Journal of Multivariate Analysis 131 (2014) 32–50

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Comparison, utility, and partition of dependence under absolutely continuous and singular distributions

Nader Ebrahimi [a], Nima Y. Jalali [b], Ehsan S. Soofi [b,*]

[a] *Division of Statistics, Northern Illinois University, DeKalb, IL, 60115, United States*
[b] *Sheldon B. Lubar School of Business, University of Wisconsin-Milwaukee, Milwaukee, WI, 53201, United States*

HIGHLIGHTS

- We show that the popular association indices fail to detect and rank dependence.
- The popular association indices under-represent the dependence of elliptical models.
- The mutual information detects and ranks dependence of absolutely continuous models.
- The mutual information measures the utility of dependence between random variables.
- We use a generalized information index to rank dependence of singular distributions.

ARTICLE INFO

Article history:
Received 21 January 2014
Available online 21 June 2014

AMS 2010 subject classifications:
62B10
62H20
94A15
94A17

Keywords:
Convolution
Copula
Entropy
Elliptical distribution
Marshall–Olkin
Mutual information
Predictability
Sarmanov families
Shock models
Utility

ABSTRACT

This paper first illustrates that a mutual information index detects and ranks dependence of a wide variety of absolutely continuous families, but the popular association and variance reduction indices fail to serve as such "common metrics". We then elaborate on some theoretical merits of the mutual information and give several results. The mutual information provides a notion of the utility of dependence for predicting random variables and quantifies how much the joint distribution is more informative about the variables than the independent model. We present insightful partitions of dependence among the components of a random vector, for a class of models recently proposed for dependence of uncorrelated variables, and for the elliptical families. We also recall that the mutual information is not applicable to singular distributions and give some results for a generalized information index for these models. The generalized index is derived for the Marshall–Olkin copula and for a new singular copula that represents the dependence of the consecutive terms of the exponential autoregressive and related processes.

1. Introduction

Independence between random variables or vectors stipulates sharp relationships between their joint, marginal, and conditional distributions. Any divergence from the independence leads to a probabilistic relationship where the random variables provide information about each other. Divergence measures between the joint distribution and the independent model capture all forms of dependence; Micheas and Zografos [45] provide several results and many references. The most

* Corresponding author.
E-mail addresses: nader@math.niu.edu (N. Ebrahimi), nima@uwm.edu (N.Y. Jalali), esoofi@uwm.edu (E.S. Soofi).

http://dx.doi.org/10.1016/j.jmva.2014.06.014

well-known and important divergence measure of dependence is the mutual information [45]. We study this measure, denoted here as M, via the four objectives given as follows.

First, we address the problem of measuring dependence as a practical issue. For studying and modeling dependence "a common metric" [55] is needed to compare various families of distributions or their copulas. Common metrics "should be as general as possible to permit modeling dependence of a wide variety of situations" [13], detect dependence in any form, exist for large collections of distributions, and order their dependence [50]. We illustrate that a mutual information index for the absolutely continuous distributions, denoted here as δ^2, provides such a common metric. This index was introduced by Linfoot [42] and is well-known [8,36]. We provide a comparison showing that δ^2 detects and ranks dependence within and between some widely used families of models and some models that recently have been studied [22,24]. The popular association indices (the product moment correlation, Spearman's rank correlation, Kendall's τ) and the fraction of expected variance reduction by conditioning fail to reveal and rank various forms of dependence.

Second, we address the conceptual issue of the utility of dependence. The virtue of dependence materializes through the probabilistic information that the variables provide about each other. The usual notions of association and variance reduction are not sufficiently general to recognize the usefulness of dependence beyond some specific forms. We argue that the utility interpretation of M used in the Bayesian framework [4,41] extends more generally to the expected utility of dependence in terms of gain in predicting a random variable through its conditional distributions instead of its marginal distribution. Other divergence functions, while detect and rank dependence [34,35,45], generally lack the expected utility representations in terms of the uncertainty reduction.

