X^n = (X_1, ..., X_n)$ of $\mathbf{X}$ will be indicated by E, and the minimum element of $\mathbf{X}$ is $\varnothing$. The marginal distributions of X^n are given by

$$p(X_{a_1} = i_{a_1}, ..., X_{a_k} = i_{a_k}) = \sum_{i_{b_1}, \ldots, i_{b_{n-k}}} p(X^n = i^n), \tag{2.1}$$

where $\{b_1, ..., b_{n-k}\}$ is the complement of $\{a_1, ..., a_k\}$ in D. The support of a marginal distribution $p(X_{a_1} = i_{a_1}, ..., X_{a_k} = i_{a_k})$ is the set of all $(i_{a_1}, ..., i_{a_k})$'s such that $p(X_{a_1} = i_{a_1}, ..., X_{a_k} = i_{a_k}) \neq 0$, which will be denoted by $\mathbf{S}_p(X_{a_1}, ..., X_{a_k})$ or briefly by $\mathbf{S}(X_{a_1}, ..., X_{a_k})$. The set of all n-dimensional distributions with supports which are finite subsets of Φ^n will be called the set of *admissible* distributions, denoted by $\mathbf{P}^n$. The complement $\bar{\alpha}$ of $\alpha \, (\in \mathbf{X})$ is the element that satisfies the relation

$$\bar{\alpha} \vee \alpha = E, \qquad \bar{\alpha} \wedge \alpha = \varnothing. \tag{2.2}$$

The partial order induced from the lattice structure of X will be denoted by $\leqslant$: $\alpha \leqslant \beta \, (\alpha, \beta \in \mathbf{X})$ if and only if $\alpha \wedge \beta = \alpha$ or equivalently, $\alpha \vee \beta = \beta$. The rank $r(\alpha)$ of $\alpha \, (\in \mathbf{X})$ is the number of distinct atoms X_a's such that $X_a \leqslant \alpha$. For any $\alpha, \beta \, (\in \mathbf{X})$ such that $\alpha \leqslant \beta$, the interval $[\alpha, \beta]$ is defined as the set of all elements γ's such that $\alpha \leqslant \gamma \leqslant \beta$, where $r(\beta) - r(\alpha)$ is called the length of the interval $[\alpha, \beta]$. It is easily seen that the interval $[\alpha, \beta]$ forms an $(r(\beta) - r(\alpha))$-dimensional Boolean sublattice of $\mathbf{X}$.

2.2. *Entropy Space and Correlative Entropy Space*

When an admissible distribution $p(X^n = i^n)$ is given, we can assign to each element α of $\mathbf{X}$ a nonnegative real value $h(\alpha)$ of entropy in Shannon's sense:

$$h(\alpha) = - \sum_{i_{a_1}, \ldots, i_{a_k}} p(X_{a_1} = i_{a_1}, ..., X_{a_k} = i_{a_k}) \log p(X_{a_1} = i_{a_1}, ..., X_{a_k} = i_{a_k}), \tag{2.3}$$ [2]

where $\alpha = X_{a_1} \vee \cdots \vee X_{a_k}$.[3] In particular, for the zero-dimensional variable $\varnothing$, we put

$$h(\varnothing) = 0. \tag{2.4}$$

[2] More precisely, we should write the left-hand side as $h_p(\alpha)$ because it depends also on the probability distribution p under consideration.

[3] By definition, $X_{a_1} \vee \cdots \vee X_{a_k} = (X_{a_1}, ..., X_{a_k})$.

Since each $h(\alpha)$ may be regarded as a functional which makes a real value correspond to an admissible distribution, we can have the vector space generated by such functionals. We thus introduce the following concept.

DEFINITION 2.1. $\bar{H}(\mathbf{X})$ is the vector space generated by the set $\{h(\alpha) \mid \alpha \in \mathbf{X}\}$:

$$\bar{H}(\mathbf{X}) = \{c_1 h(\alpha_1) + \cdots + c_k h(\alpha_k)\}, \tag{2.5}$$ [4]

where k is an arbitrary positive integer; $\alpha_1, \ldots, \alpha_k$ are arbitrary elements of $\mathbf{X}$; and $c_1, \ldots, c_k$ are arbitrary constants. We shall call $\bar{H}(\mathbf{X})$ the entropy space for the lattice $\mathbf{X}$ and elements of $\bar{H}(\mathbf{X})$ are called entropy vectors.

The dimension of $\bar{H}(\mathbf{X})$ is $2^n - 1$, where the set $\mathbf{J} = \{h(\alpha) \mid r(\alpha) \geqslant 1\}$ is a basis of $\bar{H}(\mathbf{X})$.

DEFINITION 2.2. The entropy vectors $h_1, \ldots, h_k$ are said to be linearly independent if the relation

$$c_1 h_1 + \cdots + c_k h_k = 0 \tag{2.6}$$

implies $c_1 = 0, \ldots, c_k = 0$, and otherwise is said to be linearly dependent.

DEFINITION 2.3. An entropy vector of $\bar{H}(\mathbf{X})$ such that its value of entropy vanishes for every independent admissible distribution

$$p(X^n = i^n) = p(X_1 = i_1) \cdots p(X_n = i_n) \tag{2.7}$$

is called *correlative*. We shall denote by $\bar{H}_0(\mathbf{X})$ the set of all correlative entropy vectors. It is evident that $\bar{H}_0(\mathbf{X})$ forms a linear subspace of $\bar{H}(\mathbf{X})$ so that we shall call $\bar{H}_0(\mathbf{X})$ the correlative entropy space.

The dimension of $\bar{H}_0(\mathbf{X})$ is $2^n - (n + 1)$ (see Han, 1975).

2.3. *Correlativity, Nonnegativity, and Symmetricity*

As has been pointed out by a number of researchers (e.g., Watanabe, 1954, 1960, 1960; McGill, 1954; Garner, 1958; Garner and McGill, 1956), it is in many cases "effective" to measure correlation or interdependence in terms of information-theoretic quantities. These quantities are correlative in the sense of Definition 2.3. Conversely, every correlative entropy may be considered as measuring a certain kind of correlation because it vanishes for every independent distribution. In other words, given any correlative entropy $c_0 \in \bar{H}_0(\mathbf{X})$, a certain kind of independence relation $R(c_0)$ among a set of variables is associated and

[4] Obviously, every element of $\bar{H}(\mathbf{X})$ is a functional in the sense mentioned.

then the value of c_0 is to be interpreted as the quantitative measure of invalidity of $R(c_0)$. However, all correlative entropies may not necessarily be "effective" from the viewpoint of multivariate analysis. Thus, we introduce the concept of nonnegativity:

DEFINITION 2.4. An element h of $\bar{H}(\mathbf{X})$ is said to be *nonnegative* if

$$h \geqslant 0 \tag{2.8}$$

for every admissible distribution.

If an entropy c_0 is not only correlative but also nonnegative, as we show below, we can in many cases restate the above "independence relation $R(c_0)$" in terms of "*probabilistic* independence relation $PR(c_0)$." For instance, if we consider Shannon's mutual information $c_0 = I(X_1, X_2)$, $PR(c_0)$ is

$$p(X_1 = i_1, X_2 = i_2) = p(X_1 = i_1)\, p(X_2 = i_2). \tag{2.9}$$

In the following sections, we deal with the problem of determining the complete set of nonnegative entropies in $\bar{H}(\mathbf{X})$ and $\bar{H}_0(\mathbf{X})$ as well as that of interpreting each element of this set in terms of probabilistic independence relations. In doing so, we restrict ourselves to the class of "symmetric" entropies:

DEFINITION 2.5. An element h of $\bar{H}(\mathbf{X})$ is said to be *symmetric* if the form of h in $X_1, \ldots, X_n$ is invariant for every permutation among $X_1, \ldots, X_n$.

