Journal of the Korean
Statistical Society
Vol. 26, No. 1, 1997

On Information Theoretic Index for Measuring the Stochastic Dependence Among Sets of Variates †

Hea-Jung Kim [1]

Abstract

In this paper the problem of measuring the stochastic dependence among sets of random variates is considered, and attention is specifically directed to forming a single well-defined measure of the dependence among sets of normal variates. A new information theoretic measure of the dependence called dependence index(DI) is introduced and its several properties are studied. The development of DI is based on the generalization and normalization of the mutual information introduced by Kullback(1968). For data analysis, minimum cross entropy estimator of DI is suggested, and its asymptotic distribution is obtained for testing the existence of the dependence. Monte Carlo simulations demonstrate the performance of the estimator, and show that it is useful not only for evaluation of the dependence, but also for independent model testing.

Key Words : Mutual information; Kullback-Leibler cross entropy; Dependence index; Minimum cross entropy estimator; Box's approximation; Monte Carlo simulation.

†This paper was supported by a research grant from Korea Science & Engineering Foundation, 961-0105-033-1.

[1] Department of Statistics, Dongguk University, Seoul, 100-715, Korea

1. INTRODUCTION

If p components, $X_1, X_2, \ldots, X_p$, of a $p \times 1$ random vector $\mathbf{X}$ are stochastically independent, then joint distribution of the random vector, $f(x_1, \ldots, x_p)$, gives no more information than is given by the p component distributions, $f_i(x_i)$ $i = 1, 2, \ldots, p$, separately. However, in the case of dependent components, the joint distribution gives not only the distribution of the p components, but it also gives us some additional information that can be used to measure the degree of dependence. Following Kullback(1968), Theil and Fiebig(1984), Soofi(1994), and among others, we can define the additional information about the degree of dependence between p components by a specific form of Kullback-Leibler(1951) cross entropy:

$$D = \int \cdots \int f(x_1, x_2, \ldots, x_p) \ln \frac{f(x_1, x_2, \ldots, x_p)}{f_1(x_1) f_2(x_2) \cdots f_p(x_p)} dx_1 dx_2 \cdots dx_p, \quad (1.1)$$

which is known as the mutual information, or as the strength of structure due to Watanabe(1985). The mutual information has attracted the attention of researchers who are intrigued by weakness of correlation coefficient in measuring the dependency between variables(see, for examples, Bozdogan 1990 and Harris 1978). They have used it as an alternative to correlation coefficient which contains following weakness(cf. Kapur and Kesavan, 1992): (i) the $p(p+1)/2$ correlation coefficients do not provide a single measure of dependence among p random variates; (ii) even if correlation coefficient between two variates is zero, the variates need not be independent; (iii) the correlation coefficient measure is satisfactory if the variates are normally distributed and linearly dependent, but it is not so satisfactory in other cases; (iv) the dependence between two attributes in a contingency table can not be expressed in terms of correlation coefficient. The weakness mentioned above is inherently attached to the correlation type dependence measures between sets of random variables such as canonical correlation, multiple correlation, and among others. The purpose of this paper is to suggest a new dependence measure which rectifies the weakness and to unify the correlation type dependence measures in a single well-defined one. A new information-theoretic measure of stochastic dependency among sets of random variables is introduced and studied. This dependence measure, i.e. dependence index(DI), is based on the generalization and normalization of the mutual information defined in (1.1).

This paper is organized as follows. Section 2 presents a generalized version of the mutual information (1.1) which measures the degree of dependence among any subsets of the random vector $\mathbf{X}$. Afterward, dependence index(DI) is suggested by using a specific function which normalizes the generalized mutual information. Section 3 elaborates on DI among sets of normal variates and examines several properties of DI. Section 4 suggests minimum cross entropy estimator of DI and its asymptotic distribution which enables one to test the existence of the dependence among sets of normal variates. Section 5 presents some numerical results that show the usefulness of DI as a dependence measure as well as a independent model testing criterion. Finally, Section 6 points out some conclusions.

2. MEASURE OF STOCHASTIC DEPENDENCE

The relation in (1.1) can be easily generalized for finding the mutual information among sets of random variates. Let the joint probability density function of p dimensional random vector $\mathbf{X}$ be $f(\mathbf{x})$, and let the vector is partitioned into m subvectors $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m$ so that

$$\mathbf{X}' = (\mathbf{X}_1', \mathbf{X}_2', \ldots, \mathbf{X}_m'),$$

where $\mathbf{X}_i : p_i \times 1$ and $\sum_{i=1}^{m} p_i = p$. If they are statistically independent, then the true population distribution

$$f(\mathbf{x}) \;=\; g(\mathbf{x}), \tag{2.1}$$

where $g(\mathbf{x}) = \prod_{i=1}^{m} g_i(\mathbf{x}_i)$; product of marginal joint density functions of $\mathbf{X}_i$, $i = 1, \ldots, m$. In this case, Kullback-Leibler(1951) cross entropy of the joint distribution $f(\mathbf{x})$ with respect to the product of marginal distribution $g(\mathbf{x})$ is zero. When the m sets of variates are not necessarily statistically independent, the cross entropy is given by

$$D(f:g) \;=\; \int f(\mathbf{x}) \ln \frac{f(\mathbf{x})}{g(\mathbf{x})} d\mathbf{x}. \tag{2.2}$$

