On Mutual Information for Elliptical Distributions: A Case of Nonlinear Dependence of ‘n’ Vectors

Soyinka A. Taiwo\textsuperscript{1,*}, Olosunde A. Akin\textsuperscript{2} and Asiribo E. Osebi\textsuperscript{3}

\textsuperscript{1}Department of Research and Training, Federal Neuropsychiatric Hospital Aro, Abeokuta PMB 2002, Nigeria.

\textsuperscript{2}Department of Mathematics, Obafemi Awolowo University, Ile-ife 220005, Nigeria.

\textsuperscript{3}Department of Statistics, Federal University of Agriculture, Abeokuta PMB 2022, Nigeria.

\*Corresponding Author: soyinkataiwo@yahoo.ie

Received: 3\textsuperscript{rd} November 2016/ Revised: 17\textsuperscript{th} May 2017/ Accepted: 11\textsuperscript{th} July 2017

©IAppStat-SL2017

\textbf{ABSTRACT}

In this paper, we modeled dependent categorical data via mutual information concept to obtain the measure of statistical dependence. We first derive the entropy and mutual information index for exponential power distribution. These concepts are important and were developed by Shannon in the context of information theory. Several literatures are already published in the case of the multivariate normal distribution. Then we extend these tools to the special case of a full symmetric of multivariate elliptical distributions. The upper bound for the entropy which is attained for the normal density is established. We further derived the nonlinear joint model for dependent random vectors that spans an elliptical vector space to enhance multivariate relationships among non-empty subsets of vectors via multivariate mutual information; based on the assumption that the subsets of each vector and their interactions can be represented in discrete form. To illustrate its application, the multivariate dependency among various sites based on dominance of some attributes were investigated.

\textbf{Keywords}: Dependent vectors, mutual information, matrix equivalent, covariance structures, nonlinear dependency, elliptical space.

\section{1. Introduction}

Information theory is a branch of mathematical theory of probability and statistics introduced by Shannon (1948) who used mathematical theory in communication to describe logarithmic measures of information and has
stimulated a tremendous amount of study in engineering fields on the subject of information theory. It is also the branch of applied probability and statistics that is relevant to statistical inference and therefore should be of basic interest to statisticians Kullback, (1978). Information theory seeks the quantification of information. One goal of information theory is the development of coding schemes that provide good performance in comparison with the optimal performance given by the theory. It works under the assumption of a strongly stationary random process to define an information quantity contained in a multivariate probability density function, for example, the multivariate normal distribution (Kullback, 1978; Silva & Quiroz, 2003; Misra et al., 2005; Cover & Thomas, 1991). This quantity allows to measure the cumulative information of a multivariate data set, or more specifically, to quantify the mutual information between two random variables or vectors. On the other hand, the entropy is a notion of information provided by a random process about itself and it is sufficient to study the reproduction of a marginal process through a noiseless environment. For a systematic and comprehensive account of these and related concepts, see, for example, Cover & Thomas (1991). Also comprehensive review of mathematical expectation and properties of single entropy, joint entropy, conditional entropy and maximum entropy $H(\cdot)$ in discrete and continuous case are available in various texts see Shannon (1948), Tho Pharm et. al. (2012), Reginald (2015) etc. It is worth noting that there are various measures of dependence with the sole aim of obtaining an estimate of the degree of dependence between two or more random vectors. However dependence relationship among random vectors which could be linear or nonlinear has been a frequent but often incomplete goal because of its complexity. Despite this, the frequency of its necessity in cluster analysis, cryptography, data mining, networking and imaging and so on has made its further development inevitable. Though linear correlation as a measure of degree of dependence has a clear interpretation on linear dependence, it’s under estimation of the nonlinear dependence is a setback that has restricted its usage. Among the various measures of dependence, mutual information (MI) is the most comprehensive of all due to its non-parametric nature; but its inability to give a clear cut interpretation has made its usage in data analysis relatively rare. Note that since most real life problems have varying vectors (variables) whose individual outcomes has collective conditional joint effect on each other due to complicated interactions ‘e.g molecular interactions in biological networks’; then a clear multivariate representation of vector interactions must be used to define in clear terms total and conditional multivariate mutual information (MMI) dependence [Tho Pharm et. al. (2012), Reginald (2015)]. Therefore, though there is no straight forward
interpretation of mutual information; there is need to develop a “nonlinear mutual information measures of dependence on well defined non parametrics assumptions”. Hence this study starts by developing the properties of entropy and expressing multivariate mutual information (MMI) in matrix form to reveal its analogy with symmetric covariance matrix. The study then obtain dependent vectors multivariate model via squared radial part of the elliptical contour distribution. The obtained model accommodated for varying degrees of linear and nonlinear dependence within a closed interval \([0,1]\) while establishing independence when \(p(\chi^2_{\text{nonlinear}}) = p(\chi^2_{\text{linear}})\).

