Methodol Comput Appl Probab (2014) 16:747–763
DOI 10.1007/s11009-013-9348-5

Bayesian Nonparametric Inference for a Multivariate Copula Function

Juan Wu · Xue Wang · Stephen G. Walker

Received: 25 July 2012 / Revised: 30 April 2013 /
Accepted: 28 May 2013 / Published online: 9 June 2013
© Springer Science+Business Media New York 2013

Abstract The paper presents a general Bayesian nonparametric approach for estimating a high dimensional copula. We first introduce the skew–normal copula, which we then extend to an infinite mixture model. The skew–normal copula fixes some limitations in the Gaussian copula. An MCMC algorithm is developed to draw samples from the correct posterior distribution and the model is investigated using both simulated and real applications.

Keywords Bayesian nonparametric estimation · Copula · Infinite mixture skew–normal copula model · Metropolis–Hastings algorithm

AMS 2000 Subject Classification 62H12

1 Introduction

Copula models have been investigated quite extensively in recent years. Applications are across the board; from financial risk and the insurance industry to hydrologic engineering and medical applications, where a wide variety of complex dependent structures of random variables are typically high dimensional. A copula offers a flexible tool that demonstratively allows an experimenter to divide the cumulative distribution function into two parts; the marginal distributions and a copula function. The copula can completely characterize the statistical dependence of multiple variables. Although bivariate copula have been widely discussed and applied, see for

J. Wu (✉)
School of Mathematics and Statistics,
Huazhong University of Science and Technology, Wuhan, China
e-mail: wujuan@mail.hust.edu.cn

X. Wang · S. G. Walker
School of Mathematics, Statistics and Actuarial Science,
University of Kent at Canterbury, Canterbury, UK

example Genest et al. (2009) and Nelsen (2006), the application of copula for higher dimensional data remains relatively few. The reason is that it is not straightforward to find flexible families of distributions on $[0, 1]^d$ for $d > 2$.

Our approach is to concentrate on the modeling of the copula function alone. In one respect it can be seen as a Bayesian nonparametric approach to the ideas set out in Genest et al. (1995). In this paper the data are transformed to the unit interval via the empirical distribution function. That is, if $(x_1, \ldots, x_n)$ is a continuous sample, $x_i \neq x_j$ for $i \neq j$, then first define the empirical distribution function

$$F_n(x) = n^{-1} \sum_{i=1}^{n} \mathbf{1}(x_i \leq x),$$

and then set the appropriate transformed data as $u_{ni} = F_n(x_i)$. Hence, u_{ni} will be in the unit interval and the set $(u_{n1}, \ldots, u_{nn})$ coincides with the set $(1/n, 2/n, \ldots, 1)$. The 1 at the end may cause concern for modeling and hence Genest et al. (1995) propose the use of $(1/(n+1), 2/(n+1), \ldots, n/(n+1))$ instead. In the case of bivariate data (and while we are discussing multivariate data sets, for the purpose of this introduction we will demonstrate things in the bivariate case) then, the likelihood function is given, for a sample $((x_1, y_1), \ldots, (x_n, y_n))$, by

$$\prod_{i=1}^{n} c_\theta(u_{ni}, \upsilon_{ni})$$

where

$$u_{ni} = \frac{n}{n+1} F_{nX}(x_i) \quad \text{and} \quad \upsilon_{ni} = \frac{n}{n+1} F_{nY}(y_i)$$

and c_θ is a parametric copula density function.

While we use the same transformed data as Genest et al. (1995), we instead develop a Bayesian nonparametric approach to the modeling and estimation of the copula density function. The idea is to use infinite mixture models, based on the Gaussian copula, to construct such a flexible family of copula densities. Hence, our approach follows the well known infinite mixture model whereby weights are assigned to components. The choice of the Gaussian copula to model each component is highly appropriate since it can assign arbitrary dependence, pairwise, to each of the variables. The Gaussian copula is fully characterized by a correlation matrix.