Assuming that the ranges of $\mathbf{X}_i$ are independent of one another,

$$D(f:g) \;=\; \int f(\mathbf{x}) \ln f(\mathbf{x}) d\mathbf{x} - \sum_{i=1}^{m} \int g_i(\mathbf{x}_i) \ln g_i(\mathbf{x}_i) d\mathbf{x}_i. \tag{2.3}$$

This can be thought as a generalized version of the mutual information defined in (1.1). Therefore, $D(f:g)$ is interpreted as information of variables in $g(\mathbf{x})$ for predicting the variables in $f(\mathbf{x})$, or as a measure of information loss due to change from $f(\mathbf{x})$ to $g(\mathbf{x})$. Since $D(f:g) \geq 0$(cf. Sakamoto, Ishiguro and Kitagawa, 1986) and it vanishes if and only if $\mathbf{X}_i$'s are statistically independent, $D(f:g)$ can therefore be used to measure the dependence among them. If they are dependent, $D(f:g) > 0$. Also, from a property of Kullback-Leibler cross entropy(cf. Kullback, 1968), the greater the value of $D(f:g)$, the greater would be the dependence among them. For practical purpose, we normalize the dependence measure to get dependence index.

Definition 1. Dependence index is a normalized measure of the generalized mutual information $D(f:g)$, such that

$$DI(f:g) = 1 - \exp\{-D(f:g)\}. \tag{2.4}$$

Remark 1. Since $D(f:g) \geq 0$ then $0 \leq DI(f:g) \leq 1$.

Remark 2. $DI(f:g) = 0$ if and only if $f(\mathbf{x})$ is not distinguishable from $g(\mathbf{x})$, i.e., if and only if the subsets of variates, $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m$, are stochastically independent.

Remark 3. $DI(f:g)$ approaches to 1 as the cross entropy $D(f:g)$ gets larger where the referenced distribution $g(\mathbf{x})$ is constrained to be independent among the subsets of variates, $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m$.

Relying on these, one can use $DI(f:g)$ as a normalized measure of dependence among $\mathbf{X}_i$'s. $DI(f:g)$ among subsets of a random vector of the continuous type can be straightforwardly extended to that of random vector of the discrete type.

3. DEPENDENCE INDEX FOR MULTIVARIATE NORMAL DISTRIBUTION

Let p random vector $\mathbf{X}$ be distributed according to $N_p(\mu, \Sigma)$ with joint probability density function $f(\mathbf{x}|\mu, \Sigma)$, and let the random vector $\mathbf{X}$ be partitioned into subvectors; $\mathbf{X}' = (\mathbf{X}_1', \mathbf{X}_2', \ldots, \mathbf{X}_m')$, where $\mathbf{X}_i$ is $p_i \times 1$ subvector of $\mathbf{X}$, $i = 1, 2, \ldots, m$. Suppose μ and Σ have been partitioned correspondingly into submatrices:

$$\mu' = (\mu_1', \mu_2', \ldots, \mu_m'), \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1m} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \Sigma_{m1} & \Sigma_{m2} & \cdot & \Sigma_{mm} \end{bmatrix}$$

with $cov(\mathbf{X}_i, \mathbf{X}_j) = \Sigma_{ij};\ \Sigma_{ii} > 0\ i,j = 1,\ldots,m$. Then the random vector $\mathbf{X}_i$ has $N_{p_i}(\mu_i, \Sigma_{ii})$ distribution with marginal joint probability density function $g_i(\mathbf{x}_i|\mu_i, \Sigma_{ii})$, $i = 1,\ldots,m$.

The cross entropy of the true distribution $f(\mathbf{x}|\mu, \Sigma)$ with respect to the reference distribution $g(\mathbf{x}|\mu, \Sigma) = \prod_{i=1}^m g_i(\mathbf{x}_i|\mu_i, \Sigma_{ii})$ which constraints the independence of $\mathbf{X}_i$'s is given by

$$\begin{aligned} D(f:g) &= \int f(\mathbf{x}|\mu,\Sigma)\ln f(\mathbf{x}|\mu,\Sigma)d\mathbf{x} \\ &- \sum_{i=1}^m \int g_i(\mathbf{x}_i|\mu_i,\Sigma_{ii})\ln g_i(\mathbf{x}_i|\mu_i,\Sigma_{ii})d\mathbf{x}_i \\ &= \sum_{i=1}^m \left\{\frac{p_i}{2}\ln 2\pi + \frac{1}{2}\ln|\Sigma_{ii}| + \frac{p_i}{2}\right\} - \frac{p}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma| - \frac{p}{2} \\ &= \frac{1}{2}\sum_{i=1}^m \ln|\Sigma_{ii}| - \frac{1}{2}\ln|\Sigma|. \end{aligned} \tag{3.1}$$

