

Information Measures via Copula Functions

G. R. Mohtashami Borzadaran* M. Amini

Ferdowsi University of Mashhad

Abstract. In applications of differential geometry to problems of parametric inference, the notion of divergence is often used to measure the separation between two parametric densities. Among them, in this paper, we will verify measures such as Kullback-Leibler information, J-divergence, Hellinger distance, α -Divergence, ... and so on. Properties and results related to distance between probability distributions derived via copula functions. Some inequalities are obtained in view of the dependence and information measures.

Keywords. Information measures; Fisher information; Kullback-Leibler information; Hellinger distance; α -divergence.

1 Introduction

The study of copulas and the role they play is important in probability, statistics and stochastic processes. Sklar (1959) provided a uniform representation of bivariate distribution F on the unit square and defined copula based on it. Many research papers and monographs due to copula aspect are published after Sklar (1959), such as Nelsen (2006), Joe (1997), Cherubini et al. (2004) and Mari and Kotz (2004) and their references in. Frees and Valdez (1998) introduced the concept of copulas as a tool for understanding relationships among multivariate outcomes. Also, dependence and copulas have linked with each other.

The concept of the entropy originated in the nineteenth century by C.E. Shannon (1948). During the last sixty years or so, a number of publications

* Corresponding author

discussing and extending Shannon's original work have appeared. Among them, Ali Ahmed et al. (1989), Darbellay and Vajda (1998, 2000), Dragomir (2003), Blyth (1994), Torkkola (2003), Kapur (1989, 1994), Borovkov (1998), Kagan, Linnik and Rao (1973) and Kullback (1959) are mentioned in this research. Ma and Sun (2008) and Calsaverini and Vicente (2009) are two recent notes related to the links between copula and information measures.

In this paper, various measures are obtained in view of copulas for bivariate distributions. Properties of information measures and their links with copula is another direction of this research.

2 Preliminaries and Some Information Measures

Let $(\Omega, \mathcal{B}, \mu)$ be a measure space and f be a measurable function from Ω to $[0, \infty)$, such that $\int_{\Omega} f d\mu = 1$. The Shannon entropy (or simply the entropy) of f relative to μ , is defined by

$$H(f, \mu) = - \int_{\Omega} f \ln f d\mu, \quad (\text{with } f \ln f = 0 \text{ if } f = 0), \quad (1)$$

and assumed to be defined for which $f \ln f$ is integrable. If X is an r.v. with pdf f , then we refer to H as the entropy of X and denotes it by the notation H_X as well. In the case μ is a version of counting measure, (1) leads us to a specialized version that introduced by Shannon (1948) as $H_X = - \sum_{i=1}^n p_i \ln p_i$ where $p_i \geq 0$ and $\sum_{i=1}^n p_i = 1$. One of the important issues in many applications of the probability theory is finding an appropriate measure of distance between two probability distributions. A number of divergence measures for this purpose have been studied by Kullback and Leibler (1951), Renyi (1961) and lot of references related to the various type of information measures can be find in Dragomir (2003).

Assume that the set χ be the suitable support of distributions and the σ -finite measure μ are given such that $\Omega = \{f : \chi \times \chi \rightarrow \mathbb{R}, f(x, y) > 0, \int_{\chi \times \chi} f(x, y) d\mu = 1\}$ is used earlier for some other purpose. Consider F and G be two bivariate distributions which are absolutely continuous w.r.t. measure μ and $\frac{dF}{d\mu} = f$ and $\frac{dG}{d\mu} = g$. Here, we introduce shortly the form of some familiar information measures based on bivariate distributions.

