成城大学経済研究所
研究報告 No. 27

Empirical Copulas and Some Applications

Hideatsu Tsukahara

2000年12月

The Institute for Economic Studies
Seijo University

6–1–20, Seijo, Setagaya
Tokyo 157-8511, Japan

Empirical Copulas and Some Applications

Hideatsu Tsukahara
Department of Economics, Seijo University

December 6, 2000

Abstract

This paper is for the most part a survey of various results on the copulas scattered in the literature, in a unified fashion. We shall present some asymptotic results on the empirical copula process and use them to study the asymptotic properties of tests of independence and measures of dependence based on the empirical copula.

1 Introduction

Let $F(x_1, \ldots, x_d)$ be a d-dimensional distribution function, and F_i be the ith marginal distribution function of F. It is known that there exists a distribution function C on $[0,1]^d$ with uniform marginals such that

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)), \quad \text{for all } (x_1, \ldots, x_d).$$

See Sklar (1959, 1973), Schweizer (1991), Moore and Spruill (1975), Deheuvels (1979, 1980). C is called the *copula associated with* F (some authors call it the *dependence function*). In general, any distribution function on $[0,1]^d$ with uniform marginals is called a d-dimensional *copula*. When F is continuous, it is easy to see that the copula associated with F is uniquely determined and is given by

$$F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)), \ 0 \le u_i \le 1, \ i = 1, \ldots, d.$$

Copulas have recently been drawing some attention mainly as a tool to model various dependence among random variables, including the fields of financial risk management and multivariate survival analysis, and Joe (1997) and Nelson (1999) both successfully present their use. However, the empirical copulas have not been carefully studied despite the usefulness in some situations with low dimension. Testing independence is one of those situations and is introduced in Section 2. In Section 3, we define the empirical copula and some related statistics, and Section 4 presents some asymptotic results on the empirical copula process. In particular, we give an asymptotic representation of the empirical copula process, which is useful in deriving limit distributions of many statistics based on the empirical

copula. In Section 5, we examine two well-known rank correlation coefficients and other measures of dependence from the viewpoint of copulas, and use the results in Section 4 to derive the limiting distributions of the statistics which generalize the rank correlation. Section 6 concludes the paper with some remarks.

2 Tests of Independence

Let $X^k = (X_1^k, \ldots, X_d^k)$, $k = 1, \ldots, n$ be independent and identically distributed random variables with a d-dimensional continuous distribution function F. Consider the following *Hypothesis of Independence*:

$$H_0 : F = F_1 \otimes \cdots \otimes F_d$$

In the literature, several rank tests are suggested; see Hájek and Šidák (1967), Ruymgaart (1973), Kendall and Gibbons (1990). In this paper, we concentrate on tests based on empirical distribution functions. Let

$$\mathbb{F}_n(x_1, \ldots, x_d) \triangleq \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{X_1^k \le x_1, \ldots, X_d^k \le x_d\}} = \frac{1}{n} \sum_{k=1}^{n} \prod_{i=1}^{d} \mathbf{1}_{\{X_i^k \le x_i\}} \tag{2.1}$$

$$\mathbb{F}_{ni}(x_i) \triangleq \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{X_i^k \le x_i\}} \tag{2.2}$$

The following test statistics based on $\mathbb{F}_n$ and $\mathbb{F}_{ni}$ have been proposed.

$$A_n \triangleq \sup_{x \in \mathbb{R}^d} \left| \mathbb{F}_n(x) - \prod_{i=1}^{d} \mathbb{F}_{ni}(x_i) \right|,$$

$$B_n \triangleq \int_{\mathbb{R}^d} \left[\mathbb{F}_n(x) - \prod_{i=1}^{d} \mathbb{F}_{ni}(x_i) \right]^2 d\mathbb{F}_n(x),$$

$$B_n' \triangleq \int_{\mathbb{R}^d} \left[\mathbb{F}_n(x) - \prod_{i=1}^{d} \mathbb{F}_{ni}(x_i) \right]^2 \prod_{i=1}^{d} d\mathbb{F}_{ni}(x_i),$$

$$D_n \triangleq \int_{\mathbb{R}^d} \left[\mathbb{F}_n(x) - \prod_{i=1}^{d} \mathbb{F}_{ni}(x_i) \right]^2 \prod_{i=1}^{d} w_i(\mathbb{F}_{ni}(x_i))\, d\mathbb{F}_n(x),$$

$$D_n' \triangleq \int_{\mathbb{R}^d} \left[\mathbb{F}_n(x) - \prod_{i=1}^{d} \mathbb{F}_{ni}(x_i) \right]^2 \prod_{i=1}^{d} w_i(\mathbb{F}_{ni}(x_i)) \prod_{i=1}^{d} d\mathbb{F}_{ni}(x_i),$$

where w_i in D_n and D_n' is some positive weight function on $(0, 1)$. A_n is proposed by Blum, Kiefer and Rosenblatt (1961), but it is not very tractable. B_n and B_n' are due to Hoeffding (1948) for the case $d = 2$, and to Blum, Kiefer and Rosenblatt (1961) for the general case. They also mentioned the use of D_n', and De Wet (1980) studies D_n' when $d = 2$. The statistic A_n is Kolmogorov-Smirnov type, B_n and B_n' are Cramér-von Mises type, and D_n and D_n' are Anderson-Darling type, and so the motivation of introducing them is fairly obvious. Note also that each of the above statistics is distribution-free.

3 Empirical Copulas

Let C be the copula associated with a d-dimensional continuous distribution function F. For convenience, we use the following representation of the X^k. Let $\xi^k = (\xi_1^k, \ldots, \xi_d^k)$, $k = 1, \ldots, n$ be independently and identically distributed with distribution function C, and set

$$\mathbb{G}_n(u) \triangleq \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{\xi_1^k \le u_1, \ldots, \xi_d^k \le u_d\}}, \quad \mathbb{G}_{ni}(u_i) \triangleq \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{\xi_i^k \le u_i\}}.$$

Put $X^k = (X_1^k, \ldots, X_d^k) \triangleq (F_1^{-1}(\xi_1^d), \ldots, F_d^{-1}(\xi_d^k))$. Then the distribution function of X^k is F for $k = 1, \ldots, n$. As in (2.1) and (2.2), we denote by $\mathbb{F}_n$ and $\mathbb{F}_{ni}$ the empirical joint and marginal distribution functions based on the X^k. We then have

$$\mathbb{F}_n(x) = \mathbb{G}_n(F_1(x_1), \ldots, F_d(x_d))$$

and

$$\prod_{i=1}^{d} \mathbb{F}_{ni}(x_i) = \prod_{i=1}^{d} \mathbb{G}_{ni}(F_i(x_i)).$$

Define

$$\mathbb{T}_n(u) \triangleq \mathbb{G}_n(u) - \prod_{i=1}^{d} \mathbb{G}_{ni}(u_i).$$