Thus, from Equation (2.4), the dependence index for multivariate normal distribution is

$$\begin{aligned} DI(f:g) &= 1 - \exp\{-D(f:g)\} \\ &= 1 - (|\Sigma| / \prod_{i=1}^m |\Sigma_{ii}|)^{1/2}. \end{aligned} \tag{3.2}$$

Since the correlation coefficients $\rho_{\alpha\beta} = \sigma_{\alpha\beta}/(\sigma_{\alpha\alpha}\sigma_{\beta\beta})^{1/2}$, $\alpha, \beta = 1,\ldots,p$, we have the relation between covariance matrices and correlation matrices; $|\Sigma| = |R|\prod_{\alpha=1}^p \sigma_{\alpha\alpha}$ and $|\Sigma_{ii}| = |R_{ii}|\prod_{\beta=p_1+\ldots+p_{i-1}+1}^{p_1+\ldots+p_i} \sigma_{\beta\beta}$. Therefore, the dependence index can be expressed entirely in terms of correlation coefficients:

$$DI(f:g) = 1 - (|R| / \prod_{i=1}^m |R_{ii}|)^{1/2}. \tag{3.3}$$

It can be easily seen that the dependence index becomes zero when $\mathbf{X}_i$'s are independent and that it takes value 1 if $|\Sigma| = 0$ for $|\Sigma_{ii}| \neq 0$, or if determinant of correlation matrix of the random vector is zero: $|R| = 0$ for $|R_{ii}| \neq 0$.

$DI(f : g)$, denoted by DI in the sequel, has the following properties.

Property 1. DI yields a single well-defined measure of dependence among m sets of multivariate normal variates:

$$DI = 1 - (|\Sigma| / \prod_{i=1}^{m} |\Sigma_{ii}|)^{1/2}, \quad \Sigma_{ii} > 0. \tag{3.4}$$

Familiar correlation measures of dependency between normally distributed variates are related to DI. For examples, if $p = 2$, $p_1 = p_2 = 1$, $DI = 1 - (1 - \rho^2)^{1/2}$; if $p_1 = 1, p_2 = p - 1$, $DI = 1 - (1 - \rho^2_{1.2\cdots p})^{1/2}$, where $\rho_{1.2\cdots p}$ denotes multiple correlation between $\mathbf{X}_1 : 1 \times 1$ normal variate and the other $p - 1$ normal variates; if $p_1 + p_2 = p, p_1 \leq p_2$, and $p_1, p_2 \neq 1$, $DI = 1 - (\prod_{j=1}^{p_1} (1 - \rho^2_{j(c)})^{1/2}$, where $\rho_{j(c)}$ is j-th largest canonical correlation between two sets of variates, $\mathbf{X}_1 : p_1 \times 1$ and $\mathbf{X}_2 : p_2 \times 1$; and if $p_1 = \ldots = p_m = 1$ so that $m = p$, $DI = 1 - |R|^{1/2}$. These examples show that, when we have p normal variates, DI always gives a single measure of dependence among any subsets of them.

Property 2. $0 \leq DI \leq 1$ with $DI = 0$ if and only if m sets of multivariate normal variates are independent. When m sets of the variates satisfy at least one linear relationship, $DI = 1$.

Proof. From the definition of DI, it is easy to see that the necessary sufficient condition for $DI = 0$ is the independence of the m sets of multivariate normal variates. When three sets of the variates, say $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$, are linearly dependent:

$$\mathbf{X}_3 = A\mathbf{X}_1 + B\mathbf{X}_2 + d = [A\ B][\mathbf{X}_1'\ \mathbf{X}_2']' + d, \tag{3.5}$$

where A, B and d are $p_3 \times p_1$, $p_3 \times p_2$ and $p_3 \times 1$ matrices of fixed value, respectively. Then, for $\Sigma_{ii} > 0$, $|\Sigma|$ in (3.4) can be expressed as

$$|\Sigma| = |\Sigma_{(11)}||\Sigma_{(22)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\Sigma_{(12)}|, \tag{3.6}$$

where

$$\Sigma = \begin{bmatrix} \Sigma_{(11):\sum_{i=1}^{3} p_i \times \sum_{i=1}^{3} p_i} & \Sigma_{(12):\sum_{i=1}^{3} p_i \times \sum_{i=4}^{m} p_i} \\ \Sigma_{(21):\sum_{i=4}^{m} p_i \times \sum_{i=1}^{3} p_i} & \Sigma_{(22):\sum_{i=4}^{m} p_i \times \sum_{i=4}^{m} p_i} \end{bmatrix},$$

and

$$\Sigma_{(11)} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{bmatrix}.$$

Using the linear relation (3.5), we see that

$$[\Sigma'_{13}\ \Sigma'_{23}]' = [\Sigma_{31}\ \Sigma_{32}]' = \Sigma^*[A\ B]', \quad \text{where } \Sigma^* = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

