Bayesian Nonparametric Estimation of a Copula

Juan Wu
School of Mathematics and Statistics,
Huazhong University of Science and Technology, Wuhan, China
Xue Wang, and Stephen G. Walker
School of Mathematics, Statistics and Actuarial Science,
University of Kent at Canterbury, Canterbury, UK

March 1, 2012

Abstract
A copula can characterize the complete dependence of multi-variables separately from the univariate marginals. The purpose of this paper is to provide a Bayesian nonparametric methodology to estimate a copula. We show that any bivariate copula density can be approximated by an infinite mixture of the Gaussian copula densities that are the dependence structures of the pairs with standard normal marginals. A slice sampling idea is introduced for this infinite structure that can estimate the number of occupied clusters in a sampler. The estimation procedure is proposed by the Gibbs sampling algorithm. Simulation and real data application illustrate the rational of the proposed approach.
Key words: Copula; Gaussian copula; copula density; Bayesian nonparametric estimation; Gibbs sampling; Slice sampling.

1 Introduction

Copulas have recently become popular as a modeling tool for accounting for the dependence structure of multivariate data. The aim in this paper is to contribute by presenting a new method to estimate a copula density function using Bayesian nonparametric techniques. A Copula is a multivariate distribution function with uniform marginals on $I = [0, 1]$. The copula density function, c, exists when the copula distribution function, C, is absolutely continuous and then, in the bivariate case,

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

The foundations can be found in the works of Nelsen (2002, 2006), based on the original Sklar's Theorem (Sklar, 1959), and which may explain the popularity

of copulas in statistical modeling. Suppose $H(x, y)$ is a joint cumulative distribution function of X and Y, with marginal distribution functions $F(x)$ and $G(y)$, respectively, and let

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}, \quad G^{-1}(v) = \inf\{y : g(y) \geq v\}.$$

There then exists a copula function $C : I^2 \to I$, such that

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)). \tag{2}$$

The copula C is unique if the marginal distributions F and G are continuous. Consequently, the copula approach characterizes the complete dependence structure of (X, Y) separately from the marginal distributions.

In practice, the joint distribution function H is unknown and it can be estimated by eq. (2). That is, for any x, y, $H(x, y) = C(F(x), G(y))$. It is essential to use an appropriate copula function. Choroś et al. (2010) reviewed parametric, semi–parametric and nonparametric approaches to copula estimators. We briefly describe copula inference methods for independent and identically distributed (i.i.d) random samples with dependent components.

Emphasis is typically placed on a proper copula chosen from some predetermined parametric family. Estimation approaches are obtained either by maximum likelihood, see Genest et al. (1993), pseudo-likelihood estimation, Kim et al. (2007), the method of moments or Bayesian techniques, Silva et al. (2008) and Huard et al. (2006). But there is a strong need to check whether the dependence structure is suitable for the data, i.e. model adequacy, and such issues have been considered for the Gaussian and Archimedean copulas, see Genest et al. (2009a).

More flexible approaches are obtained using nonparametric methods, which work by avoiding any assumption of a particular copula family. Classical estimators are based on the empirical copula function, developed by Deheuvels (1979). Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from the unknown distribution H and denote by F_n and G_n the empirical distribution functions associated with the F and G samples, respectively. Then $H_n(x, y)$ is taken to be the empirical distribution function, given by

$$H_n(x, y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(X_i \leq x, Y_i \leq y),$$

and let the corresponding empirical marginal distribution functions be

$$F_n(x) = H_n(x, +\infty), \quad G_n(y) = H_n(+\infty, y).$$

Then the empirical copula function can be obtained as

$$C_n(u, v) = H_n(F_n^{-1}(u), G_n^{-1}(v)), \qquad u, v \in I. \tag{3}$$

This is a rank–based estimator of a copula function. The standardized ranks $(F_n(X_i), G_n(Y_i))$ replace the unobservable pairs $(F(X_i), G(Y_i))$ and then form

a random sample $(U_1, V_1), \ldots, (U_n, V_n)$ from the copula C. Deheuvels (1979) also established consistency and asymptotic normality of the empirical copula process for i.i.d. observations of random vectors. However, the empirical copula function is not continuous, and some smoothing techniques are needed to obtain realistic estimators, using kernels, wavelets, and splines.

Fermanian et. al. (2004) proposed a kernel estimator. They construct the kernel estimator as

$$\widehat{H}_n(x, y) = \frac{1}{n} \sum_{i=1}^{n} K_n(x - X_i, y - Y_i),$$

where $K_n(x, y) = K(a_n^{-1}x, a_n^{-1}y)$ and

$$K(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} k(u, v) du dv,$$

for some bivariate kernel function $k : R^2 \to R$, with $\int k(x, y) dx dy = 1$, and a sequence of bandwidths $a_n \downarrow 0$ as $n \to \infty$. Similar smoothed marginal distributions $\widehat{F}_n$ and $\widehat{G}_n$ are obtained using univariate kernels. A smoothed kernel empirical copula estimator is given as

$$\widehat{C}_n(u, v) = \widehat{H}_n(\widehat{F}_n^{-1}(u), \widehat{G}_n^{-1}(v)), \qquad u, v \in I. \tag{4}$$

Fermanian et al. (2004) showed that, under certain regularity conditions, asymptotic normality holds for this smoothed copula process. The local linear version of the kernel was suggested by Chen and Huang (2007) and a mirror-reflection kernel estimator was also a choice. However, the kernel-based estimator is often not effective along the borders of I^2, and for good performance requires a large amount of data.