3. SYMMETRIC ENTROPY SPACE AND FUNDAMENTAL SYMMETRIC ENTROPIES

Let the set of all symmetric entropies of $\bar{H}(\mathbf{X})$ be denoted by $\bar{S}(\mathbf{X})$ and the set of all symmetric correlative entropies by $\bar{S}_0(\mathbf{X})$. Then, it is evident that $\bar{S}(\mathbf{X})$ and $\bar{S}_0(\mathbf{X})$ form linear subspaces of $\bar{H}(\mathbf{X})$ and $\bar{H}_0(\mathbf{X})$, respectively. Therefore, we shall call $\bar{S}(\mathbf{X})$ the symmetric entropy space, and $\bar{S}_0(\mathbf{X})$ the symmetric correlative entropy space.

DEFINITION 3.1. The fundamental symmetric entropy $\mathbf{e}_i^{(n)}$ is the sum of all distinct $h(\alpha)$'s over α's of rank i (Fig. 1):

$$\begin{aligned} \mathbf{e}_i^{(n)} &= \sum_{r(\alpha)=i} h(\alpha) \\ &= \sum_{(k_1,\ldots,k_i)} h(X_{k_1}, \ldots, X_{k_i}) \qquad (i = 0, 1, 2, \ldots, n). \end{aligned} \tag{3.1}$$

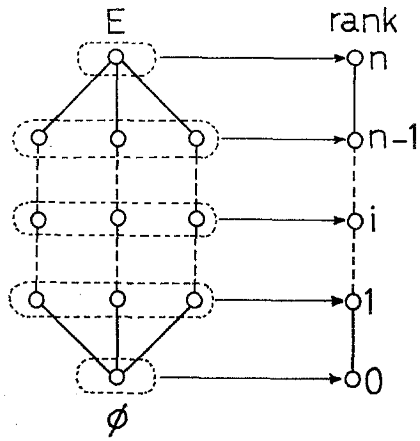

FIG. 1. Hierarchy of $\mathbf{e}_i^{(n)}$'s.

In particular, we have

$$\mathbf{e}_n^{(n)} = h(X_1, \ldots, X_n), \qquad \mathbf{e}_1^{(n)} = \sum_{a=1}^{n} h(X_a),$$

$$\mathbf{e}_0^{(n)} = \sum_{r(\alpha)=0} h(\alpha) = h(\varnothing) = 0.$$

LEMMA 3.1. *A necessary and sufficient condition for $h \in \bar{H}(\mathbf{X})$ to be symmetric is that h be expressed as a linear combination of* $\mathbf{e}_i^{(n)}$'s $(i = 1, \ldots, n)$.

Proof. Since the set $\mathbf{J} = \{h(\alpha) \mid r(\alpha) \geqslant 1\}$ is a basis of $\bar{H}(\mathbf{X})$, we can put

$$h = \sum_{r(\alpha)\geqslant 1} c_\alpha h(\alpha), \tag{3.2}$$

where c_α's are constants. Suppose that h is symmetric, then, by definition, the form of h is invariant for every permutation among $X_1, \ldots, X_n$. Summing Eq. (3.2) over all permutations, we have

$$(n!)h = \sum_{r(\alpha)\geqslant 1} \frac{n!}{\binom{n}{r(\alpha)}} c_\alpha \mathbf{e}_{r(\alpha)}^{(n)},$$

which proves the necessity. The sufficiency is obvious. Q.E.D.

THEOREM 3.1. *A necessary and sufficient condition for $h \in \bar{S}(X)$ to be correlative $(h \in \bar{S}_0(\mathbf{X}))$ is that*

$$\sum_{i=1}^{n} h_i \binom{n-1}{n-i} = 0, \tag{3.3}$$

where

$$h = \sum_{i=1}^{n} h_i \mathbf{e}_i^{(n)}. \tag{3.4}$$

Proof. For an independent distribution $p(X^n = i^n) = p(X_1 = i_1) \cdots p(X_n = i_n)$, we have

$$\begin{aligned} \mathbf{e}_i^{(n)} &= \frac{i}{n} \binom{n}{i} \sum_{a=1}^{n} h(X_a) \\ &= \binom{n-1}{n-i} \sum_{a=1}^{n} h(X_a). \end{aligned} \tag{3.5}$$

Hence, substituting (3.5) into (3.4),

$$h = \left\{ \sum_{i=1}^{n} h_i \binom{n-1}{n-i} \right\} \sum_{a=1}^{n} h(X_a).$$

Therefore, for h to vanish for every independent distribution, condition (3.3) is necessary and sufficient. Q.E.D.

EXAMPLE 3.1. Putting $h_j = 0$ $(j = 1, 2, \ldots, n-2)$ in (3.3), we have

$$h_n \binom{n-1}{0} + h_{n-1} \binom{n-1}{1} = 0.$$

Hence,

$$\begin{aligned} D_n &= \mathbf{e}_{n-1}^{(n)} - (n-1)\, \mathbf{e}_n^{(n)} \\ &= \sum_{a=1}^{n} h(X_1, \ldots, X_{a-1}, X_{a+1}, \ldots, X_n) - (n-1)\, h(X_1, \ldots, X_n) \end{aligned} \tag{3.6}$$

is a symmetric correlative entropy. This is called the *dual total correlation* (Han, 1975).

EXAMPLE 3.2. Putting $h_j = 0$ $(j = 2, 3, \ldots, n-1)$, we have

$$h_1 \binom{n-1}{n-1} + h_n \binom{n-1}{0} = 0.$$

Hence,

$$\begin{aligned} S_n &= \mathbf{e}_1^{(n)} - \mathbf{e}_n^{(n)} \\ &= \sum_{a=1}^{n} h(X_a) - h(X_1, \ldots, X_n) \end{aligned} \tag{3.7}$$

is a symmetric correlative entropy. This is the *total correlation* (Watanabe, 1960).

EXAMPLE 3.3. Let all the coefficients except for h_i and h_{i+1} be zero; then

$$h_i \binom{n-1}{n-i} + h_{i+1} \binom{n-1}{n-i-1} = 0.$$

Hence,

$$\Delta \mathbf{e}_i^{(n)} = \mathbf{e}_i^{(n)} \Big/ \left[i \binom{n}{i} \right] - \mathbf{e}_{i+1}^{(n)} \Big/ \left[(i+1) \binom{n}{i+1} \right] \tag{3.8}$$

is a symmetric correlative entropy ($i = 1,..., n-1$). In particular, putting $i = n-1$ in (3.8), we have

$$D_n = n(n-1)\, \Delta \mathbf{e}_{n-1}^{(n)} . \tag{3.9}$$

Since the number of the terms $h(X_{a_1},..., X_{a_i})$'s appearing in the defining equation of $\mathbf{e}_i^{(n)}$ is $\binom{n}{i}$, $\mathbf{e}_i^{(n)}/[i\binom{n}{i}]$ in the right-hand side of (3.8) is the averaged amount of entropies of i-dimensional marginal random vectors per variable. Therefore, $\Delta \mathbf{e}_i^{(n)}$ is the difference of the averaged entropies between the consecutive dimensions i and $i + 1$. The entropies $\Delta \mathbf{e}_i^{(n)}$'s ($i = 1,..., n-1$) are all nonnegative, as is shown below. The multivariate-analytic implication of the entropies will be given in Section 6.