A full Bayesian analysis using the Gaussian copula has been reported in Pitt et al. (2006). Here the authors use the full likelihood, including both marginal and copula model;

$$\prod_{i=1}^{n} f_X(x_i|\psi)\, f_Y(y_i|\psi)\, c_\theta\Big(F_X(x_i|\psi),\, F_Y(y_i|\psi)\Big).$$

In particular, these authors use a Gaussian copula and assign a prior to the correlation matrix. This is based on the Wishart distribution; and for a sampling definition of the prior we would sample a covariance matrix Σ from a Wishart distribution and then obtain the correlation matrix R through $R = D\Sigma D$, where D is a diagonal matrix, to be defined later, and is fully determined by Σ. We will also be adopting this prior. Dalla Valle (2009) used this idea involving both the Gaussian and Student-t copula, and applied it to operational risk management for modeling the dependency of a high number of marginals. Also, Diers et al. (2012) illustrated the modeling of dependence structures of non-life insurance risks using the Bernstein copula, where the advantages of the Bernstein copula, including its flexibility in mapping inhomogeneous dependent structures and its easy use in a simulation context was highlighted.

Within Bayesian nonparametric methodology, attempts have been made to construct distributions on $[0, 1]^d$ directly without the explicit use of copulas. This involves the use of tree–structure mixtures, Kirshner (2007), also employed by Silva and Gramacy (2009), who presented an estimator for the copula density via a Markov chain Monte Carlo (MCMC) algorithm. We would find it difficult to develop a full Bayesian nonparametric model based on copula and marginal densities, since in

$$f_X(x)\, f_Y(y)\, c\Big(F_X(x), F_Y(y)\Big)$$

we would need to model all of f_X, f_Y and c using infinite mixture models; and this would stretch any inference plan via MCMC methods to the limit. Hence, Wu et al. (2013) to used the data transform idea and concentrated solely on the copula function estimation, and we follow this idea.

The layout of the article is as follows. Section 2 contains a brief description of a copula model, and is where we also present the infinite mixture Gaussian copula model. The Metropolis–Hastings algorithm for sampling the model, in particular the correlation matrices, is described in Section 3, and the numerical illustrations involving simulated and real data sets are provided in Section 4.

2 The Copula Model

A copula is a cumulative distribution function defined on $[0, 1]^d$ such that every marginal is uniform on $[0, 1]$. The well known Sklar Theorem (Sklar 1959), provides the theoretical foundation for a copula which allows the separation of the marginal distributions of X_m, for $m = 1, \ldots, d$, for any d–vector X, and the dependence structure between these variables. The basic theory of a copula is introduced, for example, in Nelsen (2006).

Let $(U_1, \ldots, U_d)$ be real random variables with uniform marginal distributions on $[0, 1]$. A copula $C : [0, 1]^d \to [0, 1]$ is a joint distribution function

$$C(u_1, \ldots, u_d) = P\Big(U_1 \le u_1, \ldots, U_d \le u_d\Big).$$

Let $d \ge 2$ and H be any d–dimensional cumulative distribution function and F_m be the marginal distribution function for X_m. Then there exists a d–dimensional copula, C, such that

$$H(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)), \quad \forall\ (x_1, \ldots, x_d) \in \mathbb{R}^d. \tag{1}$$

Furthermore, if each marginal distribution F_m of H is continuous, then C is unique.