Genest et al. (2009b) proposed to estimate the copula density function using linear wavelets to smooth the empirical copula function, and a smoothed result for every level of resolution can be generated automatically. On the other hand, Autin et al. (2010) obtained an estimator using two shrinkage procedures based on thresholding.

Pitt et al. (2006) developed a Bayesian estimation procedure for multivariate Gaussian copula models in which the marginals were constrained to fixed parametric families. Hoff (2007) provided a semi–parametric Bayesian estimator using the Gaussian copula to construct a dependence structure. The estimators of the dependence parameters are based on a pseudo-likelihood, and the model avoids the specification of the marginal distributions, and thus suitable for mixed continuous and discrete data. Estimating the parameters of the Gaussian copula family are available via Markov Chain Monte Carlo methods using Gibbs sampling. However, their approach are confined to an elliptical copula family such as the Gaussian copula. Müller et al. (2004) stated that constraining inference to a specific parametric form may limit the scope and type of inference.

Recently, Bayesian nonparametric methodology has been considered to estimate the copula function. Ideas have been recently proposed by Guillotte et al. (2008, 2009), Silva et al. (2009) and Elidan (2010). Guillotte et al. (2008) focused on a kind of copula family, the extreme-value copulas, which can be characterized by a univariate dependence function. In this way, the high-dimensional problem is reduced to one variable. They construct a nonparametric model and propose Bayesian estimation of the dependence function whether the marginal distributions are known or not. The estimation approach is facilitated by jump MCMC algorithms. However, applications are restricted to the specificity of a copula family.

A further Bayesian framework for copula estimation is provided by Guillotte et al. (2009). First, they construct an approximation subspace $\mathcal{A}$ on the space $\mathcal{B}$ of all copula functions. This subspace $\mathcal{A}$ is finite dimensional and determined by a doubly stochastic matrix P. They prove that for any copula $C \in \mathcal{B}$ and $\epsilon > 0$, there exists a copula $A \in \mathcal{A}$ such that $\|C - A\| \leq \epsilon$, where $\|\cdot\|$ is a proper norm. Secondly, they consider any copula function C as a finite mixture of distributions and the weights are associated with the matrix P. In order to obtain the Bayesian copula estimator $\hat{C}$, they assume the Jeffreys prior on the subspace $\mathcal{A}$. Consequently, $\hat{C}$ is obtained by the Bayesian posterior mean. Numerical evaluation is based on Markov Chain Monte Carlo methods. Finally, they state their Bayesian approach outperforms traditional nonparametric estimators, such as kernel estimators, in many cases. However, this result is heavily dependent on the selection of the prior. Furthermore, the approach is not fully nonparametric Bayesian. Elidan (2010) developed a copula Bayesian network model to estimate flexible high-dimensional continuous distributions. But this work also depends on the choice of copula.

This paper proposes a Bayesian nonparametric procedure to estimate any copula density function c associated with a copula C. We focus on the estimator of the copula density $c(u, v)$ (assuming it exists). Our approach utilizes the well known infinite mixture style model, and in particular we justify the use of an infinite mixture of Gaussian copulas.

The layout of the article is as follows. Section 2 contains a brief description of copula and includes our Bayesian nonparametric model. We present the Gibbs sampling algorithm for estimating the model in Section 3. Illustrations and numerical results are described in Section 4 and we conclude with a brief discussion in Section 5. A proof of a theorem is given in the Appendix.

2 Copulas and the Nonparametric Model

Formally, see Nelsen (2006), we have the following definition of a copula:
Definition 1. A (two-dimensional) *copula* is a function $C : I^2 \to I$, such that:

(i) $C(0, x) = C(x, 0) = 0$ and $C(1, x) = C(x, 1) = x$ for all $x \in I$;

(ii) C is 2-*increasing*: for every $a, b, c, d \in I$ with $a \leq b$ and $c \leq d$ if

$$V_C([a,b] \times [c,d]) = C(a,c) - C(a,d) - C(b,c) + C(b,d) \geq 0.$$

Here, V_C is called the *C-volume* of the rectangle $[a,b] \times [c,d]$. Note that a copula C induces a probability measure on I^2 via $V_C([0,u] \times [0,v]) = C(u,v)$. For instance, the function $\pi(u,v) = uv$ is a copula that yields an independent structure. It is also easy to show that, as a consequence of the 2-increasing property in Definition, for any copula C we have C is nondecreasing in each variable, and C satisfies the Lipschitz condition; i.e. for every a, b, c, d in I,

$$|C(b,d) - C(a,c)| \leq |b-a| + |d-c|.$$

A Copula is thus uniformly continuous. The Fréchet–Hoeffding bounds inequality is an important exercise,

$$\max\{F(x) + G(y) - 1, 0\} \leq H(x,y) \leq \min\{F(x), G(y)\}, x, y \in R;$$

or

$$W(u,v) = \max\{u + v - 1, 0\} \leq C(u,v) \leq \min\{u,v\} = M(u,v), u, v \in I.$$

Here, the functions W and M are known as the Fréchet–Hoeffding lower and upper bounds. Furthermore, M and W are themselves copulas.

