

L-Bivariate and L-Multivariate Association Coefficients

*Nan Kong
Charles Lewis*

July 2008

ETS RR-08-40



***L*-Bivariate and *L*-Multivariate Association Coefficients**

Nan Kong and Charles Lewis
ETS, Princeton, NJ

July 2008

As part of its nonprofit mission, ETS conducts and disseminates the results of research to advance quality and equity in education and assessment for the benefit of ETS' constituents and the field.

ETS Research Reports provide preliminary and limited dissemination of ETS research prior to publication. To obtain a PDF or a print copy of a report, please visit:

<http://www.ets.org/research/contact.html>

Copyright © 2008 by Educational Testing Service. All rights reserved.

ETS, the ETS logo, and LISTENING. LEARNING.
LEADING. are registered trademarks of Educational Testing
Service (ETS).



Abstract

Given a system of multiple random variables, a new measure called the L -multivariate association coefficient is defined using (conditional) entropy. Unlike traditional correlation measures, the L -multivariate association coefficient measures the multiassociations or multirelations among the multiple variables in the given system; that is, the L -multivariate association coefficient measures the degree of the association for the given system. The L -multivariate association coefficient for the system of two random variables is also called the L -bivariate association coefficient. The association measured by the L -multivariate association coefficient is a general type of association, not any specific type of a linear or nonlinear association. Unlike the K -dependence coefficient, which is an asymmetrical measure, the L -multivariate association coefficient is a symmetrical measure. A direct application of the L -multivariate association coefficient is in variables selection or variables reduction. This paper also explores the relationship between the L -multivariate association coefficient and the K -dependence coefficient.

Key words: Entropy, conditional entropy, L -bivariate association coefficient, L -multivariate association coefficient, K -dependence coefficient

Acknowledgments

The authors would like to thank Helena Jia, Yi-Hsuan Lee, and Daniel Eignor for carefully reviewing the manuscript. Their constructive suggestions and comments were very helpful in the completion of this paper. The authors also thank Kim Fryer for her editorial assistance.

1 Introduction

Entropy is a measure of uncertainty for a random system or a group of random variables. Let $\vec{X} = (X_1, X_2, \dots, X_n)^T$ and $\vec{Y} = (Y_1, Y_2, \dots, Y_m)^T$ be two groups of discrete random variables with N and M possible states respectively denoted by $\{\vec{x}_1, \dots, \vec{x}_N\}$ and $\{\vec{y}_1, \dots, \vec{y}_M\}$. Joint probabilities of these possible states are denoted by $p(\vec{x}_i, \vec{y}_j)$ ($i = 1, \dots, N; j = 1, \dots, M$), while marginal probabilities are denoted by $p(\vec{x}_i)$ ($i = 1, \dots, N$) and $p(\vec{y}_j)$ ($j = 1, \dots, M$). The following are the mathematical expressions for the entropy and conditional entropy (Shannon, 1948).

Definition 1.

Entropy in terms of a group of discrete random variables is defined as

$$H(\vec{X}) = - \sum_{i=1}^N p(\vec{x}_i) \ln(p(\vec{x}_i)). \quad (1)$$

Consider that $0 \leq p(\vec{x}_i) \leq 1$ for $i = 1, 2, \dots, N$, $H(\vec{X})$ in (1) is greater than or equal to zero; that is,

$$H(\vec{X}) \geq 0. \quad (2)$$

Definition 2.

Conditional entropy of \vec{X} given \vec{Y} is defined by

$$H(\vec{X}|\vec{Y}) = - \sum_{i=1}^N \sum_{j=1}^M p(\vec{x}_i, \vec{y}_j) \ln(p(\vec{x}_i|\vec{y}_j)), \quad (3)$$

where $p(\vec{x}_i, \vec{y}_j)$ is the joint distribution of (\vec{x}, \vec{y}) and $p(\vec{x}_i|\vec{y}_j)$ is the conditional probability of \vec{x}_i given \vec{y}_j .

Similar to (2),

$$H(\vec{X}|\vec{Y}) \geq 0. \quad (4)$$

Conditional entropy $H(\vec{X}|\vec{Y})$ measures the conditional uncertainty of \vec{X} given \vec{Y} . The conditional entropy $H(\vec{X}|\vec{Y})$ is less than or equal to the entropy $H(\vec{X})$ with equality if and only if \vec{X} and \vec{Y} are independent which is,

$$H(\vec{X}|\vec{Y}) \leq H(\vec{X}). \quad (5)$$

Equality holds if and only if \vec{X} and \vec{Y} are independent.

Definition 3.

$$K(\vec{X} : \vec{Y}) = \frac{H(\vec{X}) - H(\vec{X}|\vec{Y})}{H(\vec{X})}, \quad (6)$$

where

$H(\vec{X})$ is the entropy of \vec{X} and

$H(\vec{X}|\vec{Y})$ is the conditional entropy of \vec{X} given \vec{Y} .

$K(\vec{X} : \vec{Y})$ in (6) is called the K -dependence coefficient, or dependence coefficient, of \vec{X} to \vec{Y} , or \vec{X} dependence on \vec{Y} .

The K -dependence coefficient in (6) measures the degree to which \vec{X} depends on \vec{Y} with a range between zero and one. Following are the two fundamental theorems of the K -dependence coefficient (Kong, 2007a, 2007b).

Theorem 1

$$K(\vec{X} : \vec{Y}) = 0 \quad (7)$$

if and only if \vec{X} and \vec{Y} are independent from each other.

Theorem 2

Assume that $p(\vec{x}_i) > 0$ and $p(\vec{y}_j) > 0$ ($i = 1, 2, \dots, N; j = 1, 2, \dots, M$). Then

$$K(\vec{X} : \vec{Y}) = 1$$

if and only if for each \vec{y}_j , $\exists \vec{x}_{i_j}$ such that $p(\vec{X} = \vec{x}_{i_j} | \vec{Y} = \vec{y}_j) = 1$ and $p(\vec{X} = \vec{x}_k | \vec{Y} = \vec{y}_j) = 0$, where $k \neq i_j$.

Theorems 1 and 2 state that $K(\vec{X} : \vec{Y}) = 0$ is equivalent to the independence between \vec{X} and \vec{Y} while $K(\vec{X} : \vec{Y}) = 1$ is equivalent to \vec{X} 's total dependence on \vec{Y} .

***L*-Bivariate Association Coefficient**

A measure frequently used for the relation between two random variables is the linear correlation coefficient. Because the linear correlation coefficient is defined from the first and second moments of the (joint) distribution of the random variables, the linear correlation coefficient only partially utilizes the information of the random variables to measure the relation between them. In this section, a new measure called the *L*-bivariate association coefficient is defined from (conditional) entropy to measure the relation between two groups of the random variables. The basic idea behind the *L*-bivariate association coefficient is to fully utilize the information of the random variables to measure the relation among them. Therefore, the *L*-bivariate association coefficient measures both linear and nonlinear relationships between two groups of random variables.