Parametric copula models have been extensively studied. There are numerous classes of parametric copulas, such as the elliptic family, which contains the Gaussian copula and the Student t copula; and the Clayton copula; the Gumbel copula and the Frank copula, which belong to the Archimedean family. For inference, it is important to select an appropriate parametric copula, which is far from straightforward. See for example Genest and Favre (2007) and Genest et al. (2009). Assuming the continuous marginal distributions as $F_1, \ldots, F_d$, the standard form for the copula density is given by

$$c(u_1, \ldots, u_d) = \frac{\partial^d}{\partial u_1 \cdots \partial u_d} C(u_1, \ldots, u_d) = \frac{h(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \cdots f_d(F_d^{-1}(u_d))}, \tag{2}$$

where h is the joint density of $(X_1, \ldots, X_d)$, $F_p^{-1}(u_p) = \inf\{x \in \mathbb{R} : F_p(x) \geq u_p\}$, $1 \leq p \leq d$ and $u = (u_1, \ldots, u_d) \in [0, 1]^d$. This would be a classical copula density if the margins, and thus the observations, $(U_{i1} = F_1(X_{i1}), \ldots, U_{id} = F_d(X_{id}))$ for $i = 1, \ldots, n$, are known.

From the standard normal distribution $\mathrm{N}_d(\mathbf{0}, R)$, where R is a correlation matrix, we obtain the d–dimensional Gaussian copula function:

$$C_R(u_1, \ldots, u_d) = \Phi_R^d\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right).$$

where Φ_R^d is the cumulative distribution function of $\mathrm{N}_d(\mathbf{0}, R)$, and Φ is the distribution function of N(0, 1). The density of the Gaussian copula is thus given by

$$c_R(u_1, \ldots, u_d) = |R|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{x}^T(R^{-1} - I)\mathbf{x}\right), \tag{3}$$

where $u_j = \Phi(x_j)$, for $j = 1, \ldots, d$.

However, this copula has a serious drawback, illustrated in the bivariate case, which is that $c_R(u_1, u_2) = c_R(1 - u_1, 1 - u_2)$. This can be seen by the fact that $\mathbf{x}^T\mathbf{x}$ is identical for $\mathbf{x} = (x_1, x_2)$ and $\mathbf{x} = (-x_1, -x_2)$. In the next section we introduce a new copula model which generalises the Gaussian copula and which also removes this symmetric constraint.

2.1 The Skew–Normal Copula

The Gaussian copula is one of the most widely used copulas because of its attractive properties and mathematical tractability. However, the alluded to symmetric property of the Gaussian copula makes it difficult to deal with data set with skewness; a situation often occurring in practical problems. Figure 1, plots (a) and (b), show the contour plots of the bivariate Gaussian copula with correlation coefficients 0.5 and −0.5, respectively. As we can see in both situations, the plots are fully symmetric; i.e. $c_R(u_1, u_2) = c_R(1 - u_1, 1 - u_2)$.

Instead of the Gaussian copula, we want to generate a class of copulas, which includes the Gaussian copula, and which can deal with a wide range of skewness, while at the same time maintain mathematical tractability. We base the new copula on the multivariate skew–normal distribution. Following Azzalini (1985), a random variable Z has a skew–normal distribution with a skewness parameter λ, written $Z \sim \mathcal{SN}(\lambda)$, if its density function is given by

$$sn_1(z; \lambda) = 2\phi_1(z)\Phi(\lambda z), \qquad z \in \mathbb{R}, \tag{4}$$

where sn_1 denotes the density function of the skew–normal and $\phi_1(x)$ and $\Phi(x)$ denote the $N(0, 1)$ density and distribution functions, respectively. The parameter