2. Mutual Information for Exponential Power Distribution

Following the concepts of entropy and mutual information index, starting with the unions of two or more sets; the MMI to capture the amount of information shared between two sets (input and output) and multiple sets (multiple inputs and single output) can be expressed mathematically.

**Definition 2.1:** The mutual information between two random vectors \(X \in \mathbb{R}^n\) and \(Y \in \mathbb{R}^m\) with joint and probability density functions \(f_{X,Y}(x,y)\), \(f_X(x)\) and \(f_Y(y)\) respectively is

\[
I(X,Y) = H(X) + H(Y) - H(X,Y)
\]

The mutual information index between \(X\) and \(Y\) is defined by

\[
\int_{\mathbb{R}^n} \int_{\mathbb{R}^m} f_{X,Y}(x,y) \ln \frac{f_{X,Y}(x,y)}{f_X(x)f_Y(y)} \, dx \, dy = E \left[ \ln \left( \frac{P_{X,Y}(x,y)}{f_X(x)f_Y(y)} \right) \right]
\]

**Definition 2.2:** By extension Sunil (2003); the MMI among three random vectors \(X, Y\) and \(Z\) is

\[
I(X,Y,Z) = H(X) + H(Y) + H(Z) - H(X,Y) - H(X,Z) - H(Y,Z) + H(X,Y,Z)
\]

**Definition 2.3:**[Han (1980)] The conditional multivariate mutual information (CMMI) for \(n\) dependent vectors

\[
I(X_1,X_2,...,X_{n-1}|X_n) = I(X_1,X_2,...,X_{n-1},X_n) - I(X_1,X_2,...,X_{n-1})
\]

Implying that for one input and one output MI structure we have \(I(X|Y) = I(X) - I(X,Y)| = H(X) - I(X,Y) = H(X,Y) - H(Y)\) while two inputs and one output MI has the structure \(I(X,Y|Z) = I(X,Y) - I(X,Y,Z) = H(X,Z) + H(Y,Z) - H(Z) - H(X,Y,Z)\).

2.1 The Exponential Power Distribution

**Definition 2.4:** The random variable \(X\) is said to have univariate exponential power if
Where β is the shape parameter, µ and σ are location and scale parameters respectively, such that when β=2 we have the usual normal distribution and β, Laplace distribution. Therefore, we could say that the exponential power distribution has Laplace and normal distribution as special cases. The shape parameter β regulates the tail of the distribution which made it more flexible symmetric model.

Subsequently we have the following proposition;

**Proposition 2.1**: A random variable X defined as in (5) above has

\[ H(X) = \ln \left\{ 2\sigma^\beta \beta^{1/\beta} \Gamma\left(1 + \frac{1}{\beta}\right) \right\} + \frac{1}{\beta} \]