and $\Sigma_{33} = [A\ B]\Sigma^*[A\ B]'$. These relations give

$$\begin{aligned} |\Sigma_{(11)}| &= |\Sigma^*||\Sigma_{33} - [\Sigma_{31}\ \Sigma_{32}]\Sigma^{*\,-1}[\Sigma'_{13}\ \Sigma'_{23}]| \\ &= |\Sigma^*||[A\ B]\Sigma^*[A\ B]' - [A\ B]\Sigma^*\Sigma^{*\,-1}\Sigma^*[A\ B]'| = 0, \end{aligned}$$

and $|\Sigma| = 0$. Thus the linear relation (3.5) of the three sets of variates leads to $DI = 0$. Similar proof holds for all the case when two or more sets of variates are linearly related.

Note that DI gives not only measure for linear dependence among m sets of multivariate normal variates, but also gives meaningful measure for nonlinear dependence among them. Typical example for this property is to measure the dependence between X_1 and $X_1X_2(X_1^2+X_2^2)^{-1/2}+X_1$, where X_1 and X_2 are independent standard normal variates. In this case, $X_1X_2(X_1^2 + X_2^2)^{-1/2} + X_1$ is also normally distributed(Shepp, 1962). Since it is not linear function of X_1, the correlation coefficient $\rho = (2^{1/2}/(1 + 2^{1/2}))^{1/2}$ implying linear dependence is not satisfactory for this example. Instead, the dependence index $DI = 1 - (1 + 2^{1/2})^{-1/2}$ would be more appropriate to show the dependence(nonlinear dependence) between X_1 and $X_1X_2(X_1^2+X_2^2)^{-1/2}+X_1$ than the correlation coefficient.

Property 3. DI is invariant with respect to linear transformation within each set of variates.

Proof. Let C_i be an arbitrary nonsingular matrix of order p_i and let

$$\mathbf{C} = \begin{bmatrix} C_1 & O & \cdots & O \\ O & C_2 & \cdots & O \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ O & O & \cdots & C_m \end{bmatrix}$$

Let $\mathbf{CX} + \mathbf{b} = \mathbf{X}^*$ so that $C_i\mathbf{X}_i + \mathbf{b}_i = \mathbf{X}_i^*$, $i = 1, \ldots, m$, where $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2', \ldots, \mathbf{X}_m')'$ as defined above and $\mathbf{b} = (\mathbf{b}_1', \mathbf{b}_2', \ldots, \mathbf{b}_m')'$ is a correspondingly partitioned arbitrary $p \times 1$ vector with constant components. If $\Omega = E(\mathbf{X}^* - E\mathbf{X}^*)(\mathbf{X}^* - E\mathbf{X}^*)' = \mathbf{C}\Sigma\mathbf{C}'$ be partitioned into submatrices $\Omega_{ij} = E(\mathbf{X}_i^* - E\mathbf{X}_i^*)(\mathbf{X}_j^* - E\mathbf{X}_j^*)' = C_j\Sigma_{ij}C_j'$. Then the respective distributions of $\mathbf{X}^*$ and $\mathbf{X}_i^*$ are $N_p(\mu^*, \Omega)$ and $N_{p_i}(\mu_i^*, \Omega_{ii})$, $i = 1, \ldots, m$, where $\mu^* = \mathbf{C}\mu + \mathbf{b}$ and $\mu_i^* = C_i\mu_i + \mathbf{b}_i$ Under the transformation, the cross entropy in (3.1) can be obtained as

$$\begin{aligned} D(f^* : g^*) &= \int f^*(\mathbf{x}^*|\mu^*, \Omega) \ln f^*(\mathbf{x}^*|\mu^*, \Omega) d\mathbf{x}^* \\ &- \sum_{i=1}^{m} \int g_i^*(\mathbf{x}_i^*|\mu_i^*, \Omega_{ii}) \ln g_i^*(\mathbf{x}_i^*|\mu_i^*, \Omega_{ii}) d\mathbf{x}_i^* \\ &= \frac{1}{2}\sum_{i=1}^{m}\{\ln|\Sigma_{ii}| + 2\ln|C_i|\} - \frac{1}{2}\{\ln|\Sigma| + \ln|\mathbf{C}|\} \\ &= D(f : g), \end{aligned}$$

for $|\mathbf{C}| = \prod_{i=1}^{m} |\mathbf{C}_i|$. The above result and Definition 1 yield that the dependence index in terms of $\mathbf{X}^*$ is identical with that in terms of $\mathbf{X}$.

4. MINIMUM CROSS ENTROPY ESTIMATOR

Now suppose that the multivariate normal population parameters are assumed to be unknown, as is usually with the case. Also suppose that N independent p-variate normal observations from the population, $O_N = (x_1, x_2, \ldots, x_N)$, are available. In this situation, we need to estimate DI. This section derives a minimum cross entropy estimator of DI via the minimum cross entropy procedure by Kullback(1968).