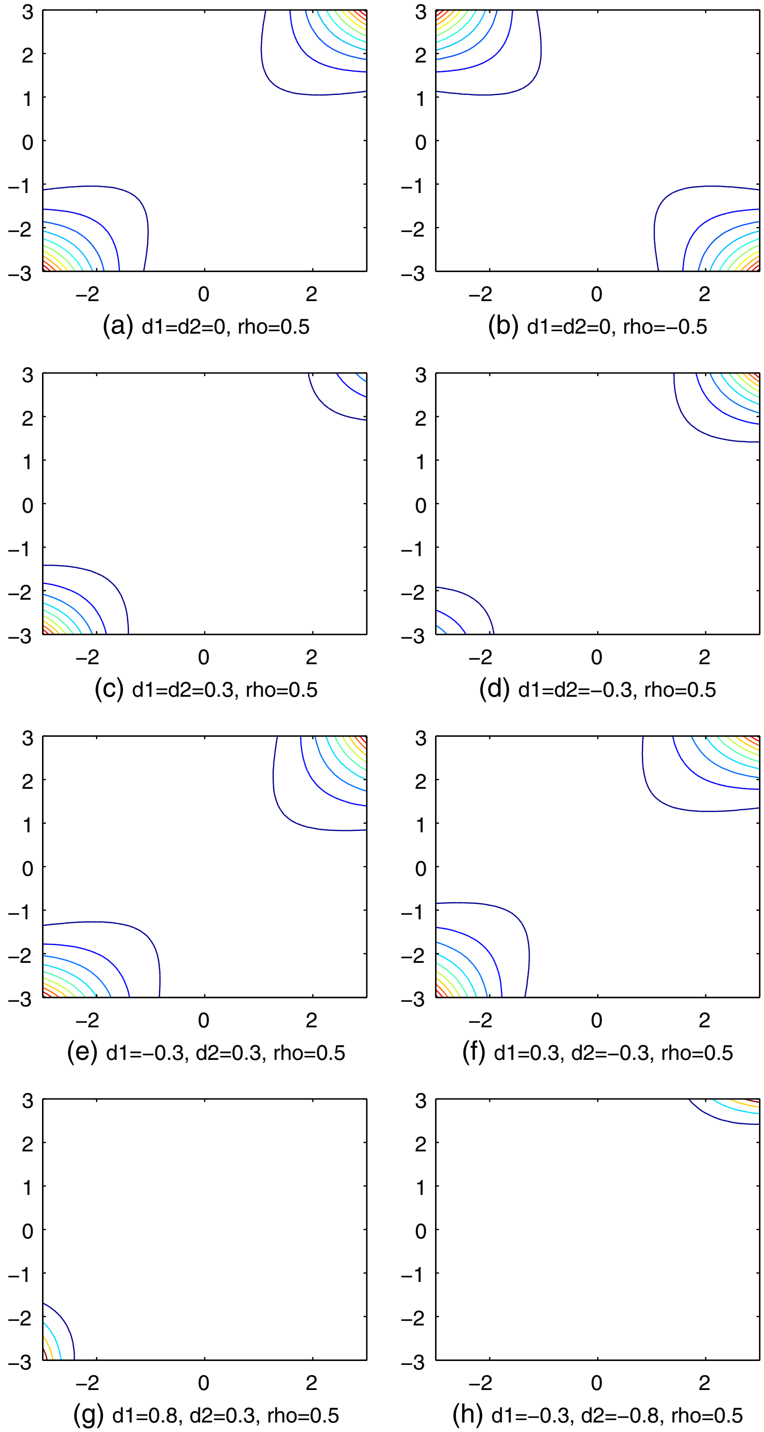


Fig. 1 Contour plots of the Gaussian copula density (plots **a** and **b**) and the skew–normal copula density with different combinations of ρ, δ_1 and δ_2 (plots **c** to **h**)

λ, which regulates the skewness, varies in $(-\infty, \infty)$ and $\lambda = 0$ corresponds to the $N(0, 1)$ density.

A further representation of Z, included in Azzalini and Dalla Valla (1996), shows a way to transform from a normal random variable to a skew–normal random variable. It states that: If Y_0 and Y_1 are independent $N(0, 1)$ variables and $\delta \in (-1, 1)$, then

$$Z = \delta|Y_0| + (1 - \delta^2)^{1/2} Y_1 \tag{5}$$

follows the skew–normal distribution, denoted as $Z \sim \mathcal{SN}(\lambda(\delta))$, where $\lambda(\delta) = \delta/(1 - \delta^2)^{1/2}$.