Definition 4.

$$L(\vec{X}, \vec{Y}) = \frac{H(\vec{X}) - H(\vec{X} | \vec{Y})}{H(\vec{X}, \vec{Y})}. \quad (8)$$

$L(\vec{X}, \vec{Y})$ in (8) is called the *L*-bivariate association coefficient for the random variables \vec{X} and \vec{Y} .

First, the *L*-bivariate association coefficient is symmetrical, which is $L(\vec{X}, \vec{Y}) = L(\vec{Y}, \vec{X})$.

This is directly from

$$H(\vec{X}) - H(\vec{X} | \vec{Y}) = H(\vec{Y}) - H(\vec{Y} | \vec{X}).$$

Second, the L -bivariate association coefficient ranges between zero and one when $H(\vec{X}, \vec{Y}) > 0$.

Theorem 3

$$0 \leq L(\vec{X}, \vec{Y}) \leq 1. \quad (9)$$

Proof: First, according to (2) and (5), there are

$$\begin{aligned} H(\vec{X}, \vec{Y}) &\geq 0, \\ H(\vec{X}) - H(\vec{X}|\vec{Y}) &\geq 0. \end{aligned}$$

From Definition 4,

$$L(\vec{X}, \vec{Y}) \geq 0.$$

Also, there is

$$H(\vec{X}) - H(\vec{X}|\vec{Y}) \leq H(\vec{X}) \leq H(\vec{X}, \vec{Y}).$$

According to Definition 4, this implies

$$L(\vec{X}, \vec{Y}) \leq 1.$$

Therefore, Theorem 3 is proved.

Theorem 4

$$L(\vec{X}, \vec{Y}) = 0 \quad \text{if and only if } \vec{X} \text{ and } \vec{Y} \text{ are independent.} \quad (10)$$

Proof: According to Definition 4,

$$L(\vec{X}, \vec{Y}) = 0$$

if and only if

$$H(\vec{X}) - H(\vec{X}|\vec{Y}) = 0. \quad (11)$$

Also from (5), the equation in (11) holds if and only if \vec{X} and \vec{Y} are independent. Thus, Theorem 4 is proved.

Theorem 5

$$L(\vec{X}, \vec{Y}) = 1 \tag{12}$$

if and only if

$$K(\vec{X} : \vec{Y}) = 1 \quad \text{and} \quad K(\vec{Y} : \vec{X}) = 1. \tag{13}$$

Proof: First, assume $L(\vec{X}, \vec{Y}) = 1$. From Definition 4,

$$H(\vec{X}, \vec{Y}) = H(\vec{X}) - H(\vec{X}|\vec{Y}).$$

Therefore,

$$H(\vec{X}, \vec{Y}) = H(\vec{X}) - H(\vec{X}|\vec{Y}) \leq H(\vec{X}) \leq H(\vec{X}, \vec{Y}),$$

and consequently,

$$H(\vec{X}) = H(\vec{X}, \vec{Y}) = H(\vec{X}) - H(\vec{X}|\vec{Y}),$$

which implies

$$H(\vec{X}|\vec{Y}) = 0.$$

Similarly, the following can be obtained:

$$H(\vec{Y}|\vec{X}) = 0.$$

From Definition 3,

$$K(\vec{X} : \vec{Y}) = 1,$$

$$K(\vec{Y} : \vec{X}) = 1.$$

Second, assume (13). From Definition 3,

$$H(\vec{X}|\vec{Y}) = 0,$$

$$H(\vec{Y}|\vec{X}) = 0.$$

Therefore,

$$H(\vec{X}, \vec{Y}) = H(\vec{X}) + H(\vec{Y}|\vec{X}) = H(\vec{X}).$$

By Definition 4,

$$L(\vec{X}, \vec{Y}) = 1.$$

Theorem 5 has proved that $L(\vec{X}, \vec{Y}) = 1$ if and only if \vec{X} totally depends on \vec{Y} ($K(\vec{X} : \vec{Y}) = 1$) and \vec{Y} totally depends on \vec{X} ($K(\vec{Y} : \vec{X}) = 1$). Theorems 4 and 5 show the two extreme cases for the L -bivariate association coefficient. For general cases, the L -bivariate association coefficient should be between zero and one.

One application of the L -bivariate association coefficient is in variables selection. Consider a system of n random variables denoted by

$$S_n = \{\vec{X}_1, \dots, \vec{X}_n\},$$

and a subsystem of S_n denoted by

$$S(i_1, \dots, i_k) = \{\vec{X}_{i_1}, \dots, \vec{X}_{i_k}\}, \quad \text{where } \{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}.$$

From Definition 4,

$$\begin{aligned} L((S_n), (S(i_1, \dots, i_k))) &= \frac{H(S_n) - H(S_n|S(i_1, \dots, i_k))}{H(S_n, S(i_1, \dots, i_k))} \\ &= \frac{H(S_n) + H(S(i_1, \dots, i_k)) - H(S_n, S(i_1, \dots, i_k))}{H(S_n)} \\ &= \frac{H(S_n) + H(S(i_1, \dots, i_k)) - H(S_n)}{H(S_n)} \\ &= \frac{H(S(i_1, \dots, i_k))}{H(S_n)} \\ &= \frac{H(\vec{X}_{i_1}, \dots, \vec{X}_{i_k})}{H(\vec{X}_1, \dots, \vec{X}_n)}, \end{aligned}$$

Therefore,

$$L((S_n), (S(i_1, \dots, i_k))) = \frac{H(\vec{X}_{i_1}, \dots, \vec{X}_{i_k})}{H(\vec{X}_1, \dots, \vec{X}_n)}. \quad (14)$$

Given a level of $\alpha : 1 \geq \alpha > 0$, say $\alpha=0.99$, select a subsystem $S(i_1, \dots, i_k)$ such that

$$L(S_n, S(i_1, \dots, i_k)) \geq \alpha.$$

Under the given level α , the selected subsystem $S(i_1, \dots, i_k)$ carries at least $\alpha\%$ of the information of the S_n , but the number of variables in $S(i_1, \dots, i_k)$ is less than n in general. If $\alpha = 1$, then all of the variables will be selected and no variable reduction takes place unless some variables totally depend on others.