Then one can easily see that the statistics introduced in Section 2 have the following representation:

$$A_n = \sup_{u \in [0,1]^d} |\mathbb{T}_n(u)| =: \|\mathbb{T}_n\|,$$

$$B_n = \int_{[0,1]^d} [\mathbb{T}_n(u)]^2 \, d\mathbb{G}_n(u), \qquad B_n' = \int_{[0,1]^d} [\mathbb{T}_n(u)]^2 \prod_{i=1}^{d} d\mathbb{G}_{ni}(u_i),$$

$$D_n = \int_{[0,1]^d} [\mathbb{T}_n(u)]^2 \prod_{i=1}^{d} w_i(\mathbb{G}_{ni}(u_i)) \, d\mathbb{G}_n(u),$$

$$D_n' = \int_{[0,1]^d} [\mathbb{T}_n(u)]^2 \prod_{i=1}^{d} w_i(\mathbb{G}_{ni}(u_i)) \prod_{i=1}^{d} d\mathbb{G}_{ni}(u_i).$$

Since F is assumed to be continuous, it is easy to see that $F = F_1 \otimes \cdots \otimes F_d$ if and only if $C = \lambda$, where $\lambda(u_1, \ldots, u_d) \triangleq \prod_{i=1}^{d} u_i$. This fact indicates the use of statistics based on empirical copula in testing the hypothesis of independence. Let us define the *empirical copula* by

$$\mathbb{C}_n(u) \triangleq \mathbb{F}_n(\mathbb{F}_{n1}^{-1}(u_1), \ldots, \mathbb{F}_{nd}^{-1}(u_d)),$$

and set

$$\widetilde{A}_n \triangleq \sup_{u\in[0,1]^d} |\mathbb{C}_n(u) - \lambda(u)|,$$

$$\widetilde{B}_n \triangleq \int_{[0,1]^d} [\mathbb{C}_n(u) - \lambda(u)]^2 \, d\mathbb{C}_n(u), \qquad \widetilde{B}'_n \triangleq \int_{[0,1]^d} [\mathbb{C}_n(u) - \lambda(u)]^2 \, d\lambda(u),$$

$$\widetilde{D}_n \triangleq \int_{[0,1]^d} [\mathbb{C}_n(u) - \lambda(u)]^2 \prod_{i=1}^{d} w_i(u_i) \, d\mathbb{C}_n(u),$$

$$\widetilde{D}'_n \triangleq \int_{[0,1]^d} [\mathbb{C}_n(u) - \lambda(u)]^2 \prod_{i=1}^{d} w_i(u_i) \, d\lambda(u).$$

Using the representation introduced earlier,

$$\begin{aligned}\mathbb{C}_n(u) &= \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{\xi_1^k \le F_1\circ\mathbb{F}_{n1}^{-1}(u_1),\dots,\xi_d^k \le F_d\circ\mathbb{F}_{nd}^{-1}(u_d)\}} \\ &= \mathbb{G}_n(F_1\circ\mathbb{F}_{n1}^{-1}(u_1),\dots,F_d\circ\mathbb{F}_{nd}^{-1}(u_d)).\end{aligned}$$

By the definition of $\mathbb{F}_{ni}^{-1}$, it follows that $\mathbb{F}_{ni}^{-1}(u_i) = F_i(\mathbb{G}_{ni}^{-1}(u_i))$, and hence

$$\mathbb{C}_n(u) = \mathbb{G}_n(\mathbb{G}_{n1}^{-1}(u_1),\dots,\mathbb{G}_{nd}^{-1}(u_d)).$$

This implies that the law of $\mathbb{C}_n$ is the same for all F having the same associated copula C. Note that

$$A_n = \sup_{\substack{1\le k_i\le n\\ 1\le i\le d}} \left|\mathbb{C}_n\left(\frac{k_1}{n},\cdots,\frac{k_d}{n}\right) - \prod_{i=1}^{d}\frac{k_i}{n}\right| = \widetilde{A}_n.$$

We also have

$$B_n = \int_{[0,1]^d} \left[\mathbb{C}_n(u) - \prod_{i=1}^{d} \mathbb{G}_{ni}(\mathbb{G}_{ni}^{-1}(u_i))\right]^2 d\mathbb{C}_n(u),$$

so B_n is not exactly the same as $\widetilde{B}_n$, but later we show that the difference between them is of order $n^{-3/2}$.

For the computation of the test statistics, one may use the representation

$$B_n = \frac{1}{n}\sum_{k=1}^{n}\left[\frac{1}{n}\sum_{l=1}^{n}\mathbf{1}_{\{X_1^l\le X_1^k,\dots,X_d^l\le X_d^k\}} - \frac{\prod_{i=1}^{d}R_i^k}{n^d}\right],$$

where R_i^k is the rank of X_i^k among $X_i^1,\dots,X_i^n$. For $d=2$, it holds that

$$B_n = \frac{1}{n^5}\sum_{k=1}^{n}[N_1(k)N_4(k) - N_2(k)N_3(k)]^2,$$

where

$$N_1(k) \triangleq \#\text{ of } \{(X_1^l, X_2^l) \text{ such that } X_1^l \le X_1^k,\ X_2^l \le X_2^k\},$$
$$N_2(k) \triangleq \#\text{ of } \{(X_1^l, X_2^l) \text{ such that } X_1^l \le X_1^k,\ X_2^l > X_2^k\},$$
$$N_3(k) \triangleq \#\text{ of } \{(X_1^l, X_2^l) \text{ such that } X_1^l > X_1^k,\ X_2^l \le X_2^k\},$$
$$N_4(k) \triangleq \#\text{ of } \{(X_1^l, X_2^l) \text{ such that } X_1^l > X_1^k,\ X_2^l > X_2^k\}.$$

Analogously, one obtains

$$B_n' = \frac{1}{n}\sum_{k_1=1}^{n}\cdots\sum_{k_d=1}^{n}\left[\frac{1}{n}\sum_{l=1}^{n}\mathbf{1}_{\{R_1^l\le k_1,\ldots,R_d^l\le k_d\}} - \prod_{i=1}^{d}\frac{k_i}{n}\right].$$

and

$$A_n = \sup_{\substack{1\le k_i\le n\\ 1\le i\le d}}\left|\frac{1}{n}\sum_{l=1}^{n}\mathbf{1}_{\{R_1^l\le k_1,\ldots,R_d^l\le k_d\}} - \prod_{i=1}^{d}\frac{k_i}{n}\right|$$

4 Asymptotic Theory

To study the asymptotic behavior of the statistics in the preceding section, we first need to introduce several processes which appear in the limiting random variables.

Let C be any copula. The d-dimensional *C-Brownian sheet* $\mathbb{W}^C$ is a continuous Gaussian random field with

$$\mathbf{E}(\mathbb{W}^C(u)) = 0, \quad \mathbf{E}(\mathbb{W}^C(u)\mathbb{W}^C(v)) = C(u\wedge v),$$

where $u\wedge v = (u_1\wedge v_1,\ldots,u_d\wedge v_d)$. The d-dimensional *pinned C-Brownian sheet* $\mathbb{U}^C$ is a continuous Gaussian random field with

$$\mathbf{E}(\mathbb{U}^C(u)) = 0, \quad \mathbf{E}(\mathbb{U}^C(u)\mathbb{U}^C(v)) = C(u\wedge v) - C(u)C(v).$$

Its version may be given by

$$\mathbb{U}^C(u) = \mathbb{W}^C(u) - C(u)\mathbb{W}^C(\mathbf{1})$$

The d-dimensional *Brownian pillow* [Piterbarg (1996)], or *completely tucked Brownian sheet* [Van der Vaart and Wellner (1996)] $\mathbb{V}$ is a continuous Gaussian random field with

$$\mathbf{E}(\mathbb{V}(u)) = 0, \quad \mathbf{E}(\mathbb{V}(u)\mathbb{V}(v)) = \prod_{i=1}^{d}(u_i\wedge v_i - u_iv_i).$$

See Adler (1990) and Piterbarg (1996) for more information on these processes.