As mentioned in Azzalini (1985) and Azzalini and Dalla Valla (1996), the density (Eq. 4) enjoys a number of formal properties which resemble those of the normal distribution and are also suitable for the analysis of data exhibiting a unimodal distribution, but with skewness present.

Multivariate extensions of Eq. 4 were first proposed by Azzalini (1985) and expanded further by Azzalini and Dalla Valla (1996). For the $d-$dimensional extension of Eq. 4, we consider here the transformation method mentioned in Azzalini and Dalla Valla (1996), using the idea of Eq. 5. Consider a $d-$dimensional normal random variable $\mathbf{Y} = (Y_1, \ldots, Y_d)$ with standard normal marginals, independent of $Y_0 \sim N(0, 1)$; thus

$$\begin{pmatrix} Y_0 \\ \mathbf{Y} \end{pmatrix} \sim \mathcal{N}_{d+1} \left\{ \mathbf{0}, \begin{pmatrix} 1 & 0 \\ 0 & R \end{pmatrix} \right\}, \tag{6}$$

where R is a $d \times d$ correlation matrix. If $(\delta_1, \ldots, \delta_d)$ are in $(-1, 1)^d$, define

$$Z_j = \delta_j |Y_0| + \left(1 - \delta_j^2\right)^{1/2} Y_j, \qquad j = 1, \ldots, d,$$

so that $Z_j \sim \mathcal{SN}(\lambda(\delta_j))$. Then $\mathbf{Z} = (Z_1, \cdots, Z_d)^T$ follows the multivariate skew–normal distribution and its density function can be written as

$$sn_d(\mathbf{z}) = 2\phi_d(\mathbf{z}; \mathbf{\Omega})\Phi(\alpha^T \mathbf{z}), \qquad \mathbf{z} \in \mathbb{R}^d,$$

where $\phi_d(\mathbf{z}; \mathbf{\Omega})$ denotes the density function of $d-$dimentional normal distribution with the covariance matrix $\mathbf{\Omega}$ and

$$\alpha^T = \frac{\lambda^T R^{-1} \Delta^{-1}}{(1 + \lambda^T R^{-1} \lambda)^{1/2}},$$

$$\Delta = \text{diag}\left(\left(1 - \delta_1^2\right)^{1/2}, \ldots, \left(1 - \delta_d^2\right)^{1/2} \right),$$

$$\lambda = (\lambda(\delta_1), \ldots, \lambda(\delta_d))^T,$$

and

$$\mathbf{\Omega} = \Delta(R + \lambda\lambda^T)\Delta.$$

Following Eqs. 1 and 2, we can write the $d-$dimensional skew–normal copula C as

$$C(u_1, \ldots, u_d) = SN_d\left(SN_1^{-1}(u_1), \ldots, SN_1^{-1}(u_d)\right),$$

where SN_d and SN_1^{-1} are the $d-$dimentional skew–normal distribution function and the inverse of the univariate skew–normal distribution function, respectively. The corresponding skew–normal copula density is given by

$$c(u_1, \ldots, u_d) = \frac{\partial^d}{\partial u_1 \cdots \partial u_d} C(u_1, \ldots, u_d) \tag{7}$$

$$= \frac{sn_d(\mathbf{z})}{sn_1(z_1) \cdots sn_1(z_d)}, \tag{8}$$

where $z_i = SN_1^{-1}(u_i)$ for $i = 1, \ldots, d$. Note that when $\lambda_1 = \cdots = \lambda_d = 0$, the skew–normal copula density will degenerate to a Gaussian copula density in Eq. 3.

