International Journal of Approximate Reasoning 68 (2016) 15–26

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Measures of the functional dependence of random vectors ☆

Santi Tasena *, Sompong Dhompongsa

Department of Mathematics, Faculty of Sciences, Chiang Mai University, Chiang Mai, 50200, Thailand

ARTICLE INFO

Article history:
Received 21 March 2015
Received in revised form 15 July 2015
Accepted 4 October 2015
Available online 22 October 2015

Keywords:
Copula
Complete dependence
Functional dependence
Measure of dependence
Sobolev norm

ABSTRACT

In this work, we define a set of properties that any measure of functional dependence that exists between random vectors should possess. We also construct measures of functional dependence and show that they satisfy the properties mentioned above. Relationships between these measures and previously defined measures of functional dependence between random variables are discussed.

1. Introduction

How can we determine whether a random vector Y is a function of a random vector X? More generally, how can we order random vectors in terms of how much they depend on a specific random vector? For continuous random variables, two related measures have been proposed: one measure was proposed by Dette et al. [1] (see also Siburg and Stoimenov [4]), while another measure was proposed by Trutschnig [7]. When written using the copula associated with these random variables, both measures were similar. Both are based on modified Sobolev norms, where the former is based on the L^2-Sobolev norm while the latter is based on the L^1-Sobolev norm. Note that Dette's measure was extended to the case of a continuous random vector Y and a continuous random variable X by [6].

All of these measures have one common property: they can be written in terms of the first partial derivative of the copula associated with (X, Y). An intuitive extension of these measures to the case of continuous random vectors is to use higher order derivatives of copulas. This is not applicable, however, since higher order derivatives of copulas generally do not exist, not even in the weak sense. A correct extension is obtained by interpreting the first partial derivatives of a copula in probabilistic terms: in this sense, they then represent a conditional distribution of uniform random variables when given another uniform random variable. Based on this idea, we propose a family of measures of functional dependence in the case of random vectors X and Y. Where applicable, we will discuss relationships between these measures and previously defined measures.

The organization of this work is as follows. In Section 2, we discuss terminologies and notations used throughout this work. In Section 3, we discuss the basic properties of measures of functional dependence. In Section 4, we give an example

☆ This research is jointly supported by the Development and Promotion of Science and Technology Talents Project (DPST) and the Institute for the Promotion of Teaching Science and Technology (IPST).

* Corresponding author.
E-mail addresses: santi.tasena@cmu.ac.th (S. Tasena), sompongd@chiangmai.ac.th (S. Dhompongsa).

http://dx.doi.org/10.1016/j.ijar.2015.10.002

construction of these measures. In Section 5, we provide proofs that these measures satisfy the properties discussed in Section 3. Sections 3–5 can be read in any particular order, however.

2. Preliminaries

In this work, the space $\mathbb{R}^n$ will be regarded as the product lattice of the real line $\mathbb{R}$. For example, the statement $(x_1, \dots, x_n) \le (y_1, \dots, y_n)$ means $x_i \le y_i$ for all $i = 1, \dots, n$.

Since this work deals mainly with random vectors, the terminology *joint*, as in *joint distribution*, will have a slightly different meaning than usual. For any (n-dimensional) random vector X, the (n-dimensional) *distribution* of X is the function F_X defined by $F_X(x) = \mathbb{P}(X \le x)$. The ($n$-dimensional) *density* of an (n-dimensional) distribution F_X, if exists, is a function f_X such that $F_X(x) = \int_{(-\infty,x]} f_X(t)\mathrm{d}t$ for all x, that is, $f_X = \frac{\partial^n F_X}{\partial x_1 \cdots \partial x_n}$. To simplify the notation, we will write $\frac{\partial F_X}{\partial x}$ instead of $\frac{\partial^n F_X}{\partial x_1 \cdots \partial x_n}$. For any random vectors X and Y, the *joint distribution* of X and Y is the function $F_{X,Y}$ defined as $F_{X,Y}(x, y) = \mathbb{P}(X \le x, Y \le y)$. If we identify the space $\mathbb{R}^n \times \mathbb{R}^k$ with the space $\mathbb{R}^{n+k}$, then the joint distribution of X and Y is simply the distribution of (X, Y). The term *joint*, however, helps us to differentiate whether we are considering two vectors separately. Moreover, the distributions of X and Y are marginals of the joint distribution of X and Y. If the joint distribution $F_{X,Y}$ has a density $f_{X,Y}$, then its marginals F_X and F_Y also have densities f_X and f_Y, which are related via the formula $f_X(x) = \int f_{X,Y}(x, y)\mathrm{d}y$ and $f_Y(y) = \int f_{X,Y}(x, y)\mathrm{d}x$, respectively.

