Journal of Data Science **10**(2012), 1-17

A New Family of Bivariate Copulas Generated by Univariate Distributions[1]

Xiaohu Li* and Rui Fang
Xiamen University

Abstract: A new family of copulas generated by a univariate distribution function is introduced, relations between this copula and other well-known ones are discussed. The new copula is applied to model the dependence of two real data sets as illustrations.

Key words: Archimedean copula, upper tail dependence coefficient, Weibull distribution.

1. Introduction

In financial engineer, insurance risk management, economics, climatology and many other fields of social science, researchers are usually confronted with multivariate data without independence. Hence, to model the dependence is becoming more and more essential in understanding and interpreting the data. Pearson's correlation coefficient, Kendall's tau and Spearman's rho are indices often used to measure the dependence behind the data, however, they fail to characterize the dependence structure in a complete manner. In addition, since the normality assumption of data usually does not provide an adequate approximation to data sets with heavy tail, non-normal multivariate distributions are used in practice (see Johnson and Kotz, 1992; Kotz, Balakrishnan and Johnson, 2000). However, there are possible drawbacks of these distributions. For example, each family may have its own marginal distributions, which vary from family to family, and measures of dependence often rely on the marginal distributions themselves. These shortcomings make the statistical modeling rather complicated and hence unwieldy.

Sklar (1959) first introduced the copula to model dependence structure. A non-decreasing and right-continuous bivariate function $C : [0, 1] \times [0, 1] \longrightarrow [0, 1]$ is called a copula if it satisfies

$$C(u_1, 0) = C(0, u_2) = 0, \qquad C(u_1, 1) = u_1, \qquad C(1, u_2) = u_2,$$

[1]Supported by National Science Foundation of China (11171278).
*Corresponding author.

$$C(v_1, v_2) + C(u_1, u_2) - C(v_1, u_2) - C(u_1, v_2) \geq 0, \quad u_1 \geq v_1, u_2 \geq v_2.$$

The concept of copula is relatively simple, the construction does not constrain the choice of marginal distributions and it provides a good way to impose a dependence structure on predetermined marginal distributions. Thus, it is particularly useful for modeling dependence in practical applications. In risk management, finance and econometrics, copula has become more or less a standard tool. A large number of applications of copulas can be found in various areas. For instance, Frees and Wang (2006), Chen and Fan (2006).

In this note, we introduce a new family of copulas generated by a univariate distribution function, relations between the new copula and other well-known ones are discussed, and some dependence indices of this copula are studied. As an illustration, the new copula is employed to fit two real data sets as well.

2. A New Family of Copulas

Let $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\}$ be the standard normal density, one may easily verify that

$$f(x, y) = \phi(x)\phi(y)(1 + \sin x \sin y)$$

forms a bivariate probability density. It is not difficult to obtain the copula of this distribution as follows:

$$C_\Phi(u, v) = uv + \int_0^u \sin(\Phi^{-1}(t))\,\mathrm{d}t \int_0^v \sin(\Phi^{-1}(t))\,\mathrm{d}t, \tag{2.1}$$

where Φ is the standard normal distribution function and the right continuous inverse $\Phi^{-1}(t) = \sup\{x : \Phi(x) \leq t\}$ for $t \in [0, 1]$. Naturally, one may wonder whether the resulted bivariate function

$$C_\Psi(u, v) = uv + \int_0^u \sin(\Psi^{-1}(t))\,\mathrm{d}t \int_0^v \sin(\Psi^{-1}(t))\,\mathrm{d}t, \text{ for } (u, v) \in [0, 1]^2, \tag{2.2}$$

still serves as a copula when some distribution function Ψ substitutes for Φ in (2.1)? The first theorem addresses a necessary and sufficient condition for C_Ψ to be a copula.

Theorem 1 C_Ψ in (2.2) is a copula if and only if

$$\int_0^1 \sin(\Psi^{-1}(t))\,\mathrm{d}t = 0. \tag{2.3}$$

Proof: Necessity: Since $C_\Psi(u, v)$ is a copula, then $C_\Psi(1, 1) = 1$. Namely,

$$1 + \left(\int_0^1 \sin(\Psi^{-1}(t))\,\mathrm{d}t\right)^2 = 1.$$

Hence, it holds that $\int_0^1 \sin(\Psi^{-1}(t))\,\mathrm{d}t = 0$.

Sufficiency: As a distribution function, Ψ and hence $C_\Psi(u,v)$ is right-continuous. Obviously, it holds that

$$C_\Psi(1,1) = 1, \quad C_\Psi(u,0) = C_\Psi(0,v) = 0, \text{ for } (u,v) \in [0,1]^2,$$

$$\lim_{v\to 1} C_\Psi(u,v) = u, \qquad \lim_{u\to 1} C_\Psi(u,v) = v, \text{ for } (u,v) \in [0,1]^2.$$

Note that, for $u \in [0,1]$,

$$\begin{aligned} \frac{\partial C_\Psi(u,v)}{\partial v} &= u + \sin(\Psi^{-1}(v)) \int_0^u \sin(\Psi^{-1}(t))\,\mathrm{d}t \\ &\geq u - \left| \int_0^u \sin(\Psi^{-1}(t))\,\mathrm{d}t \right| \\ &\geq 0. \end{aligned}$$