Actually, (14) can also be obtained in a similar way from the K -dependence coefficient in Definition 3. In the following application, only the L -bivariate association coefficient can be applied. Assume S_1 and S_2 are the two subsystems of S_n . Given a level of $\alpha : 1 \geq \alpha > 0$, say $\alpha = 0.99$, if

$$L(S_1, S_2) \geq \alpha. \quad (15)$$

The association between S_1 and S_2 is very great, and therefore, the S_1 and S_2 are redundant (one of these subsystems can be removed from the system).

Although the L -bivariate association coefficient is defined to measure the degree of the association for the system of two (groups of) random variables, it can also measure the degree of the association for a system of multiple (groups of) random variables. First, examine the system of three random variables denoted by

$$S_3 = \{\vec{X}_1, \vec{X}_2, \vec{X}_3\}.$$

Theorem 6

$$L((\vec{X}_1, \vec{X}_2), \vec{X}_3) = 0 \quad \text{and} \quad L(\vec{X}_1, \vec{X}_2) = 0$$

if and only if

$$p(\vec{x}_{i_1}, \vec{x}_{i_2}, \vec{x}_{i_3}) = p(\vec{x}_{i_1}) \times p(\vec{x}_{i_2}) \times P(\vec{x}_{i_3}),$$

where $p(\vec{x}_{i_1}, \vec{x}_{i_2}, \vec{x}_{i_3})$ ($i_1 = 1, \dots, m_1, i_2 = 1, \dots, m_2, i_3 = 1, \dots, m_3$) are the joint probabilities of the random variables \vec{X}_1, \vec{X}_2 , and \vec{X}_3 . $p(\vec{x}_i)$ are the probabilities of the random variable \vec{X}_i ($i = 1, 2, 3$).

Proof: First, from Theorem 4,

$$L((\vec{X}_1, \vec{X}_2), \vec{X}_3) = 0$$

if and only if (\vec{X}_1, \vec{X}_2) and \vec{X}_3 are independent:

$$p(\vec{x}_{i_1}, \vec{x}_{i_2}, \vec{x}_{i_3}) = p(\vec{x}_{i_1}, \vec{x}_{i_2}) \times p(\vec{x}_{i_3}).$$

Second,

$$L(\vec{X}_1, \vec{X}_2) = 0$$

if and only if \vec{X}_1 and \vec{X}_2 are independent:

$$p(\vec{x}_{i_1}, \vec{x}_{i_2}) = p(\vec{x}_{i_1}) \times p(\vec{x}_{i_2}).$$

It is obvious that

$$p(\vec{x}_{i_1}, \vec{x}_{i_2}, \vec{x}_{i_3}) = p(\vec{x}_{i_1}) \times p(\vec{x}_{i_2}) \times p(\vec{x}_{i_3})$$

if and only if

$$p(\vec{x}_{i_1}, \vec{x}_{i_2}, \vec{x}_{i_3}) = p(\vec{x}_{i_1}, \vec{x}_{i_2}) \times p(\vec{x}_{i_3}), \quad \text{and}$$

$$p(\vec{x}_{i_1}, \vec{x}_{i_2}) = p(\vec{x}_{i_1}) \times p(\vec{x}_{i_2}).$$

Therefore, Theorem 6 is proved.

Theorem 6 shows that, for a given system of three random variables, if \vec{X}_3 is independent from the system $\{\vec{X}_1, \vec{X}_2\}$ and \vec{X}_1 is independent from \vec{X}_2 , then this given system is in a (jointly) independent state. Similarly, the following two propositions are also true:

$$L((\vec{X}_1, \vec{X}_3), \vec{X}_2) = 0 \quad \text{and} \quad L(\vec{X}_1, \vec{X}_3) = 0$$

if and only if

$$p(\vec{x}_{i_1}, \vec{x}_{i_2}, \vec{x}_{i_3}) = p(\vec{x}_{i_1}) \times p(\vec{x}_{i_2}) \times p(\vec{x}_{i_3}),$$

and

$$L((\vec{X}_2, \vec{X}_3), \vec{X}_1) = 0 \quad \text{and} \quad L(\vec{X}_2, \vec{X}_3) = 0$$

if and only if

$$p(\vec{x}_{i_1}, \vec{x}_{i_2}, \vec{x}_{i_3}) = p(\vec{x}_{i_1}) \times p(\vec{x}_{i_2}) \times p(\vec{x}_{i_3}).$$

Theorem 6 measures the state of independence for a system. The following theorem measures a state of perfect association for the same system.

Theorem 7(a)

$$L((\vec{X}_1, \vec{X}_2), \vec{X}_3) = 1 \quad \text{and} \quad L(\vec{X}_1, \vec{X}_2) = 1$$

if and only if

$$K(\vec{X}_i : \vec{X}_j) = 1 \quad (i \neq j \text{ and } i, j = 1, 2, 3).$$

Proof: First, assume

$$L(\vec{X}_1, \vec{X}_2) = 1 \quad \text{and} \quad L((\vec{X}_1, \vec{X}_2), \vec{X}_3) = 1.$$

According to Theorem 5,

$$K(\vec{X}_1 : \vec{X}_2) = 1 \quad \text{and} \quad K(\vec{X}_2 : \vec{X}_1) = 1$$

and

$$K(\vec{X}_3 : \vec{X}_1, \vec{X}_2) = 1 \quad \text{and} \quad K(\vec{X}_1, \vec{X}_2 : \vec{X}_3) = 1.$$

By Definition 3,

$$H(\vec{X}_3 | \vec{X}_1, \vec{X}_2) = 0 \quad \text{and} \quad H(\vec{X}_1, \vec{X}_2 | \vec{X}_3) = 0.$$

Also,

$$0 \leq H(\vec{X}_1 | \vec{X}_3) \leq H(\vec{X}_1, \vec{X}_2 | \vec{X}_3) = 0,$$

$$0 \leq H(\vec{X}_2 | \vec{X}_3) \leq H(\vec{X}_1, \vec{X}_2 | \vec{X}_3) = 0.$$

Therefore,

$$H(\vec{X}_1 | \vec{X}_3) = 0 \quad \text{and} \quad H(\vec{X}_2 | \vec{X}_3) = 0. \tag{16}$$

On the other hand,

$$\begin{aligned} & H(\vec{X}_1, \vec{X}_2, \vec{X}_3) \\ &= H(\vec{X}_1, \vec{X}_2) + H(\vec{X}_3 | \vec{X}_1, \vec{X}_2) \\ &= H(\vec{X}_2) + H(\vec{X}_1 | \vec{X}_2) = H(\vec{X}_2). \end{aligned}$$