Kullback Leibler information:

$$D_{\text{KL}}(F, G) = \int \int_{\chi \times \chi} \ln \frac{f(x, y)}{g(x, y)} f(x, y) d\mu, \quad (2)$$

χ^2 - divergence:

$$D_{\chi^2}(F, G) = \int \int_{\chi \times \chi} \frac{[f(x, y) - g(x, y)]^2}{f(x, y)} d\mu, \quad (3)$$

Hellinger distance:

$$D_H(F, G) = \int \int_{\chi \times \chi} \left[\sqrt{f(x, y)} - \sqrt{g(x, y)} \right]^2 d\mu, \quad (4)$$

α -divergence:

$$D_\alpha(F, G) = \frac{1}{1 - \alpha^2} \int \int_{\chi \times \chi} \left\{ 1 - \frac{g^{\frac{1+\alpha}{2}}(x, y)}{f^{\frac{1+\alpha}{2}}(x, y)} \right\} f(x, y) d\mu, \quad (5)$$

Jeffery's distance (J-divergence):

$$D_J(F, G) = \int \int_{\chi \times \chi} [f(x, y) - g(x, y)] \ln \frac{f(x, y)}{g(x, y)} d\mu, \quad (6)$$

Combination of version of α -Divergence:

$$D_{C\alpha}(F, G) = \frac{4}{\beta^2} \int \int_{\chi \times \chi} \frac{\left\{ g^{\frac{\beta}{2}}(x, y) - f^{\frac{\beta}{2}}(x, y) \right\}^2}{g^{\beta-1}(x, y)} d\mu, \quad (7)$$

Bhattacharyya distance:

$$D_{Bh}(F, G) = \int \int_{\chi \times \chi} \sqrt{g(x, y)f(x, y)} d\mu, \quad (8)$$

Harmonic distance:

$$D_{Ha}(F, G) = \int \int_{\chi \times \chi} \frac{2g(x, y)f(x, y)}{g(x, y) + f(x, y)} d\mu, \quad (9)$$

Triangular discrimination:

$$D_\Delta(F, G) = \int \int_{\chi \times \chi} \frac{[f(x, y) - g(x, y)]^2}{g(x, y) + f(x, y)} d\mu, \quad (10)$$

Lei and Wang divergence:

$$D_{\text{LW}}(F, G) = \int \int_{\chi \times \chi} f(x, y) \ln \frac{2f(x, y)}{g(x, y) + f(x, y)} d\mu. \quad (11)$$

Relative information generating function:

The relative information generating function of f given the reference measure g is defined by Guiasu and Reischer (1985) as,

$$R(F, G, t) = \int \int_{\chi \times \chi} \left[\frac{f(x, y)}{g(x, y)} \right]^{t-1} f(x, y) d\mu, \quad (12)$$

where $t \geq 1$ and the integral is convergent. We note that $R(F, G, 1) = 1$.

Power divergence measures:

Cressie and Read (1984) proposed the power divergence measure (PWD) which gathers most of the interesting specification. This measure is defined as

$$\text{PWD}(F, G) = \frac{1}{\lambda(\lambda + 1)} \int \int_{\chi \times \chi} \left\{ \left[\frac{f(x, y)}{g(x, y)} \right]^\lambda - 1 \right\} f(x, y) d\mu. \quad (13)$$

The power divergence measure leads to different well-known divergence measures for different values of λ . PWD for $\lambda = -2, -1, -0.5, 0, 1$, implies Neyman Chi-square, Kullback Leibler, squared Hellinger distance, Likelihood disparity and Pearson Chi-square divergence respectively. Note that $\text{PWD}(F, G) = \frac{1}{\lambda(\lambda+1)} [R(F, G, \lambda + 1) - 1]$.

3 Information Measures in View of Copula Distributions

The copula function $C(u, v)$ is a bivariate distribution function with uniform marginal on $[0, 1]$, such that

$$F(x, y) = C_F\{F_1(x), F_2(y)\}.$$

By Sklar's Theorem (Sklar, 1959), this copula exists and is unique if F_1 and F_2 are marginal continuous distribution functions. Thus, we can construct bivariate distributions $F(x, y) = C_F\{F_1(x), F_2(y)\}$ with given univariate marginal F_1 and F_2 by using copula C_F , (Nelsen, 2006). We have the following properties:

- (Nelsen, 2006) Let $F(x, y)$ be a joint distribution function with marginal $F_1(x)$ and $F_2(y)$, then
 (i) The copula C_F is given by

$$C_F(u, v) = F\{F_1^{-1}(u), F_2^{-1}(v)\}, \quad \forall u, v \in [0, 1],$$

where, F_1^{-1} and F_2^{-1} are quasi-inverses of F_1 and F_2 respectively.