Many copula can be found in the literature, and among them, the most popular one is the Gaussian copula. Durante et al. (2010) discussed some general properties that a copula should have for being considered valuable in statistical applications. A 'good' copula should have some probabilistic interpretation, be flexible and have a wide range of dependence. The Gaussian copula satisfies such requirements. Let $N_\rho(x,y)$ denote the standard bivariate normal distribution function with correlation coefficient ρ. Then C_ρ, the copula corresponding to N_ρ, is given by

$$C_\rho(u,v) = N_\rho(\Phi^{-1}(u), \Phi^{-1}(v)).$$

where Φ denotes the standard normal distribution function. Since there is no closed form expression for Φ^{-1}, there is no closed form expression for N_ρ. However, N_ρ can be evaluated approximately in order to construct bivariate distribution functions with the same dependence structure as the standard bivariate normal distribution function but with non-normal marginals.

The Gaussian copula density $c_\rho(u,v)$ is given by eq. (1), that is

$$c_\rho(u,v) = \frac{1}{\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2} \left(\Phi^{-1}(u), \Phi^{-1}(v)\right) \left(\Sigma^{-1} - \mathbf{I}\right) \begin{pmatrix} \Phi^{-1}(u) \\ \Phi^{-1}(v) \end{pmatrix} \right\} \quad (5)$$

where

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

and $\mathbf{I}$ is the identity matrix.

We observe that any copula density function can be approximated to an arbitrary precision by a mixture of the Gaussian copula densities. The details are presented later in this section, and here is a simple example of $c(u, v)$ as a mixture of five Gaussian copula densities:

$$c(u,v) = \sum_{j=1}^{5} w_j c_{\rho_j}(u,v), \tag{6}$$

where $\rho_1 = -0.9, \rho_2 = -0.5, \rho_3 = 0, \rho_4 = 0.2, \rho_1 = 0.75$, and $w_1 = 0.1, w_2 = 0.15, w_3 = 0.05, w_4 = 0.5, w_5 = 0.2$. It is easy to show that $c(u, v)$ is a copula density.

Four copula density graphs are presented in Figure 1. The panel (a) is the Gaussian copula density and the shape is symmetric. The above mixture of five Gaussian copula densities is shown in the panel (b). It is asymmetric with some high peaks at the corners, not like the Gaussian copula density. The panel (c) shows the t copula density, that is,

$$c_{\rho,\nu}(u,v) = \frac{\Gamma(\frac{\nu+2}{2})\Gamma(\frac{\nu}{2})}{[\Gamma(\frac{\nu+1}{2})]^2\sqrt{1-\rho^2}} \frac{[1+\frac{q_\rho(t_\nu^{-1}(u),t_\nu^{-1}(v))}{\nu}]^{-\frac{\nu+2}{2}}}{[1+\frac{t_\nu^{-1}(v))^2}{\nu}]^{-\frac{\nu+1}{2}}}$$

with $q_\rho(x, y) = (x^2 + y^2 - 2\rho xy)/(1 - \rho^2)$ and $t_\nu^{-1}(u)$ denoting the inverse of the distribution of the univariate student t density with ν degrees of freedom. The shape changes with the different parameters. It is symmetric with the heavy tails on the borders. We rebuilt this dependence structure using our predictive data later. The panel (d) shows the Gumbel copula density with a heavy tail on the top. In general, any copula density function with arbitrary shape can be considered as the mixture of the Gaussian copula densities.

The plan now is to generate a complete class of copula models by mixing the correlation parameter; yielding

$$c(u,v) = \sum_{j=1}^{\infty} w_j \, c_{\rho_j}(u,v), \tag{7}$$

where $c_{\rho_j}(u, v)$ is the Gaussian copula density in eq. (5), the weights (w_j) sum to one and the (ρ_j) vary in $(-1, +1)$.

The motivation for using the Gaussian copula is as follows: taking our cue from the result that bivariate density functions on the real plain can be arbitrarily well approximated with respect to the L_1 distance by a mixture of bivariate normal distributions; of the type

$$f(x,y) = \sum_{j=1}^{\infty} w_j \, \mathrm{N}\Big((x,y)|\mu_j, \Sigma_j\Big).$$

In order to recover the copula, by transforming $U = \Phi(X)$ and $V = \Phi(Y)$, we need to obtain the marginal distributions for X and Y as standard normal.

Theorem 1. For X and Y to be marginally standard normal, it is necessary and sufficient for $\mu_j = 0$ for all j and

$$\Sigma_j = \begin{pmatrix} 1 & \rho_j \\ \rho_j & 1 \end{pmatrix}.$$

Proof. See Appendix.

Hence, our nonparametric copula model is given by (7) having appropriately transformed X and Y. This is clearly a copula density function as it is a mixture of such densities. Moreover, due to its derivation we contend it can arbitrarily well approximate any copula with respect to the L_1 distance. Undertaking full Bayesian posterior analysis may not be have been possible with an infinite mixture of copulas; however, it is quite feasible using a slice sampling algorithm recently introduced to the literature.

3 The Gibbs sampling algorithm

In order to transform the infinite model (7) to be finite, we introduce a standard latent variable z_i for each i such that the joint density of (u_i, v_i, z_i), is given by

$$c(u_i, v_i, z_i) = \sum_{j=1}^{\infty} \mathbf{1}(z_i < \xi_j)\,(w_j/\xi_j)\,c_{\rho_j}(u_i, v_i)$$

for some deterministic decreasing sequence (to 0) (ξ_j). See Kalli et al. (2010) for details of this approach for estimating infinite mixture models. The model in the presence of the (z_i) becomes finite since the number of j such that $\xi_j > z_i$ is finite and so, conditional on z_i, the number of parameters is finite. Furthermore, another latent variable d_i is introduced, which allocates each observation to one component of the mixture model. Therefore the joint density of (u_i, v_i, z_i, d_i) is

$$c(u_i, v_i, z_i, d_i) = \mathbf{1}(z_i < \xi_{d_i})\,(w_{d_i}/\xi_{d_i})\,c_{\rho_{d_i}}(u_i, v_i),$$

where $d_i \in \{1, 2, 3, \ldots\}$. This form omits any sums.