Therefore,

$$H(\vec{X}_1, \vec{X}_2, \vec{X}_3) = H(\vec{X}_2). \quad (17)$$

Also, there is

$$\begin{aligned} & H(\vec{X}_1, \vec{X}_2, \vec{X}_3) \\ &= H(\vec{X}_2, \vec{X}_3) + H(\vec{X}_1|\vec{X}_2, \vec{X}_3) \\ &= H(\vec{X}_2) + H(\vec{X}_3|\vec{X}_2). \end{aligned}$$

The above equation is directly from

$$0 \leq H(\vec{X}_1|\vec{X}_2, \vec{X}_3) \leq H(\vec{X}_1|\vec{X}_2) = 0.$$

This implies

$$H(\vec{X}_1|\vec{X}_2, \vec{X}_3) = 0.$$

Therefore,

$$H(\vec{X}_1, \vec{X}_2, \vec{X}_3) = H(\vec{X}_2) + H(\vec{X}_3|\vec{X}_2). \quad (18)$$

From (17) and (18),

$$H(\vec{X}_2) = H(\vec{X}_2) + H(\vec{X}_3|\vec{X}_2).$$

Therefore,

$$H(\vec{X}_3|\vec{X}_2) = 0. \quad (19)$$

Similarly,

$$H(\vec{X}_3|\vec{X}_1) = 0. \quad (20)$$

By (16), (19), and (20),

$$\begin{aligned} & K(\vec{X}_1 : \vec{X}_3) = 1 \text{ and } K(\vec{X}_3 : \vec{X}_1) = 1, \\ & K(\vec{X}_2 : \vec{X}_3) = 1 \text{ and } K(\vec{X}_3 : \vec{X}_2) = 1. \end{aligned}$$

This is the proof of necessity.

Second, assume

$$K(\vec{X}_i : \vec{X}_j) = 1 \quad (i \neq j \text{ and } i, j = 1, 2, 3).$$

From Theorem 5,

$$L(\vec{X}_1, \vec{X}_2) = 1. \tag{21}$$

Also, by Definition 3,

$$H(\vec{X}_i | \vec{X}_j) = 0 \quad (i \neq j \text{ and } i, j = 1, 2, 3).$$

There is

$$0 \leq H(\vec{X}_3 | \vec{X}_1, \vec{X}_2) \leq H(\vec{X}_3 | \vec{X}_1) = 0.$$

Therefore,

$$H(\vec{X}_3 | \vec{X}_1, \vec{X}_2) = 0.$$

By Definition 3,

$$K(\vec{X}_3 : (\vec{X}_1, \vec{X}_2)) = 1. \tag{22}$$

Similarly,

$$0 \leq H(\vec{X}_1 \vec{X}_2 | \vec{X}_3) \leq H(\vec{X}_1 | \vec{X}_3) + H(\vec{X}_2 | \vec{X}_3) = 0 + 0 = 0.$$

Therefore,

$$H(\vec{X}_1, \vec{X}_2 | \vec{X}_3) = 0.$$

According to Definition 3,

$$K((\vec{X}_1, \vec{X}_2) : \vec{X}_3) = 1. \tag{23}$$

From (22) and (23),

$$L((\vec{X}_1, \vec{X}_2), \vec{X}_3) = 1. \tag{24}$$

The proof of sufficiency is (21) and (24).

From Theorem 5, Theorem 7(a) can be rewritten as Theorem 7(b).

Theorem 7(b)

$$L((\vec{X}_1, \vec{X}_2), \vec{X}_3) = 1 \quad \text{and} \quad L(\vec{X}_1, \vec{X}_2) = 1$$

if and only if

$$L(\vec{X}_1, \vec{X}_2) = 1 \quad L(\vec{X}_1, \vec{X}_3) = 1 \quad L(\vec{X}_2, \vec{X}_3) = 1.$$

Proof: This is obvious from Theorem 5.

Similar to Theorem 6, the following compositions are also true:

$$L((\vec{X}_1, \vec{X}_3), \vec{X}_2) = 1 \quad \text{and} \quad L(\vec{X}_1, \vec{X}_3) = 1,$$

if and only if

$$L(\vec{X}_1, \vec{X}_2) = 1 \quad L(\vec{X}_1, \vec{X}_3) = 1 \quad L(\vec{X}_2, \vec{X}_3) = 1,$$

and

$$L((\vec{X}_2, \vec{X}_3), \vec{X}_1) = 1 \quad \text{and} \quad L(\vec{X}_2, \vec{X}_3) = 1,$$

if and only if

$$L(\vec{X}_1, \vec{X}_2) = 1 \quad L(\vec{X}_1, \vec{X}_3) = 1 \quad L(\vec{X}_2, \vec{X}_3) = 1.$$

Theorems 6 and 7(a) or 7(b) measure the statistical association of three random variables. The idea in Theorems 6 and 7(a) or 7(b) can be extended to the case of multiple variables. The following theorems measure the statistical association for a system of n random variables.

Theorem 8

$$\begin{aligned}
 L(\vec{X}_1, (\vec{X}_2, \dots, \vec{X}_n)) &= 0, \\
 L(\vec{X}_2, (\vec{X}_3, \dots, \vec{X}_n)) &= 0, \\
 &\dots\dots\dots \\
 L(\vec{X}_{n-1}, \vec{X}_n) &= 0,
 \end{aligned}$$

if and only if

$$p(\vec{x}_{i_1}, \dots, \vec{x}_{i_n}) = p(\vec{x}_{i_1}) \times \dots \times p(\vec{x}_{i_n})$$

where $p(\vec{x}_{i_1}, \dots, \vec{x}_{i_n})$ ($i_1 = 1, \dots, m_1, \dots, i_n = 1, \dots, m_n$) are the joint probabilities of the random variables $\vec{X}_1, \dots, \vec{X}_n$. $p(\vec{x}_i)$ ($i = 1, \dots, m_i$) are the probabilities of the random variable \vec{X}_i .

Proof: Proof is similar to the proof in Theorem 6.

Theorem 9(a)

$$\begin{aligned}
 L(\vec{X}_1, (\vec{X}_2, \dots, \vec{X}_n)) &= 1, \\
 L(\vec{X}_2, (\vec{X}_3, \dots, \vec{X}_n)) &= 1, \\
 &\dots\dots\dots \\
 L(\vec{X}_{n-1}, \vec{X}_n) &= 1
 \end{aligned}$$

if and only if

$$K(\vec{X}_i : \vec{X}_j) = 1 \quad (i \neq j \text{ and } i, j = 1, 2, \dots, n).$$

Proof: Proof is similar to the proof in Theorem 7(a).

Theorem 9(a) can also be rewritten in terms of the L -bivariate association coefficient.