Definition 2.(Kullback, 1968) The minimum cross entropy $I(\ell^* : h)$ is the minimum value of

$$I(\ell : h) = \int \ell(x) \log \frac{\ell(x)}{h(x)} dx \tag{4.1}$$

for a given true density $h(x)$ and all $\ell(x)$ such that

$$\theta = \int T(x)\ell(x)dx.$$

The minimum value $I(\ell^* : h) = \theta\tau(\theta) - \log M(\tau(\theta))$ occurs for the conjugate distribution with generalized density given by

$$\ell^*(x) = \frac{\exp\{\tau T(x)\}h(x)}{M(\tau)},$$

where $M(\tau) = \int \exp\{\tau T(x)\}h(x)dx$ and $\theta = \frac{d}{d\tau}\log M(\tau)$.

Shore and Johnson(1980) showed that above constrained minimization is the unique correct method for inductive inference. When $h(x)$ is the genralized density of N independent observations, $O_N = (x_1, \ldots, x_N)$, we can estimate $I(\ell^*; h)$ by using the observed value of $T(x)$ in the sample O_N as an estimate of θ, $\hat{\theta}$, and a related estimate of τ, $\hat{\tau} = \tau(\hat{\theta})$, such that

$$T(x) = \hat{\theta} = \frac{d}{d\tau}\log M(\tau)|_{\tau=\tau(\hat{\theta})}.$$

This leads to the estimate of $I(\ell^* : h; O_N)$ given by

$$\hat{I}(\ell^* : h; O_N) = \hat{\theta}\tau(\hat{\theta}) - \log M(\tau(\hat{\theta})). \tag{4.2}$$

Definition of the estimate can be easily generalized to the case of multivariate normal density(cf. Kullback, 1968). Consider a sample O_N of N independent observation from a p-variate normal population $N_p(\mu, \Sigma)$, the moment generating function of the sample mean $\bar{X}$ and the unbiased sample covariance matrix S with $n = N - 1$ degrees of freedom is known to be(cf. Anderson, 1984)

$$M(\tau, T) = |I_p - 2\Sigma T/n|^{-1/2}\exp\{\tau'\mu + \tau'\Sigma\tau/(2N)\},$$

where $\tau' = (\tau_1, \ldots, \tau_p)$ and $T = (\tau_{ij}),\ i,j = 1, \ldots, p$. With $T(x) = (\bar{X}, S)$ and letting $f(x|\mu, \Sigma)$(equivalent to $h(x)$ in Definition 2) as the density function of $N_p(\mu, \Sigma)$, we have, from (4.2), the minimum cross entropy estimate of $I(f^* : f; O_N) = I(f^* : f; \bar{X}) + I(f^* : f; S)$ as

$$\hat{I}(f^* : f; O_N) = \frac{N}{2}(\bar{X} - \mu)'\Sigma^{-1}(\bar{X} - \mu) + \frac{n}{2}\left(\log\frac{|\Sigma|}{|S|} - p + \mathrm{tr}S\Sigma^{-1}\right). \tag{4.3}$$

Partition the sample mean vector and the sample covariance matrix as

$$\bar{X} = (\bar{X}_1', \ldots, \bar{X}_m')', \quad S = \{S_{ij}\}, \quad i,j = 1, \ldots, m,$$

where $\bar{x}_i$ is $p_i \times 1$ vector and $S_{ij} : p_i \times p_j$ is (ij)-th submatrix of S, then the minimum cross entropy estimate $\hat{I}(f^* : f; O_N)$ in (4.3) yields the following result.

Theorem 1. Minimum cross entropy estimator of $DI(f : g) = DI$ in (3.2) is

$$\hat{DI} = 1 - \left(\frac{|S|}{\prod_{i=1}^{m} |S_{ii}|}\right)^{1/2}. \tag{4.4}$$

Proof. By minimizing $\hat{I}(f^* : f; O_N)$ in (4.3), we obtain the minimum cross entropy estimators of true population parameters of $N_p(\mu, \Sigma)$ with density $f(x|\mu, \Sigma)$. Differentiation of $\hat{I}(f^* : f; O_N)$ with respect to μ and Σ gives the minimum cross entropy estimators, $\hat{\mu} = \bar{X}$ and $\hat{\Sigma} = S$. The minimum cross entropy procedure can be straightforwardly applied for estimating the parameters of the marginal densities in (3.1), $g_i(\mathbf{x}_i|\mu_i, \Sigma_{ii})$, $i = 1, \ldots, m$. The procedure yields the minimum cross entropy estimators, $\hat{\mu}_i = \bar{X}_i$ and $\hat{\Sigma}_{ii} = S_{ii}$. Substituting the estimators for the corresponding parameters in $D(f : g)$, we get

$$\begin{aligned}\hat{D}(f : g) &= \int f(\mathbf{x}|\bar{X}, S) \ln f(\mathbf{x}|\bar{X}, S) d\mathbf{x} \\ &\quad - \sum_{i=1}^{m} \int g_i(\mathbf{x}_i|\bar{X}_i, S_{ii}) \ln g_i(\mathbf{x}_i|\bar{X}_i, S_{ii}) d\mathbf{x}_i \\ &= \frac{1}{2}\sum_{i=1}^{m} \ln |S_{ii}| - \frac{1}{2} \ln |S|. \end{aligned} \tag{4.5}$$

This gives the desired estimator, $\hat{DI} = 1 - \exp\{-\hat{D}(f : g)\}$.