- (ii) The partial derivatives $\frac{\partial C_F(u, v)}{\partial u}$ and $\frac{\partial C_F(u, v)}{\partial v}$ exist and $c(u, v) = \frac{\partial^2 C_F(u, v)}{\partial u \partial v}$ is density function of $C_F(u, v)$.

Ma and Sun (2008) defined copula entropy as follows:

Definition 1. Let X be a two dimensional random vector with copula density $c(u, v)$. Copula entropy of X is defined as

$$H_c(X) = - \int \int_{u, v} c(u, v) \ln c(u, v) du dv.$$

Kullback Leibler information:

$$D_{KL}(F, F_1 F_2) = \int_0^1 \int_0^1 c(u, v) \ln c(u, v) du dv. \quad (14)$$

In this case, Kullback Leibler information is called mutual information.

Theorem 1. *Mutual information of the random variable is equal to the negative entropy of their corresponding copula function,*

$$D_{KL}(F, F_1 F_2) = -H_c(X).$$

Proof. Via $f(x, y) = c\{F_1(x), F_2(y)\}f_1(x)f_2(y)$ easily derived. \square

- On noting Theorem 1, difference of the information contained in joint distribution and marginal densities is equal to copula entropy. Hence,

$$H(X) = \sum_{i=1}^2 H(X_i) + H_c(X),$$

where $X = (X_1, X_2)$ and independence X_1 and X_2 implies $H_c(X) = 0$.

- Normalizing D_{KL} index, Joe (1997) defined

$$\delta^* = [1 - \exp\{-2D_{KL}(F, F_1F_2)\}]^{\frac{1}{2}},$$

where δ^* is confined to the interval $[0, 1]$. When the dependence is maximal, D_{KL} tends to infinity and δ^* can be considered as a measure of dependence. As an example, let $X \sim N(\mu, \Sigma)$ where $\mu = [\mu_1, \mu_2]$ and $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$ then, $\delta^* = |\rho|$ that is a suitable measure for finding correlation coefficient in this case.

- Joe (1989) obtained measure of dependence in multivariate cases as

$$\rho_{X_1, X_2, \dots, X_n}^* = \left[\frac{\delta_{X_1, X_2, \dots, X_n}}{\sum_{i=1}^n H(X_i) - \text{Max}_j H(X_j)} \right]^{\frac{1}{2}},$$

where

$$\begin{aligned} \delta_{X_1, X_2, \dots, X_n} &= \int \int_{\mathbb{R}^n} \dots \int \left[\ln \frac{f(x_1, x_2, \dots, x_n)}{\prod_{i=1}^n f_i(x_i)} \right] \\ &\quad \times f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n \end{aligned}$$

and $H(X_i) = -\int_{\mathbb{R}} [\ln f_i(x_i)] f_i(x_i) dx_i$. In bivariate case, let X and Y be identically distributed but not necessarily independent, then $\rho_{X_1, X_2}^* = \frac{D_{KL}(F, F_1F_2)}{H(X)}$ where $0 \leq \rho_{X_1, X_2}^* \leq 1$ and $\rho_{X_1, X_2}^* = 0$ implies independence. So, $\rho_{X_1, X_2}^* = 1$ implies X_1 and X_2 are perfectly correlated.

Also, the linear correlation for standardized variables in terms of the copula densities is $\rho(X, Y) = \int_0^1 \int_0^1 c(u, v) F_1^{-1}(u) F_2^{-1}(v) du dv$. If X and Y are independent, then, $c(u, v) = 1$ and consequently $\rho(X, Y) = 0$. A better alternative for measuring correlation concordance would be the rank correlation that is known as Spearman's ρ as $\rho_{\text{rank}}(X, Y) = 12 \int_0^1 \int_0^1 uv c(u, v) du dv - 3$. Note that this measure for a pair of continuous random variables X and Y is identical to Pearson's correlation coefficient (linear correlation) of $U = F_1(X)$ and $V = F_2(Y)$.