With all the n observations, we can write down the full likelihood function as

$$\prod_{i=1}^{n} c(u_i, v_i, z_i, d_i) = \prod_{i=1}^{n} \mathbf{1}(z_i < \xi_{d_i}) \frac{w_{d_i}}{\xi_{d_i}} c_{\rho_{d_i}}(u_i, v_i).$$

Therefore the full posterior distribution can be expressed as

$$f(\rho|\mathbf{u}, \mathbf{v}, \mathbf{z}, \mathbf{d}) \propto \pi(\rho) * \prod_{i=1}^{n} c(u_i, v_i, z_i, d_i) = \pi(\rho) * \prod_{i=1}^{n} \mathbf{1}(z_i < \xi_{d_i}) \frac{w_{d_i}}{\xi_{d_i}} c_{\rho_{d_i}}(u_i, v_i).$$

To be able to implement a Gibbs sampler to the model we discussed above, a set of full conditional density functions is required. The chain can be initialized

in the following way. We initialize $\{d_i = i, i = 1, \cdots, n\}$ and then simulate $\{z_i, i = 1, \cdots, n\}$ from a uniform distribution between 0 and $\xi_{d_i} = e^{-\beta d_i}$. Then we define $N_i = \{-\llcorner \log z_i \lrcorner\}$, where $\llcorner X \lrcorner$ defines the largest integer less than or equal to X. Define, also, $N_{\max} = \max_{1 \le i \le n}\{N_i\}$.

Step 1: Updating w. The conditional distribution of the parameter, w_j, given $d_i, i = 1, \cdots, n$, is proportional to

$$w_j|\cdots \sim \text{beta}\left(a + \#\{d_i = j\}, b + \#\{d_i > j\}\right),$$

where $\#\{d_i = j\}$ registers the number of d_i, which equals j and $\#\{d_i > j\}$ registers the number of $\{d_i > j\}$.

Step 2: Updating z. The z_j follows the uniform distribution with support $(0, \xi_{d_i})$.

Step 3: Updating d_i. The values of d_i can take between 0 and N_i, which is derived from the value of z_i. We have the density of d_i as proportional to

$$Pr(d_i = j|\cdots) \propto \frac{w_j}{\xi_j} c_{\rho_j}(u_i, v_i).$$

Step 4: Updating ρ. The full conditional distribution of the parameter ρ_j is

given as following

$$f(\rho_j|\cdots) \propto \pi(\rho_j) * \prod_{d_i = j} c_{\rho_j}(u_i, v_i),$$

where $\pi(\rho_j)$ is the prior distribution for ρ_j, which we assume it following the uniform distribution with support $(-1, 1)$. The detailed sampling procedure for $c_{\rho_j}(u_i, v_i)$ will be discussed later. Once we can sample from $c_{\rho_j}(u_i, v_i)$, the chain is completed.

To be able to sample from $c_{\rho_j}(u_i, v_i)$, we introduce the slice sampling idea. Substituting $u = \Phi(x)$ and $v = \Phi(y)$ into (5), we can obtain the dependence structure of (X, Y), c, the copula density as

$$c(\Phi(x), \Phi(y)) = \frac{1}{\sqrt{1-\rho^2}} \exp\Big\{ -\frac{1}{2(1-\rho^2)}(\rho^2 x^2 + \rho^2 y^2 - 2\rho xy)\Big\},$$

with $-1 < \rho < 1$.

Then, the full likelihood can be presented as

$$\begin{aligned} f(\rho; x, y) &= \prod_{i=1}^{n} \frac{1}{\sqrt{1-\rho^2}} \exp\Big\{ -\frac{1}{2(1-\rho^2)}(\rho^2 x_i^2 + \rho^2 y_i^2 - 2\rho x_i y_i)\Big\} \\ &= \left(\frac{1}{1-\rho^2}\right)^{\frac{n}{2}} \exp\left\{ -\frac{1}{2(1-\rho^2)}\left[\rho^2 \sum_{i=1}^{n}(x_i^2 + y_i^2) - 2\rho \sum_{i=1}^{n} x_i y_i\right]\right\}. \end{aligned}$$

Let $A = \sum_{i=1}^{n}(x_i^2 + y_i^2)$ and $B = 2\sum_{i=1}^{n} x_i y_i$, posterior distribution of ρ, given the prior $\pi(\rho)$, is proportional to

$$f(\rho|\cdots) \propto \pi(\rho)\left(\frac{1}{1-\rho^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(A\rho^2 - B\rho)\right\}.$$

We introduce two latent variables λ and η in such way that the posterior distribution can be represented as

$$f(\rho, \lambda, \eta) \propto \pi(\rho) * e^{-\frac{\lambda}{2}}\mathbf{1}\left(\lambda > \frac{A\rho^2 - B\rho}{1-\rho^2}\right) * \mathbf{1}\left(\eta < \left(\frac{1}{1-\rho^2}\right)^{\frac{n}{2}}\right).$$

It is straightforward to see that integrated out the latent variables λ and η, we will obtain the posterior distribution of ρ. Now we can use the Gibbs sampler. We start initializing by simulating $\rho \sim \mathrm{U}(-1, 1)$. Then we update following the steps:

$$f(\eta|\cdots) \propto \mathrm{U}\left[0, \left(\frac{1}{1-\rho^2}\right)^{\frac{n}{2}}\right],$$

$$f(\lambda|\cdots) \propto e^{-\frac{\lambda}{2}}\mathbf{1}\left(\lambda > \frac{A\rho^2 - B\rho}{1-\rho^2}\right)$$

and

$$f(\rho|\lambda, \eta) \propto \pi(\rho) * \mathbf{1}\left(\{\lambda > \frac{A\rho^2 - B\rho}{1-\rho^2}\} \cap \{\eta < \left(\frac{1}{1-\rho^2}\right)^{\frac{n}{2}}\}\right).$$

In the next section we run this code on some predictive data sets and also a real data set.