Theorem 9(b)

$$L(\vec{X}_1, (\vec{X}_2, \dots, \vec{X}_n)) = 1,$$

$$L(\vec{X}_2, (\vec{X}_3, \dots, \vec{X}_n)) = 1,$$

.....

$$L(\vec{X}_{n-1}, \vec{X}_n) = 1$$

if and only if

$$L(\vec{X}_i, \vec{X}_j) = 1 \quad (i < j \text{ and } i, j = 1, 2, \dots, n).$$

Proof: This is obvious from Theorem 5.

In this section, the L -bivariate association coefficient is defined that measures the association among two (groups of) random variables. In order to measure the association among three random variables, two L -bivariate association coefficients are needed. In general, $n-1$ L -bivariate association coefficients are needed to measure the association among n random variables. As shown, many different combinations of the L -bivariate association coefficients can measure the n (≥ 3) random variables in which each combination of the L -bivariate association coefficients measures a different aspect of the given system of n random variables. When the systems of multiple random variables are compared, these multivalued measures may cause confusion. It is obvious that the L -bivariate association coefficient needs to be improved upon or extended to measure the association for the given system of the multiple random variables. In the next section, a new concept called the L -multivariate association coefficient is defined that measures the degree of the association for a system of multiple random variables.

2 L -Multivariate Association Coefficient

The central issue in multivariate statistics is to study the relation among random variables. Traditionally the relation among the random variables means the relation between two random variables or two groups of random variables. In the real world, the relation not only exists between two objects, but it also exists for multiple objects. Sometimes people need to measure the relation in a system of multiple objects, which is called the association among the multiple variables. One example of this sort of system is the solar system, which has ten big objects - *one star and nine*

planets. In this section, a new concept called the L -multivariate association coefficient is defined that measures the association among the multiple random variables, and then some fundamental properties of the L -multivariate association coefficient are explored.

Definition 5.

$$L(\vec{X}_1, \dots, \vec{X}_n) = \frac{H(\vec{X}_1) + \dots + H(\vec{X}_n) - H(\vec{X}_1, \dots, \vec{X}_n)}{(n-1)H(\vec{X}_1, \dots, \vec{X}_n)}. \quad (n \geq 2) \quad (25)$$

The L -multivariate association coefficient for the system of the random variables $\vec{X}_1, \dots, \vec{X}_n$ is represented by (25).

It is obvious that, in case of $n = 2$, the L -multivariate association coefficient is actually the L -bivariate association coefficient. Also, from Definition 5, the L -multivariate association coefficient does not depend on the order of the sequence of the random variables $\vec{X}_1, \dots, \vec{X}_n$:

$$L(\vec{X}_1, \dots, \vec{X}_n) = L(\vec{X}_{i_1}, \dots, \vec{X}_{i_n})$$

where

$$(i_1, \dots, i_n) \text{ is a permutation of } (1, 2, \dots, n).$$

Similar to the L -bivariate association coefficient, the L -multivariate association coefficient ranges between zero and one.

Theorem 10

$$0 \leq L(\vec{X}_1, \dots, \vec{X}_n) \leq 1. \quad (26)$$

Proof: First, there are

$$H(\vec{X}_1, \dots, \vec{X}_n) \leq H(\vec{X}_1) + \dots + H(\vec{X}_n),$$

and

$$H(\vec{X}_1, \dots, \vec{X}_n) \geq 0.$$

By Definition 5,

$$L(\vec{X}_1, \dots, \vec{X}_n) = \frac{H(\vec{X}_1) + \dots + H(\vec{X}_n) - H(\vec{X}_1, \dots, \vec{X}_n)}{(n-1)H(\vec{X}_1, \dots, \vec{X}_n)} \geq 0.$$

On the other hand, for any $i \in \{1, \dots, n\}$,

$$H(\vec{X}_i) \leq H(\vec{X}_1, \dots, \vec{X}_n).$$

There is

$$\sum_{i=1}^n H(\vec{X}_i) \leq n \times H(\vec{X}_1, \dots, \vec{X}_n).$$

Therefore,

$$\sum_{i=1}^n H(\vec{X}_i) - H(\vec{X}_1, \dots, \vec{X}_n) \leq (n-1) \times H(\vec{X}_1, \dots, \vec{X}_n).$$

From Definition 5,

$$L(\vec{X}_1, \dots, \vec{X}_n) = \frac{\sum_{i=1}^n H(\vec{X}_i) - H(\vec{X}_1, \dots, \vec{X}_n)}{(n-1)H(\vec{X}_1, \dots, \vec{X}_n)} \leq 1.$$

Theorem 10 shows that the L -multivariate association coefficient is a normalized measure. This makes it possible for to compare the association among multiple random variables in the different systems.

Theorem 11

$$L(\vec{X}_1, \dots, \vec{X}_n) = 0 \tag{27}$$

$$\tag{28}$$

if and only if

$$p(\vec{x}_{i_1}, \dots, \vec{x}_{i_n}) = p(\vec{x}_{i_1}) \times \dots \times p(\vec{x}_{i_n}) \tag{29}$$

where $p(\vec{x}_{i_1}, \dots, \vec{x}_{i_n})$ ($i_1 = 1, \dots, m_1, \dots, i_n = 1, \dots, m_n$) are the joint probabilities of the random variables $\vec{X}_1, \dots, \vec{X}_n$ and $p(\vec{x}_i)$ ($i = 1, \dots, m_i$) are the probabilities of the random variable \vec{X}_i .

Proof:

$$L(\vec{X}_1, \dots, \vec{X}_n) = 0,$$

\Leftrightarrow

$$H(\vec{X}_1) + \dots + H(\vec{X}_n) - H(\vec{X}_1, \dots, \vec{X}_n) = 0,$$

\Leftrightarrow

$$\sum_{i_1} \dots \sum_{i_n} p(\vec{x}_{i_1}, \dots, \vec{x}_{i_n}) \ln \frac{p(\vec{x}_{i_1}, \dots, \vec{x}_{i_n})}{p(\vec{x}_{i_1}) \dots p(\vec{x}_{i_n})} = 0. \quad (30)$$

By Kullback-Leibler divergence theory (Kullback, 1959), (30) holds if and only if

$$p(\vec{x}_{i_1}, \dots, \vec{x}_{i_n}) = p(\vec{x}_{i_1}) \dots p(\vec{x}_{i_n}).$$

This is the proof of Theorem 11.

Theorem 11 shows that $L(\vec{X}_1, \dots, \vec{X}_n) = 0$ if and only if $\vec{X}_1, \dots, \vec{X}_n$ are (jointly) independent. Therefore, $L(\vec{X}_1, \dots, \vec{X}_n) = 0$ is equivalent to the independent state of the system of random variables $\vec{X}_1, \dots, \vec{X}_n$.