χ^2 — divergence:

$$D_{\chi^2}(F, F_1F_2) = \int_0^1 \int_0^1 \frac{[c(u, v) - 1]^2}{c(u, v)} du dv. \quad (15)$$

Hellinger distance:

$$D_H(F, F_1 F_2) = \int_0^1 \int_0^1 \left[\sqrt{c(u, v)} - 1 \right]^2 dudv. \quad (16)$$

α -divergence:

$$D_\alpha(F, F_1 F_2) = \frac{1}{1 - \alpha^2} \int_0^1 \int_0^1 \left[1 - \{c(u, v)\}^{-\frac{(\alpha+1)}{2}} \right] c(u, v) dudv. \quad (17)$$

Jeffery's distance (J-divergence):

$$D_J(F, F_1 F_2) = \int_0^1 \int_0^1 [c(u, v) - 1] \ln c(u, v) dudv. \quad (18)$$

Combination of version of α -divergence:

$$D_{C\alpha}(F, F_1 F_2) = \frac{4}{\beta^2} \int_0^1 \int_0^1 [1 - \{c(u, v)\}^{\frac{\beta}{2}}]^2 dudv. \quad (19)$$

Bhattacharyya distance:

$$D_{Bh}(F, F_1 F_2) = \int_0^1 \int_0^1 \sqrt{c(u, v)} dudv. \quad (20)$$

Harmonic distance:

$$D_{Ha}(F, F_1 F_2) = \int_0^1 \int_0^1 \left[\frac{2c(u, v)}{c(u, v) + 1} \right] dudv. \quad (21)$$

Triangular discrimination:

$$D_\Delta(F, F_1 F_2) = \int_0^1 \int_0^1 \left[\frac{\{c(u, v) - 1\}^2}{1 + c(u, v)} \right] dudv. \quad (22)$$

Lei and Wang divergence:

$$D_{LW}(F, F_1 F_2) = \int_0^1 \int_0^1 c(u, v) \ln \left[\frac{2c(u, v)}{1 + c(u, v)} \right] dudv. \quad (23)$$

Relative information generating function:

$$R(F, F_1 F_2, t) = \int_0^1 \int_0^1 [c(u, v)]^t dudv. \quad (24)$$

where $t \geq 1$ and the integral is convergent.

Power divergence measures :

$$\text{PWD}(F, F_1 F_2) = \frac{1}{\lambda(\lambda + 1)} \int_0^1 \int_0^1 [\{c(u, v)\}^\lambda - 1] c(u, v) dudv. \quad (25)$$

- It is easy to see that $D_H(F, F_1 F_2) = 2[1 - D_{Bh}(F, F_1 F_2)] \leq 2$. Via Taylor expansion and approximation, we can get, $D_{KL}(F, F_1 F_2) \approx \frac{1}{2} D_{\chi^2}(F, F_1 F_2)$, $D_J(F, F_1 F_2) \approx \frac{1}{2} [D_{\chi^2}(F, F_1 F_2) + D_{\chi^2}(F_1 F_2, F)]$, $D_{\chi^2}(F, F_1 F_2) \approx 4D_H(F, F_1 F_2)$ and $D_{\chi^2}(F, F_1 F_2) \geq D_H(F, F_1 F_2)$. The $D_{C\alpha}(F, F_1 F_2)$ and $D_\alpha(F, F_1 F_2)$ are linked via the following identity: $D_{C\alpha}(F, F_1 F_2) = 16(\frac{2}{\beta} - 1)A + 16(1 - \frac{1}{\beta})B$ where A and B are $D_\alpha(F, F_1 F_2)$ with $\alpha = 1 - \beta$ and $\alpha = 1 - 2\beta$ respectively. The chi-squared divergence $D_{\chi^2}(F, F_1 F_2) = D_{C\alpha}(F, F_1 F_2)$ and $D_{\chi^2}(F, F_1 F_2) = 2D_\alpha(F, F_1 F_2)$ on taking $\beta = 2$ in (7) and $\alpha = -3$ in (5) respectively. Also, the Hellinger distance $D_H(F, F_1 F_2) = \frac{1}{4} D_{C\alpha}(F, F_1 F_2)$ and $D_H(F, F_1 F_2) = \frac{1}{2} D_\alpha(F, F_1 F_2)$ on taking $\beta = 1$ in (7) and $\alpha = 0$ in (5) respectively. The Hellinger distance is symmetric and has all properties of a metric. Also, $D_{Bh}(F, F_1 F_2)$, $D_{Ha}(F, F_1 F_2)$, $D_\Delta(F, F_1 F_2)$ are symmetric and $D_{LW}(F, F_1 F_2) + D_{LW}(F_1 F_2, F) \leq D_\Delta(F, F_1 F_2)$.
- Let C_0 and C_1 be two copula functions. Cuadras (2009) defined mixture copula C_θ as

$$C_\theta(u, v) = (1 - \theta)C_0(u, v) + \theta C_1(u, v), \quad \theta \in [0, 1].$$