By symmetry, C_Ψ is non-decreasing in $(u,v) \in [0,1]^2$. And, for $0 \leq u_i \leq v_i \leq 1$, $i = 1,2$,

$$\begin{aligned} &C_\Psi(u_1,u_2) + C_\Psi(v_1,v_2) - C_\Psi(u_1,v_2) - C_\Psi(v_1,u_2) \\ &= (v_2 - u_2)(v_1 - u_1) + \int_{u_1}^{v_1} \sin(\Psi^{-1}(t))\,\mathrm{d}t \int_{u_2}^{v_2} \sin(\Psi^{-1}(t))\,\mathrm{d}t \\ &\geq (v_2 - u_2)(v_1 - u_1) - \left| \int_{u_1}^{v_1} \sin(\Psi^{-1}(t))\,\mathrm{d}t \right| \left| \int_{u_2}^{v_2} \sin(\Psi^{-1}(t))\,\mathrm{d}t \right| \\ &\geq 0. \end{aligned}$$

That is, C_Ψ possesses the supermodularity. So, C_Ψ is a copula. □

For ease of reference, we call the copula well defined in (2.2) as a sine copula with generator Ψ from now on. Denote X one random variable with distribution function Ψ. Then, the equality (2.3) is equivalent to $\mathrm{E}[\sin(X)] = 0$, which is usually easy to be verified in some occasions. In practice, for the sake of convenience and tractability in mathematics, one may choose those symmetric distribution functions to serve as generators.

Suppose a random variables X's with lattice distribution on $\{k\pi\colon\ k = 0,\ \pm 1,\ \pm 2, \cdots\}$. That is, its distribution function Ψ satisfies

$$\sum_{k=-\infty}^{\infty} P(X = k\pi) = 1.$$

Then, $\sin(\Psi^{-1}(t)) = 0$ and hence

$$C_\Psi(u,v) \equiv uv = C_I(u,v), \quad \text{for any } (u,v) \in [0,1]^2.$$

Namely, C_I is a sine copula.

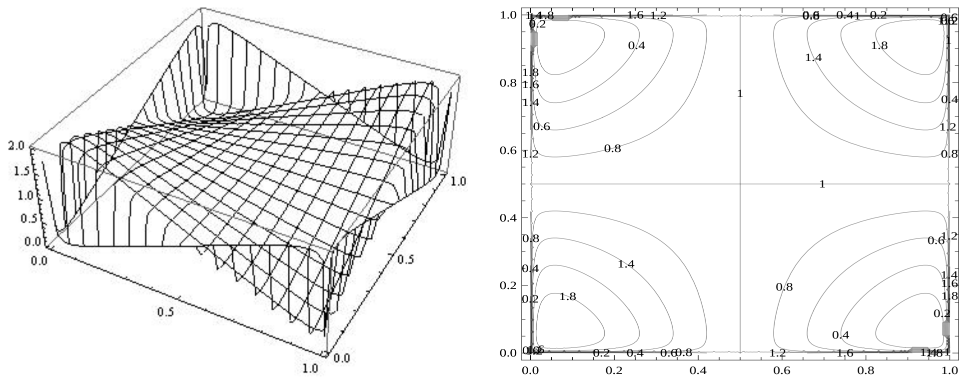
Before proceeding to other main conclusions, let us have a look at the following examples of the sine copula in (2.2), and it should be remarked here that the density of a sine copula usually is not unimodal due to periodicity of the sine operator.

Example 1 (Uniform generators) Let Ψ be a distribution function of a random variable X uniformly distributed on $[c, c+2\pi]$ with any constant c. Then,

$$\mathrm{E}[\sin(X)] = \int_c^{c+2\pi} \sin x \,\mathrm{d}\Psi(x) = 0.$$

Hence, the corresponding C_Ψ in (2.2) serves as a sine copula. □

Example 2 (Symmetric generators) Let Ψ be a distribution function symmetric with respect to its expectation $n\pi$, $n \in \mathbb{Z}$. It may be verified that (2.3) holds for any $n \in \mathbb{Z}$. Hence, by Theorem 1, C_Ψ is a copula. Figure 1 plots the density of C_Ψ with Ψ being the standard normal distribution.



(a) Copula density (b) Level curves

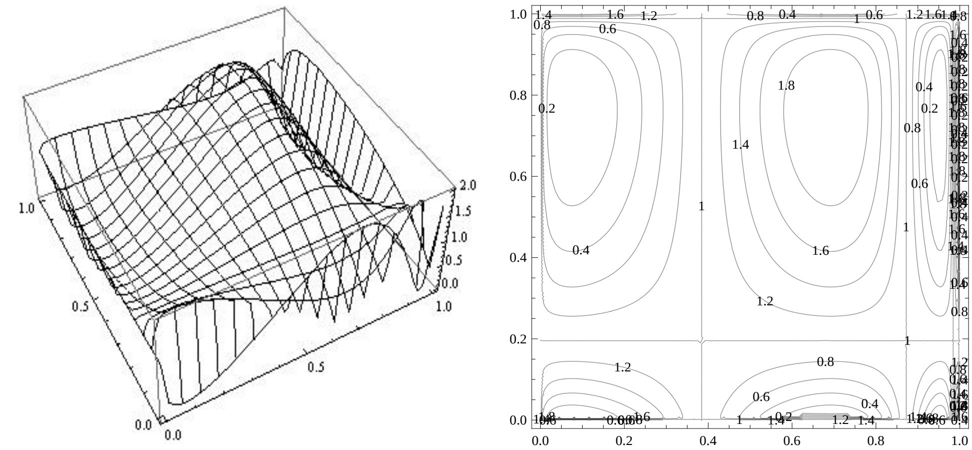
Figure 1: Copula generated by $N(0,1)$ distribution function

Note that both the normal distribution with expectation $n\pi$ and the student distribution satisfy (2.3), they serve as generators of the copulas in (2.2). □

Example 3 (Gamma generators) Let Ψ be the Gamma distribution function with parameters $(4n, 1)$, $n \in \mathbb{Z}^+$, then

$$\int_0^1 \sin(\Psi^{-1}(t))\,\mathrm{d}t = \int_0^\infty \frac{x^{4n-1}e^{-x}\sin x}{(4n-1)!}\,\mathrm{d}x = 0. \tag{2.4}$$

Figure 2 plots the density of C_Ψ with Ψ being the $\Gamma(4,1)$ distribution function.