4.1 Special Construction

Neuhaus (1971) and Bickel and Wichura (1971) showed

$$\mathbb{U}_n^C \triangleq \sqrt{n}(\mathbb{G}_n - C) \xrightarrow{\mathcal{L}} \mathbb{U}^C,$$

for $J_1^{(d)}$-topology on D_d, as $n \to \infty$. For the precise definition of the space D_d and the $J_1^{(d)}$-topology, and for more on multivariate empirical processes, we refer to Einmahl (1987). By Skorokhod-Dudley-Wichura representation theorem, there exist a triangular array $\xi^{nk} = (\xi_1^{nk}, \ldots, \xi_d^{nk})$, $k = 1, \ldots, n$, $n \in \mathbb{N}$ of row-independent random vectors with distribution function C and a pinned C-Brownian sheet $\mathbb{U}^C$, all defined on some probability space, such that

$$\|\mathbb{U}_n^C - \mathbb{U}^C\| \to 0, \quad \text{a.s.}$$

where $\mathbb{U}_n^C$ is defined as above with empirical distribution function based on $\xi^{n1}, \ldots, \xi^{nn}$. The merit of this construction is that it is easier to deal with random variables directly than their laws. When $C = \lambda$, we simply write $\mathbb{U}_n$ and $\mathbb{U}$ for $\mathbb{U}_n^\lambda$ and $\mathbb{U}^\lambda$ respectively. All the asymptotics below is based on the above special construction.

4.2 Asymptotics for $\mathbb{T}_n$ when $C = \lambda$

In order to find the null distributions of test statistics, we need to study the asymptotic behavior of $\mathbb{T}_n$ when $C = \lambda$. In this case, we can write

$$\sqrt{n}\mathbb{T}_n(u) = \mathbb{U}_n(u) - \sqrt{n}\left[\prod_{i=1}^d \mathbb{G}_{ni}(u_i) - \lambda(u)\right].$$

Using the identity

$$\prod_{i=1}^d a_i - \prod_{i=1}^d b_i = \sum_{i=1}^d (a_i - b_i) \prod_{j=1}^{i-1} b_j \prod_{h=i+1}^d a_h, \tag{4.1}$$

we get

$$\sqrt{n}\mathbb{T}_n(u) = \mathbb{U}_n(u) - \sum_{i=1}^d \mathbb{U}_n(\mathbf{1}, u_i, \mathbf{1}) \prod_{j=1}^{i-1} u_j \prod_{h=i+1}^d \mathbb{G}_{nh}(u_h),$$

where $\mathbf{1}$ is the vector consisting of 1 with appropriate dimension. We define a random field $\mathbb{T}$ by

$$\mathbb{T}(u) \triangleq \mathbb{U}(u) - \sum_{i=1}^d \mathbb{U}(\mathbf{1}, u_i, \mathbf{1}) \prod_{j \neq i} u_j.$$

Its covariance is given by

$$\mathbf{E}(\mathbb{T}(u)\mathbb{T}(v)) = \prod_{i=1}^d (u_i \wedge v_i) + (d-1) \prod_{i=1}^d u_i v_i - \sum_{i=1}^d (u_i \wedge v_i) \prod_{j \neq i} u_j v_j$$

Notice that if $|a_i| \leq 1$ and $|b_i| \leq 1$ in the identity (4.1), then

$$\left| \prod_{i=1}^{d} a_i - \prod_{i=1}^{d} b_i \right| = \sum_{i=1}^{d} |a_i - b_i|. \tag{4.2}$$

This gives

$$\begin{aligned}
&\|\sqrt{n}\mathbb{T}_n(u) - \mathbb{T}(u)\| \\
\leq &\|\mathbb{U}_n - \mathbb{U}\| + \sum_{i=1}^{d} \left[\|\mathbb{U}_n - \mathbb{U}\| + \|\mathbb{U}_n\| \left| \prod_{j=1}^{i-1} u_j \prod_{h=i+1}^{d} \mathbb{G}_{nh}(u_h) - \prod_{j \neq i} u_j \right| \right] \\
\leq &(d+1)\|\mathbb{U}_n - \mathbb{U}\| + \|\mathbb{U}_n\| \sum_{i=1}^{d} \sum_{h=i+1}^{d} \sup_{u \in [0,1]} |\mathbb{G}_{nh}(u) - u|.
\end{aligned}$$

Two terms on the right clearly converge in probability to 0, hence we conclude that $\|\sqrt{n}\mathbb{T}_n - \mathbb{T}\| \xrightarrow{\mathrm{P}} 0$.

When $d = 2$, the covariance function of $\mathbb{T}$ is equal to $(u_1 \wedge v_1 - u_1 v_1)(u_2 \wedge v_2 - u_2 v_2)$. This shows that the limiting random field of $\sqrt{n}\mathbb{T}_n$ is a Brownian pillow. However, $\mathbb{T}$ is no longer a Brownian pillow for $d \geq 3$.

4.3 Asymptotics for the Empirical Copula Process

We define the *empirical copula process* $\mathbb{D}_n^C$ by

$$\mathbb{D}_n^C(u) \triangleq \sqrt{n}(\mathbb{C}_n(u) - C(u)).$$

The following Bahadur-Kiefer type asymptotic representation of the empirical copula process was stated in Stute (1984), p. 371, and he indicated the outline of a proof. We shall now write down the details of the proof.

4.1 Theorem *Assume that the copula C associated with F is twice continuously differentiable on $(0,1)^d$ and the second derivative is continuous on $[0,1]^d$. Then with probability 1, we have*

$$\mathbb{D}_n^C(u) = \mathbb{U}_n^C(u) + \sum_{i=1}^{d} C^i(u)\mathbb{U}_n^C(\mathbf{1}, u_i, \mathbf{1}) + O\left(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}\right),$$

uniformly in u. Here we use the following notation:

$$C^i(u) = \frac{\partial C}{\partial u_i}(u), \quad i = 1, 2, \ldots, d.$$

Proof. Write

$$\mathbb{D}_n^C(u) = \mathbb{U}_n^C(u) + J_n(u) + K_n(u),$$

where

$$J_n(u) = \sqrt{n}[C(\mathbb{G}_{n1}^{-1}(u_1), \ldots, \mathbb{G}_{nd}^{-1}(u_d)) - C(u)],$$
$$K_n(u) = \mathbb{U}_n^C(\mathbb{G}_{n1}^{-1}(u_1), \ldots, \mathbb{G}_{nd}^{-1}(u_d)) - \mathbb{U}_n^C(u).$$