On noting that the copula density of mixture copula is

$$c_\theta(u, v) = (1 - \theta)c_0(u, v) + \theta c_1(u, v), \quad \theta \in [0, 1],$$

the relation information generating function (case integer t) for it, can be find via,

$$\begin{aligned} \int_0^1 \int_0^1 [c_\theta(u, v)]^t dudv &= \sum_{j=0}^t \frac{t!}{j!(t-j)!} (1 - \theta)^{t-j} (\theta)^j \\ &\times \int_0^1 \int_0^1 [c_0(u, v)]^{t-j} [c_1(u, v)]^j dudv. \end{aligned}$$

So, for some of the copula density c_0 and c_1 , we can find easily via the calculation of the integrals, the information generating function for mixture copula of c_0 and c_1 . For other measures such as power divergence, Hellinger distance and α -divergence it is also applicable.

- Let X and Y be continuous random variables with copula C_{XY} . If α and β are strictly increasing on range of X and range of Y respectively, then, $C_{X^*Y^*}(u, v) = C_{XY}(u, v), \forall u, v \in [0, 1]$, where $X^* = \alpha(X)$ and $Y^* = \beta(Y)$, (Nelsen, 2006). Because of invariance of C_{XY} under strictly increasing transformations of X and Y , all of the information measures of the transformation of X and Y are the same as the information measures that are defined based on C_{XY} . Also, the following results are noticeable:

- (1) If α is strictly increasing (decreasing) and β is strictly decreasing (increasing) on range of X and range of Y respectively, then, the information measures of $C_{X^*Y^*}$ are obtained via the integrals that are expressed the information measures based on it via copuls density with $c(u, 1-v)\{c(1-u, v)\}$ in place of $c(u, v)$ in integrals.
- (2) If α and β are strictly decreasing on range of X and range of Y respectively, then, the information measures of $C_{X^*Y^*}$ are obtained via the integrals that are expressed the information measures based on it via copuls density with $c(1-u, 1-v)$ in place of $c(u, v)$ in integrals.
- (3) Let X_1 and Y_1 be random variables with continuous distribution functions F_1 and G_1 respectively and copula C_{XY} . Let F_2 and G_2 be another pair of distribution functions such that $X_2 = F_2^{-1}\{F_1(X_1)\}$ and $Y_2 = G_2^{-1}\{G_1(Y_1)\}$, then, all of the information measures of that introduced here based on $C_{X_2Y_2}$ can be achieved via the integrals of copula based on $C_{X_1Y_1}$.

4 Inequalities of Information Measures for Weakly Negative Dependence

Ranjbar et al. (2008) presented a new definition of dependence which is discussed in this section and obtained inequalities due to information measures based on it.

Definition 2. The random variables X and Y are said weakly negatively de-

pendent (WND) if there exist a $\gamma > 1$ such that, $f(x_1, x_2) \leq \gamma \cdot f_1(x_1) \cdot f_2(x_2)$ where $f(x_1, x_2)$, $f_1(x_1)$ and $f_2(x_2)$ are joint density and marginal densities of X and Y , respectively.