Proof. Given \( H(X) = -\int_R f(x) \ln f(x) dx \) then taking f(x) to be (5) we have that

\[ H(X) = E \left\{ \frac{-|x-\mu|^{\beta}}{\beta \sigma^{\beta}} \right\} - E \left\{ \ln \left( 2\sigma^\beta \beta^{1/\beta} \Gamma\left(1 + \frac{1}{\beta}\right) \right) \right\} \]

and the result (6) follows

**2.2 The Multivariate Extension to Exponential Power Distribution**

Definition 2.5} {(Elliptical Multivariate Density)}: Let \( X \sim EC_{(n)} = [\mu, \sigma, \beta, h^n] \) be an elliptical contour (EC) random vector in \( \mathbb{R}^n \) with location \( \mu \in \mathbb{R}^n \) dispersion matrix \( \Sigma \in \mathbb{R}^{n \times n} \), shape parameter \( \beta \in \mathbb{R} \) and density generator \( h^n \in \mathbb{R}^n \); then the probability density function of the elliptical contour has the general form

\[ c_{d,f} \left| \Sigma \right|^{-\frac{1}{2}} h^n \left( \left(x - \mu \right)^\beta \Sigma^{-1} \left(x - \mu \right)^{\beta/2} \right); \ X \in \mathbb{R}^n \]. Here, the density generator \( h^n \) is a non-negative real valued function from \( c_{d,f} h^n(s) \) for \( s \succ 0 \) which makes the function a valid probability density function (Arellano-valley (2012)). Therefore, the multivariate extension of (5) in the absolute continuous case has the density
Nonlinear dependence measure of ‘n’ vectors

\[ f(x_1, x_2, ..., x_n) = \left| \Sigma \right|^{-\frac{1}{2}} \frac{\Gamma\left(\frac{n+\frac{1}{2}}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{\pi^{n/2}}{\beta^n} e^{-\frac{1}{\beta} \sum \left(x - \mu^T \Sigma^{-1} x - \mu\right)^2} \]  \quad (7)

The univariate normal density has been well known to have the maximum entropy subject to the condition that mean and variance are fixed. The similar result is true for its multivariate extension. Now, let f(x) be the density function of a n-dimensional random variable f(x) from EPD. We maximize

\[- \int f(x) \ln f(x) dv \]  \quad (8)

where dv stands for the volume element, subject to the conditions that the mean and the dispersion matrix have given values; E(X)= \( \mu \) and Var(X)= \( \Sigma \) where \( \beta \) determine the kurtosis (Gomez, et. al., 1998). Thus the correlation structure can be obtained directly from \( \Sigma \) in the usual way. Let \( g(\{x\}) \) be alternative density from n-dimensional normal density given as;

\[ g(x) = (2\pi)^{-\frac{n}{2}} \left| \Sigma \right|^{-\frac{1}{2}} \exp\left\{ \frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\} \]  \quad (9)

From the information theory inequality we have for any of the two alternative densities \(- \int f(x) \ln f(x) dv \leq - \int f(x) \ln g(x) dv \) the inequality is attained when \( f(x) = g(x) \) almost everywhere dv; so we have

\[- \int f(x) \ln f(x) dv \leq - \int f(x) \left[ - \frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln \left| \Sigma \right| - \frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right] dv \]

\[= \frac{n}{2} \ln(2\pi) + \frac{1}{2} \ln \left| \Sigma \right| + \frac{1}{2} \text{trace} \left\{ \Sigma^{-1} \int (x - \mu)^T (x - \mu)^T f(x) dv \right\} = \]

\[= \frac{n}{2} \ln(2\pi) + \frac{1}{2} \ln \left| \Sigma \right| + \frac{2^{-\frac{1}{2}} \Gamma\left(\frac{n+\frac{1}{2}}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} trace \left\{ \Sigma^{-1} \Sigma \right\} \]  \quad (10)

The expression is the upper bound to entropy which is attained for the normal density. The entropy of the distribution so determined is then;
Proposition 2.2: The multivariate joint entropy for n vectors \( H(x_1, x_2, ..., x_n) \) is

\[
H(x_1, x_2, ..., x_n) = \left\{ 2K\beta^{\frac{1}{\beta}}|\Sigma|^{\frac{1}{2}}\Gamma\left( 1 + \frac{1}{\beta} \right) \right\} \left\{ \frac{1}{2} \ln|\Sigma| + \ln \left( \frac{n^{\beta/2} \Gamma(1+n/\beta) \beta^{n/\beta}}{\pi^{n/2} \Gamma(1+n/\beta)} \right) \right\} + \frac{1}{\beta} \right)
\]

where \( K = |\Sigma|^{-\frac{1}{2}} \frac{\Gamma(1+n/\beta)}{\pi^{n/2} \beta^{n/\beta}} \).