By the differentiability assumption on C, we can use Taylor's theorem to get

$$J_n(u) = \sum_{i=1}^{d} C^i(u)\sqrt{n}(\mathbb{G}_{ni}^{-1}(u_i) - u_i) + \frac{\sqrt{n}}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\partial^2 C}{\partial u_i \partial u_j}(v^*)(\mathbb{G}_{ni}^{-1}(u_i) - u_i)(\mathbb{G}_{nj}^{-1}(u_j) - u_j),$$

where v^* lies in the interior of the line segment joining $(\mathbb{G}_{n1}^{-1}(u_1), \ldots, \mathbb{G}_{nd}^{-1}(u_d))$ and $(u_1, \ldots, u_d)$. We know that

$$\sqrt{n}(\mathbb{G}_{ni}^{-1}(u_i) - u_i) = -\mathbb{U}_n^C(\mathbf{1}, \mathbb{G}_{ni}^{-1}, \mathbf{1}) + \sqrt{n}(\mathbb{G}_{ni} \circ \mathbb{G}_{ni}^{-1}(u_i) - u_i).$$

Using $|\mathbb{G}_{ni} \circ \mathbb{G}_{ni}^{-1}(u_i) - u_i| \leq 1/n$ and the Smirnov-Chung law of the iterated logarithm for the empirical distribution functions

$$\|\mathbb{G}_{ni} - I\| = \|\mathbb{G}_{ni}^{-1} - I\| = O\left(n^{-1/2}(\log\log n)^{1/2}\right),$$

we obtain

$$J_n(u) = -\sum_{i=1}^{d} C^i(u)\mathbb{U}_n^C(\mathbf{1}, \mathbb{G}_{ni}^{-1}(u_i), \mathbf{1}) + O(n^{-1/2}\log\log n), \quad \text{a.s.}$$

uniformly in u. Furthermore, Stute (1982) (p. 99) shows that

$$\sup_{u_i} |\mathbb{U}_n^C(\mathbf{1}, \mathbb{G}_{ni}^{-1}(u_i), \mathbf{1}) - \mathbb{U}_n^C(\mathbf{1}, u_i, \mathbf{1})| = O(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}), \quad \text{a.s.}$$

It thus follows that uniformly in u,

$$J_n(u) = -\sum_{i=1}^{d} C^i(u)\mathbb{U}_n^C(\mathbf{1}, u_i, \mathbf{1}) + O(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}), \quad \text{a.s.}$$

Next we consider $K_n(u)$. For $a_n \in \mathbb{R}$, put

$$\omega(a_n) = \sup\left\{\left|\mathbb{U}_n^C\left((x_1, y_1] \times \cdots \times (x_d, y_d]\right)\right| : y_i - x_i \leq a_n,\ 1 \leq i \leq d\right\};$$

here we regard $\mathbb{G}_n$ and C as measures on $[0,1]^d$. By Stute (1984), Theorem 1.7, we can find two positive constants c_1 and c_2 such that

$$P(\omega(a) > s) \leq c_1 a_n^{-d} \exp\left[-\frac{c_2 s^2}{a_n}\right].$$

We take $a_n = n^{-1/2}(\log\log n)^{1/2}$, suggested by the Smirnov-Chung law of the iterated logarithm. For this a_n, we choose $s = s_n = n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}$, so that

$$\sum_{n=1}^{\infty} a_n^{-d} \exp\left[-\frac{c_2 s_n^2}{a_n}\right] < \infty.$$

This implies, by Borel-Cantelli lemma, that

$$\sup_u |K_n(u)| \le \text{const} \cdot \omega_n(a_n) = O(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}), \quad \text{a.s.}$$

proving the theorem. ■

As an immediate corollary to this theorem, we obtain

$$\|\mathbb{D}_n^C - \mathbb{D}^C\| \xrightarrow{\text{a.s.}} 0,$$

where

$$\mathbb{D}^C(u) \triangleq \mathbb{U}^C(u) - \sum_{i=1}^{d} C^i(u)\mathbb{U}^C(\mathbf{1}, u_i, \mathbf{1}).$$

(In fact, to get the invariance principle $\mathbb{D}_n^C \xrightarrow{\mathcal{L}} \mathbb{D}^C$, we only need to assume that C is continuously differentiable on $(0,1)^d$). Clearly $\mathbb{D}^C$ is a centered Gaussian random field. It is possible to write down its covariance function, but it gains us little. In the special case where $C = \lambda$, $C^i(u) = \prod_{j\neq i} u_j$, so we have $\mathbb{D}^\lambda \stackrel{\mathcal{L}}{=} \mathbb{T}$.

4.4 Asymptotic Distributions of the Test Statistics under H_0

By the results in Section 4.2, one immediately sees that

$$\sqrt{n}A_n = \sqrt{n}\widetilde{A}_n = \|\sqrt{n}\mathbb{T}_n\| \xrightarrow{\text{P}} \|\mathbb{T}\|,$$

For only a handful of Gaussian random fields, the exact distribution of the maximum is known (Adler (1990)), but $\mathbb{T}$ is not one of them. Piterbarg (1996) (Theorem 9.2) obtained an approximation of the tail probability for the maximum of Brownian pillow $\mathbb{V}$. When $d = 2$, $\mathbb{T}$ is a Brownian pillow, so his result may be used to get approximate critical values for $\|\mathbb{T}\|$: when $d = 2$,

$$P(\|\mathbb{T}\| > u) = 32\pi u^2(1 - \Phi(4u))(1 + o(1)), \quad \text{as } u \to \infty,$$

where Φ is the distribution function of the standard normal distribution.

As for the Cramér-von Mises type statistics, it is straightforward to show that

$$n\widetilde{B}_n' = \int_{[0,1]^d} [\mathbb{D}_n^\lambda(u)]^2\, du \xrightarrow{\text{P}} \int_{[0,1]^d} \mathbb{T}^2(u)\, du.$$

Also, since $\|\mathbb{D}_n^\lambda\| = O_p(1)$, it is easy to see that

$$|\widetilde{B}_n - \widetilde{B}'_n| \le \frac{1}{n}\left|\int_{[0,1]^d} [\mathbb{D}_n^\lambda(u)]^2\, d(\mathbb{C}_n(u) - \lambda(u))\right| \xrightarrow{\mathrm{P}} 0.$$

By the definition of B_n,

$$\begin{aligned}|B_n - \widetilde{B}_n| &= \left|\int \left[\mathbb{C}_n(u) - \prod_{i=1}^d \mathbb{G}_{ni}(\mathbb{G}_{ni}^{-1}(u_i))\right]^2 d\mathbb{C}_n(u) - \widetilde{B}_n\right| \\ &\le 2\|\mathbb{C}_n - \lambda\| \int \left|\prod_{i=1}^d \mathbb{G}_{ni}(\mathbb{G}_{ni}^{-1}(u_i)) - \lambda(u)\right| d\mathbb{C}_n(u) \\ &\quad + \int \left[\prod_{i=1}^d \mathbb{G}_{ni}(\mathbb{G}_{ni}^{-1}(u_i)) - \lambda(u)\right]^2 d\mathbb{C}_n(u)\end{aligned}$$

Using $\|\mathbb{C}_n - \lambda\| = O_p(n^{-1/2})$, $|\mathbb{G}_{ni} \circ \mathbb{G}_{ni}^{-1}(u_i) - u_i| \le 1/n$ and the identity (4.1), one finds $|B_n - \widetilde{B}_n| = O_p(n^{-3/2})$, whence

$$n\widetilde{B}_n \xrightarrow{\mathrm{P}} \int_{[0,1]^d} \mathbb{T}^2(u)\, du.$$

B'_n can be handle analogously, so all four test statistics B_n, B'_n, $\widetilde{B}_n$ and $\widetilde{B}'_n$ have the same limiting distribution as we naturally expect, namely, that of $\int \mathbb{T}^2(u)\, du$. To find this distribution, we need some general theory.