(a) Copula density (b) Level curves

Figure 2: Copula generated by $\Gamma(4,1)$ distribution function

In fact, it can be calculated by standard calculus that (2.4) holds for $n = 1$. By iteratively integrating by parts on (2.4), we have

$$\begin{aligned}\int_0^\infty \frac{x^{4n+3}e^{-x}\sin x}{(4n+3)!}\,\mathrm{d}x &= \frac{1}{2}\int_0^\infty \frac{x^{4n+1}e^{-x}\cos x}{(4n+1)!}\,\mathrm{d}x \\ &= -\frac{1}{4}\int_0^\infty \frac{x^{4n-1}e^{-x}\sin x}{(4n-1)!}\,\mathrm{d}x \\ &= 0.\end{aligned}$$

By induction, this invokes (2.4). □

Recall that

$$C_\varphi(u,v) = \begin{cases} \varphi^{-1}(\varphi(u)+\varphi(v)), & \varphi(u)+\varphi(v) \le \varphi(0), \\ 0, & \text{otherwise}, \end{cases}$$

is called an Archimedean copula generated by φ, which is a strictly increasing and convex function $\varphi(t)$ such that $\varphi(1) = 0$ and $\varphi'(t) \le 0$ for all $t \in (0,1)$. Archimedean copulas are very popular in actuarial science, finance and insurance risk management. It is well-known that Fréchet upper bound $C_U(u,v) = \min\{u,v\}$ does not belong to Archimedean family while Fréchet lower bound $C_L(u,v) = \max\{0, u+v-1\}$ does. Although sine copulas include the independent copula as one special case, Example 4 tells that neither Fréchet upper bound copula nor Fréchet lower bound copula belong to this new family (2.2), and thus, the family of sine copulas is inconsistent to Archimedean family.

Example 4 For a copula $C(u,v)$, define $\delta_C(u) = C(u,u)$. Then, $\delta_{C_U}(u) = u$ and $\delta_{C_L}(u) = \max\{0, 2u-1\}$. Let

$$f(u) = \delta_{C_\Psi}(u) - \delta_{C_U}(u) = u^2 - u + \left[\int_0^u \sin\left(\Psi^{-1}(t)\right) \mathrm{d}t\right]^2,$$

then $f(0) = f(1) = 0$, and

$$f'(u) = 2u - 1 + 2\sin\left(\Psi^{-1}(u)\right) \int_0^u \sin\left(\Psi^{-1}(t)\right) \mathrm{d}t.$$

It is easy to verify that $f'(1) > 0$ and $f'(u) < 0$ for some $u \in (0, \frac{1}{4})$. Then, $f'(\alpha) = 0$ for some $\alpha \in (\frac{1}{4}, 1)$. Note that $f''(u) \geq 0$, $f'(u) \leq 0$ for $u \in [0, \alpha]$ and $f'(u) > 0$ for $u \in (\alpha, 1]$, we have

$$\delta_{C_\Psi}(u) < u = \delta_{C_U}(u), \text{ for } u \in (0,1).$$

That is, C_U is not in the family (2.2).

Obviously, $\delta_{C_L}(u) = 0$ for $u \in [0, 1/2]$. Note that, for any sine copula C_Ψ,

$$\delta_{C_\Psi}(u) = u^2 + \left(\int_0^u \sin(\Psi^{-1}(t))\mathrm{d}t\right)^2 \neq 0, \quad \text{for } u \in [0, 1/2],$$

C_L is not a sine copula, either. □

In fact, with the help of the following proposition, we may further clarify the substantial difference between the sine family and Archimedean family.

Proposition 1 (Nelsen, 2006) For an Archimedean copula with generator φ, it holds that

$$C_\varphi(C_\varphi(u,v), w) = C_\varphi(u, C_\varphi(v,w)), \qquad u,\ v,\ w \in [0,1].$$

According to the following example, the sine copula does not possess the above property.

Example 5 Consider a discrete random variable X,

$$P(X = \pi/2) = 1/4, \qquad P(X = \pi) = 1/2, \qquad P(X = 3\pi/2) = 1/4.$$

The distribution function and its right continuous inverse are

$$\Psi(x) = \begin{cases} 0, & 0 \leq x < \pi/2, \\ 1/4, & \pi/2 \leq x < \pi, \\ 3/4, & \pi \leq x < 3\pi/2, \\ 1, & x \geq 3\pi/2, \end{cases} \qquad \Psi^{-1}(t) = \begin{cases} \pi/2, & 0 \leq t \leq 1/4, \\ \pi, & 1/4 < t \leq 3/4, \\ 3\pi/2, & 3/4 < t \leq 1. \end{cases}$$

It may be evaluated that $\int_0^1 \sin(\Psi^{-1}(t))\,\mathrm{d}t = 0$. By Theorem 1, C_Ψ is a copula.