The class of WND random variables is well defined and a large class of these random variables can be find. The following examples are evidence of WND random variables:

Example 1. (i) Suppose that X_1 and X_2 have half-normal distribution, then

$$f_{X_1, X_2}(x_1, x_2) = \frac{2}{\pi \sqrt{1 - \rho^2}} \exp \left[-\frac{1}{2(1 - \rho^2)} \{x_1^2 + x_2^2 - 2\rho x_1 x_2\} \right], \quad x_1, x_2 > 0,$$

$$f_{X_i}(x_i) = \sqrt{\frac{1}{\pi}} \exp \left\{ -\frac{1}{2} x_i^2 \right\}; \quad i = 1, 2.$$

If $-1 < \rho \leq 0$, then X_1 and X_2 are negative quadrant dependence (NQD) random variables. Moreover,

$$\frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_1}(x_1)f_{X_2}(x_2)} = \frac{1}{\sqrt{1 - \rho^2}} \exp \left[\frac{-\rho^2}{2(1 - \rho^2)} (x_1^2 + x_2^2) + \frac{\rho}{1 - \rho^2} x_1 x_2 \right] \leq \frac{1}{\sqrt{1 - \rho^2}}.$$

Then $f(x_1, x_2) \leq \gamma \cdot f_1(x_1) \cdot f_2(x_2)$, where $\gamma = 1/\sqrt{1 - \rho^2} \geq 1$. So, X_1 and X_2 are WND.

(ii) Let X and Y be two random variables with joint Farlie-Gumbel-Morgenstern (FGM) distribution, we have

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) [1 + \alpha\{1 - 2F_X(x)\}\{1 - 2F_Y(y)\}].$$

On the other hand, it's obvious that

$$|1 + \alpha\{1 - 2F_X(x)\}\{1 - 2F_Y(y)\}| \leq 1 + |\alpha|,$$

and

$$f_{X,Y}(x, y) \leq [1 + |\alpha|]f_X(x)f_Y(y).$$

Therefore, the random variables X and Y are WND with $\gamma = 1 + |\alpha| \geq 1$. In addition, we know if $-1 < \alpha \leq 0$, then X and Y are negative quadrant dependence (NQD).

So, we have the following inequalities for the information measures:

$$D_{\text{KL}}(F, F_1 F_2) \leq \ln \gamma,$$

$$D_{\alpha}(F, F_1 F_2) \leq \frac{1}{1 - \alpha^2} \left[1 - \gamma^{-\frac{\alpha+1}{2}} \right],$$

$$D_{\text{J}}(F, F_1 F_2) \leq (\gamma - 1) \ln \gamma, \quad D_{\text{Bh}}(F, F_1 F_2) \leq \sqrt{\gamma},$$

$$D_{\text{Ha}}(F, F_1 F_2) \leq \frac{2\gamma}{1 + \gamma},$$

$$D_{\text{LW}}(F, F_1 F_2) \leq \frac{\gamma - 1}{1 + \gamma}.$$

Also,

$$D_{\chi^2}(F, F_1 F_2) \leq \left[1 - \frac{1}{\gamma} \right]^2,$$

$$D_{\text{H}}(F, F_1 F_2) \leq (\sqrt{\gamma} - 1)^2,$$

$$D_{\text{C}\alpha}(F, F_1 F_2) \leq \frac{4}{\beta^2} \left[1 - \gamma^{\frac{\beta}{2}} \right]^2,$$

$$D_{\Delta}(F, F_1 F_2) \leq \frac{(\gamma - 1)^2}{\gamma(1 + \gamma)}.$$

Note that $|1 - \frac{f_1(x)f_2(y)}{f(x,y)}| > 1 - \frac{1}{\gamma}$ implies $f(x, y) \geq \gamma f_2(y)f_1(x)$ or $f(x, y) \leq \frac{1}{2 - \frac{1}{\gamma}} f_2(y)f_1(x)$ that is contradicted with considering $f(x, y) \leq \gamma f_2(y)f_1(x)$ for $\gamma \geq 1$.

- We know that power divergence measure implies different well-known divergence measures for different values of λ such as Neyman Chi-square, Kullback Leibler, squared Hellinger distance, Likelihood disparity and Pearson Chi-square divergence. The following inequalities are noticeable for the above cases:

$$R(F, F_1 F_2, t) \leq \gamma^{t-1}, \quad t \geq 1,$$

$$\text{PWD}(F, G) \leq \frac{1}{\lambda(\lambda + 1)} \{\gamma^{\lambda} - 1\}, \quad \lambda \geq -1.$$

If $\gamma = 1$ implies that these two random variables X and Y are independent in most of the cases that is mentioned in this note.