THEOREM 3.2. *The dimension of $\bar{S}(\mathbf{X})$ is n, whereas the dimension of $\bar{S}_0(\mathbf{X})$ is $n - 1$. The set $\{\mathbf{e}_i^{(n)} \mid i = 1,..., n\}$ is a basis of $\bar{S}(\mathbf{X})$, and the set $\mathbf{J}_0 = \{\Delta \mathbf{e}_i^{(n)} \mid i = 1,..., n-1\}$ is a basis of $\bar{S}_0(\mathbf{X})$.*

Proof. Since, as was shown in the proof of Lemma 3.1, the set $\mathbf{J} = \{h(\alpha) \mid r(\alpha) \geqslant 1\}$ is a basis of $\bar{H}(\mathbf{X})$, $h(\alpha)$'s ($r(\alpha) \geqslant 1$) are linearly independent. Then, it follows from (3.1) that $\mathbf{e}_i^{(n)}$'s are linearly independent. Therefore, we can establish the result for $\bar{S}(\mathbf{X})$ by virtue of Lemma 3.1. Similarly for $\bar{S}_0(\mathbf{X})$.
Q.E.D.

Let $\mathbf{B}_2$ denote the set of all intervals $[\sigma, \alpha]$'s of length 2 of $\mathbf{X}$ and consider the entropy function $h([\sigma, \alpha])$ on $\mathbf{B}_2$ as defined by

$$h([\sigma, \alpha]) = -h(\alpha) + h(\sigma \vee X_a) + h(\sigma \vee X_b) - h(\sigma), \tag{3.10}$$

where $\alpha = \sigma \vee X_a \vee X_b$ and $r(\alpha) - r(\sigma) = 2$ (Fig. 2). The right-hand side of (3.10) is Shannon's conditional mutual information between X_a and X_b given σ. Therefore, every Shannon's conditional mutual information between two atomic variables is in one-to-one correspondence with an interval of length 2. Obviously, $h([\sigma, \alpha])$ is a nonnegative correlative entropy which vanishes if and only if X_a and X_b are conditionally independent given σ. In order to describe the nonnegativity of elements in $\bar{S}_0(\mathbf{X})$ in terms of fundamental symmetric entropies, we shall show here the relation between $\Delta \mathbf{e}_i^{(n)}$'s and $h([\sigma, \alpha])$'s.

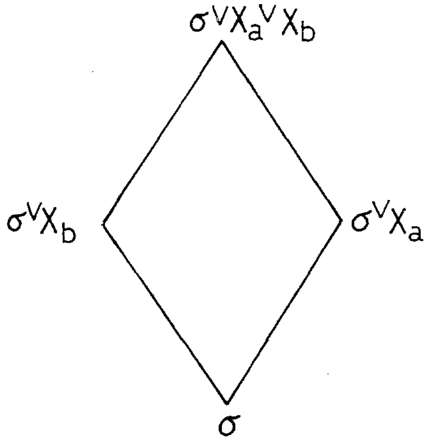

FIG. 2. Interval of length 2.

Putting

$$\Delta^2 \mathbf{e}_i^{(n)} = (i+1)\,\Delta \mathbf{e}_i^{(n)} - (i-1)\,\Delta \mathbf{e}_{i-1}^{(n)} \qquad (i = 1,\dots, n-1), \tag{3.11}$$

we have the following lemma.

LEMMA 3.2. *For $i = 1,\dots, n-1$,*

$$\Delta^2 \mathbf{e}_i^{(n)} = \left(\binom{n}{i+1}\binom{i+1}{2}\right)^{-1} \sum_{\substack{[\sigma,\alpha]\in \mathbf{B}_2 \\ r(\alpha)=i+1}} h([\sigma, \alpha]), \tag{3.12}$$

where the sum in the right-hand side of (3.12) *is taken over all the intervals of length* 2 *with rank* $r(\alpha) = i+1$.

Proof. From (3.1), (3.8), (3.10), and (3.11),

$$\sum_{\substack{[\sigma,\alpha]\in \mathbf{B}_2 \\ r(\alpha)=i+1}} h([\sigma, \alpha]) = -\frac{i(i+1)}{2}\,\mathbf{e}_{i+1}^{(n)} + i(n-i)\,\mathbf{e}_i^{(n)} - \frac{(n-i)(n-i+1)}{2}\,\mathbf{e}_{i-1}^{(n)}$$

$$= \binom{n}{i+1}\binom{i+1}{2}\,\Delta^2 \mathbf{e}_i^{(n)}. \qquad \text{Q.E.D.}$$

Since the number of all intervals of length 2 with $r(\alpha) = i+1$ is

$$\binom{n}{i+1}\binom{i+1}{2},$$

$\Delta^2 \mathbf{e}_i^{(n)}$ is the averaged amount of all Shannon's (conditional) mutual informations per interval with the maximum element of rank $i + 1$. From this lemma, the nonnegativity

$$\Delta^2 \mathbf{e}_i^{(n)} \geqslant 0 \qquad (i = 1,\dots, n-1) \tag{3.13}$$

follows.

THEOREM 3.3. *The symmetric correlative entropies $\Delta \mathbf{e}_k^{(n)}$'s $(k = 1,\dots, n-1)$ can be expressed as linear combinations of $\Delta^2 \mathbf{e}_i^{(n)}$'s with nonnegative coefficients*:

$$\Delta \mathbf{e}_k^{(n)} = \frac{1}{k(k+1)} \sum_{i=1}^{k} i \Delta^2 \mathbf{e}_i^{(n)}. \tag{3.14}$$

Proof. By multiplying both sides of (3.11) by i and summing it from $i = 1$ to $i = k$, (3.14) results. Q.E.D.

From this theorem, we have

$$\Delta \mathbf{e}_k^{(n)} \geqslant 0 \qquad (k = 1,\dots, n-1). \tag{3.15}$$

Inequalities quite analogous to (3.13) and (3.15) have already been obtained for a "non-symmetric" case by Watanabe (1960) in analyzing the range of probabilistic influence between variables forming a stationary random process. The inequalities here are essentially the symmetric counterparts of Watanabe's. It should be noted that either of these inequalities has been derived using only the nonnegativity of Shannon's informations. This implies that the multivariate-analytic interpretation can be made based on that for Shannon's informations.

COROLLARY 3.1. *The set $\{\Delta^2 \mathbf{e}_i^{(n)} \mid i = 1,\dots, n-1\}$ is a basis of $\bar{S}_0(\mathbf{X})$.*

Proof. Obvious from Theorem 3.2 and relations (3.11), (3.14). Q.E.D.

4. NONNEGATIVITY OF SYMMETRIC ENTROPIES

4.1. *The Necessary and Sufficient Condition of Nonnegativity*

In this section, we describe the main result concerning the nonnegativity property of symmetric correlative entropies, i.e., the necessary and sufficient condition for an element of $\bar{S}_0(\mathbf{X})$ to be nonnegative.

LEMMA 4.1. *Let k be a positive integer $(1 \leqslant k \leqslant n)$. Then, there exists an admissible distribution p_k which satisfies the following conditions.*

(1) *All k variables taken from $\{X_1, \dots, X_n\}$ are independent,*

(2) *The values of all $(n-k)$ variables are uniquely determined when the values of the remaining k variables are fixed.*

Proof. It is sufficient to show that a required distribution exists among the admissible distributions such that each $X_1 ,..., X_n$ has the support $\Phi_0 = \{1,..., N\}$ (N is a prime number to be determined below) and such that

$$\begin{aligned} p(X_1 = i_1 ,..., X_n = i_n) &= 1/N^k \qquad \text{if} \quad (i_1 ,..., i_n) \in Q_0 , \\ &= 0 \qquad\quad \text{otherwise,} \end{aligned} \tag{4.1}$$

where Q_0 is a subset of $\Phi_0{}^n$ with the cardinal number N^k.