It is noted from (4.3) that the estimator $\hat{DI}$ is invariant under nonsingular linear transformations of the sample space. Moreover, via Hardamard inequality, we see that $0 \leq \hat{DI} \leq 1$. Thus we can easily see that $\hat{DI}$ has the same properties(Property 2 and Property 3) as DI. Concerning Property 1 of $\hat{DI}$, in the next section, we will conduct a simulation study to check it.

Given the independent p-variate observations $O_N = (x_1, x_2, \ldots, x_N)$ from $N_p(\mu, \Sigma)$, it is sometimes of interest to study whether the sample index $\hat{DI}$ actually reflects dependence among subsets of $\mathbf{X}$ or is merely the result of sampling variation. That is, it is necessary to test the hypothesis $H : DI = 0$ versus $A : DI > 0$, where DI denotes the population dependence index among m sets of normal variates. In other words, we need to test that the subvectors

$\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m$ of $\mathbf{X}$ are independent. Using Theorem 9.3.4 in Anderson(1984), we can show that, under H, the h-th moment of $1-\hat{DI} = (|S|/\prod_{i=1}^m |S_{ii}|)^{1/2}$ is

$$E\{(1-\hat{DI})\}^h = \prod_{i=2}^{m}\prod_{j=1}^{p_i} \frac{\Gamma\{(N-\bar{p}_i-j)/2+h/2\}}{\Gamma\{(N-j)/2\}} \frac{\Gamma\{(N-\bar{p}_i-j)/2\}}{\Gamma\{(N-j)/2+h/2\}}, \tag{4.6}$$

where $\bar{p}_i = p_1 + p_2 + \ldots + p_{i-1}$. Thus we have the following asymptotic result.

Corollary 1. If $\hat{DI} = 1 - (|S|/\prod_{i=1}^m |S_{ii}|)^{1/2}$ denotes the sample dependence index, based on N observations from $N_p(\mu, \Sigma)$, the test for $H : DI = 0$ versus $A : DI > 0$ is to reject H if $\hat{DI} > c$, where c is found for any given level of significance from

$$\begin{aligned} Pr\{-\alpha \ln\{1-\hat{DI})\} \le v\} &= Pr\left\{\chi_h^2 \le v\right\} &(4.7)\\ &+ \frac{\beta}{\alpha^2}\left[Pr\left\{\chi_{h+4}^2 \le v\right\} - Pr\left\{\chi_h^2 \le v\right\}\right] + O(\alpha^{-3}), \end{aligned}$$

where

$$h = \frac{1}{2}(p^2 - \sum_{i=1}^m p_i^2), \tag{4.8}$$

$$\alpha = 2N - 3 - \frac{2(p^3 - \sum_{i=1}^m p_i^3)}{3(p^2 - \sum_{i=1}^m p_i^2)}, \tag{4.9}$$

$$\beta = \frac{p^4 - \sum_{i=1}^m p_i^4}{48} - \frac{5(p^2 - \sum_{i=1}^m p_i^2)}{96} - \frac{(p^3 - \sum_{i=1}^m p_i^3)^2}{72(p^2 - \sum_{i=1}^m p_i^2)}. \tag{4.10}$$

Proof. The proof is the same as p.386 of Anderson(1984), except for the constant term α.

5. SIMULATION STUDY

Two Monte Carlo simulations are performed to evaluate the performance of $\hat{DI}$. The simulations examine the usefulness of $\hat{DI}$ as a well-defined single valued dependence measure. The simulations consist of 4 blocks($N = 10, 20, 50, 100$) of 100 realizations with N observations in each realization. For the first simulation, an observation consists of four components $(x_{1\ell}, x_{2\ell}, x_{3\ell}, x_{4\ell})$, $\ell = 1, 2, \ldots, N$, and it is generated from 4 dimensional normal distribution having two independent sets of variates, $\{X_1, X_2\}$, $\{X_3, X_4\}$, so that $\mathbf{X} = (X_1, X_2, X_3, X_4)' \sim N_4(0, \Sigma)$, where

$$\Sigma = \begin{bmatrix} 8 & 3 & 0 & 0 \\ 3 & 4 & 0 & 0 \\ 0 & 0 & 6 & 4 \\ 0 & 0 & 4 & 5 \end{bmatrix}.$$

The second simulation differs from the first in that Σ is standardized and changed systematically:

$$\Sigma = \begin{bmatrix} 1 & d & 0 & 0 \\ d & 1 & 0 & 0 \\ 0 & 0 & 1 & 1-|d| \\ 0 & 0 & 1-|d| & 1 \end{bmatrix}, \quad \text{where } d = \pm.5,\ \pm.6,\ \pm.7,\ \pm.8,\ \pm.9.$$

The results of the first and second simulations are summarized in Table I and Table II, respectively.