Let $\{Y(t)\}_{t\in[0,1]^d}$ be a centered Gaussian random field with paths in the space of square integrable functions on $[0,1]^d$ and with covariance function $K(t,u)$. Also let $\lambda_1, \lambda_2, \ldots$, and $f_1, f_2, \ldots$ be the eigenvalues and normalized eigenfunctions of the kernel K. This means that the λ_j and f_j satisfy

$$\int_{[0,1]^d} K(t,u) f_j(t)\, dt = \lambda_j f_j(t), \quad u \in [0,1]^d$$

and

$$\int_{[0,1]^d} f_j(t) f_k(t)\, dt = \begin{cases} 0 & \text{for } j \ne k \\ 1 & \text{for } j = k \end{cases}$$

We assume that the following Kac-Siegert decomposition (Kac and Siegert (1947)) holds:

$$K(t,u) = \sum_{j=1}^\infty \lambda_j f_j(t) f_j(u). \tag{4.3}$$

This holds whenever K is continuous by Mercer's theorem. Consider the process $\mathbb{U}(u)/\sqrt{u(1-u)}$, where $\mathbb{U}$ is a 1-dimensional Brownian bridge. This process appears in the asymptotic distribution of the classical Anderson-Darling statistics,

and the covariance function of this process gives an interesting example in which (4.3) holds while K is not continuous. See Anderson and Darling (1952).

Under the above assumptions, the following Karhunen-Loève expansion of $Y(t)$ holds:

$$Y(t) = \sum_{j=1}^{\infty} Z_j f_j(t), \quad \text{(in } L^2 \text{ uniformly in } t\text{)}.$$

where the Z_j's are uncorrelated zero-mean normal rv's with $\text{var}(Z_j) = \lambda_j$. See Ash and Gardner (1975). Setting $Z_j^* = Z_j/\sqrt{\lambda}$, which has a standard normal distribution, it follows that

$$\int_0^1 Y^2(t)\,dt = \sum_{j=1}^{\infty} \lambda_j Z_j^{*2},$$

Thus we obtains the characteristic function of $\int_0^1 Y^2(t)\,dt$:

$$\mathbf{E}\left(\exp\left\{ir\int_0^1 Y_t^2\,dt\right\}\right) = \prod_{j=1}^{\infty}(1-2i\lambda_j r)^{-\frac{1}{2}}.$$

We now apply the above result with $Y = \mathbb{T}$. In the case with $d = 2$, we know that $K(u,v) = (u_1 \wedge v_1 - u_1v_1)(u_2 \wedge v_2 - u_2v_2)$. Blum, Kiefer and Rosenblatt (1961) showed that the characteristic function of $\int \mathbb{T}^2(u)du$ is given by

$$\prod_{j,k=1}^{\infty}\left(1-\frac{2ir}{\pi^4 j^2 k^2}\right)^{-\frac{1}{2}}.$$

Then inverting this characteristic function numerically gives the probabilities we need.

When $d \geq 3$, the form of $K(u,v)$ is not very tractable. Deheuvels (1981) and Cotterill and Csörgő (1985) both claim to find the characteristic function of $\int \mathbb{T}^2(u)\,du$, but the two forms are apparently different.

In a comletely analogous fashion, one can see that the Anderson-Darling type test statistics D_n, D_n', $\widetilde{D}_n$ and $\widetilde{D}_n'$ are all distributed asymptotically as $\int \mathbb{T}^2(u)\prod w_i(u_i)\,du$ under appropriate assumptions on the weight functions w_i. Its distribution has not yet been found except for the case $d = 2$. In that case, De Wet (1980) used a different method to derive the asymptotic distribution of D_n' directly. For the general case, we would need to develop the asymptotic theory of weighted empirical copula processes.

5 Rank Correlations and Other Measures of Dependence

5.1 Population versions of rank correlations

Let X and Y be random variables with joint distribution function $F(x,y)$. We denote by F_X and F_Y the marginal distribution functions of X and Y respectively.

The simplest ordinal measure of association between X and Y may be given by

$$P[(X - x_0)(Y - y_0) > 0] = 1 - F_X(x_0) - F_Y(y_0) + 2F(x_0, y_0)$$

for some fixed (x_0, y_0). If we take $x_0 = \operatorname{med} X$ and $y_0 = \operatorname{med} Y$, we get $2F(\operatorname{med} X, \operatorname{med} Y)$. The *Blomqvist's q* is a symmetrized version of this, and is defined by

$$\begin{aligned} q &\triangleq P[(X - \operatorname{med} X)(Y - \operatorname{med} Y) > 0] - P[(X - \operatorname{med} X)(Y - \operatorname{med} Y) < 0] \\ &= 2P[(X - \operatorname{med} X)(Y - \operatorname{med} Y) > 0] - 1 \\ &= E[\operatorname{sgn}(X - \operatorname{med} X)\operatorname{sgn}(Y - \operatorname{med} Y)] \end{aligned}$$

The above choice of (x_0, y_0) is quite arbitrary, so it seems more natural to take average over all (x_0, y_0) with weights given by $F(x, y)$. This amounts to considering

$$\begin{aligned} \tau &\triangleq P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] \\ &= 2P[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1, \end{aligned}$$

where (X_1, Y_1) and (X_2, Y_2) are independent randam vectors with distribution function F. Since $E[\operatorname{sgn}(X_1 - X_2)] = E[\operatorname{sgn}(Y_1 - Y_2)] = 0$, we have

$$\tau = E[\operatorname{sgn}(X_1 - X_2)\operatorname{sgn}(Y_1 - Y_2)] = \operatorname{cov}[\operatorname{sgn}(X_1 - X_2)\operatorname{sgn}(Y_1 - Y_2)],$$

which is also equal to the correlation coefficient between $\operatorname{sgn}(X_1 - X_2)$ and $\operatorname{sgn}(Y_1 - Y_2)$ because $\operatorname{var}[\operatorname{sgn}(X_1 - X_2)] = \operatorname{var}[\operatorname{sgn}(Y_1 - Y_2)] = 1$. This τ is known as *Kendall's tau*. In terms of F, it can be written as (see Schweizer and Wolff (1981))

$$\tau = 4 \iint F(x, y)\, dF(x, y) - 1. \tag{5.1}$$

Using the copula C associated with F, we have

$$\tau = 4 \iint C(u, v)\, dC(u, v) - 1. \tag{5.2}$$

If we take average with weights $F_X(x)F_Y(y)$ above instead of $F(x, y)$, then we get

$$P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0],$$

where (X_1, Y_1), (X_2, Y_2) and (X_3, Y_3) are independent randam vectors with distribution function F. A similar argument as above shows that this quantity equals