Proof. \( H(x_1, x_2, ..., x_n) = -E\{\ln f(x_1, x_2, ..., x_n)\} \).

For simplicity let

\[
H(x_1, x_2, ..., x_n) = A \left\{ \frac{1}{2} \ln|\Sigma| + \ln D + \frac{1}{\beta} \right\}
\]

Where

\[
A = \left\{ 2K\beta^{\frac{1}{\beta}}|\Sigma|^{\frac{1}{2}}\Gamma\left( 1 + \frac{1}{\beta} \right) \right\} = \frac{\Gamma(n/2)\Gamma(1/\beta)}{\pi^{n/2} \beta^{(n-1)/\beta} \Gamma(n/\beta)}
\]

and

\[
D = \left\{ \frac{n^{\beta/2} \Gamma(1+n/\beta) \beta^{n/\beta}}{\pi^{n/2} \Gamma(1+n/\beta)} \right\}.
\]

Note that from (13) we can obtain the MI for two vectors

\[
I(x_1, x_2) = H(x_1) + H(x_2) - H(x_1, x_2) = A \left\{ -\frac{1}{2} \ln \frac{\Sigma_{11} \Sigma_{22} - (\Sigma_{12})^2}{\Sigma_{11} \Sigma_{22}} + \ln D + \frac{1}{\beta} \right\} = A \left\{ -\frac{1}{2} \ln(1 - \rho_{12}^2) + \ln D + \frac{1}{\beta} \right\}
\]
Likewise for three vectors we have

\[
I(x_1, x_2, x_3) = A \left\{ \frac{1}{2} \ln(1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2) + (\rho_{12}\rho_{13})^2 + (\rho_{12}\rho_{23})^2 + (\rho_{13}\rho_{23})^2 \right\}
\]

\[
+ \ln D + \frac{1}{\beta}
\]

So for the conditional mutual information index in definition 2.3 we can obtain the best linear hierarchical relationship among input vectors and output vector say

\[
I(x_1, x_2 \mid x_3) = I(x_1, x_2) - I(x_1, x_2, x_3) = \left\{ \frac{1}{2} \ln(1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2) + (\rho_{12}\rho_{13})^2 + (\rho_{12}\rho_{23})^2 + (\rho_{13}\rho_{23})^2 \right\}
\]

\[
+ \ln D + \frac{1}{\beta}
\]

Hence the MMI index for n exponential power distributed vectors is

\[
I(x_1, x_2, \ldots, x_n) = A \left\{ \frac{1}{2} \ln|\Sigma_{n\times n}| \right\} + \ln D + \frac{1}{\beta}
\]

(14)

and it converges to the closed interval \( \{H(x), 1\} \) as n increases from 1 to \( \infty \). Thus \( I(x_1, x_2, \ldots, x_n) \) is always finite.

Note that (14) has two components, the linear part

\[
I(x_1, x_2, \ldots, x_n)_{\text{Linear}} = A \left\{ \frac{1}{2} \ln|\Sigma_{n\times n}| \right\}
\]

which is a measure of linear dependence/similarity/correlation coefficient and the nonlinear part

\[
I(x_1, x_2, \ldots, x_n)_{\text{Nonlinear}} = A \left\{ \ln D + \frac{1}{\beta} \right\}
\]

So the nonlinear dependent measure can be expressed as the difference between total dependence and the linear dependence

\[
I(x_1, x_2, \ldots, x_n)_{\text{Nonlinear}} = I(x_1, x_2, \ldots, x_n)_{\text{Total}} - I(x_1, x_2, \ldots, x_n)_{\text{Linear}}
\]

(Arellano-valley (2012); Reginald (2015); Brillinger and Guha (2006)).