Table I gives population DI, the mean and the standard deviation of estimates of DI(denoted by $\hat{DI}$) for each of the following sets of variates.

M_1 :	$\{X_1\},\{X_2\},\{X_3\},\{X_4\}$	M_8 :	$\{X_1X_2\},\{X_3X_4\}$
M_2 :	$\{X_1X_2\},\{X_3\},\{X_4\}$	M_9 :	$\{X_1X_3\},\{X_2X_4\}$
M_3 :	$\{X_1X_3\},\{X_2\},\{X_4\}$	M_{10} :	$\{X_1X_4\},\{X_2X_3\}$
M_4 :	$\{X_1X_4\},\{X_2\},\{X_3\}$	M_{11} :	$\{X_1X_2X_3\},\{X_4\}$
M_5 :	$\{X_2X_3\},\{X_1\},\{X_4\}$	M_{12} :	$\{X_1X_2X_4\},\{X_3\}$
M_6 :	$\{X_2X_4\},\{X_1\},\{X_3\}$	M_{13} :	$\{X_1X_3X_4\},\{X_2\}$
M_7 :	$\{X_3X_4\},\{X_1\},\{X_2\}$	M_{14} :	$\{X_2X_3X_4\},\{X_1\}$.

For each realization, the test in Theorem 2 was conducted under the null hypothesis that the sets of variables in M_i are independent, $i = 1, 2, \ldots, 14$. The mean and the standard deviation of p-values resulting from the tests are given in the table. All the standard deviations are noted in the parentheses.

Several points were noted from Table I. The minimum cross entropy estimator($\hat{DI}$) is estimating the dependence index(DI) satisfactorily: Comparing with the variance of $\hat{DI}$, we note that the bias of $\hat{DI}$ is not significant. Moreover, as would be expected, the variance decreases as the sample size N increases. Results of the asymptotic test for the independence of sets of variates defined in M_i, $i = 1, 2, \ldots, 14$, are listed in terms of p-values.

Table 1. RESULTS OF SIMULATION 1, WITH MEAN AND STANDARD DEVIATION OF $\hat{DI}$ AND *Pro.* (= P-VALUE x 10^2).

		$N = 10$		$N = 20$		$N = 50$		$N = 100$	
Sets	DI	$\hat{DI}$	*Pro.*	$\hat{DI}$	*Pro.*	$\hat{DI}$	*Pro.*	$\hat{DI}$	*Pro.*
M_1	.4280	.4945	1.45	.4489	.00	.4310	.00	.4339	.00
		(.1253)	(4.10)	(.0681)	(.00)	(.0523)	(.00)	(.0373)	(.00)
M_2	.3169	.3959	3.03	.3472	.00	.3311	.00	.3276	.00
		(.1147)	(3.02)	(.0607)	(.01)	(.0461)	(.00)	(.0357)	(.00)
M_3	.4280	.4791	1.06	.4444	.00	.4283	.00	.4321	.00
		(.1258)	(2.96)	(.0690)	(.00)	(.0527)	(.00)	(.0372)	(.00)
M_4	.4280	.4805	1.19	.4440	.00	.4283	.00	.4315	.00
		(.1277)	(3.59)	(.0697)	(.00)	(.0523)	(.00)	(.0373)	(.00)
M_5	.4280	.4806	1.17	.4430	.00	.4279	.00	.4321	.00
		(.1271)	(3.67)	(.0670)	(.00)	(.0533)	(.00)	(.0372)	(.00)
M_6	.4280	.4817	1.08	.4429	.00	.4282	.00	.4318	.00
		(.1263)	(3.19)	(.0680)	(.00)	(.0525)	(.00)	(.0375)	(.00)
M_7	.1522	.2497	1.97	.1891	1.79	.1657	.06	.1703	.00
		(.1141)	(2.29)	(.0621)	(4.58)	(.0452)	(.30)	(.0392)	(.00)
M_8	.0000	0996	54.4*	.0389	53.1*	.0187	54.4*	.0150	45.3*
		(.0663)	(26.9)	(.0267)	(28.7)	(.0142)	(30.1)	(.0100)	(29.0)
M_9	.4208	.4658	0.74	.4383	.00	.4255	.00	.4300	.00
		(.1270)	(2.15)	(.0690)	(.00)	(.0529)	(.00)	(.0374)	(.00)
M_{10}	.4208	.4661	0.96	.4379	.00	.4252	.00	.4297	.00
		(.1300)	(3.64)	(.0717)	(.00)	(.0533)	(.00)	(.0372)	(.00)
M_{11}	.3168	.3603	1.56	.3342	.00	.3237	.00	.3231	.00
		(.1141)	(3.54)	(.0603)	(.00)	(.0470)	(.00)	(.0356)	(.00)
M_{12}	.3168	.3619	1.72	.3347	.00	.3247	.00	.3224	.00
		(.1181)	(3.71)	(.0606)	(.00)	(.0454)	(.00)	(.0357)	(.00)
M_{13}	.1522	.2103	1.54	.1734	.99	.1581	.02	.1540	.00
		(.1132)	(2.01)	(.0632)	(2.75)	(.0451)	(.13)	(.0386)	(.00)
M_{14}	.1522	.2126	1.40	.1723	.97	.1579	.02	.1544	.00
		(.1116)	(1.86)	(.0611)	(2.91)	(.0454)	(.11)	(.0390)	(.00)