Let $K(N)$ be a Galois field induced from the integer ring $\{0, \pm 1, \pm 2,...\}$ in modulo N. Let distinct (column) vectors in the k-dimensional vector space V_k over $K(N)$ be denoted by $\mathbf{a}_1 ,..., \mathbf{a}_M$ ($M = N^k$), and let the ($M \times k$)-matrix A be defined by

$$A = (\mathbf{a}_1 ,..., \mathbf{a}_M)^t. \tag{4.2}$$

Furthermore, let the ($k \times n$)-matrix B be defined by

$$B = \begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 2^1 & \cdots & (n-1) \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdot & \cdot & & \cdot \\ 1 & 1 & 2^{k-1} & \cdots & (n-1)^{k-1} \end{pmatrix}. \tag{4.3}$$

We now define Q_0 as the set such that $(i_1 ,..., i_n) \in Q_0$ if and only if $(i_1 ,..., i_n)$ is a row vector of $C = AB$ (C is an ($M \times n$)-matrix). The distribution (4.1) with the support so defined is one that satisfies the condition of the lemma. In fact, let $\alpha = (X_{a_1} ,..., X_{a_k})$ be an arbitrary k-dimensional marginal variable of X^n. Accordingly, let the submatrix of C formed of $(1,..., M)$-rows and $(a_1 ,..., a_k)$-columns be denoted by $C\alpha$, and the submatrix of B formed of $(1,..., k)$-rows and $(a_1 ,..., a_k)$-columns by B_α. Then, we have

$$C_\alpha = AB_\alpha . \tag{4.4}$$

Since B_α is a van der Monde's matrix, its determinant is

$$1^{l_1} 2^{l_2} \cdots (n-2)^{l_{n-1}} \tag{4.5}$$

to within a sign. This is not zero in modulo N if N is chosen so that $N \geqslant n - 1$. Hence, B_α is nonsingular. Therefore, the correspondence between the set of row vectors of C_α and the set of row vectors of A is one-to-one. Thus, every $(i_{a_1} ,..., i_{a_k}) \in \Phi_0{}^k$ appears just one time as a row vector in C_α; i.e., $X_{a_1} ,..., X_{a_k}$ are independent. Suppose that the values of $(X_{a_1} ,..., X_{a_k})$ are arbitrarily fixed, say, to $(i_{a_1} ,..., i_{a_k})$. Let the row vector of A associated with the row vector $(i_{a_1} ,..., i_{a_k})$ of C_α by the one-to-one correspondence (4.4) be denoted by

$(j_{a_1}, ..., j_{a_k})$. Then, the value of $(X_1, ..., X_n)$ is uniquely determined by the correspondence $C = AB$, i.e.,

$$(s_1, ..., s_n) = (j_{a_1}, ..., j_{a_k})B, \tag{4.6}$$

where $(s_{a_1}, ..., s_{a_k}) = (i_{a_1}, ..., i_{a_k})$. Q.E.D.

LEMMA 4.2. *Let k be a positive integer* $(1 \leqslant k \leqslant n-1)$. *Then, there exists an admissible distribution under which*

$$\begin{aligned} \Delta^2 \mathbf{e}_i^{(n)} &= h_0 \quad \text{if} \quad i = k, \\ &= 0 \quad \text{otherwise,} \end{aligned}$$

where h_0 is a positive constant.

Proof. We show that the distribution p_k specified in Lemma 4.1 satisfies the above requirement. Let $[\sigma, \alpha]$ be an interval of length 2 with $r(\alpha) = k+1$, and let $\alpha = \sigma \vee X_a \vee X_b$ (Fig. 3). Then, we have

$$\begin{aligned} h([\sigma, \alpha]) &= -h(\sigma \vee X_a \vee X_b) + h(\sigma \vee X_a) + h(\sigma \vee X_b) - h(\sigma) \\ &= h(\sigma \vee X_a) - h(\sigma) = \log N^k - \log N^{k-1} = \log N. \end{aligned} \tag{4.7}$$

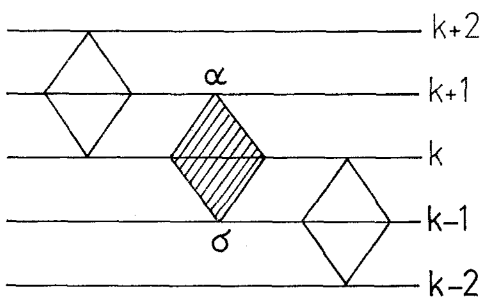

FIG. 3. Hierarchy of intervals of length 2.

On the other hand, it is evident that

$$h([\sigma, \alpha]) = 0 \qquad \text{for } r(\alpha) \neq k+1. \tag{4.8}$$

Then, (3.12), (4.7), and (4.8) establish the required result. Q.E.D.

THEOREM 4.1 (fundamental theorem 1). *A necessary and sufficient condition for an element h of $\bar{S}_0(\mathbf{X})$,*

$$h = \sum_{i=1}^{n-1} a_i \Delta^2 \mathbf{e}_i^{(n)}, \tag{4.9}$$

to be nonnegative is that

$$a_i \geqslant 0 \qquad (i = 1,..., n - 1). \tag{4.10}$$

Proof. Since the set $\{\Delta^2 \mathbf{e}_i^{(n)} \mid i = 1,..., n - 1\}$ is a basis of $\bar{S}_0(\mathbf{X})$ (Corollary 3.1), we can put h as (4.9). Then, (4.10) results from Lemma 4.2. Q.E.D.

COROLLARY 4.1. *The entropy $\Delta^2 \mathbf{e}_k^{(n)}$ $(k = 1,..., n - 1)$ cannot be expressed as a sum of two linearly independent nonnegative entropies of $\bar{S}_0(\mathbf{X})$.*

Proof. Suppose that, for two nonnegative entropies $\mathbf{v}_1$ and $\mathbf{v}_2$ of $\bar{S}_0(\mathbf{X})$,

$$\Delta^2 \mathbf{e}_k^{(n)} = \mathbf{v}_1 + \mathbf{v}_2 , \tag{4.11}$$

where

$$\mathbf{v}_1 = \sum_{i=1}^{n-1} a_{1i} \Delta^2 \mathbf{e}_i^{(n)},$$

$$\mathbf{v}_2 = \sum_{i=1}^{n-1} a_{2i} \Delta^2 \mathbf{e}_i^{(n)}.$$

Then, by comparing the coefficients in both sides of (4.11), we have

$$a_{1i} + a_{2i} = 0 \qquad (i \neq k). \tag{4.12}$$

Hence, by virtue of Theorem 4.1 and (4.12),

$$a_{1i} = a_{2i} = 0 \qquad (i \neq k),$$

which implies that $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly dependent. Q.E.D.

4.2. *Entropy Measures of Multivariate Symmetric Correlations*

We shall here give the multivariate-analytic interpretation for the nonnegative entropies thus far derived as well as that for the necessary and sufficient condition (4.10). In view of Lemma 3.2, the entropy $\Delta^2 \mathbf{e}_i^{(n)}$ is a nonnegative "symmetric" entropy and it vanishes if and only if every two atomic variables X_a , X_b are conditionally independent given the values of other arbitrary $(i - 1)$ variables. Therefore, $\Delta^2 \mathbf{e}_i^{(n)}$ can be used as an information-theoretic measure of such a multivariate "symmetric" correlation. In other words, $\Delta^2 \mathbf{e}_i^{(n)}$ is associated with the conjunction H_i of probabilistic conditional independence relations

$$H_i = \bigcap_{X_a, X_b, \sigma} [P_\sigma(X_a , X_b) = p_\sigma(X_a)\, p_\sigma(X_b)], \tag{4.13}$$

where $P_\sigma(X_a, X_b)$ is the conditional probability of (X_a, X_b) given σ; X_a, X_b run over all pairs of atomic variables and σ over all $(i-1)$-dimensional random variables. Then, the value of $\Delta^2\mathbf{e}_i^{(n)}$ may be considered as measuring the invalidity of the hypothesis H_i.