$$E[\operatorname{sgn}(X_1 - X_2)\operatorname{sgn}(Y_1 - Y_3)];$$

it also equals the covariance and correlation coefficient between $\operatorname{sgn}(X_1 - X_2)$ and $\operatorname{sgn}(Y_1 - Y_3)$). This lies between $-1/3$ and $1/3$, so we define

$$\rho \triangleq 3E[\operatorname{sgn}(X_1 - X_2)\operatorname{sgn}(Y_1 - Y_2)],$$

so that we have $-1 \leq \rho \leq 1$. This ρ is called *Spearman's rho*. A straightforward calculation shows that

$$\rho = 12 \iint [F(x,y) - F_X(x)F_Y(y)]\, dF_X(x) dF_Y(y) \tag{5.3}$$

$$= 3\left[4 \iint F(x,y)\, dF_X(x) dF_Y(y) - 1\right] \tag{5.4}$$

One should contrast (5.1) with (5.4) to see the similarity between τ and $\frac{1}{3}\rho$. Another expression for ρ is (Joag-Dev (1984))

$$\rho = 3 \iint [1 - 2F_X(x)][1 - 2F_Y(y)]\, dF(x,y). \tag{5.5}$$

Using the copula C, we obtain

$$\rho = 12 \iint [C(u,v) - uv)]\, dudv = 3 \iint (1-2u)(1-2v)\, dC(u,v). \tag{5.6}$$

5.2 Sample rank correlations

As is well known, there are sample versions of the measures of association discussed in the preceding subsection. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample of size n from the distribution F. As for Blomqvist's q, $\widehat{q}$ is defined to be the number of (X_i, Y_i)'s in the first and third quadrants around the sample medians minus the number of (X_i, Y_i)'s in the second and fourth quadrants around the sample medians, divided by n. If n is even, this is unambiguously defined, but for odd n, some modification is necessary; see Kruskal (1958). Kendall's sample rank correlation coefficient is defined by

$$\widehat{\tau} \triangleq \frac{1}{n(n-1)} \sum\sum_{i \neq j} \text{sgn}(X_i - X_j)\text{sgn}(Y_i - Y_j)$$

Let K and L be the number of concordant and discordant pairs $((X_i, Y_i), (X_j.Y_j))$, $i < j$ respectively (A pair $((X_i, Y_i), (X_j.Y_j))$ is called concordant if $(X_i - X_j)(Y_i - Y_j) > 0$; it is called discordant if $(X_i - X_j)(Y_i - Y_j) < 0$). Then one can show that

$$\widehat{\tau} = \frac{2(K-L)}{n(n-1)} = 1 - \frac{4L}{n(n-1)} = \frac{4K}{n(n-1)} - 1.$$

Spearman's sample rank correlation coefficient is by definition given by

$$\widehat{\rho} \triangleq \frac{1}{n(n^2-1)} \left[12 \sum_{i=1}^{n} R_i^X R_i^Y - 3n(n+1)^2\right] = 1 - \frac{6\sum_i (R_i^X - R_i^Y)^2}{n(n^2-1)},$$

where R_i^X is the rank of X_i in the X_j's, and similarly for R_i^Y.

Replacing F_X, F_Y and F with the corresponding empirical distribution functions $\mathbb{F}_{X,n}$, $\mathbb{F}_{Y,n}$ and $\mathbb{F}_n$ in (5.1), (5.3) and (5.5) does not yield the sample rank correlations τ and ρ exactly. In fact, an easy computation gives

$$4\iint \mathbb{F}_n(x,y)\,d\mathbb{F}_n(x,y) - 1 = \frac{4(K+n)}{n^2} - 1,$$

so

$$4\iint \mathbb{F}_n(x-,y-)\,d\mathbb{F}_n(x,y) - 1 = \frac{4K}{n^2} - 1$$

is much closer to $\widehat{\tau}$. For the Spearman's rank correlation, two natural empirical analogues of (5.3) and (5.5) are

$$12\iint [\mathbb{F}_n(x,y) - \mathbb{F}_{X,n}(x)\mathbb{F}_{Y,n}(y)]\,d\mathbb{F}_{X,n}(x)d\mathbb{F}_{Y,n}(y) = \frac{1}{n^3}\left[12\sum_{i=1}^{n} R_i^X R_i^Y - 3n(n+1)^2\right]$$

and

$$3\iint [1-2\mathbb{F}_{X,n}(x)][1-2\mathbb{F}_{Y,n}(y)]\,d\mathbb{F}_n(x,y) = \frac{1}{n^3}\left[12\sum_{i=1}^{n} R_i^X R_i^Y - 3n^2(n+2)\right].$$

They are both very close to $\widehat{\rho}$.

5.3 Empirical Copulas and Rank Correlations

Let $\mathbb{C}_n$ denote the empirical copula based on the sample $(X_1,Y_1),\ldots,(X_n,Y_n)$, and define

$$\widetilde{\rho}_n = 12\iint [\mathbb{C}_n(u,v) - uv]\,dudv$$
$$\widetilde{\tau}_n = 4\iint \mathbb{C}_n(u,v)\,d\mathbb{C}_n(u,v) - 1$$

These are empirical versions of (5.2) and (5.6) respectively and are natural estimators of ρ and τ, based on the empirical copula. Thanks to the results in Section 4.3, we can derive asymptotic properties of these statistics with ease. We assume throughout this subsection that $C(u,v)$ is continuously differentiable with the partial derivatives

$$C^1(u,v) = \frac{\partial C(u,v)}{\partial u}, \quad C^2(u,v) = \frac{\partial C(u,v)}{\partial v}.$$

Then,

$$\sqrt{n}(\widetilde{\rho}_n - \rho) = 12\iint \sqrt{n}[\mathbb{C}_n(u,v) - C(u,v)]\,dudv$$

converges in law to $12 \iint \mathbb{D}^C(u,v)\, dudv$, which has a centered normal distribution. Its variance is cumbersome, but if $C = \lambda$, i.e., the X_i's and Y_i's are independent, then $\mathbb{D}^\lambda$ is a Brownian pillow and a straightforward calculation shows that the variance of $12 \iint \mathbb{D}^C(u,v)\, dudv$ under $C = \lambda$ is 1. This agrees with the known asymptotic variance of Spearman's rank correlation coefficient (Kendall and Gibbons (1990)). More generally, let us define, for any function J on $[0,1]^3$,

$$S(C) \triangleq \iint J(u,v,C(u,v))\, dudv.$$

The corresponding sample quatity $S(\mathbb{C}_n)$ may be called Spearman type rank statistic (see Gaenssler and Stute (1987)). The above $\widetilde{\rho}_n$ corresponds to the case $J(u,v,w) = 12(w - uv)$. The asymptotic distribution for this type of statistics has been obtained by Gaenssler and Stute (1987):

5.1 Theorem *Assume that J has a continuous partial derivative $J^3(u,v,w)$ with respect to w such that $\sup_{u,v,w} |J^3(u,v,w)| < \infty$. Then $\sqrt{n}(S(C_n) - S(C))$ converges in distribution to a normal distribution with mean 0 and variance*

$$\sigma(S) \triangleq \operatorname{var}\left(\iint J^3(u,v,w) Z(u,v)\, dudv \right),$$

provided $\sigma(S) > 0$. The process $Z(u,v)$ is defined by

$$\begin{aligned} Z(u,v) \triangleq & \mathbf{1}_{\{\xi \le u,\, \eta \le v\}} - C(u,v) \\ & - C^1(u,v)(\mathbf{1}_{\{\xi \le u\}} - u) - C^2(u,v)(\mathbf{1}_{\{\eta \le v\}} - v), \end{aligned}$$

and (ξ, η) is a random vector with distribution function C.