Denote, for $u \in [0,1]$ and for $u, v, w \in [0,1]$,

$$H(u) = \int_0^u \sin(\Psi^{-1}(t))\,\mathrm{d}t,$$

$$\begin{aligned} G(u,v,w) = {} & wH(u)H(v) + H(w)H(uv + H(u)H(v)) \\ & -uH(v)H(w) - H(u)H(vw + H(v)H(w)). \end{aligned}$$

Note that $G(u,v,w) \neq 0$ for some $u, v, w \in [0,1]$ if and only if

$$C_\Psi(C_\Psi(u,v),w) \neq C_\Psi(u, C_\Psi(v,w)),$$

for some u, v and w in $[0,1]$, from $G(1/2, 3/4, 3/4) = 1/64 \neq 0$ and Proposition 2, it may be concluded that C_Ψ is not an Archimedean copula. □

3. Some Dependence Indices

This section presents some dependence indices of the sine copula, for example, the upper tail dependence coefficient λ_U, the lower tail dependence coefficient λ_L, Kendall's τ and Spearman's ρ. Readers may refer to Kaas *et al.* (2001), Joe (1993) and Denuit *et al.* (2005) for more details on these measures.

Proposition 2 For any random vector (X_1, X_2) with copula (2.2), both the upper tail dependence coefficient $\lambda_U = 0$ and the lower tail dependence coefficient $\lambda_L = 0$ whatever the generator Ψ is.

Proof: By using L'Hospital's rule,

$$\begin{aligned} \lambda_U &= \lim_{v \to 0} \frac{1 - 2(1-v) + C(1-v, 1-v)}{v} = \lim_{s \to 1} \frac{1 - 2s + C_\Psi(s,s)}{1-s} \\ &= \lim_{s \to 1} \frac{-2 + 2s + 2\sin(\Psi^{-1}(s)) \int_0^s \sin(\Psi^{-1}(t))dt}{-1} = 0, \end{aligned}$$

and

$$\begin{aligned} \lambda_L &= \lim_{u \to 0} \frac{C(u,u)}{u} = \lim_{u \to 0} \frac{u^2 + \left(\int_0^u \sin(\Psi^{-1}(t))\mathrm{d}t\right)^2}{u} \\ &= \lim_{u \to 0} 2u + 2\sin(\Psi^{-1}(u)) \int_0^u \sin(\Psi^{-1}(t))\mathrm{d}t = 0. \end{aligned}$$

□

Proposition 2 tells us that any pair of random variables (X_1, X_2) with sine copula (2.2) is asymptotically independent.

In the past decades, some researchers had spent a lot of time in studying the dependence between the sample maximum and minimum due to various practical requirement. Recently, Schmitz (2004) investigated this dependence structure through determining their copula and conjectured that the ratio of Spearman's ρ to Kendall's τ tends to 3/2 as the sample size tends to infinity. Afterward, Li and Li (2007) provided a theoretical proof for this conjecture. Proposition 3 below reveals that Kendall's τ and Spearman's ρ of the sine copula achieve this limit whatever the generator is. In fact, in Archimedean family, both spearman's ρ and Kendall's τ are usually closely related to the generator. This fascinating fact makes sine copula rather flexible in modeling the dependence between sample minimum and maximum.

Proposition 3 For any random vector (X_1, X_2) with copula (2.2), $\rho(X_1, X_2) = \frac{3}{2}\,\tau(X_1, X_2)$.

Proof: By integrating by parts, we have

$$\begin{aligned}
\tau(X_1, X_2) &= 4\int_0^1\int_0^1 C_\Psi(u,v)\,\mathrm{d}C_\Psi(u,v) - 1\\
&= 4\int_0^1\int_0^1 \left(uv + \int_0^u \sin\left(\Psi^{-1}(t)\right)\,\mathrm{d}t \int_0^v \sin\left(\Psi^{-1}(t)\right)\,\mathrm{d}t\right)\\
&\quad\cdot\left[1+\sin(\Psi^{-1}(u))\sin(\Psi^{-1}(v))\right]\,\mathrm{d}u\,\mathrm{d}v - 1\\
&= 8\left(\int_0^1 u\sin(\Psi^{-1}(u))\,\mathrm{d}u\right)^2,
\end{aligned}$$

and

$$\begin{aligned}
\rho(X_1, X_2) &= 12\int_0^1\int_0^1 uv\,\mathrm{d}C_\Psi(u,v) - 3\\
&= 12\int_0^1\int_0^1 uv\left[1+\sin(\Psi^{-1}(u))\sin(\Psi^{-1}(v))\right]\,\mathrm{d}u\,\mathrm{d}v - 3\\
&= 12\left(\int_0^1 u\sin(\Psi^{-1}(u))\,\mathrm{d}u\right)^2.
\end{aligned}$$

Thus, $\rho(X_1, X_2) = \frac{3}{2}\tau(X_1, X_2)$. □

Since

$$\tau(X_1, X_2) = 8\left[\int_0^1 u\sin(\Psi^{-1}(u))\,\mathrm{d}u\right]^2 \le \frac{1}{2},$$

Kendall's τ of a sine copula is always bounded below 2^{-1}. This also helps to determine whether the sine copula is suitable for a certain data set in practice.

Actually, when

$$\Psi^{-1}(u) = \begin{cases} -\pi/2, & u \in [0, 1/2], \\ \pi/2, & u \in (1/2, 1], \end{cases} \tag{3.1}$$

τ achieves the maximum 2^{-1}.