Third, we give new results on partitioning of dependence in three contexts. (a) We give a partition of dependence of a random vector which implies the *super-additivity* of M proved by [45]. (b) We extend the result of [6] on the M between the sum of two independent random variables and each summand to the case when the summands are not independent. This result gives a decomposition of the dependence for the families of models recently defined in terms of the convolution property [24]. (c) We show that M of each elliptical family decomposes into two parts: one part is the M of the Gaussian family (a function of the correlation matrix), and the other part is M of the model in the family with orthogonal scale. Application to the Student-t family provides insights about the dependence of the family.

Fourth, we recall that M is not applicable to singular distributions because there is a positive probability for a functional relationship between the variables. We give some results for a generalization of δ^2 used by [25]. We derive the generalized index for the Marshall–Olkin copula and for a new singular copula that represents the dependence of the consecutive terms of the exponential autoregressive process [32] and related processes [58].

This paper is organized as follows. Section 2 defines notations, the information functions, and three broad families of distributions used in the paper. Section 3 compares the aforementioned indices for serving as "common metrics" and presents the interpretation of M as the utility of dependence. Section 4 presents several results for M. Section 5 illustrates the inapplicability of M to singular distributions and presents some results for a generalization of δ^2. Section 6 summarizes the paper and provides some concluding remarks. An Appendix shows derivations of some indices.

2. Preliminaries

We denote by F and G the cumulative distribution functions and by f and g their probability density functions (pdf's). For most of the paper, we consider bivariate distributions of two random variables, X_1 and X_2, and denote their joint pdf by $f(x_1, x_2)$, the marginal distribution of X_i by F_i and its pdf by f_i, the conditional distribution of X_i given $X_j = x_j$ by $F_{i|j},\ i \neq j$, and its pdf by $f_{i|j}$. Multivariate dependence is presented when there is little need for additional notations.

Several examples and results are provided for the following three broad families of the absolutely continuous distributions.

1. The families with d-dimensional elliptical pdf's:

$$f_h(\boldsymbol{x}|\Sigma, \boldsymbol{\mu}) = k|\Sigma|^{-1/2} h\Big((\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\Big), \in \Re^d, \quad |\Sigma| > 0, \tag{1}$$

   where $\boldsymbol{\mu}$ is the location vector, $\Sigma = [\sigma_{ij}]$ is a positive-definite scale matrix, $|\Sigma|$ denotes the determinant, and $h(\cdot) = h_d(\cdot)$ is referred to as the density generator which generally includes a parameter (vector), in addition to $(\boldsymbol{\mu}, \Sigma)$ [31].
2. The families with bivariate pdf's in the following form:

$$f_q(x_1, x_2|\beta) = f_1(x_1) f_2(x_2)[1 + \beta q(x_1, x_2)], \quad (x_1, x_2) \in \Re^2,\ |q(x_1, x_2)| \le B,\ |\beta| \le B^{-1}, \tag{2}$$

   where $f_i(x_i),\ i = 1, 2$ are the marginal pdf's, and the link function $q(x_1, x_2)$ is a measurable bounded function such that $\int_{\Re^2} q(x_1, x_2) f_1(x_1) f_2(x_2) dx_1 dx_2 = 0$. For $q(x_1, x_2) = q_1(x_1) q_2(x_2)$, (2) gives Sarmanov families, so we refer to families with pdf's in form of (2) as the Generalized Sarmanov families.
3. The families related to the bivariate Pareto Type II distribution with survival function

$$\bar{F}(x_1, x_2|\alpha) = (1 + x_1 + x_2)^{-\alpha}, \quad x_i \ge 0,\ \alpha > 0. \tag{3}$$

We also provide results for singular distributions where a functional relationship is probable, $0 < P[X_i = \phi(X_j)] = \pi < 1$, $i \neq j$. These distributions do not have bivariate density relative to the Lebesgue measure which are needed for computing