Theorem 12(a)

$$L(\vec{X}_1, \dots, \vec{X}_n) = 1, \quad (31)$$

$$(32)$$

if and only if

$$K(\vec{X}_i : \vec{X}_j) = 1 \quad (i \neq j \text{ and } i, j = 1, 2, \dots, n). \quad (33)$$

Proof: First, assume

$$L(\vec{X}_1, \dots, \vec{X}_n) = 1.$$

\Leftrightarrow

$$\sum_{i=1}^n H(\vec{X}_i) = n \times H(\vec{X}_1, \dots, \vec{X}_n).$$

 \Leftrightarrow

$$\sum_{i=1}^n (H(\vec{X}_1, \dots, \vec{X}_n) - H(\vec{X}_i)) = 0. \quad (34)$$

Also, for any $i = 1, \dots, n$, there is

$$H(\vec{X}_1, \dots, \vec{X}_n) - H(\vec{X}_i) \geq 0.$$

By (34), there is

$$H(\vec{X}_1, \dots, \vec{X}_n) = H(\vec{X}_i).$$

Therefore, for any $i \neq j$,

$$H(\vec{X}_j) \leq H(\vec{X}_i, \vec{X}_j) \leq H(\vec{X}_1, \dots, \vec{X}_n) = H(\vec{X}_j).$$

 \Rightarrow

$$H(\vec{X}_j) = H(\vec{X}_i, \vec{X}_j).$$

 \Leftrightarrow

$$H(\vec{X}_i | \vec{X}_j) = 0.$$

 \Leftrightarrow

$$K(\vec{X}_i : \vec{X}_j) = 1.$$

Second, assume

$$K(\vec{X}_i : \vec{X}_j) = 1 \quad \text{for any } i \neq j.$$

 \Leftrightarrow

$$H(\vec{X}_i | \vec{X}_j) = 0 \quad \text{for any } i \neq j.$$

Therefore, for any $i = 1, \dots, n$,

$$\begin{aligned}
& H(\vec{X}_1, \dots, \vec{X}_{i-1}, \vec{X}_{i+1}, \dots, \vec{X}_n | \vec{X}_i) \\
& \leq H(\vec{X}_1 | \vec{X}_i) + \dots + H(\vec{X}_{i-1} | \vec{X}_i) + H(\vec{X}_{i+1} | \vec{X}_i) + \dots + H(\vec{X}_n | \vec{X}_i) \\
& = 0 + \dots + 0 + 0 + \dots + 0 = 0.
\end{aligned}$$

This implies

$$H(\vec{X}_1, \dots, \vec{X}_{i-1}, \vec{X}_{i+1}, \dots, \vec{X}_n | \vec{X}_i) = 0.$$

\implies

$$\begin{aligned}
& H(\vec{X}_1, \dots, \vec{X}_n) \\
& = H(\vec{X}_i) + H(\vec{X}_1, \dots, \vec{X}_{i-1}, \vec{X}_{i+1}, \dots, \vec{X}_n | \vec{X}_i) \\
& = H(\vec{X}_i).
\end{aligned}$$

This is

$$H(\vec{X}_1, \dots, \vec{X}_n) = H(\vec{X}_i).$$

\implies

$$H(\vec{X}_1) + \dots + H(\vec{X}_n) = n \times H(\vec{X}_1, \dots, \vec{X}_n).$$

Therefore,

$$L(\vec{X}_1, \dots, \vec{X}_n) = \frac{H(\vec{X}_1) + \dots + H(\vec{X}_n) - H(\vec{X}_1, \dots, \vec{X}_n)}{(n-1)H(\vec{X}_1, \dots, \vec{X}_n)} = 1.$$

This is the proof of Theorem 12(a).

Theorem 12(a) shows that $L(\vec{X}_1, \dots, \vec{X}_n) = 1$ if and only if the random variables $\vec{X}_1, \dots, \vec{X}_n$ totally depend on each other. In the case of $L(\vec{X}_1, \dots, \vec{X}_n) = 1$, any variable in $\vec{X}_1, \dots, \vec{X}_n$ totally determines all of the other variables, or all of the other variables totally depend on this variable. In terms of information, the random variables $\vec{X}_1, \dots, \vec{X}_n$ can be thought to be a single random variable in the case of $L(\vec{X}_1, \dots, \vec{X}_n) = 1$. One special case is $L(\vec{X}, \dots, \vec{X}) = 1$, which is from

$$\begin{aligned}
& L(\vec{X}, \dots, \vec{X}) \\
&= \frac{H(\vec{X}) + \dots + H(\vec{X}) - H(\vec{X}, \dots, \vec{X})}{(n-1) \times H(\vec{X}, \dots, \vec{X})} \\
&= \frac{n \times H(\vec{X}) - H(\vec{X})}{(n-1) \times H(\vec{X})} = 1.
\end{aligned}$$

Theorem 12(a) can also be expressed in terms of the L -bivariate association coefficient.

Theorem 12(b)

$$L(\vec{X}_1, \dots, \vec{X}_n) = 1, \tag{35}$$

$$\tag{36}$$

if and only if

$$L(\vec{X}_i, \vec{X}_j) = 1 \quad (i < j \text{ and } i, j = 1, 2, \dots, n). \tag{37}$$

Proof: This is obvious from Theorem 5.

Now, consider a system of n random variables, which is denoted by

$$S = \{\vec{X}_1, \dots, \vec{X}_n\}. \tag{38}$$

The system S is divided into two subsystems, which are

$$S_1 = \{\vec{X}_{i_1}, \dots, \vec{X}_{i_{n_1}}\} \subseteq S,$$

$$S_2 = \{\vec{X}_{j_1}, \dots, \vec{X}_{j_{n_2}}\} \subseteq S$$

where

$$n_1 + n_2 = n,$$

$$S_1 \cup S_2 = S,$$

$$S_1 \cap S_2 = \phi.$$

Theorem 13(a)

$$L((S_1), (S_2)) = 0 \quad L(S_1) = 0 \quad L(S_2) = 0, \quad (39)$$

if and only if

$$p(\vec{x}_1, \dots, \vec{x}_n) = p(\vec{x}_1) \times \dots \times p(\vec{x}_n). \quad (40)$$

Proof: First, assume (39) holds. From Theorem 4,

$$L((S_1), (S_2)) = 0$$

\iff

$$\begin{aligned} & p(\vec{x}_1, \dots, \vec{x}_n) \\ &= p(\vec{x}_{i_1}, \dots, \vec{x}_{i_{n_1}}, \vec{x}_{j_1}, \dots, \vec{x}_{j_{n_2}}) \\ &= p(\vec{x}_{i_1}, \dots, \vec{x}_{i_{n_1}}) \times p(\vec{x}_{j_1}, \dots, \vec{x}_{j_{n_2}}). \end{aligned}$$