However from

\[
A = \frac{\Gamma(n/2)\Gamma(1/\beta)}{\pi^{n/2} \beta^{(n-1)/2} \Gamma\left(\frac{n}{\beta}\right)}
\]

where n is the number of interacting vectors, then as n increases from 1 to \( \infty \); then A reveals a decreasing function
which converges to the limit \( \frac{1}{n} \) with properties \( A>0 \) and \( A \leq 1 \). So for simplicity if \( A=1 \) then the nonlinear dependence measure becomes
\[
I(x_1, x_2, \ldots, x_n)_{\text{Nonlinear}} = I(x_1, x_2, \ldots, x_n)_{\text{Total}} - \frac{1}{2} \ln|\Sigma_{nn}| \tag{15}
\]

3. Nonlinear Multivariate Model

3.1 Relationship between Covariance and MMI matrix structures

Definition 3.1: Given two random vectors \((X,Y)\) the covariance matrix \( \Sigma \) is given as
\[
\begin{pmatrix}
\Sigma_{xx} & \Sigma_{xy} \\
\Sigma_{yx} & \Sigma_{yy}
\end{pmatrix}.
\]
Likewise the MI structure for vectors \((X,Y)\) from definition (2.1) can be re-expressed in matrix form as
\[
e^{I(X,Y)} =
\begin{pmatrix}
e^{H(X)-\frac{1}{2}H(X,Y)} & 0 \\
0 & e^{H(Y)-\frac{1}{2}H(X,Y)}
\end{pmatrix}.
\]

Definition 3.2: Given three random vectors \((X,Y,Z)\); the covariance matrix \( \Sigma \) is given as
\[
\begin{pmatrix}
\Sigma_{xx} & \Sigma_{xy} & \Sigma_{xz} \\
\Sigma_{yx} & \Sigma_{yy} & \Sigma_{yz} \\
\Sigma_{zx} & \Sigma_{zy} & \Sigma_{zz}
\end{pmatrix}
\]
while its corresponding MMI matrix structure from definition (2.2) is
\[
e^{I(X,Y,Z)} =
\begin{pmatrix}
e^{H(X)-H(Y,Z)+\frac{1}{3}H(X,Y,Z)} & 0 & 0 \\
0 & e^{H(Y)-H(X,Z)+\frac{1}{3}H(X,Y,Z)} & 0 \\
0 & 0 & e^{H(Z)-H(X,Y)+\frac{1}{3}H(X,Y,Z)}
\end{pmatrix}.
\]

Remark: Note that the MMI in the matrices above are the index of the exponential.

Hence MMI and \( \Sigma \) has the exponential relationship \( e^{I(x_1, x_2, \ldots, x_p, x_p)} \propto \Sigma \) \([\text{Reginald (2015); Adrenallo et al.,(2012)}]\).

Given that the squared radial function in (7) is dependent on the in-built shape parameter \( \beta \) within the density generator function,
\[
s = \left| (x - \mu)^T \Sigma^{-1} (x - \mu) \right|^{\beta/2} \sim \Re(h^\beta) \tag{16}
\]
Hence for multivariate normally distributed vectors $\beta = 2$, we have the chi-squared radial part

$$s = (x - \mu)^T \Sigma^{-1} (x - \mu)^{\beta/2} \sim \chi^2_n$$

(17)

So re-writing (15) and substituting into (17) we obtain the absolute nonlinear squared radial function for multivariate normally distributed discrete interacting vectors as

$$s = \left| I(x_1, x_2, ..., x_n) - \frac{1}{2} \ln|\Sigma_{n_n}| \right| \sim \left| \ln \left( \chi^2_n \right) \left( (x - \mu)^T (x - \mu) \right)^{-1} \right|$$

(18)

Hence supposing $X$ and $Y$ are random vectors with discrete variates $j$ and $k$ respectively and total degrees of freedom $df = jk - 1$ in a $j \times k$ contingency table; then the nonlinear chi-square partitioning is given as

$$\chi^2_{Nonlinear,1-\alpha} = -\left( \ln \chi^2_{Total,jk-1-\alpha} - \ln \chi^2_{Independen,j+k-2,1-\alpha} - \ln \chi^2_{Linear,(j-1)(k-1),1-\alpha} \right)$$