According to Theorem 4.1, every nonnegative entropy in $\bar{S}_0(\mathbf{X})$ needs to be expressed as a linear combination of $\Delta^2\mathbf{e}_i^{(n)}$'s with nonnegative coefficients. This implies that multivariate "symmetric" correlations which can be expressed in terms of information-theoretic measures are necessarily confined to the class of those specified by conjunctions of a certain number of H_i's. More precisely, let h_0 be an arbitrary nonnegative element of $\bar{S}_0(\mathbf{X})$ and put

$$h_0 = \sum_{i=1}^{n-1} a_i \Delta^2\mathbf{e}_i^{(n)} \qquad (a_i \geqslant 0); \tag{4.14}$$

then h_0 vanishes if and only if the independence relations $H_{i_1}, \ldots, H_{i_k}$ are all valid where $i_1, \ldots, i_k$ are indices such that $a_{i_1}, \ldots, a_{i_k}$ are positive coefficients in (4.14). Here, the values of these positive coefficients designate the weights specifying which parts of the H_i's to emphasize. Hence, we obtain various multivariate correlation measures according to the method of assigning the weights.

Thus, the set $\{\Delta^2\mathbf{e}_1^{(n)}, \ldots, \Delta^2\mathbf{e}_{n-1}^{(n)}\}$ forms the measuring basis of possible "symmetric" correlations from the information-theoretic point of view. We say that $\Delta^2\mathbf{e}_i^{(n)}$ is a measure at *higher* level than $\Delta^2\mathbf{e}_j^{(n)}$ when $i > j$. In this connection, the measuring basis $\{\Delta^2\mathbf{e}_1^{(n)}, \ldots, \Delta^2\mathbf{e}_{n-1}^{(n)}\}$ is endowed with a "hierarchical structure" of correlation analysis from the lowest $\Delta^2\mathbf{e}_1^{(n)}$ to the highest $\Delta^2\mathbf{e}_{n-1}^{(n)}$. The conventional correlation analysis based on dispersion matrix (nonconditional pairwise correlation) coincides in the symmetric case with that using the information carried only by the lowest $\Delta^2\mathbf{e}_1^{(n)}$.

According to Corollary 4.1, the set of all nonnegative entropies in $\bar{S}_0(\mathbf{X})$ forms a convex cone with $k\,\Delta^2\mathbf{e}_i^{(n)}$'s $(k \geqslant 0)$ as its extreme lines. Since no $\Delta^2\mathbf{e}_i^{(n)}$ can be decomposed into two linearly independent nonnegative entropies, $\Delta^2\mathbf{e}_i^{(n)}$'s express the *minimal* correlations or *minimal* independence relations which cannot be decomposed into more fundamental ones. Therefore, the above hierarchical structure is "complete."

EXAMPLE 4.1. Consider the three-dimensional case $n = 3$. The measuring basis is

$$\begin{aligned}
\Delta^2\mathbf{e}_1^{(3)} &= 2\Delta\mathbf{e}_1^{(3)} \\
&= \tfrac{2}{3}[h(X_1) + h(X_2) + h(X_3)] - [h(X_1, X_2) + h(X_2, X_3) + h(X_3, X_1)], \\
\Delta^2\mathbf{e}_2^{(3)} &= 3\Delta\mathbf{e}_2^{(3)} - \Delta\mathbf{e}_1^{(3)} \\
&= -\tfrac{1}{3}[h(X_1) + h(X_2) + h(X_3)] - h(X_1, X_2, X_3) \\
&\quad + \tfrac{2}{3}[h(X_1, X_2) + h(X_2, X_3) + h(X_3, X_1)].
\end{aligned}$$

(1) Put $a_1 = 1$, $a_2 = 2$; then

$$h_0 = \Delta^2 \mathbf{e}_1^{(3)} + 2\Delta^2 \mathbf{e}_2^{(3)} = 6\Delta \mathbf{e}_2^{(3)}$$
$$= h(X_1, X_2) + h(X_2, X_3) + h(X_3, X_1) - 2h(X_1, X_2, X_3) = D_3. \quad (4.15)$$

(2) Put $a_1 = 2$, $a_2 = 1$; then

$$h_0 = 2\Delta^2 \mathbf{e}_1^{(3)} + \Delta^2 \mathbf{e}_2^{(3)}$$
$$= h(X_1) + h(X_2) + h(X_3) - h(X_1, X_2, X_3) = S_3. \quad (4.16)$$

(3) Put $a_1 = a_2 = 1$; then

$$\begin{aligned} h_0 &= \Delta^2 \mathbf{e}_1^{(3)} + \Delta^2 \mathbf{e}_2^{(3)} \\ &= \tfrac{1}{3}[h(X_1) + h(X_2) + h(X_3)] - h(X_1, X_2, X_3) \\ &\quad + \tfrac{1}{3}[h(X_1, X_2) + h(X_2, X_3) + h(X_3, X_1)]. \end{aligned} \quad (4.17)$$

Equations (4.15), (4.16), and (4.17) all vanish if and only if X_1, X_2, X_3 are independent, so that we may use any one of them to measure "total" correlation among X_1, X_2, X_3 (see Sect. 5). However, D_n emphasizes rather *conditional* correlations whereas S_n emphasizes rather *nonconditional* correlations, and (4.17) has an intermediate character.

EXAMPLE 4.2. Consider the same case as above.

(1) Put $a_1 = 1$, $a_2 = 0$; then $h_0 = \Delta^2 \mathbf{e}_1^{(3)}$. This measures only nonconditional correlations, omitting conditional correlations.

(2) Put $a_1 = 0$, $a_2 = 1$; then $h^0 = \Delta^2 \mathbf{e}_2^{(3)}$. This measures only conditional correlations, omitting nonconditional correlations. In both cases, only "partial" correlations are measured.

5. Dual Measures of Symmetric Correlations

Han (1975) showed in analyzing the linear dependence structure of $\bar{H}(\mathbf{X})$ and many of the results obtained for the lattice $\mathbf{X}$ have their dual counterparts which are described in terms of the dual lattice $\mathbf{X}^*$ of $\mathbf{X}$. Such a dual correspondence is also the case for $\bar{S}(\mathbf{X})$ as well as for $\bar{S}_0(\mathbf{X})$.

DEFINITION 5.1. Let β and γ be arbitrary elements of $\mathbf{X}$ such that $\beta \wedge \gamma = \varnothing$ and let $h_\beta(\gamma)$ denote Shannon's conditional entropy of γ given β. The dual fundamental symmetric entropy $\mathbf{c}_i^{(n)}$ is the sum of all distinct $h_{\bar{\alpha}}(\alpha)$'s over α's of rank i, where $\bar{\alpha}$ is the complement of α in the lattice $\mathbf{X}$:

$$\begin{aligned}\mathbf{c}_i^{(n)} &= \sum_{r(\alpha)=i} h_{\bar{\alpha}}(\alpha)\\ &= \sum_{(k_1,\ldots,k_i)} h_{X_{l_1}\ldots X_{l_{n-i}}}(X_{k_1},\ldots, X_{k_i}) \qquad (i = 0, 1,\ldots, n),\end{aligned} \tag{5.1}$$

where $\{l_1,\ldots, l_{n-i}\} = \{1,\ldots, n\} - \{k_1,\ldots, k_i\}$. In particular, we have

$$\mathbf{c}_n^{(n)} = h(X_1,\ldots, X_n), \qquad \mathbf{c}_0^{(n)} = 0,$$

$$\mathbf{c}_{n-1}^{(n)} = \sum_{a=1}^{n} h_{X_a}(X_1,\ldots, X_{a-1}, X_{a+1},\ldots, X_n).$$