According to Theorem 11, there are

$$L(S_1) = 0.$$

\iff

$$\begin{aligned} & p(\vec{x}_{i_1}, \dots, \vec{x}_{i_{n_1}}) \\ &= p(\vec{x}_{i_1}) \times \dots \times p(\vec{x}_{i_{n_1}}). \end{aligned}$$

And,

$$L(S_2) = 0.$$

\iff

$$\begin{aligned} & p(\vec{x}_{j_1}, \dots, \vec{x}_{j_{n_2}}) \\ &= p(\vec{x}_{j_1}) \times \dots \times P(\vec{x}_{j_{n_2}}). \end{aligned}$$

Therefore,

$$\begin{aligned} & p(\vec{x}_1, \dots, \vec{x}_n) \\ &= p(\vec{x}_{i_1}, \dots, \vec{x}_{i_{n_1}}) \times P(\vec{x}_{j_1}, \dots, \vec{x}_{j_{n_2}}) \\ &= p(\vec{x}_{i_1}) \times \dots \times P(\vec{x}_{i_{n_1}}) \times p(\vec{x}_{j_1}) \times \dots \times p(\vec{x}_{j_{n_2}}) \\ &= p(\vec{x}_1) \times \dots \times p(\vec{x}_n). \end{aligned}$$

Second, assume

$$p(\vec{x}_1, \dots, \vec{x}_n) = p(\vec{x}_1) \times \dots \times p(\vec{x}_n)$$

\Leftrightarrow

$$p(\vec{x}_1, \dots, \vec{x}_n) = p(\vec{x}_{i_1}, \dots, \vec{x}_{i_{n_1}}) \times p(\vec{x}_{j_1}, \dots, \vec{x}_{j_{n_2}})$$

$$p(\vec{x}_{i_1}, \dots, \vec{x}_{i_{n_1}}) = P(\vec{x}_{i_1}) \times \dots \times P(\vec{x}_{i_{n_1}})$$

$$p(\vec{x}_{j_1}, \dots, \vec{x}_{j_{n_2}}) = P(\vec{x}_{j_1}) \times \dots \times P(\vec{x}_{j_{n_2}}).$$

\Leftrightarrow

$$L((S_1), (S_2)) = 0,$$

$$L(S_1) = 0,$$

$$L(S_2) = 0.$$

This is the proof of Theorem 13(a).

Theorem 13(a) can be rewritten in terms of the L -multivariate association coefficient.

Theorem 13(b)

$$L((S_1), (S_2)) = 0 \quad L(S_1) = 0 \quad L(S_2) = 0, \tag{41}$$

if and only if

$$L(S) = 0. \tag{42}$$

Proof: This is obvious from Theorem 11.

Theorem 14(a)

$$L((S_1), (S_2)) = 1 \quad L(S_1) = 1 \quad L(S_2) = 1 \tag{43}$$

if and only if

$$K(\vec{X}_i : \vec{X}_j) = 1 \quad (i \neq j \text{ and } i, j = 1, 2, \dots, n). \tag{44}$$

Proof: First, assume (43) holds. According to Theorem 12(b), for any \vec{X}_i and $\vec{X}_j \in S_1$ where $i \neq j$, there is

$$L(S_1) = 1,$$

\Leftrightarrow

$$K(\vec{X}_i : \vec{X}_j) = 1.$$

Similarly, for any \vec{X}_i and $\vec{X}_j \in S_2$ where $i \neq j$, there is

$$L(S_2) = 1,$$

\Leftrightarrow

$$K(\vec{X}_i : \vec{X}_j) = 1.$$

By Definition 4,

$$L((S_1), (S_2)) = 1,$$

\Leftrightarrow

$$H(S_1|S_2) = 0,$$

$$H(S_2|S_1) = 0.$$

\Leftrightarrow

$$H(\vec{X}_1, \dots, \vec{X}_n) = H(S_1) + H(S_2|S_1) = H(S_1),$$

$$H(\vec{X}_1, \dots, \vec{X}_n) = H(S_2) + H(S_1|S_2) = H(S_2).$$

Also,

$$L(S_1) = 1,$$

\Rightarrow

$$H(S_1) = H(\vec{X}_{i_1}) = \dots = H(\vec{X}_{i_{n_1}}).$$

Similarly, there is

$$H(S_2) = H(\vec{X}_{j_1}) = \cdots = H(\vec{X}_{j_{n_2}}).$$

Therefore,

$$\begin{aligned} H(\vec{X}_1, \dots, \vec{X}_n) &= H(S_1) = H(\vec{X}_{i_1}) = \cdots = H(\vec{X}_{i_{n_1}}), \\ H(\vec{X}_1, \dots, \vec{X}_n) &= H(S_2) = H(\vec{X}_{j_1}) = \cdots = H(\vec{X}_{j_{n_2}}). \end{aligned}$$

This implies

$$H(\vec{X}_1, \dots, \vec{X}_n) = H(\vec{X}_1) = \cdots = H(\vec{X}_n).$$

Therefore, for any $i \neq j$,

$$H(\vec{X}_j) \leq H(\vec{X}_i, \vec{X}_j) \leq H(\vec{X}_1, \dots, \vec{X}_n) = H(\vec{X}_j),$$

\implies

$$H(\vec{X}_j) = H(\vec{X}_i, \vec{X}_j),$$

\iff

$$H(\vec{X}_i | \vec{X}_j) = 0,$$

\iff

$$K(\vec{X}_i : \vec{X}_j) = 1.$$

Second, assume

$$K(\vec{X}_i, \vec{X}_j) = 1 \quad (i \neq j \text{ and } i, j = 1, 2, \dots, n).$$

For any \vec{X}_i and $\vec{X}_j \in S_1$ where $i \neq j$, there is

$$K(\vec{X}_i : \vec{X}_j) = 1,$$

\iff

$$L(S_1) = 1.$$

Similarly, for any \vec{X}_i and $\vec{X}_j \in S_2$ where $i \neq j$, there is

$$K(\vec{X}_i : \vec{X}_j) = 1,$$

$$\iff$$

$$L(S_2) = 1.$$

For the case that $\vec{X}_i \in S_1$ and $\vec{X}_j \in S_2$,

$$K(\vec{X}_i : \vec{X}_j) = 1,$$

$$K(\vec{X}_j : \vec{X}_i) = 1.$$

$$\iff$$

$$H(\vec{X}_i | \vec{X}_j) = 0,$$

$$H(\vec{X}_j | \vec{X}_i) = 0.$$

Therefore,

$$\begin{aligned} H(S_1 | S_2) &\leq H(\vec{X}_{i_1} | S_2) + \cdots + H(\vec{X}_{i_{n_1}} | S_2) \\ &\leq H(\vec{X}_{i_1} | \vec{X}_j) + \cdots + H(\vec{X}_{i_{n_1}} | \vec{X}_j) \\ &= 0 + \cdots + 0 = 0. \end{aligned}$$

$$\implies$$

$$H(S_1 | S_2) = 0.$$

Similarly, there is

$$H(S_2 | S_1) = 0,$$

$$\iff$$

$$K(S_1 : S_2) = 1,$$

$$K(S_2 : S_1) = 1.$$

$$\iff$$

$$L((S_2), (S_1)) = 1.$$

This is the proof of Theorem 14(a).