4 Numerical Results

In order to examine the numerical performance of the proposed model, we perform two examples below. As a first example, consider two groups of data that are generated from the extremely strong positive correlation and the t copula, respectively. The copula density estimator is obtained for every data set. Secondly, we apply our Bayesian method to the real data. The evidences illustrate that our Bayesian nonparametric estimators give effective approaches in general.

For simulation, two common dependence structures are considered. For each copula, we generate a data set of (U, V) using Matlab. That is, an i.i.d. standard uniform bivariate random sample is got from the specified copula. The sample sizes considered are $n = 500$. Based on copula theory, $U_1, U_2, \ldots, U_n$ is a sample from uniform $U(0, 1)$ and so is $V_1, V_2, \ldots, V_n$. Therefore, the bivariate copula density $c(u, v)$ can be considered as the joint density (pdf) of (U, V). The predictive data (U', V') is estimated by the Gibbs sampling in Section 3. With 4,000 iterations, 500 pairs are used after burning in the first 3,500 ones.

The bivariate copula density $c(u, v)$ is obtained by the classic kernel method. It is the same to (U', V').

The first dependence model is the extremely strong positive correlation. The Gaussian copula in eq.(5) is chosen with the correlation coefficient $\rho = 0.99$. Four graphical representations of this group of data are provided in Figure 2. In the top row, the panels (a) and (b) show the generated data (U, V) on the scatter and kernel-based copula density plots, respectively. Those of predictive data (U', V') are in the panel (c) and (d). The approach illustrates that it is easy to estimate this linear structure.

The second dependence structure is the t copula. The scatter plot in the panel (a) of Figure 3 shows the feature of this copula. The parameters are especially chosen with the correlation coefficient $\rho = -0.5$ and the degree of freedom $\nu = 1$ because the t copula with low freedom produces greater upper and lower tail dependence. See Figure 3. This dependence structure can also be mixed with the Gaussian copula densities. The predictive approaches in panel (c) and (d). It approximately returns the real copula density.

The last example uses the medical data from the Framingham Heart study database. Qu et al. (2012) analyzed similar data using MPLE-TV method. Based on the systolic (SBP) and diastoloc (DBP) blood pressures (in mm*Hg) measured on 468 subjects, we focus on the dependence structure via our infinite mixture model. In the simulation, let X be SBP and Y be DBP, the marginals $F(x)$ and $G(y)$ are unknown. We then follow Genest et al. (2009b) and consider the observations $(\hat{U}_i, \hat{V}_i) = (R_i/n, S_i/n) = (F_n(X_i), G_n(Y_i)), i = 1, \ldots, n$. Here, R_i is the rank of X_i among $X_1, \ldots, X_n$ and S_i is the rank of Y_i among $Y_1, \ldots, Y_n$, denoted by F_n and G_n the empirical distribution. These standardized ranks of the sample closely approximate to the pairs $(U_i, V_i) = (F(X_i), G(Y_i))$. The scatter plot of log(SBP) and log(DBP) is in the panel (a) of Figure 4, and panel (b) shows the plot of rank (log(SBP))/n and rank (log(DBP))/n. It illustrates that there is a positive correlation between the two responses, but not very strong.

The Monte Carlo sample size is 10,000 with the initial 9,000 used as a burn in period. See Figure 5. The scatter plot of the predictive data in the panel (c)-(e) broadly characterizes the dependence structure of real data in panel (a) and (b). They show a strong agreement between the real data and our predictive sampler using the Gaussian mixture structure.

In this infinite model, the key is to find out the exact number of occupied clusters in a sample. That is, how slice sampling ideas in Section 3 can be applied to estimate the copula density. We are concerned with the finite number of variable to sample a valid Markov chain with correct stationary distribution. The number of clusters, k, in each iteration is shown in the panel (f). The mean of k is 7. That means the copula density is considered as the mixture of 7 Gaussian copula densities. This shape of copula is not similar to any known family model such as the Gumbel or Gaussian copula. But, more flexible copula density can be estimated by our Bayesian nonparametric model as we should expect.

5 Discussion

We present a Bayesian nonparametric methodology to estimate arbitrary bivariate copula density. There is no need for any assumption of the copula family models, so this approach extends the range of estimation. If a copula function is known, our predictive data can reconstruct the key features of a copula that are symmetry, skewness, heavy tail or linear relationship. It is also true on the phenomenon when the copula is unknown. The estimator is not obtained until the observations are transformed into standard uniform variables.

Two points need to be discussed. First, the Bayesian approaches rely on the consistency and convergence rate of the Markov chain. Much work is ahead for reliable diagnosis. Secondly, there is more challenges in higher dimensions. How to deal with "the curse of dimensionality"? The following work will be done on multi-dimensional copula density estimation using the Bayesian nonparametric method.

Appendix: Proof of Theorem 1.