An *(n-dimensional) copula* is a distribution of n random variables that have a uniform distribution in the unit interval $\mathbb{I} = [0, 1]$. An *(n-dimensional) subcopula* is a restriction of a copula into the product of closed subsets of $\mathbb{I}$. Any random vector $X = (X_1, \dots, X_n)$ is associated with a unique subcopula[A3] S_X, where the domain of S_X is the product of the ranges of $F_{X_1}, \dots, F_{X_n}$ such that

$$F_X(x_1, \dots, x_n) = S_X\left(F_{X_1}(x_1), \dots, F_{X_n}(x_n)\right) \tag{1}$$

for all $x_1, \dots, x_n \in \mathbb{R}$ [5]. If the distribution of X_i is continuous for all i, then S_X is a copula. By regarding a pair of random vectors X and Y as a random vector (X, Y), we may define the *joint subcopula* associated with X and Y as the subcopula associated with (X, Y). Moreover, we call the subcopula associated with X and the subcopula associated with Y the marginals of the joint (sub)copula associated with X and Y.

Given random vectors X and Y, we define the *conditional distribution* $F_{Y|X}$ of Y given X by letting

$$F_{Y|X}(y|x) = \lim_{h \searrow 0} \frac{\mathbb{P}(Y \le y, x - h < X \le x + h)}{\mathbb{P}(x - h < X \le x + h)} \tag{2}$$

whenever the limit exists. It is well-known that $F_{Y|X}$ has a version in which the function $F_{Y|X}(\cdot|x)$ is a distribution for all x, which is the so-called regular conditional distribution of Y given X. In cases where the joint distribution of X and Y has a density $f_{X,Y}$, the quotient $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$ is the *density* of $F_{Y|X}$, that is, $F_{Y|X}(y|x) = \int_{(-\infty,y]} f_{Y|X}(t|x)\mathrm{d}t$. Moreover, $F_{Y|X}(y|x) = \frac{\frac{\partial^n F_{X,Y}}{\partial x_1 \cdots \partial x_n}(x,y)}{\frac{\partial^n F_X}{\partial x_1 \cdots \partial x_n}(x)}$. To simplify the notation, we write $\frac{\partial F_{X,Y}}{\partial F_X}$ instead of $\frac{\frac{\partial^n F_{X,Y}}{\partial x_1 \cdots \partial x_n}}{\frac{\partial^n F_X}{\partial x_1 \cdots \partial x_n}}$ in this case.

Note that

$$\begin{aligned}\int 1_{(-\infty,x]}(s)1_{(-\infty,y]}(t)\mathrm{d}F_{X,Y}(s,t) = \mathbb{P}(X \le x, Y \le y) \\ &= \int 1_{(-\infty,x]}(s)\mathbb{P}(Y \le y|X = s)\mathrm{d}F_X(s) \\ &= \int 1_{(-\infty,x]}(s)1_{(-\infty,y]}(t)\mathrm{d}F_{Y|X}(t|s)\mathrm{d}F_X(s)\end{aligned}$$

for all x, y. Using monotone class theorems (see, e.g., Yeh [8, Theorem 1.10]), we can conclude that

$$\int f(x, y)\mathrm{d}F_{X,Y}(s,t) = \int f(x, y)\mathrm{d}F_{Y|X}(t|s)\mathrm{d}F_X(s) \tag{3}$$

for the bounded measurable function f.

For continuous random variables X and Y, Dette et al. [1] defined a measure $\omega(Y|X) = \omega(C_{X,Y})$ by letting

$$\omega(Y|X) = 6\int\int\left(\frac{\partial C_{X,Y}}{\partial u}(u, v)\right)^2 \mathrm{d}u\mathrm{d}v - 2$$

where $C_{X,Y}$ is the joint copula associated with X and Y. Note that $\omega(Y|X)$ is actually a normalization of the function $C \mapsto \int\int\left(\frac{\partial C_{X,Y}}{\partial u}(u, v) - \frac{\partial \Pi}{\partial u}(u, v)\right)^2 \mathrm{d}u\mathrm{d}v$ where $\Pi(u, v) = uv$ is the copula associated with independent continuous random