Also, Theorem 14(a) can be expressed in terms of the L -bivariate association coefficient.

Theorem 14(b)

$$L((S_1), (S_2)) = 1 \quad L(S_1) = 1 \quad L(S_2) = 1 \quad (45)$$

$$(46)$$

if and only if

$$L(\vec{X}_i, \vec{X}_j) = 1 \quad (i < j \text{ and } i, j = 1, 2, \dots, n). \quad (47)$$

Proof: This is obvious from Theorem 5.

As a normalized measure, the L -multivariate association coefficient $L(\vec{X}_1, \dots, \vec{X}_n)$ measures the degree of the association among the random variables $\vec{X}_1, \dots, \vec{X}_n$. If the random variables $\vec{X}_1, \dots, \vec{X}_n$ are considered as a system, the L -multivariate association coefficient measures the degree of the association for this system. $L(\vec{X}_1, \dots, \vec{X}_n) = 0$ means that this system is in the independent state, while $L(\vec{X}_1, \dots, \vec{X}_n) = 1$ means that this system is in a state in which the whole system is equivalent to any one of the variables in this system in terms of information. With the L -multivariate association coefficient, the different systems of the random variables are comparable to each other in terms of their associations. Therefore, these systems can be classified according to their associations.

3 Discussion

Traditionally, the concept of the correlation coefficient among the random variables is the one that measures the relation between two (groups of) random variables (Agresti, 2002; Altham, 1970; Burg & Lewis, 1988; Goodman & Kruskal, 1954; Haberman, 1980; Theil, 1970). For a system S of multiple random variables in (38), the association among these random variables $\vec{X}_1, \dots, \vec{X}_n$ includes the relations between the different individual variables, the relations

between the different subgroups or subsystems of the random variables, and the interactions of the different combinations of these variables. If the association of the system is zero, that is, there is no association among the random variables in this system, then this system is actually in the independent state. The mathematical expression for a system of independent state is

$$p(\vec{x}_1, \dots, \vec{x}_n) = p(\vec{x}_1) \times \dots \times p(\vec{x}_n),$$

or

$$L(\vec{X}_1, \dots, \vec{X}_n) = 0.$$

On the other hand, the system could be reduced to a single variable in terms of information. That is, any individual variable could totally determine any other individual variable and also totally depend on any other individual variable. In this case, the system is said to be in the state of perfect association. The mathematical expression for a system of perfect association is

$$L(\vec{X}_i, \vec{X}_j) = 1 \quad (i < j \text{ and } i, j = 1, 2, \dots, n),$$

or

$$L(\vec{X}_1, \dots, \vec{X}_n) = 1.$$

The general case for the system S should be between the perfect association and independent state with the mathematical expression of

$$0 \leq L(\vec{X}_1, \dots, \vec{X}_n) \leq 1.$$

If the system is in perfect association, the information carried by the multiple variables $\vec{X}_1, \dots, \vec{X}_n$ is nothing but the information carried by any individual variable among $\vec{X}_1, \dots, \vec{X}_n$. That is, without losing any information, only an individual variable among $\vec{X}_1, \dots, \vec{X}_n$ is needed instead of keeping all of these variables. Therefore, the L -multivariate association coefficient can be used for variables reduction. Given a threshold value, say 0.95, all of the variables are classified into several groups in which the L -multivariate association coefficient is larger than the given

threshold value, and then one variable is selected from each group. The selected variables will carry the most information of the original variables.

If the system is in an independent state, no information is shared among the variables in this system, that is, there is no relation among the variables in the system. Actually the system can be split into n independent subsystems in which each subsystem contains one variable. In practice, if the L -multivariate association coefficient is fairly small, say less than 0.01, the several groups of variables in the system can be split into the several corresponding subsystems.

It is also noted that the association measured by the L -multivariate association coefficient is a general type of association, not any specific type of a linear or nonlinear association. The L -multivariate association coefficient fully utilizes the information carried by the variables $\vec{X}_1, \dots, \vec{X}_n$ to measure the association among these variables. This is the basic difference between the L -multivariate association coefficient and linear correlation.

In numerical analysis, the L -multivariate association coefficient can be directly calculated from the joint probability distribution $p(\vec{x}_1, \dots, \vec{x}_n)$. In practice, this joint probability distribution may not be estimable in the case of small sample size. Small sample size does not mean a small amount of information (large sample size does not mean a large amount of information). How to fully utilize the information in a data set to estimate the L -multivariate association coefficient is a topic for the future study.

References

- Agresti, A. (2002). *Categorical data analysis* (pp. 36-59). New York: John Wiley & Sons, Inc.
- Altham, P. M. E. (1970). The measurement of association of rows and columns for an $r \times s$ contingency table. *Journal of the Royal Statistical Society, Series B*, 3263–73.
- Burg, W., & Lewis, C. (1988). Some properties of two measures of multivariate association. *Psychometrika*, 53, 109 – 122.
- Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classification. *Journal of the American Statistical Association*, 49, 732–764.
- Haberman, S. J. (1980). Measures of association. *Encyclopedia of statistical sciences*, (vol. 1, pp. 130–137). New York: John Wiley & Sons, Inc.
- Kong, L. N. (2007a). *An entropy-based measure of dependence between two groups of random variables* (ETS Research Rep. No. RR-07-20). Princeton, NJ: ETS.
- Kong, L. N. (2007b). An entropy-based measure of dependence of one set of random variables on another set of random variables. In *2007 JSM proceedings*, Nonparametric statistics section [CD]. Alexandria, VA: American Statistical Association.
- Kullback, S. (1959). *Information theory and statistics*. New York: John Wiley & Sons, Inc.
- Shannon, C. E. (1948). A mathematical theory of communications. *Bell System Technical Journal*, 27, 379–423, 623–656.
- Theil, H. (1970). On the estimation of relationships involving qualitative variables. *American Journal of Sociology*, 76, 103–154.