Recall that the normal or Gaussian copula

$$C_\alpha = H_\alpha\big(\Phi^{-1}(u), \Phi^{-1}(v)\big),$$

where H_α is the standard bivariate normal distribution function with correlation α and Φ is the distribution function of $N(0, 1)$ (Li, 2000). By Proposition 3, we immediately have the following corollary.

Corollary 4 The normal copula is not in the family of sine copulas.

Proof: By Exercise 5.4.6 of Denuit *et al.* (2005), the normal copula C_α has the Kendall's τ and Spearman's ρ as

$$\tau = \frac{2}{\pi} \arcsin \alpha, \qquad \rho = \frac{6}{\pi} \arcsin \frac{\alpha}{2}.$$

It is evident that $\rho/\tau = \frac{3 \arcsin(\alpha/2)}{\arcsin \alpha} \neq 3/2$. So, the normal copula is not a sine copula. □

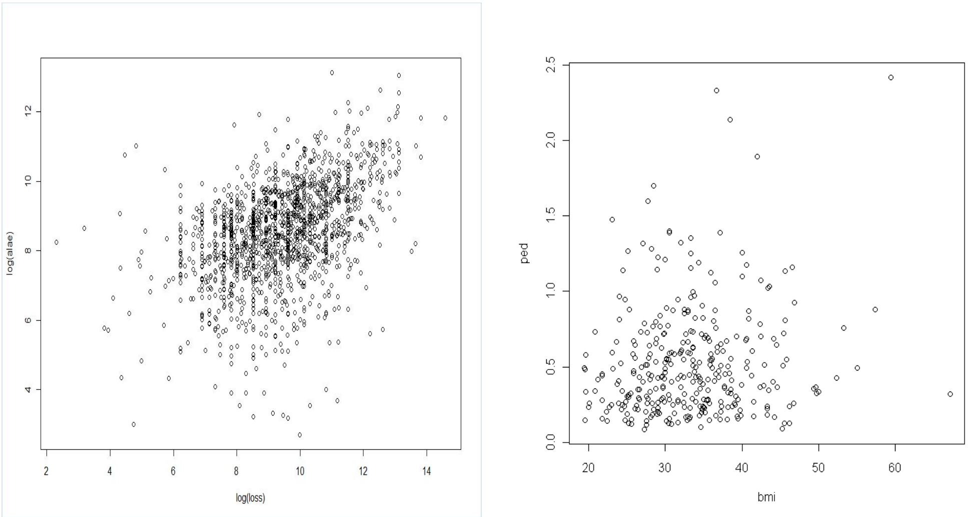
4. Two Applications

To demonstrate the new copula discussed in previous sections, in this section, we fit the sine copula to the following two real data sets:

- The first data set is (also used by Frees and Valdez, 1998) from the US Insurance Services Office and comprising general liability claims randomly chosen from late settlement lags, it contains 1500 observations, each consists of loss and allocated loss adjustment expenses (ALAE's) on a single claim.

- According to the manual of R's package MASS, the US National Institute of Diabetes and Digestive and Kidney Diseases collected a data set from a population of women (at least 21 years old, of Pima Indian heritage and living near Phoenix, Arizona) who were tested for diabetes according to World Health Organization criteria. This data set consists of 332 complete records after dropping the (mainly missing) data on serum insulin.

Scatter plots of the two data sets in Figure 3 demonstrate obvious dependence between involved random variables, their summary statistics are given in Table 1. Also from Figure 4 it may be reasonable to use the same family of some distribution to estimate both margins in the same data set.



(a) log(loss) versus log(ALAE) (b) bmi versus ped

Figure 3: Scattered plots of two real data sets

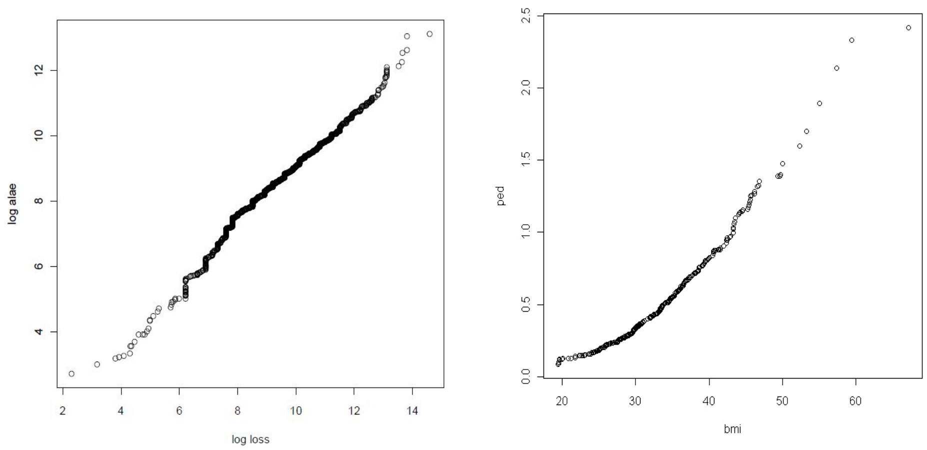
Table 1: Summary statistics of two data sets

	loss	ALAE	p-value	bmi	ped	p-value[1]
Total	1500	1500		332	332	
Minimum	10	15		19.40	0.09	
1st Quartile	4000	2333		28.17	0.27	
Median	12000	5471		32.90	0.44	
Mean	41208.4	12588.1		33.24	0.53	
3rd Quartile	35000	12572		37.20	0.68	
Maximum	2173595	501863		67.10	2.42	
Standard deviation	102747.7	28145.6		7.28	0.36	
Kendall's τ	0.31542		$< 2.2 \times 10^{-16}$	0.0642		0.0816
Spearman's ρ	0.45187		$< 2.2 \times 10^{-16}$	0.0970		0.0777
UTDC[2]	0.3716		5×10^{-5}	0.0873		0.3601
LTDC[3]	0.4481		5×10^{-5}	0.5469		0.6640