p-values show that the test successfully discriminates the independent sets of variates in M_8. This fact is asterisked in the table. In order to show the performance of the test, some numerical results under the design of simulation 2 are given in Table 2.

Table 2. THE PERCENTAGE OF CORRECT TEST

	$N=10$		$N=20$		$N=50$		$N=100$	
d	$TEST1$	$TEST2$	$TEST1$	$TEST2$	$TEST1$	$TEST2$	$TEST1$	$TEST2$
.5	93	97	93	98	94	97	96	100
−.5	93	97	93	98	94	99	94	99
.6	96	100	96	100	96	99	94	99
−.6	92	96	93	98	94	99	94	99
.7	98	100	94	100	99	100	94	99
−.7	93	97	93	98	94	99	94	99
.8	98	100	93	98	94	99	94	99
−.8	93	97	93	98	94	99	94	99
.9	93	97	93	98	94	99	94	99
−.9	93	97	93	98	94	99	94	97

Table 2 presents results of the test conducted by using 100 independent trials. $TEST1$ and $TEST2$ denote percentages of accepting null hypothesis($H: DI = 0$ for the true model(M_8) under the significance levels $\alpha = .05$ and $\alpha = .01$, respectively. We see from Table 2 that, irrespective of d and N, the test by asymptotic result nearly achieves the desired significance levels.

6. CONCLUDING REMARKS

In this paper the mutual information by Kullback(1968) is generalized and normalized to arrive at a single well-defined measure of dependence among m sets of variates. This new information-theoretic dependence measure, so called dependence index(DI), is specifically directed to forming a measure of dependence among sets of normal variates. In terms of property, advantages of DI over the usual correlation type measure are illustrated. For data analysis, a best invariant cross entropy estimator($\hat{DI}$) of DI is suggested, and its asymptotic distribution is obtained for testing the existence of dependence among sets of normal variates. It is revealed that $\hat{DI}$ can estimate and test the dependence among any combination of sets of normal variates, thereby allowing an unified inference about the dependence.

The simulations show that the dependence index can be estimated successfully by means of $\hat{DI}$, and $\hat{DI}$ is useful not only for evaluating the underlying dependence among sets of normal variates, but also as a criterion for testing independent sets of the variates. Application of DI to the multivariate distributions other than the normal case is possible. A multivariate distribution to which DI may be immediately applicable is the multivariate Pareto distribution which has a characteristic of vanishing correlation coefficients.

REFERENCES

(1) Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis,* John Wiley and Sons, New York.

(2) Bozdogan, H. (1990). On the information-based measure of covariance complexity and its application to the evaluation of multivariate linear models. *Communication in Statistics -Theory and Methods*, **19**(1), 221-278.

(3) Box, G. E. P. (1940). A general distribution theory for a class of likelihood criteria, *Biometrika*, **36**, 317-346.

(4) Harris, C. J. (1978). An information theoretic approach to estimation. In *Recent Theoretical Developments in Control*, ed. by M. J. Gregson, Academic Press. London.

(5) Kapur, J. N. and Kesavan, H. J. (1992). *Entropy Optimization Principles with Applications*, Academic Press. New York.

(6) Kullback, S. (1968). *Information Theory and Statistics,* Dover Publications, New York.

(7) Kullback, S. and Leibler, R. A. (1951). On information and sufficiency, *Annals of Mathematical Statistics,* **22**, 79-86.

(8) Murray, G. D. (1977). A note on the estimation of probability density functions, *Biometrika*, **64**, 150-152.

(9) Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1986). *Akaike Information Criterion Statistics*, D. Reidel Publishing Co., Tokyo.

(10) Shepp, L. (1962). Normal functions of normal random variables, *SIAM Review*, **4**, 255-256.

(11) Shore, J. E. and Johnson, R. W. (1980). Axiomatic deviation of the principle of maximum entropy and the principle of minimum cross entropy, *IEEE Transactions on Information Theory*, IT **26**(1), 26-37.

(12) Soofi, E. S. (1994) Capturing the intangible concept of information, *Journal of American Statistical Association*, **89**, 1243-1254.

(13) Theil, H. and Fiebig, D. G. (1984). *Exploiting Continuity; Maximum Entropy Estimation of Continuous Distributions.* Bellinger Publishing Company, Cambridge, Massachusetts.

(14) Watanabe, S. (1985). *Pattern Recognition: Human and Mechanical.* John Wiley and Sons, New York.