The duality between $\mathbf{e}_i^{(n)}$'s and $\mathbf{c}_i^{(n)}$'s is seen by writing them as

$$\mathbf{e}_i^{(n)} = \sum_{r(\alpha)=i} \{h(\alpha) - h(\varnothing)\}, \tag{5.2}$$

$$\mathbf{c}_i^{(n)} = \sum_{r(\bar{\alpha})=n-i} \{h(E) - h(\bar{\alpha})\}, \tag{5.3}$$

in which the duality is realized through the "complement operation": $\alpha \leftrightarrow \bar{\alpha}$ (particularly, $\varnothing \leftrightarrow E$). Rewriting these relations, we have

$$\mathbf{c}_i^{(n)} = \binom{n}{i} \mathbf{e}_n^{(n)} - \mathbf{e}_{n-i}^{(n)} \qquad (i = 0, 1,\ldots, n), \tag{5.4}$$

$$\mathbf{e}_i^{(n)} = \binom{n}{i} \mathbf{c}_n^{(n)} - \mathbf{c}_{n-i}^{(n)} \qquad (i = 0, 1,\ldots, n). \tag{5.5}$$

The dual counterpart of Theorem 3.1 is given in terms of dual fundamental symmetric entropies:

THEOREM 5.1. *A necessary and sufficient condition for $h \in \bar{S}(\mathbf{X})$ to be correlative ($h \in \bar{S}_0(\mathbf{X})$) is that*

$$\sum_{i=1}^{n} s_i \binom{n-1}{n-i} = 0, \tag{5.6}$$

where

$$h = \sum_{i=1}^{n} s_i \mathbf{c}_i^{(n)}. \tag{5.7}$$

Proof. The set $\{\mathbf{c}_1^{(n)},\ldots, \mathbf{c}_n^{(n)}\}$ is a basis of $\bar{S}(\mathbf{X})$ because of Theorem 3.2 and relation (5.5). Hence, expression (5.7) is validated; then, the proof is quite similar to that of Theorem 3.1. Q.E.D.

EXAMPLE 5.1. Putting $s_j = 0$ $(j = 1, 2, \ldots, n-2)$ in (5.6), we have the symmetric correlative entropy (total correlation; see Example 3.2)

$$\begin{aligned} S_n &= (n-1)\,\mathbf{c}_n^{(n)} - \mathbf{c}_{n-1}^{(n)} \\ &= \sum_{a=1}^{n} h(X_a) - h(X_1, \ldots, X_n). \end{aligned} \tag{5.8}$$

EXAMPLE 5.2. Putting $s_j = 0$ $(j = 2, 3, \ldots, n-1)$, we have the symmetric correlative entropy (dual total correlation; see Example 3.2)

$$\begin{aligned} D_n &= \mathbf{c}_n^{(n)} - \mathbf{c}_1^{(n)} \\ &= \sum_{a=1}^{n} h(X_1, \ldots, X_{a-1}, X_{a+1}, \ldots, X_n) - (n-1)\, h(X_1, \ldots, X_n). \end{aligned} \tag{5.9}$$

EXAMPLE 5.3. Letting all the coefficients except for s_i and s_{i+1} be zero, we have the symmetric correlative entropy

$$\Delta \mathbf{c}_i^{(n)} = \mathbf{c}_{i+1}^{(n)} \Big/ \left[(i+1)\binom{n}{i+1}\right] - \mathbf{c}_i^{(n)} \Big/ \left[i\binom{n}{i}\right] \tag{5.10}$$

for $i = 1, \ldots, n-1$. In particular, putting $i = n-1$ in (5.10), we have

$$S_n = n(n-1)\, \Delta \mathbf{c}_{n-1}^{(n)}. \tag{5.11}$$

The entropy $\mathbf{c}_i^{(n)}/[i\binom{n}{i}]$ is the averaged amount of *conditional* entropies of i-dimensional random vectors per variable so that $\Delta \mathbf{c}_i^{(n)}$ is the difference of the averaged conditional entropies.

COROLLARY 5.1. *The set* $\{\Delta \mathbf{c}_i^{(n)} \mid i = 1, \ldots, n-1\}$ *is a basis of* $\bar{S}_0(\mathbf{X})$, *and the set* $\{\mathbf{c}_i^{(n)} \mid i = 1, \ldots, n\}$ *is a basis of* $\bar{S}(\mathbf{X})$.

In parallel with (3.11), putting, for $i = 1, \ldots, n-1$,

$$\Delta^2 \mathbf{c}_i^{(n)} = (i+1)\, \Delta \mathbf{c}_i^{(n)} - (i-1)\, \Delta \mathbf{c}_{i-1}^{(n)}, \tag{5.12}$$

we have the dual counterpart of Theorem 3.3:

THEOREM 5.2. *The symmetric correlative entropies* $\Delta \mathbf{c}_k^{(n)}$*'s* $(k = 1, \ldots, n-1)$ *are expressed as linear combinations of* $\Delta^2 \mathbf{c}_i^{(n)}$*'s with nonnegative coefficients*:

$$\Delta \mathbf{c}_k^{(n)} = \frac{1}{k(k+1)} \sum_{i=1}^{k} i \Delta^2 \mathbf{c}_i^{(n)}. \tag{5.13}$$

Proof. Obvious from relation (5.12). Q.E.D.

Expanding $\Delta^2 \mathbf{c}_i^{(n)}$ in terms of $\mathbf{e}_j^{(n)}$'s by using relations (5.12), (5.10) and (5.4), we have the important equality

$$\Delta^2 \mathbf{c}_i^{(n)} = \Delta^2 \mathbf{e}_{n-i}^{(n)} \qquad (i = 1,..., n-1). \tag{5.14}$$

This shows a deep duality between the fundamental entropies and the dual fundamental entropies.

From (5.14), it immediately follows that

$$\Delta^2 \mathbf{c}_i^{(n)} \geqslant 0, \tag{5.15}$$

$$\Delta \mathbf{c}_i^{(n)} \geqslant 0. \tag{5.16}$$

Since the set $\{\Delta^2 \mathbf{c}_i^{(n)} \mid i = 1,..., n-1\} = \{\Delta^2 \mathbf{e}_i^{(n)} \mid i = 1,..., n-1\}$ is a basis of $\bar{S}_0(\mathbf{X})$ and $\mathbf{c}_1^{(n)}$ is not linearly dependent on this set, we see that the set $\{\mathbf{c}_1^{(n)}\} \cup \{\Delta^2 \mathbf{e}_i^{(n)} \mid i = 1,..., n-1\}$ is a basis of $\bar{S}(\mathbf{X})$. Then, we obtain, as an extension of Theorem 4.1, the necessary and sufficient condition for an element of $\bar{S}(\mathbf{X})$ to be nonnegative:

THEOREM 5.3 (fundamental theorem 2). *A necessary and sufficient condition for an element h of $\bar{S}(\mathbf{X})$,*

$$h = a_0 \mathbf{c}_1^{(n)} + \sum_{i=1}^{n-1} a_i \Delta^2 \mathbf{e}_i^{(n)}, \tag{5.17}$$

to be nonnegative is that

$$a_0 \geqslant 0, \qquad a_i \geqslant 0 \qquad (i = 1,..., n-1). \tag{5.18}$$

Proof. Choose an independent distribution with the support of which the cardinal number is more than one. Then, $\Delta^2 \mathbf{e}_i^{(n)}$'s vanish and hence the condition $a_0 \geqslant 0$ follows from the nonnegativity. The remaining part in (5.18) is obtained by noting that $\mathbf{c}_1^{(n)}$ vanishes under every distribution p_k in Lemma 4.1. Q.E.D.

6. Total Correlation and Dual Total Correlation

We now show the multivariate-analytic interpretation for $\Delta \mathbf{e}_i^{(n)}$'s and $\Delta \mathbf{c}_i^{(n)}$'s and thereby clarify the differences and similarities between the total correlation S_n and the dual total correlation D_n as entropy measures of multivariate symmetric correlation.