[1] The null hypotheses of the standard tests of Kendall's τ and Spearman's ρ are $\tau = 0$ and $\rho = 0$, respectively. And null hypotheses of the tests of upper/lower tail dependence coefficients are $\lambda_U = 0$ and $\lambda_L = 0$, respectively. Tests of tail dependence in Kojadinovic and Yan (2010) are used.

[2] upper tail dependence coefficient.

[3] lower tail dependence coefficient.



(a) log(loss) versus log(ALAE)

(b) bmi versus ped

Figure 4: QQ plots of the two data sets

For a parametric copula, Oakes (1989) described an omnibus procedure to estimate the dependence parameter in a copula model as follows: Suppose the copula C_α with density c_α is to be estimated. Let the rescaled empirical distribution function corresponding to the j-th component of the vector of observations

$$\hat{F}_{j,m}(x) = \frac{1}{m+1} \sum_{i=1}^{m} \mathbb{I}\{x_{ij \le x}\}, \qquad j = 1, 2,$$

and the pseudo-log-likelihood

$$\sum_{i=1}^{m} \ln c_\alpha \left(\hat{F}_{1,m}(x_{i1}), \hat{F}_{2,m}(x_{i2}) \right). \tag{4.1}$$

Then, the value $\hat{\alpha}$ maximizing (4.1) is the desired estimator of α.

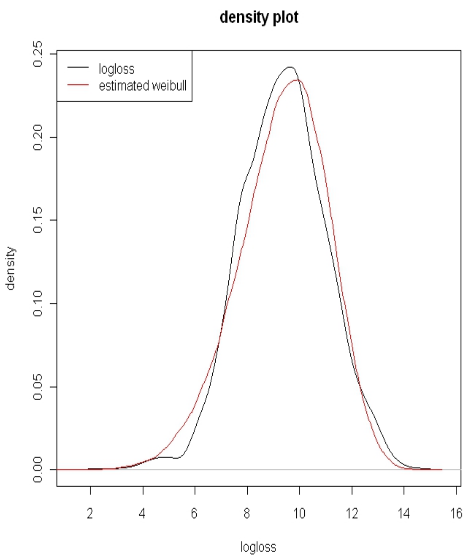
4.1 Loss and Allocated Loss Adjustment Expenses

Note that the ratio of Spearman's ρ and Kendall's τ is 1.433, which is not far from 1.5, by Proposition 3, we may have a try to fit some sine copula to this data set. By Example 2, $\Psi \sim N(0, \alpha^2)$ may serve as a generator of a sine copula. All calculations were performed in R (http://www.r-project.org). For this data set, we use the omnibus procedure to get the estimator $\hat{\alpha} = 0.84$. We use MLE method and Weibull distribution to estimate the marginal distributions of the data on log scale. The margins are estimated as Weibull distributions with scale parameters 10.05, 9.11 and shape parameters 6.34, 6.80, respectively.

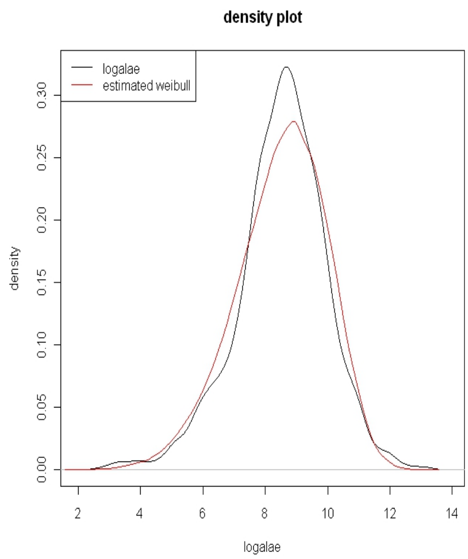
From Table 2, the estimated upper and lower tail dependence coefficients of the data are significantly different from 0 and hence the data might be fitted better by some other copula models having tail dependence, however, we can still compare the three chosen copula models. As can be seen in Figure 5, Weibull distribution fits the margins of the data well. Now, we use these estimated margins and different estimated copulas such as Frank copula, Clayton copula and the sine copula with normal generator $\Psi \sim N(0, 0.84^2)$ to fit the data. It is hard to tell which model performs better from Figure 6, however, according to Table 2 of AICs (Akaike, 1980), we may conclude that sine copula performs better than Clayton copula though it is not as good as Frank copula.