(19)

So the nonlinear multivariate model for the measure of vectors dependence is given as

$$\chi^2_{calculated} = (x - \mu)^T \exp \left| I(x_1, x_2, ..., x_n) - \frac{1}{2} \ln|\Sigma_{n_n}| \right| (x - \mu)$$

(20)

Implying that

$$\chi^2_{calculated} = \exp \left| I(x_1, x_2, ..., x_n) \right| \left( (x - \mu)^T |\Sigma_{n_n}|^{1/2} (x - \mu) \right).$$

We can extend the partitioning to accommodate for many vectors interacting simultaneously. This approach can be used to analyze entropy and mutual information significance among discrete experimental data when the outcome is in frequency and the usual analysis of variance is not applicable.

### 3.2 Measure of Mutual information significance

From (18) considering the null hypothesis

$$H_0 = \left| I(x_1, x_2, ..., x_n) - \frac{1}{2} \ln|\Sigma_{n_n}| \right| = \delta = 0$$

against the alternative hypothesis

$$H_a = \left| I(x_1, x_2, ..., x_n) - \frac{1}{2} \ln|\Sigma_{n_n}| \right| = \delta > 0$$

then the test of nonlinear dependence significance can be ascertained via the probability function
\[ \Pr \left\{ (x - \mu)^T \exp \left[ I(x_1, x_2, \ldots, x_n) - \frac{1}{2} \ln \Sigma_{\mu} \right] (x - \mu) \geq \chi^2_n \right\} \leq 1 - \alpha \]  

(21)

at specific location parameter for known observations (Johnson and Wichern, 2006). This implies that the nonlinear estimate increases from \( \delta = 0 \) (total linear dependence) at \( \chi^2_{\text{calculated}} = (x - \mu)^T (x - \mu) \), to various proportional sizes of linear and nonlinear dependence measure beyond \( \delta = 0 \) and finally as \( \delta \) increases positively to infinity the \( \chi^2_{\text{calculated}} \) approaches zero signifying that vectors are totally nonlinearly dependent.

See Figure 1 in the appendix for illustration.

A: Linear Independent \( \Pr(\chi^2_{\text{calculated}} \geq \chi^2_n) \leq \frac{\alpha}{2} \);

B: Mix model but more Linear \( \Pr(\chi^2_{\text{calculated}} \geq \chi^2_n) \leq \frac{1}{2} - \alpha \);

C: Mix model but slightly Linear \( \Pr(\chi^2_{\text{calculated}} \geq \chi^2_n) \leq \frac{1}{2} - \frac{\alpha}{2} \);

D: Independent \( \Pr(\chi^2_{\text{calculated}} \geq \chi^2_n) \leq \left\{ \frac{1 - \alpha}{2}, \frac{1 + \alpha}{2} \right\} \);

E: Mix model but slightly Nonlinear \( \Pr(\chi^2_{\text{calculated}} \geq \chi^2_n) \leq \frac{1}{2} + \frac{\alpha}{2} \);

F: Mix model but more Nonlinear \( \Pr(\chi^2_{\text{calculated}} \geq \chi^2_n) \leq \frac{1}{2} + \alpha \);

G: Nonlinearly dependent \( \Pr(\chi^2_{\text{calculated}} \geq \chi^2_n) \leq 1 - \frac{\alpha}{2} \).

So far, we have established the measures of single entropy (6), upper entropy bound (11), multivariate entropy (13) and the multivariate mutual Information index (14) of the exponential power distribution. In addition, we have obtained the nonlinear multivariate model for the measure of vectors dependence (20) along with its significance (21).

Next, we demonstrate the usage of our result in cluster analysis of objects with various attributes/variables as properties arranged in a contingency table with frequency of attributes as output.
4. Application

Considering the archaeological site profile for different types of pottery in an area of the American Southwest (Johnson and Wichern, 2007 page 719 Applied Multivariate Statistical Analysis). We apply our result to estimating the nonlinear measure of similarities/dependence among sites based on the level of attributes dominance in each site analytically and graphically.