The limiting distribution for $\widetilde{\tau}_n$ is more complicated. Corresponding to the Spearman type rank statistics, we put, for any function J on $[0,1]^3$,

$$T(C) \triangleq \iint J(u,v,C(u,v))\, dC(u,v).$$

We call $T(\mathbb{C}_n)$ a Kendall type rank statistic. Note that taking $J(u,v,w) = 4w - 1$ gives $\widetilde{\tau}_n$ defined above. The asymptotic distribution of a Kendall type rank statistic is given in the following:

5.2 Theorem *Assume that $J(u,v,w)$ is continuously differentiable. We denote the partial derivatives by*

$$J^1(u,v,w) = \frac{\partial J(u,v,w)}{\partial u},\ J^2(u,v,w) = \frac{\partial J(u,v,w)}{\partial v},\ J^3(u,v,w) = \frac{\partial J(u,v,w)}{\partial w},$$

and assume that they are uniformly bounded. Then $\sqrt{n}(T(C_n) - T(C))$ converges in distribution to a normal distribution with mean 0 and variance

$$\begin{aligned} \sigma(T) \triangleq \operatorname{var}\Bigg(& J(\xi,\eta,C(\xi,\eta)) + \iint \Big[J^3(u,v,C(u,v))(\mathbf{1}_{\{\xi \le u,\, \eta \le v\}} - C(u,v)) \\ & + J^1(u,v,C(u,v))(\mathbf{1}_{\{\xi \le u\}} - u) + J^2(u,v,C(u,v))(\mathbf{1}_{\{\eta \le v\}} - v) \Big]\, dC(u,v) \Bigg), \end{aligned}$$

provided $\sigma(T) > 0$, and (ξ, η) is defined as in Theorem 5.1.

We only indicate the outline of a proof; the argument is typical and the details can be easily filled in. Write

$$\begin{aligned}&\sqrt{n}(T(\mathbb{C}_n) - T(C))\\ &= \iint J^3(u,v,C(u,v))\mathbb{D}_n^C(u,v)\, d\mathbb{C}_n(u,v) + \iint J(u,v,C(u,v))\, d\mathbb{D}_n^C(u,v) + R_n,\end{aligned}$$

where $R_n = o_P(1)$. The first integral will converge to

$$\iint J^3(u,v,C(u,v))\mathbb{D}^C(u,v)\, dC(u,v),$$

and the second to $\iint J(u,v,C(u,v))\, d\mathbb{D}^C(u,v)$, which is just a symbol for the limiting random variable, not a stochastic integral in any sense. By the invariance principle and an straightforward calculation involving integration by parts, it can be seen that the covariance structure of these two limiting random variables is the same as that of

$$\iint J^3(u,v,C(u,v))Z(u,v)\, dC(u,v)$$

($Z(u,v)$ is defined in Theorem 5.1) and

$$\begin{aligned}&J(\xi,\eta,C(\xi,\eta)) - \iint J(u,v,C(u,v))\, dC(u,v)\\ &+ \iint [J^1(u,v,C(u,v)) + J^3(u,v,C(u,v))C^1(u,v)](\mathbf{1}_{\{\xi\le u\}} - u)\, dC(u,v)\\ &+ \iint [J^2(u,v,C(u,v)) + J^3(u,v,C(u,v))C^2(u,v)](\mathbf{1}_{\{\eta\le v\}} - v)\, dC(u,v).\end{aligned}$$

Summing up the two cancels some terms out, and the result is as given in the statement of the theorem.

5.4 Other Measures of Dependence

Besides the rank correlation coefficients, there are several measures of dependence which can be written in terms of copulas. Many of them can be classified into two classes: Spearman type and Kendall type. Examples of Spearman type measures are

$$S_p(C) = \iint [C(u,v) - uv]^p\, dudv, \quad (p > 1)$$

and

$$S_1(C) = \iint |C(u,v) - uv|\, dudv.$$

$S_2(C)$ is discussed in Yanagimoto (1970), and $S_1(C)$ in Schweizer and Wolff (1981). They claim that $S_1(C)$ has many desirable properties as a measure of

association. Examples of Kendall type measures are obtained by replacing "$dudv$" by "$dC(u,v)$" in the above: for $p > 1$,

$$T_p(C) = \iint [C(u,v) - uv]^p \, dC(u,v), \quad T_1(C) = \iint |C(u,v) - uv| \, dC(u,v).$$

$T_2(C)$ is the well-known Hoeffding's Δ (Hoeffding (1948)). Of course, there are measures of dependence which belongs to neither of the above two classes; $\sup_{u,v} |C(u,v) - uv|$ is such an example.

We can use Theorems 5.1 and 5.2 to find the asymptotic distributions of the sample versions (replacing C by $\mathbb{C}_n$) of these measures of dependence except for S_1 and T_1. However, note that $J^3(u,v,w) = p(w-uv)^{p-1}$ for S_p and T_p $(p > 1)$, so when $C(u,v) = uv$, we have $J^3(u,v,C(u,v)) = 0$. This violates the assumption that $\sigma(S)$ and $\sigma(T)$ are strictly positive. Thus, we cannot apply Theorems 5.1 and 5.2 when two random variables under consideration are actually independent. This corresponds to the null distribution of the test statistics for independence discussed before; we must find alternative ways to evaluate the asymptotic null distribution.

6 Concluding Remarks

The empirical copula is a nonparametric estimator of the copula in models for multivariate data. The present paper demonstrates that we can use the empirical copula for testing independence and for estimating several measures of dependence, and derives the asymptotic properties of those test statistics and estimators. Obviously, the empirical copulas can be used for many other purposes, and here we shall discuss some of them.

Recently, copulas have proved useful to model dependence in financial risk management (Embrechts et al. (1999a, 1999b), and Clemen and Reilly (1999)) and analysis of multivariate survival data (Hougaard (2000)). The reason is that, for the nonnormal multivariate distributions, the use of the classical correlation coefficient can be seriously misleading. One specifies a model by choosing a parametric form $C_\theta(u)$ of copula to model dependence we have in mind. It is then important to have methods to estimate the unknown parameter θ. We can use empirical copulas to estimate the unknown parameters of copula by the minimum distance method: For example, $\widehat{\theta}$ which minimizes

$$M(\theta) \triangleq \int_{[0,1]^d} [\mathbb{C}_n(u) - C_\theta(u)]^2 \, du$$

may be used as an estimator of θ. Alternatively, estimation based on ranks of the observations is another possibility, and investigation of those two methods will be our next research project.