Table 2: AIC's of all models for loss and alae

copula	Frank	sine	Clayton	data
AIC	5356.29	**5401.42**	5448.893	
ρ/τ	1.456	**1.5**	1.476	1.433
Number of parameters	1	**1**	1	
UTDC	0	**0**	0	0.3716
LTDC	0	**0**	0.2617	0.4481

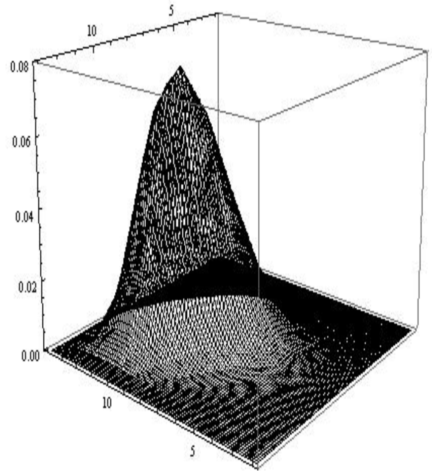


(a) log(loss) with parameters $(6.34, 10.05)$

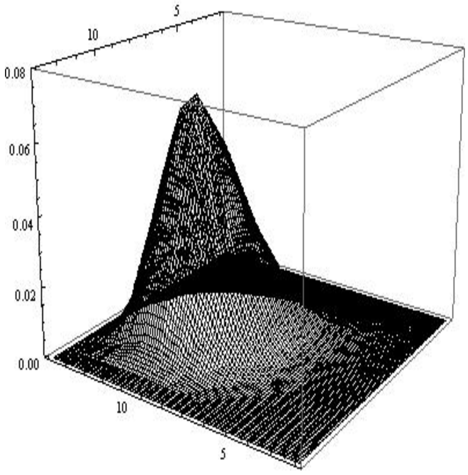


(b) log(ALAE) with parameters $(6.80, 9.11)$

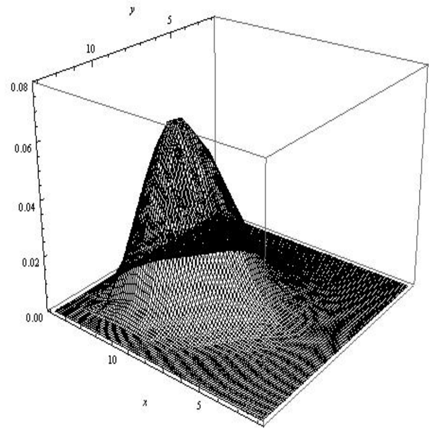
Figure 5: Estimated Weibull marginal densities



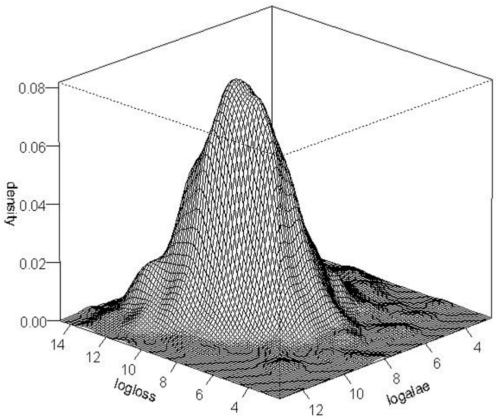
(a) Frank copula with $\alpha = 3.158$



(b) Clayton copula with $\alpha = 0.517$



(c) Sine copula with $\alpha = 0.84$

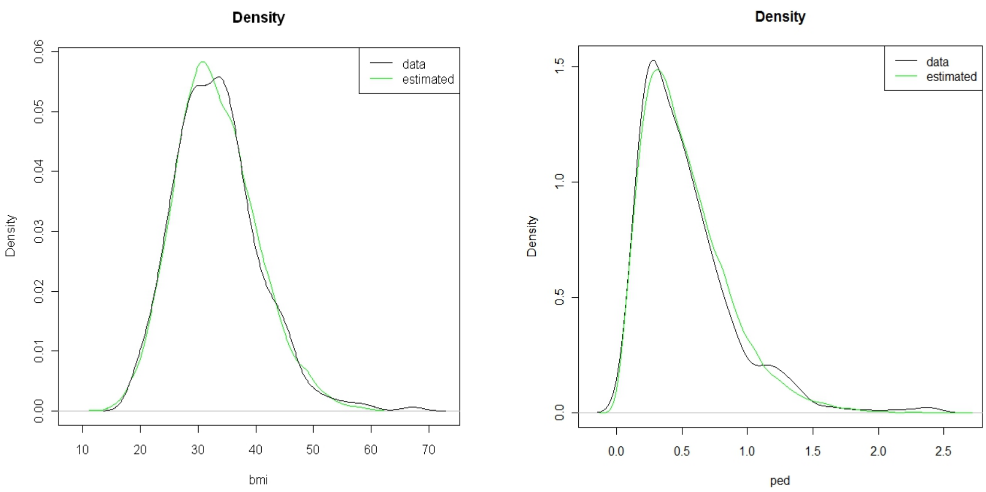


(d) Kernel density estimation

Figure 6: Densities of loss and ALAE with estimated Weibull margins

4.2 Body Mass Index and Diabetes Pedigree Function

Let us model the dependence between body mass index (bmi) and diabetes pedigree function (ped). For this data set, we employ the omnibus procedure to get the estimator $\hat{\alpha} = 0.34$, and Gamma distribution is employed to fit the margins of the data, the shape parameters are estimated as 21.82, 2.58 and rate parameters 0.66, 4.89, respectively. As can be seen in Figure 7, Gamma distribution fits the margins of the data well. Note that the estimated upper/lower tail dependence coefficient of the data are significantly close to 0, it is reasonable to employ sine copula, Frank copula and Clayton copula (when our focuses are on the upper tail behavior of the data) to model the dependence of the data. Likewise, we use the estimated margins and these copulas with estimated parameters (using omnibus procedure) to fit the data.