5 Conclusion

The mutual information is actually negative copula entropy. We derived forms of some information measures based on copula functions. Also, several results and properties are obtained via various information measures on using copula. For bivariate distributions finding characterizations and results in view of copula and information measures is the direction of continuing this research.

Acknowledgment

The authors would like to thank the referees for useful comments, and suggestions. This research was supported by a grant from Ferdowsi University of Mashhad (No. MS88121GMB).

References

- Ali Ahmed, N. and Gokhale, D.V. (1989). Entropy expressions and their estimators for multivariate distributions. *IEEE Tran. Inf. Theory*, **35**, 688-692.
- Blyth, S. (1994). Local divergence and association. *Biometrika*, **81**, 579-84.
- Borovkov, A.A. (1998). *Mathematical Statistics*. Gordon and Breach Science Publishers.
- Calsaverini, R.S. and Vicente, R. (2009). *An information theoretic approach to statistical dependence: copula information*. arXiv:0911.4207v1 [in.ST].
- Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula Methods in Finance*. Wiley, New York.
- Cressie, N. and Read, T.R.C. (1984). Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc., Ser. B* **46**, 440-464.
- Cuadras, C.M. (2009). Constructing copula functions with weighted geometric means. *J. of Stat. Plan. Inference*, **139**, 3766-3772.
- Darbellay, G.A. and Vajda, I. (1998). Estimation of the mutual information with data-dependent partitions. *Tech. Report*, **1921**, Inst. of Inf. Theory and Automation Academy of Sciences of the Czech Republic.
- Darbellay, G.A. and Vajda, I. (2000). Entropy expressions for multivariate continuous distributions. *IEEE Inf. Theory*, **46**, 709-712.

- Dragomir, S.S. (2003). *On the p - logarithmic and α - power divergence measures in information theory*. arXiv:math.PR/0304240 v1.
- Frees, E.W. and Valdez, E. (1997). Understanding relationship using copulas. *North American Actuarial Journal*, **2**, 53-77.
- Guiasu, S. and Reischer, C. (1985). The relative information generating function. *Information Sciences*, **35**, 235-241.
- Joe, H. (1989). Relative entropy measures of multivariate dependence. *JASA*, **84**, 157-164.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall, London, New York.
- Kagan, A.M., Linnik, Yu.V. and Rao, C.R. (1973). *Characterization Problems in Mathematical Statistics*. Wiley, New York.
- Kapur, J.N. (1989). *Maximum Entropy Models in Sciences and Engineering*. Wiley, New York.
- Kapur, J.N. (1994). *Measure of Information and their Applications*. Wiley, New York.
- Kullback, S. and Leibler, R.A. (1951). On information and sufficiency. *Ann. Math. Statist.*, **22**, 79-86.
- Kullback, S. (1959). *Information and Statistics*. Wiley, New York.
- Ma, Jian and Sun, Z. (2008). *Mutual information is copula entropy*. arXiv:0808.0845v1 [cs.IT].
- Mari, D.D. and Kotz, S. (2004). *Correlation and Dependence*. Imperial College Press.
- Nelsen, R.B. (2006). *An Introduction to Copulas*. Springer, New York.
- Ranjbar, V., Amini, M. and Bozorgnia, A. (2008). Asymptotic behavior of weighted sums of dependent random variables with heavy tailed distributions. *J. Sciences, Islamic Republic of Iran*. **19**, 357-363.
- Renyi, A. (1961). On measures of entropy and information. Proc. 4th Berkeley Symposium. *Statist. Probability*. **1**, 547-561.
- Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379-423.
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*. **8**, 229-231.
- Torkkola, K. (2003). Feature extraction by Non-parametric mutual information maximization. *J. of Machine Learning Research*, **3**, 1415-1438.

G. R. Mohtashami Borzadaran

Department of Statistics and
Ordered and Spatial Data Center
of Excellence
Ferdowsi University of Mashhad,
Mashhad, Iran.
email: *grmohtashami@um.ac.ir*
email: *gmb1334@yahoo.com*

M. Amini

Department of Statistics and
Ordered and Spatial Data Center
of Excellence
Ferdowsi University of Mashhad,
Mashhad, Iran.