To prove the theorem, we need to show that

$$\sum_{j=1}^{+\infty} w_j N(x|\mu_j, \sigma_j^2) = N(x|0,1) \tag{8}$$

for all the j it is necessary and sufficient for $\mu_j = 0$ and $\sigma_j^2 = 1$ for all j.

To maintain (8), we replace the both sides of equation by their characteristic functions, as follows

$$\begin{aligned} & \sum_{j=1}^{+\infty} w_j e^{it\mu_j - (1/2)\sigma_j^2 t^2} = e^{-(1/2)t^2} \\ \Leftrightarrow \quad & \sum_{j=1}^{+\infty} w_j \cos(t\mu_j) e^{-(1/2)\sigma_j^2 t^2} + i \sum_{j=1}^{+\infty} w_j \sin(t\mu_j) e^{-(1/2)\sigma_j^2 t^2} = e^{-(1/2)t^2}, \end{aligned}$$

which we obtain the following two equations

$$\sum_{j=1}^{+\infty} w_j \cos(t\mu_j) e^{-(1/2)\sigma_j^2 t^2} = e^{-(1/2)t^2} \tag{9}$$

and

$$\sum_{j=1}^{+\infty} w_j \sin(t\mu_j) e^{-(1/2)\sigma_j^2 t^2} = 0 \tag{10}$$

Applying the Taylor expansion to both sides of (9), we have

$$\begin{aligned}
& \sum_{j=1}^{+\infty} w_j \cos(t\mu_j) \Big(\sum_{k=0}^{+\infty} \frac{(-(1/2)\sigma_j^2 t^2)^k}{k!} \Big) = \sum_{k=0}^{+\infty} \frac{(-(1/2)t^2)^k}{k!} \\
\iff & \sum_{k=0}^{+\infty} \frac{(-(1/2)t^2)^k}{k!} \Big(\sum_{j=1}^{+\infty} w_j \cos(t\mu_j) \sigma_j^{2k} \Big) = \sum_{k=0}^{+\infty} \frac{(-(1/2)t^2)^k}{k!} \\
\iff & \sum_{j=1}^{+\infty} w_j \cos(t\mu_j) \sigma_j^{2k} = 1 \qquad \forall t
\end{aligned}$$

Let $t = 0$, we have

$$\begin{aligned}
& \sum_{j=1}^{+\infty} w_j \sigma_j^{2k} = 1 \\
\Rightarrow \quad & \sigma_j^2 = 1
\end{aligned}$$

for all j.

From (10), we have

$$\sum_{j=1}^{+\infty} w_j \sin(t\mu_j) = 0$$

for all j. Applying the Taylor expansion for the *sin* function, we have

$$\sum_{j=1}^{+\infty} w_j [t\mu_j - (t\mu_j)^3/3! + (t\mu_j)^5/5! - \cdots + (-1)^{k+1} (t\mu_j)^{2k-1}/(2k-1)! + \cdots] = 0.$$

For the above equation to be real for all the $t \in R$, the following equation has to be true for all the $k \in \{1, 2, \cdots, +\infty\}$

$$\sum_{j=1}^{+\infty} w_j \mu_j^{2k-1} = 0.$$

Hence $\mu_j = 0$ for all j.

References

[1] Autin F., Pennec E.L., Tribouley K., 2010. Thresholding methods to estimate copula density. Journal of Multivariate Analysis 101, 200-222.

[2] Chen S.X., Huang T., 2007. Nonparametric estimation of copula functions for dependence modelling. Canadian Journal of Statistics 35, 265-282.

[3] Choroś B., Ibragimov R., Permiakova E., 2010. Copula estimation, in: Copula Theory and Its Applications. In: Lecture Notes in Statistics, V198. Springer, 77-91.

[4] Deheuvels P., 1979. La fonction de dépendance empirique et ses propriétés, un test non paramétrique d'indépendance. Bulletin de l'Académie Royale de Belgique, Classes de Sciences 65, 274-292.

[5] Durante F., Sempi C., 2010. Copula theory: an introduction. In: F. Durante, W. Härdle, P. Jaworski, T. Rychlik (eds.) Workshop on Copula Theory and its Applications. Springer, Dortrecht.

[6] Elidan G., 2010. Copula Bayesian networks. Advances in Neural Information Processing Systems 23, 559-567.

[7] Fermanian J.D., Radulovic D., Wegkamp M., 2004. Weak convergence of empirical copulas processes. Bernoulli 10, 847-860.

[8] Genest C., Rivest L.P., 1993. Statistical inference procedures for bivariate Archimedean copulas. Journal of the American Statistical Association 88, 1034-1043.

[9] Genest C., Gendron M., Bourdeau-Brien M., 2009a. The advent of copulas in finance. The European Journal of Finance 00, 1-10.

[10] Genest C., Masiello E., Tribouley K., 2009b. Estimating copula densities through wavelets. Insurance: Mathematics and Economics 44, 170-181.

[11] Guillotte S., Perron F., 2008. A Bayesian estimator for the dependence function of a bivariate extreme-value distribution. The Canadian Journal of Statistics 36, 383-396.

[12] Guillotte S., Perron F., 2009. Nonparametric Bayesian estimation of a bivariate copula using the Jeffreys prior. 28, Retrieved from http:// arxiv.org/abs/0908.2372.

[13] Hoff P.D., 2007. Extending the rank likelihood for semiparametric copula estimation. Annals of Applied Statistics 1, 265-283.

[14] Huard D., Evin G., Favre A.C., 2006. Bayesian copula selection. Computational Statistics and Data Analysis 51, 809-822.