THEOREM 6.1. *The nonnegative entropy $\Delta \mathbf{e}_i^{(n)}$ $(i = 1,..., n-1)$ vanishes if and only if all $(i+1)$ variables taken from $\{X_1,..., X_n\}$ are independent.*

Proof. The condition $\Delta\mathbf{e}_i^{(n)} = 0$ is, by Theorem 3.3, equivalent to the following i conditions:

$$\Delta^2\mathbf{e}_k^{(n)} = 0 \qquad (k = 1,\ldots, i). \tag{6.1}$$

Among these conditions the first $\Delta^2\mathbf{e}_1^{(n)} = 0$ implies the pairwise independence

$$p(X_a = i_a\,, X_b = i_b) = p(X_a = i_a)\, p(X_b = i_b) \tag{6.2}$$

for every pair $(X_a\,, X_b)$. Likewise, the second $\Delta^2\mathbf{e}_2^{(n)} = 0$ implies the pairwise conditional independence

$$\begin{aligned} &p(X_a = i_a\,, X_b = i_b\,, X_c = i_c)\, p(X_c = i_c) \\ &\quad = p(X_a = i_a\,, X_c = i_c)\, p(X_b = i_b\,, X_c = i_c) \end{aligned} \tag{6.3}$$

for every triple $(X_a\,, X_b\,, X_c)$, and so on. Condition (6.3) together with (6.2) yields the condition of triplewise independence

$$p(X_a = i_a\,, X_b = i_b\,, X_c = i_c) = p(X_a = i_a)\, p(X_b = i_b)\, p(X_c = i_c).$$

Repeating this procedure i times, we have the required result. Q.E.D.

The following theorem is the dual counterpart of Theorem 6.1.

THEOREM 6.2. *The nonnegative entropy $\Delta\mathbf{c}_i^{(n)}$ $(i = 1,\ldots, n-1)$ vanishes if and only if all $(i+1)$ variables taken from $\{X_1\,,\ldots, X_n\}$ are conditionally independent given the values of the remaining $(n-i-1)$ variables.*

COROLLARY 6.1. *The total correlation S_n as well as the dual total correlation D_n vanishes if and only if $X_1\,,\ldots, X_n$ are independent.*

Proof. It follows from (3.9), (5.11).

The weights of $\Delta\mathbf{e}_i^{(n)}$ and $\Delta\mathbf{c}_i^{(n)}$ with respect to $\Delta^2\mathbf{e}_j^{(n)}$ are shown in Fig. 4. In particular, the weights of S_n and D_n are shown in Fig. 5. We see from these figures that the sequence of measures $\Delta\mathbf{e}_i^{(n)}$'s, D_n emphasizes *higher* correlations; on the other hand, the sequence of dual measures $\Delta\mathbf{c}_i^{(n)}$'s, S_n emphasizes *lower* correlations. Moreover, the correlations higher than the ith level are completely omitted from $\Delta\mathbf{e}_i^{(n)}$, and the correlations lower than the $(n-i)$th level are completely omitted from $\Delta\mathbf{c}_i^{(n)}$.

EXAMPLE 6.1. We consider the case $n = 3$. Let $p_0{}^*$ and p_0 be two probability

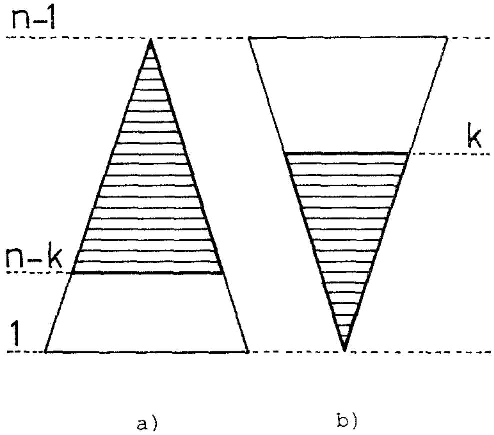

FIG. 4. Weights of $\Delta \mathbf{e}_k^{(n)}$ and $\Delta \mathbf{c}_k^{(n)}$. (a) Weights of $\Delta \mathbf{c}_k^{(n)}$. (b) Weights of $\Delta \mathbf{e}_k^{(n)}$.

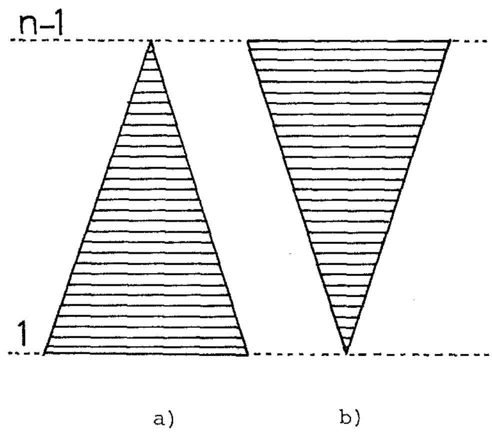

FIG. 5. Weights of S_n and D_n. (a) Weights of S_n. (b) Weights of D_n.

distributions of random variables X_1, X_2, X_3 with the supports $\mathbf{S}(X_1) = \mathbf{S}(X_2) = \mathbf{S}(X_3) = \{1, 2\}$:

$$
\begin{aligned}
p_0^*(X_1 = i_1, X_2 = i_2, X_3 = i_3) &= \tfrac{1}{2} \qquad \text{if} \quad i_1 = i_2 = i_3, \\
&= 0 \qquad \text{otherwise,}
\end{aligned}
$$

and

$$
\begin{aligned}
p_0(X_1 = i_1, X_2 = i_2, X_3 = i_3) &= \tfrac{1}{4} \qquad \text{if} \quad i_1 + i_2 + i_3 = \text{even}, \\
&= 0 \qquad \text{otherwise.}
\end{aligned}
$$

These distributions are illustrated in Fig. 6. Then, we have $S_3 = 2 \log 2 > \log 2 = D_3$ for p_0^*, and $S_3 = \log 2 < 2 \log 2 = D_3$ for p_0. Hence, the values of S_3 and D_3 are not in a definite order. If we are to measure "total" correlation in reference to S_3, p_0^* should be regarded as specifying a stronger correlation or interdependence among X_1, X_2, X_3 than p_0. On the other hand, if we are to measure "total" correlation in reference to D_3, p_0^* should be regarded as specifying weaker correlation or interdependence than p_0. Remark that, for p_0^*, X_1, X_2, X_3 are identical random variables so that the structure of depen-

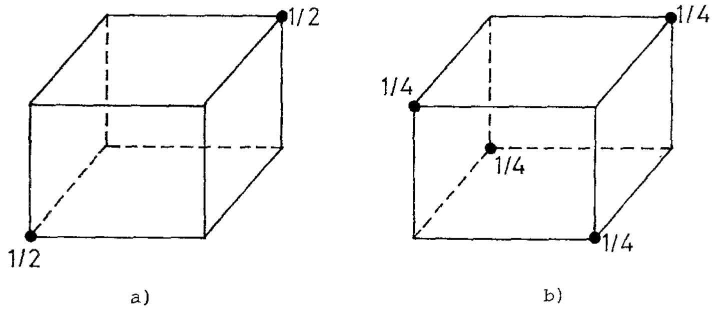

FIG. 6. Configuration of probability distributions. (a) p_0^*. (b) p_0.

dence can be described in terms of pairwise relations alone: X_1 and X_2 are identical, X_2 and X_3 are identical; on the other hand, for p_0, X_1, X_2, X_3 are pairwise independent and so the structure of dependence cannot be described without using triplewise relations. Summarizing, we may say that S_3 emphasizes local (lower) correlations and D_3 emphasizes overall (higher) correlations.