Note also that the above $M(\theta_0)$ may be regarded as a goodness-of-fit test statistic for testing validity of the presupposed multivariate model with copula $C_{\theta_0}(u)$. In any case, it seems impossible to find the exact distribution of $M(\theta)$,

and again we have to resort to asymptotics. That is, it is necessary, though by no means easy, to study

$$\int_{[0,1]^d} \left[\mathbb{D}^{C_\theta}(u)\right]^2 du$$

for true and contiguous θ.

Finally, we remark that the efficiency aspect of estimation of the parameters of many copula models is examined in Bickel et al. (1993), and that Tjøstheim (1996) studies estimation methods using Hellinger distance and Kullback-Leibler information with estimated densities, together with discussion of tests of independence.

Acknowledgement

The author would like to thank Akimichi Takemura for bringing problems on empirical copulas to my attention and posing many sharp questions.

References

[1] Adler, R. J. (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*, Institute of Mathematical Statistics, Hayward, California.

[2] Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes, *Ann. Math. Statist.*, **23**, 193–212.

[3] Ash, R. B. and Gardner, M. F. (1975). *Topics in Stochastic Processes*, Academic Press, New York.

[4] Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press, Baltimore and London.

[5] Bickel, P. J. and Wichura, M. J. (1971). Convergence for multiparameter stochastic processes and some applications, *Ann. Math. Statist.*, **42**, 1656–1670.

[6] Blum, J. R. and Kiefer, J. and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function, *Ann. Math. Statist.*, **32**, 485–498.

[7] Clemen, R. T. and Reilly, T. (1999). Correlations and copulas for decision and risk analysis, *Management Science*, **45**, No. 2, 208–224.

[8] Cotterill, D. S. and Csörgő, M. (1985). On the limiting distribution of and critical values for the Hoeffding, Blum, Kiefer, Rosenblatt independence criterion, *Statist. Decisions*, **3**, 1–48.

[9] Csörgő, M. (1979). Strong approximations of the Hoeffding, Blum, Kiefer, Rosenblatt multivariate empirical process, *J. Mult. Anal.*, **9**, 84–100.

[10] Csörgő, M. (1984). Invariance principles for empirical processes, *Handbook of Statistics, Volume 4: Nonparametric Methods*, pp. 431–462, P. R. Krishnaiah and P. K. Sen (eds.), Elsevier Science B.V., Amsterdam.

[11] Deheuvels, P. (1979). La fonction de dépendence empirique et ses propriétés, Un test non paramétrique d'indépendance, *Bulletin de la classe des sciences, Académie Royale de Belgique, 5e série*, **65**, 274–292.

[12] Deheuvels, P. (1980). Non parametric tests of independence, in: *Statistique non Paramétrique Asymptotique*, pp. 95–107, J. P. Raoult (ed.), Springer-Verlag, Berlin.

[13] Deheuvels, P. (1981). An asymptotic decomposition for multivariate distribution-free tests of independence, *J. Mult. Anal.*, **11**, 102–113.

[14] De Wet, T. (1980). Cramér-von Mises tests of independence, *J. Mult. Anal.*, **10**, 38–50.

[15] Einmahl, J. H. J. (1987). *Multivariate Empirical Processes*, Centrum voor Wiskunde en Informatica, Math. Centrum, Amsterdam.

[16] Embrechts, P., McNeil, A. and Straumann, D. (1999a). Correlation: Pitfalls and alternatives, *Risk Magazine*, May, 69–71.

[17] Embrechts, P., McNeil, A. and Straumann, D. (1999b). *Correlation and dependency in risk management: properties and pitfalls*, Preprint, ETH Zurich, at http://www.math.ethz.ch/~mcneil.

[18] Gaenssler, P. and Stute, W. (1987). *Seminar on Empirical Processes*, Birkhäuser, Basel · Boston.

[19] Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests*, Academic Press, New York.

[20] Hoeffding, W. (1948). A nonparametric test of independence, *Ann. Math. Statist.*, **19**, 546–557.

[21] Hougaard, P. (2000). *Analysis of Multivariate Survival Data*, Springer-Verlag, New York.

[22] Joag-Dev, K. (1984). Measure of Dependence, in: *Handbook of Statistics, Volume 4: Nonparametric Methods*, pp. 79–88, P. R. Krishnaiah and P. K. Sen (eds.), Elsevier Science B.V., Amsterdam.

[23] Joe, H. (1997). *Multivariate Models and Dependence Concepts*, Chapman and Hall, London.

[24] Kac, M. and Siegert, A. J. F. (1947). An explicit representation of a stationary Gaussian process, *Ann. Math. Statist.*, **18**, 438–442.

[25] Kendall, M. and Gibbons, J. D. (1990). *Rank Correlation Methods*, 5th ed., Oxford University Press, New York.

[26] Koziol, J. A. and Nemec, A. F. (1979). On a Cramér-von Mises type statistic for testing bivariate independence, *Canad. J. Statist.*, **7**, 43–52.

[27] Kruskal, W. H. (1958). Ordinal measures of association, *J. Amer. Statist. Assoc.*, **53**, 814–861.

[28] Moore, D. S. and Spruill, M. C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit, *Ann. Statist.*, **3**, 599–616.

[29] Nelsen, R. B. (1999). *An Introduction to Copulas*, Lecture Notes in Statistics, Vol. 139, Springer-Verlag, New York.

[30] Neuhaus, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter, *Ann. Math. Statist.*, **42**, 1285–1295.

[31] Piterbarg, V. I. (1996). *Asymptotic Methods in the Theory of Gaussian Processes and Fields*, American Mathematical Society, Providence, Rhode Island.

[32] Rüschendorf, L. (1976). Asymptotic distributions of multivariate rank order statistics, *Ann. Statist.*, **4**, 912–923.

[33] Ruymgaart, F. H. (1973). *Asymptotic Theory of Rank Tests for Independence*, Mathematisch Centrum, Amsterdam.

[34] Schweizer, B. (1991). Thirty years of copulas, in: *Advances in Probability Distributions with Given Marginals*, pp. 13–50, G. Dall'Aglio, S. Kotz and G. Salinetti (eds.), Kluwer Academic Publishers, Dordrecht.

[35] Schweizer, B. and Wolff, E. F. (1981). On nonparametric measures of dependence for random variables, *Ann. Statist.*, **9**, 879–885.

[36] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*, Wiley, New York.

[37] Sklar, M. (1959). Fonctions de répartition á n dimensions et leurs marges, *Publ. Inst. Statist. Univ. Paris*, **8**, 229–231.

[38] Stute, W. (1982). The oscillation behavior of empirical processes, *Ann. Probab.*, **10**, 86–107.

[39] Stute, W. (1984). The oscillation behavior of empirical processes: the multivariate case, *Ann. Probab.*, **12**, 361–379.

[40] Tjøstheim, D. (1996). Measures of dependence and tests of independence, *Statistics*, **28**, 249–284.

[41] Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer-Verlag, New York.

[42] Yanagimoto, T. (1970). On measures of association and a related problem, *Ann. Inst. Stat. Math.*, **22**, 57–63.

Empirical Copulas and Some Applications （研究報告 No. 27）

平成12年12月20日 印 刷
平成12年12月25日 発 行 非売品

著 者 塚 原 英 敦
発行所 成城大学経済研究所
〒157-8511 東京都世田谷区成城 6-1-20
電 話 03 (3482) 1181番
印刷所 白陽舎印刷工業株式会社