(a) bmi with parameters $(21.82, 0.66)$ (b) ped with parameters $(2.58, 4.89)$

Figure 7: Estimated Gamma marginal densities

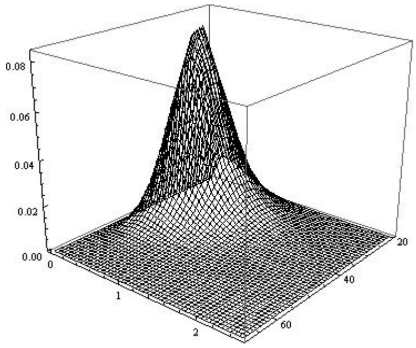
The estimation of the joint density functions are plotted in Figure 8. According to Table 3, sine copula performs better than the other copulas.

Table 3: AIC's of all models for bmi and ped

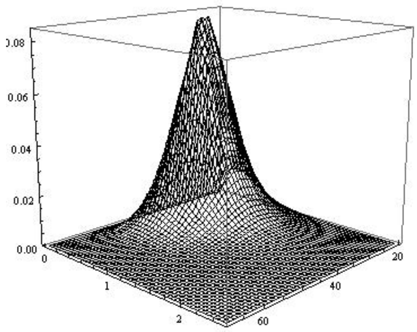
copula	Frank	Clayton	sine	data
AIC	2342.811	2342.683	**2342.125**	
ρ/τ	1.498	1.497	**1.5**	1.5096
Number of parameters	1	1	**1**	
UTDC	0	0	**0**	0.0836
LTDC	0	0.0018	**0**	0.5469

Acknowledgements

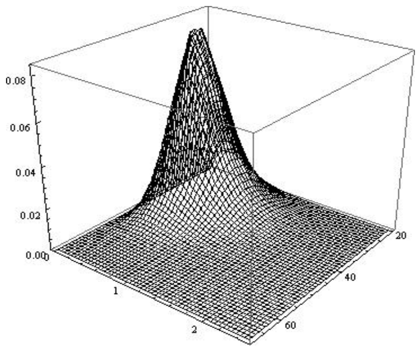
Authors would like to thank the referee for his/her insightful comments, which help to improve the earlier version of this manuscript.



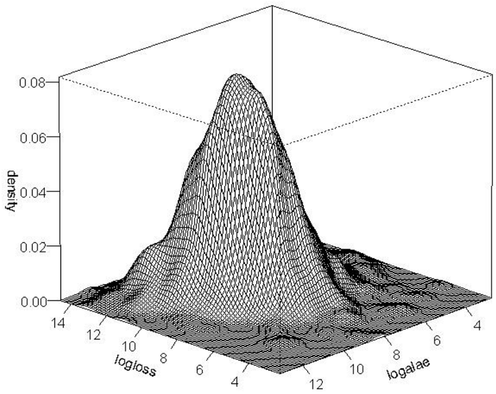
(a) Clayton copula with $\alpha = 0.11$



(b) Frank copula with $\alpha = 0.59$



(c) Sine copula with $\alpha = 0.34$



(d) density estimation

Figure 8: Densities of bmi and ped with estimated Gamma margins

References

Akaike, H. (1980). Likelihood and the Bayes procedure. In J. M. Bernardo, M. H. De Groot, D. V. Lindley and A. F. M. Smith (Eds.), *Bayesian Statistics*, pp. 143-166. University Press, Valencia, Spain.

Chen, X. and Fan, Y. (2006). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Economics* **135**, 125-154.

Denuit, M., Dhaene, J., Goovaerts, M. and Kaas, R. (2005). *Actuarial Theory for Dependent Risks*: *Measures, Orders and Models*. John Wiley & Sons, New York.

Joe, H. (1993). Parametric families of multivariate distributions with given marginals. *Journal of Multivariate Analysis* **46**, 262-282.

Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal* **2**, 1-25.

Frees, E. W. and Wang, P. (2006). Copula credibility for aggregate loss models. *Insurance*: *Mathematics and Economics* **36**, 360-373.

Johnson, N. L. and Kotz, S. (1992). *Distributions in Statistics*: *Discrete Multivariate Distributions*. John Wiley & Sons, New York.

Kojadinovic, I. and Yan, J. (2010). Nonparametric rank-based tests of bivariate extreme-value dependence. *Journal of Multivariate Analysis* **101**, 2234-2249.

Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000). *Continuous Multivariate Distributions*. John Wiley & Sons, New York.

Li, David X. (2000). On default correlation: a copula function approach. *Journal of Fixed Income* **9**, 43-54.

Li, X. and Li, Z. (2007). Proof of a conjecture on Spearma's ρ and Kendall's τ for sample minimum and maximum. *Journal of Statistical Planning and Inference* **137**, 359-361.

Nelsen, B. R. (2006). *An Introduction to Copula*. Lecture Notes in Statistics. Second Edition. Springer, New York.

Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association* **84**, 487-493.

Schmitz, V. (2004). Revealing the dependence structure between $X_{(1)}$ and $X_{(n)}$. *Journal of Statistical Planning and Inference* **123**, 41-47.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publicationis de l'Institut de Statistique de Paris* **8**, 229-231.

Received February 23, 2011; accepted August 30, 2011.

Xiaohu Li
School of Mathematical Sciences
Xiamen University
Xiamen 361005, China
mathxhli@hotmail.com

Rui Fang
School of Mathematical Sciences
Xiamen University
Xiamen 361005, China
xmushuxue@sina.com