EXAMPLE 6.2. Let p_0^* and p_0 be two probability distributions of $X_1, \ldots, X_n$ with the supports $\mathbf{S}(X_1) = \cdots = \mathbf{S}(X_n) = \{1, 2, \ldots, N\}$, where N is a prime number ($N \geqslant n - 1$), and put

$$\begin{aligned} p_0^*(X_1 = i_1, \ldots, X_n = i_n) &= 1/N \qquad \text{if } i_1 = \cdots = i_n, \\ &= 0 \qquad \text{otherwise,} \end{aligned}$$

and

$$p_0(X_1 = i_1, \ldots, X_n = i_n) = p_{n-1}, \qquad \text{defined in Section 4, Lemma 4.1.}$$

In the former case, $X_1, \ldots, X_n$ are identical and, in the latter case, every $(n - 1)$ variables are independent. We have, for p_0^*, $S_n = (n - 1) \log N$, $D_n = \log N$ so that $D_n/S_n \to 0$ $(n \to \infty)$. On the other hand, for p_0, $S_n = \log N$, $D_n = (n - 1) \log N$ and hence $D_n/S_n \to \infty$ $(n \to \infty)$. Thus, in the figurative terms, we may say that S_n is "effective" for measuring local lower correlations whereas D_n is "effective" for measuring overall higher correlations (Fig. 7). Therefore, local lower or overall higher correlations dominate in the hierarchy of correlations according to $S_n > D_n$ or $S_n < D_n$. Note that $p_0{}^*$ and p_0 are a dual elementary distribution and an elementary distribution, respectively (Han, 1975).

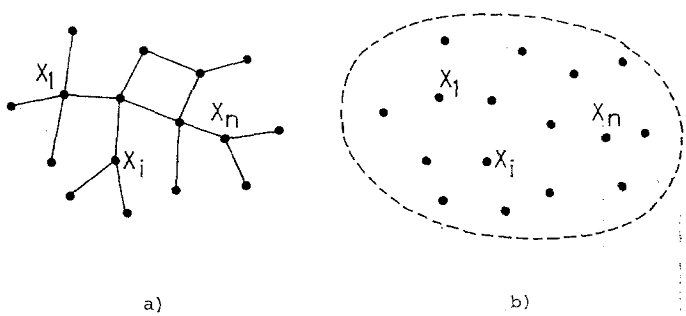

FIG. 7. Patterns of correlation. (a) Local lower correlation. (b) Overall higher correlation.

Finally, we show some property of S_n and D_n. Expressing $\Delta \mathbf{c}_{n-1}^{(n)}$ and $\Delta \mathbf{e}_{n-1}^{(n)}$ in terms of $\Delta^2 \mathbf{e}_j^{(n)}$ and comparing the coefficients, we have

$$0 \leqslant S_n \leqslant (n-1)\, D_n, \tag{6.4}$$

$$0 \leqslant D_n \leqslant (n-1)\, S_n. \tag{6.5}$$

In particular, we have $S_2 = D_2 = h(X_1) + h(X_2) - h(X_1, X_2)$.

THEOREM 6.3.

$$0 \leqslant S_2 \leqslant \cdots \leqslant S_{n-1} \leqslant S_n, \tag{6.6}$$

$$0 \leqslant D_2 \leqslant \cdots \leqslant D_{n-1} \leqslant D_n, \tag{6.7}$$

where S_i and D_i are the (dual) total correlations defined for $X_1, \ldots, X_i$.

Proof. Inequality (6.6) has been derived, e.g., by Baldwin (1966). We show here only (6.7). By definition,

$$D_i = \sum_{k=1}^{i} h(X_1 \cdots X_{k-1} X_{k+1} \cdots X_i) - (i-1)\, h(X_1, \ldots, X_i).$$

Rewriting, we have

$$D_i = -\sum_{k=1}^{i-1} h_{X_1\cdots X_{k-1}X_{k+1}\cdots X_i}(X_k) + h(X_1, \ldots, X_{i-1})$$

$$\geqslant -\sum_{k=1}^{i-1} h_{X_1\cdots X_{k-1}X_{k+1}\cdots X_{i-1}}(X_k) + h(X_1, \ldots, X_{i-1})$$

$$= \sum_{k=1}^{i-1} h(X_1 \cdots X_{k-1}X_{k+1} \cdots X_{i-1}) - (i-2)\, h(X_1 \cdots X_{i-1}) = D_{i-1} \,.\ \text{Q.E.D.}$$

7. Concluding Remarks

We have so far described the nonnegativity property of a class of information-theoretic correlation measures by introducing the concept of symmetric correlative entropy space. What we thereby intend is not to study the nonnegativity of a "particular" measure but to clarify a structural interrelationship existing in a "collection" of nonnegative measures. The concept of hierarchy of correlation measures derived in Section 4.2 is one such structural property which offers a unifying perspective for the information-theoretic multivariate analysis. With the aid of Theorem 4.1, this may be restated as follows: Any probabilistic dependence relation associated with a symmetric nonnegative (correlative) entropy is decomposed into relations associated with Shannon's (conditional) mutual informations; this decomposition is unique and minimal. In other words, the nonnegativity in the collection of symmetric correlative entropies is completely reduced to that of Shannon's (conditional) mutual informations. Considering that this collection has a range wide enough to cover many "effective" correlation measures, this result seems to give a mathematical foundation for the information-theoretic multivariate analysis to a considerable extent. Incidentally, the following question arises: Is the situation the same for the general nonsymmetric case? We conjecture that the nonsymmetric case would also be solved along almost the same line.

Another structural property is the duality, which was first pointed out by Han (1975). In view of duality, dual total correlation is a mere transform of Watanabe's total correlation, in which it should be noted that the relevant nonnegativity as well as the multivariate implication remains almost preserved under this transformation. The principle which underlies such a dual correspondence may be summarized as replacement of probability distributions such that $p(\alpha) \to p_{\bar{\alpha}}(\alpha)$. The duality concerning the set of entropies immediately follows therefrom. The set of Shannon's informations is "self-dual" in this sense. We may, therefore, say that the nonnegativity so far revealed is a self-dual property (Lemma 3.2, equality (5.14)).

Almost all the results, which have been developed in terms of discrete variables, may also be validated, with obvious modifications, for the continuous-variable case. Although there is no essential difference by which to distinguish two cases, the following aspects appertaining to the continuous case should be noted. (1) The value of entropy is not necessarily nonnegative, (2) the concept of correlativity coincides precisely with the invariancy under component-wise transformations of coordinates such that $X_1 \to f_1(X_1), \ldots, X_n \to f_n(X_n)$, (3) the nonnegativity in $\bar{S}(\mathbf{X})$ coincides completely with that in $\bar{S}_0(\mathbf{X})$.

ACKNOWLEDGMENT

The author thanks Professor Masao Iri of the University of Tokyo for his valuable advice and helpful discussions. The idea of introducing symmetricity was suggested by him.

RECEIVED: October 3, 1975; REVISED: March 24, 1977

REFERENCES

BALDWIN, J. G. (1966), The dependency capacity of finite Borel fields, *Inform. Contr.* **9**, 380–392.

GARNER, W. R. (1958), Symmetric uncertainty analysis and its implications for psychology, *Psychol. Rev.* **65**, 183–196.

GARNER, W. R., AND MCGILL, W. J. (1956), Relation between uncertainty, variance, and correlation analyses, *Psychometrika* **21**, 219–228.

HAN, T. S. (1975), Linear dependence structure of the entropy space, *Inform. Contr.* **29**, 337–368.

HAN, T. S., Multiple mutual informations and multiple interactions in frequency data, to appear.

KU, H. H., AND KULLBACK, S. (1968), Interaction in multidimensional contigency tables: An information-theoretic approach, *J. Nat. Bur. Standards USA, Ser. B.* **72**, 159–188.

KULLBACK, S. (1959), "Information Theory and Statistics," Wiley, New York.

MCGILL, W. J. (1954), Multivariate information transmission, *Psychometrika* **19**, 97–116.

WATANABE, S. (1954), A study of ergodicity and redundancy based on intersymbol corrleation of finite range, *in* "Transactions of a Symposium Information Theory, Cambridge, September 15–17," p. 85.

WATANABE, S. (1960), Information theoretical analysis of multivariate correlation, *IBM J.*, 66–81.

WATANABE, S. (1969), "Knowing and Guessing," Wiley, New York.