[15] Kalli M., Griffin J.E., Walker S.G., 2010. Slice sampling mixture models. Statistics and Computing 21, 93-105.

[16] Kim G., Silvapulle M.J., Silvapulle P., 2007. Comparison of semiparametric and parametric methods for estimating copulas. Computational Statistics and Data Analysis 51, 2836-2850.

[17] Müller P., Quintana F.A., 2004. Nonparametric Bayesian data analysis. Statistical Science 19, 95-110.

[18] Nelsen R.B., 2002. Concordance and copulas: A survey. Distributions with Given margins and Statistical Modelling. Kluwer Academic Publishers, Dordrecht, 169-178.

[19] Nelsen R.B., 2006. An introduction to copulas, Second ed. Springer, New York.

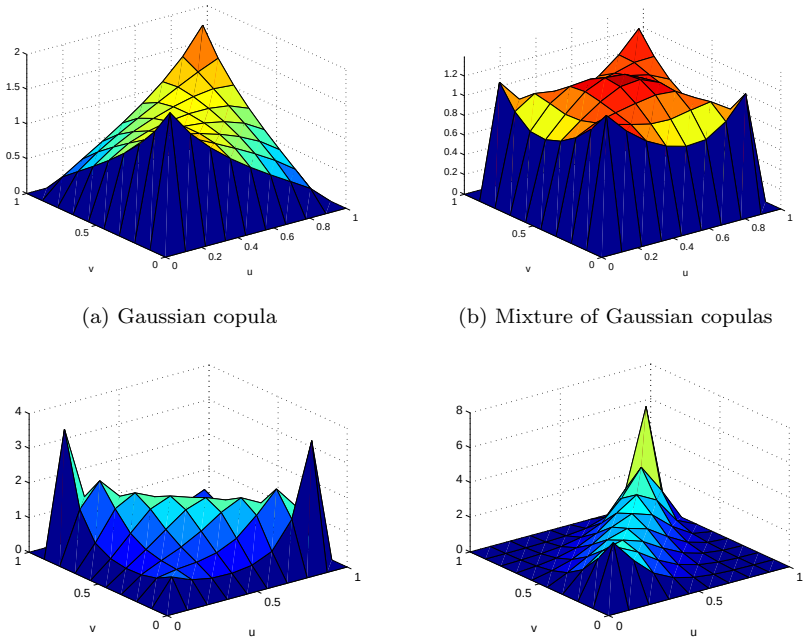
[20] Pitt M., Chan D., Kohn R., 2006. Efficient Bayesian inference for Gaussian copula regression models. Biometrika 93, 537-554.

[21] Qu L., Yin W., 2012. Copula density estimation by total variation penalized likelihood with linear equality constraints. Computational Statistics and Data Analysis 56, 384-398.

[22] Silva R., Gramacy R.B., 2009. MCMC methods for Bayesian mixtures of copulas. Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS), Clearwater Beach, Florida, USA, 5 of JMLR: W&CP 5.

[23] Silva R.S., Lopes H.F., 2008. Copula, marginal distributions and model selection: a Bayesian note. Statistics and Computing 18, 313-320.

[24] Sklar A., 1959. Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de statistique de l'Université de Paris 8, 229-231.



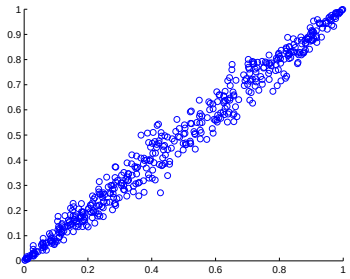
(a) Gaussian copula

(b) Mixture of Gaussian copulas

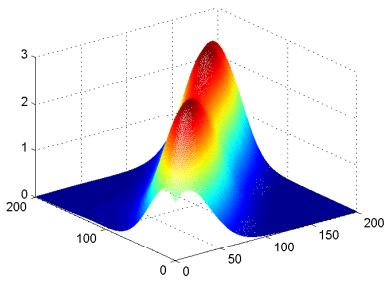
(c) t copula

(d) Gumbel copula

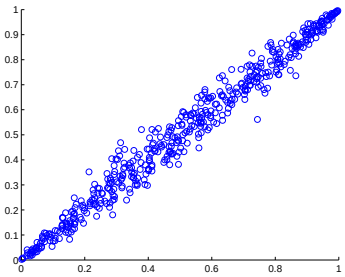
Figure 1: Plots of the copula density functions. Panel (a): the Gaussian copula density with the correlation coefficient $\rho = 0.5$; panel (b): the mixture of 5 Gaussian copula densities in eq. (5) with $\rho_1 = -0.9, \rho_2 = -0.5, \rho_3 = 0, \rho_4 = 0.2, \rho_1 = 0.75$ and the weights are $w_1 = 0.1, w_2 = 0.15, w_3 = 0.05, w_4 = 0.5, w_5 = 0.2$; panel (c): the t copula density with $\rho = -0.5$ and the degree of freedom parameter $\nu = 1$; panel (d): the Gumbel copula density with $\rho = 3$.