Solution: First ensure that the total frequency sum in the entire contingency table is equal to 1. Then obtain the following statistics

\[ H(x) = - \sum p(x_i) \log p(x_i); \]

\[ \Sigma_{x_ix_j} = \sum_{i=1} p(x_i) - \left[ \sum_{i=1} p(x_i) \right]^2; \]

\[ \Sigma_{x_ix_j} = \sum \left( p(x_i)p(x_j) \right) - \left[ \sum_{i=1} p(x_i) \sum_{j=1} p(x_j) \right]; \]

for \( i \neq j \) and its MMI estimate.

\[ \Sigma_{x_1x_2} = 0.2310541, \quad \Sigma_{x_2x_3} = 0.102941, \quad \Sigma_{x_3x_4} = 0.2445796, \]

\[ \Sigma_{x_1x_3} = 0.0857727, \quad \Sigma_{x_1x_4} = 0.03717287, \quad \Sigma_{x_2x_3} = 0.1361021, \]

\[ \Sigma_{x_2x_4} = 0.03014127, \quad \Sigma_{x_3x_4} = 0.03881232, \quad \Sigma_{x_2x_4} = 0.009251411, \]

\[ \Sigma_{x_3x_4} = 0.0364925, \quad I(x_1, x_2, x_3, x_4) = 1.578873, \quad \Sigma_{n\times n} = \ln \left( \frac{A}{B} \right) \]

where \( A \) and \( B \) are given as

\[
A = \begin{bmatrix}
\Sigma_{x_1x_1} & 0 & 0 & 0 \\
0 & \Sigma_{x_2x_2} & 0 & 0 \\
0 & 0 & \Sigma_{x_3x_3} & 0 \\
0 & 0 & 0 & \Sigma_{x_4x_4}
\end{bmatrix}
\]

\[
B = \begin{bmatrix}
\Sigma_{x_1x_1} & \Sigma_{x_1x_2} & \Sigma_{x_1x_3} & \Sigma_{x_1x_4} \\
\Sigma_{x_1x_2} & \Sigma_{x_2x_2} & \Sigma_{x_2x_3} & \Sigma_{x_2x_4} \\
\Sigma_{x_1x_3} & \Sigma_{x_2x_3} & \Sigma_{x_3x_3} & \Sigma_{x_3x_4} \\
\Sigma_{x_1x_4} & \Sigma_{x_2x_4} & \Sigma_{x_3x_4} & \Sigma_{x_4x_4}
\end{bmatrix}^{-1}
\]
Next we evaluate $\chi^2_{\text{calculated}} = \exp -\left( \frac{1}{2} \left( x - \mu \right)^T \left( \sum_{n=1}^{N} \right)^{1/2} \left( x - \mu \right) \right)$ for each pair within and between the archaeological sites based on the potsherd attributes to determine its measure of nonlinear similarity/dependence and plot the similarities among sites based on attributes dominance.

$$\chi^2_{\text{calculated}} = \exp -\left[ 1.578873 \right]$$

$$\left( x - \mu \right)^T \begin{bmatrix} 1.710402 & 1.607452 & 1.527878 & 1.567132 \\ 1.676536 & 2.013314 & 1.706059 & 1.845842 \\ 1.546382 & 1.668627 & 1.751634 & 1.655994 \\ 1.649796 & 1.847052 & 1.683549 & 1.966281 \end{bmatrix} \left( x - \mu \right)$$  \hspace{1cm} (22)
Nonlinear dependence measure of ‘n’ vectors

Table 1: Frequencies of potsherd attributes at various sites

| Sites | $X_1$ | $X_2$ | $X_3$ | $X_4$ | Total |
|-------|-------|-------|-------|-------|-------|
| $P_o$ | 30    | 10    | 10    | 39    | 89    |
| $P_1$ | 53    | 4     | 16    | 2     | 75    |
| $P_2$ | 73    | 1     | 41    | 1     | 116   |
| $P_3$ | 20    | 6     | 1     | 4     | 31    |
| $P_4$ | 46    | 36    | 37    | 13    | 132   |
| $P_5$ | 45    | 6     | 59    | 10    | 120   |
| $P_6$ | 16    | 28    | 169   | 5     | 218   |
| Total | 283   | 91    | 333   | 74    | 781   |

Table 2: Measure of Nonlinear dependence/similarity among paired sites

| Sites | $P_o$ | $P_1$ | $P_2$ | $P_3$ | $P_4$ | $P_5$ | $P_6$ |
|-------|-------|-------|-------|-------|-------|-------|-------|
| $P_o$ | 2.816 |       |       |       |       |       |       |
| $P_1$ | 2.249 | 1.916 |       |       |       |       |       |
| $P_2$ | 3.458 | 2.943 | 4.54  |       |       |       |       |
| $P_3$ | 0.955 | 0.801 | 1.227 | 0.335 |       |       |       |
| $P_4$ | 4.095 | 3.387 | 5.218 | 1.420 | 6.059 |       |       |
| $P_5$ | 3.632 | 3.034 | 4.698 | 1.258 | 5.399 | 4.89  |       |
| $P_6$ | 6.626 | 5.462 | 8.514 | 2.263 | 9.908 | 8.994 | 16.654 |

Note: Proportion of nonlinear estimate can be obtained using R code `pchisq(calcuated $\chi^2$, No of attributes=4). However see appendix for the clusters.

5. Conclusion

The measure of nonlinear dependence via partitioning of elliptical mutual information index into linear and nonlinear component has put an end to the era of difficulty in clear cut interpretation of the proportion of nonlinear dependence
in the modeling of dependent categorical data. This study has answered two unresolved issues in statistical dependence analysis. First the proportion of multivariate mutual information (MMI) total dependence that is linear along with the nonlinear proportion. Secondly the study obtained the significance of the linear and nonlinear proportional size. In addition, the study provided the methodology to obtaining the functional relationship among multi-dimensional variables with at most ordinal scale of measurement in a categorical table relationship.

References

1. Gomez, E. Gomez-Villegas, M.A., and Martin, J.M. (1998). A multivariate generalization of the exponential power family of distributions. Communications in Statistics A27, 589-600. https://doi.org/10.1080/03610929808832115

2. Kullback, S. (1978). Information theory and statistics. Dover Edition, Gloucester.

3. Lindsey J.K. (1999). Multivariate Elliptically Contoured Distributions for repeated Measurements. Biometrics 55, pp 1277-1280. https://doi.org/10.1111/j.0006-341X.1999.01277.x

4. Reinaldo B. Arellano-valley, Javier E. Contreras-reyes and Marc G. Genton (2012). Shanon

5. Reginald Smith (2015). A mutual information approach to calculating non linearity. The ISI

6. Shannon C.E. (1948). A mathematical theory of communication. Bell sys. Tech. J., 27(3):379-423.

7. Sunil Srinivasa (2003). Review on Multivariate Mutual Information. www.nd.edu/tutorial/sunil.

8. Han T.S. (1980). Multiple Mutual Information and Multiple Interactions in Frequency Data Information and Control, volume 46(1); page 26-45.

9. ThoHoan Pham, TuBao Ho, QuynhDiep Nguyen, Dang Hung Tran and Van Hoang Nguyen (2012). Multivariate Mutual Information measures for Discovering Biological Networks.

10. Thomas M. Cover and Joy A. Thomas (1991): Elements of Information Theory. John Wiley & Sons, Inc. https://doi.org/10.1002/0471200611, PMcid:PMC257837
Nonlinear dependence measure of ‘n’ vectors

11. David R. Brillinger and Apratim Guha (2006): Mutual Information in the frequency domain. Journal of Statistical Planning and Inference https://doi.org/10.1016/j.jspi.2006.06.026

Appendix

Figure 1: Arbitrary Area of the Chi-square curve for nonlinear dependence distribution
Figure 2. A dendogram for similarities among pottery sites