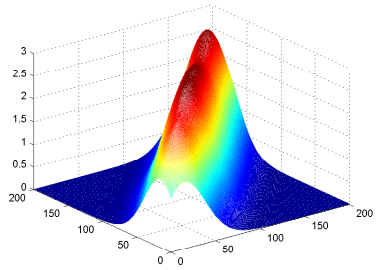
(a) Scatter of data



(b) Copula density of data

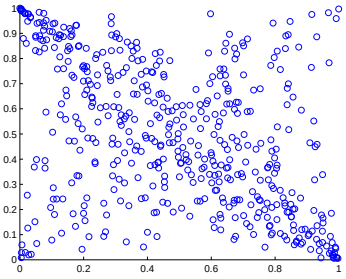


(c) Scatter of predictive sample

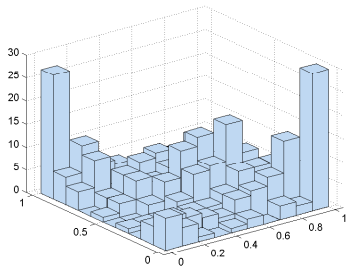


(d) Copula density of predictive sample

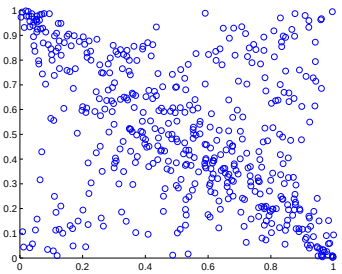
Figure 2: Gussian copula with $\rho = 0.99$, and the number $n = 500$. Panel (a): scatter of the predictive data; panel (b): kernel-based copula density of the predictive data; panel (c) and (d): scatter and kernel-based copula density of the predictive data, respectively.



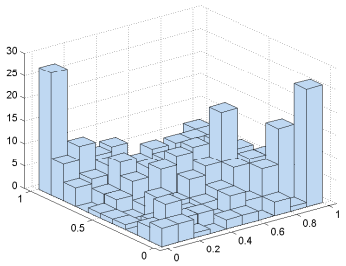
(a) Scatter of data



(b) Histogram of data

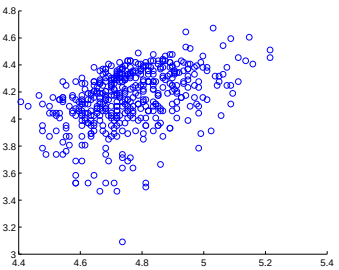


(c) Scatter of predictive sample

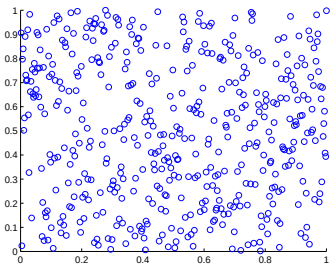


(d) Histogram of predictive sample

Figure 3: The t copula density with $\rho = -0.5$, the degree of freedom $\nu = 1$, and the number $n = 500$. Panel (a): scatter of the predictive data; panel (b): histogram of the predictive data; panel (c) and (d): scatter and histogram of the predictive data, respectively.

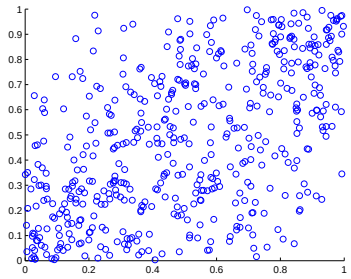


(a) Scatter of real data

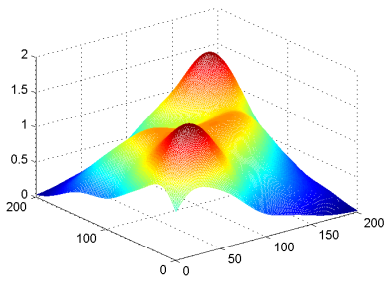


(b) Rank-rank of real data

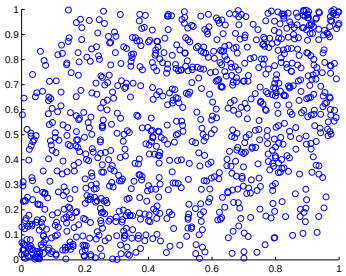
Figure 4: Panel (a): scatter plot of log(SBP) and log(DBP); Panel (b): rank-rank plot of (log(SBP))/n and (log(DBP))/n.



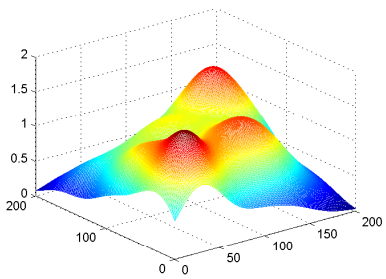
(a) Scatter of data



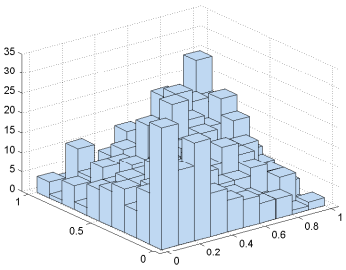
(b) Copula density of data



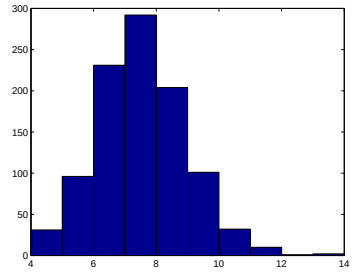
(c) Scatter of predictive sample



(d) Copula density of predictive sample



(e) Histogram of predictive sample



(f) Histogram of the number of clusters

Figure 5: Dependence structure of SBP and DBP, and 468 pairs of the data transformed by the empirical distribution function. Panel (a) and (b): scatter and kernel-based copula density plots of the data, respectively; Panel (c)-(e): scatter, kernel-based copula density and histogram plots of the predictive sampler, respectively; Panel (f): histogram of k, the number of cluster in the mixture Gaussian copula densities.