For example, let us look in detail at the bivariate skew–normal copula. As mentioned in the paper of Azzalini and Dalla Valla (1996), the bivariate skew–normal density function with parameters $(\rho, \delta_1, \delta_2)$ is given by

$$f(x, y) = 2\phi_2((x, y), \omega)\, \Phi(\alpha_1 x + \alpha_2 y) \tag{9}$$

where ϕ_2 is the bivariate normal with $\mathbf{0}$ mean and correlation matrix with off diagonal element ω, and

$$\omega = \rho\sqrt{1-\delta_1^2}\sqrt{1-\delta_2^2} + \delta_1\delta_2$$

and ρ is the off diagonal element of the correlation matrix R in Eq. 6 when $d = 2$. Here α_1 and α_2 are given as

$$\alpha_1 = \frac{\delta_1 - \delta_2\omega}{\left\{(1-\omega^2)\left(1-\omega^2-\delta_1^2-\delta_2^2+2\delta_1\delta_2\omega\right)\right\}^{1/2}}$$

and

$$\alpha_2 = \frac{\delta_2 - \delta_1\omega}{\left\{(1-\omega^2)\left(1-\omega^2-\delta_1^2-\delta_2^2+2\delta_1\delta_2\omega\right)\right\}^{1/2}}.$$

The copula density based on this skew–normal distribution definition would be

$$c_{\rho,\delta_1,\delta_2}(u, v) = \frac{\phi_2((x, y), \omega)\, \Phi(\alpha_1 x + \alpha_2 y)}{2\phi(x)\Phi(\lambda(\delta_1)x)\ \phi(y)\Phi(\lambda(\delta_2)y)}$$

where

$$\lambda(\delta) = \frac{\delta}{\sqrt{1-\delta^2}}$$

and $x = SN_1^{-1}(u)$ and $y = SN_1^{-1}(v)$.

Figure 1, plots (c) to (h), show the contours of the bivariate skew–normal copula with different combinations of δ_1 and δ_2. As we can see, the skew–normal copula is able to cover a more general class of copula.

The only reference we can find to the skew–normal copula is the Master Thesis of Elmasri (2012). While the use of the skew–normal copula is quite different to ours, the basic structure is the same. We use infinite mixture models and MCMC, whereas Elmasri uses an EM algorithm and uses the copula to model longitudinal data. The skewed t copula, see for example Demarta and McNeil (2005), is well documented in the literature, but is, however, known to be difficult to estimate.

2.2 Mixtures of Skew–Normal Copulas

Our aim here is to construct a nonparametric copula density, c, by a mixture of multivariate skew–normal copulas as follows:

$$c(u_1, \ldots, u_d) = \sum_{j=1}^{\infty} w_j c_{R_j,\delta_j}(u_1, \ldots, u_d), \tag{10}$$

where $c_{R_j,\delta_j}(u_1, \ldots, u_d)$, for all j, are skew–normal copula densities, as in Eq. 7. Additionally, R_j is a correlation matrix defined in Eq. 6, $\delta_j = (\delta_{j1}, \ldots, \delta_{jd})$ is the skewness parameter vector and the weights, $(w_j,\ j = 1, 2, \ldots)$ are described below.

We use a stick–breaking prior for the weights and this can be based on the Dirichlet process; see Ferguson (1973). Hence, for $(v_j)_{j=1}^{\infty}$, which are independent and identically distributed from beta$(1, \xi)$, for some $\xi > 0$, we have $w_1 = v_1$ and, for $j > 1$,

$$w_j = v_j \prod_{l<j} (1 - v_l). \tag{11}$$

It is easy to show that $\sum_{j=1}^{\infty} w_j = 1$ a.s. A more general idea is to use $v_j \sim$ beta(a_j, b_j); see Ishwaran and James (2001). On the other hand, the prior for each R_j is based on the Wishart distribution; see Pitt et al. (2006). If a covariance matrix, Σ, has prior Wish(k, A), with degrees of freedom k, and the scale matrix A, the density is

$$\pi(\Sigma) = \frac{1}{2^{\frac{kd}{2}} |A|^{\frac{k}{2}} \Gamma_d(\frac{k}{2})} |\Sigma|^{\frac{k-d-1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}(A^{-1}\Sigma)\right\}, \tag{12}$$

where Γ_d is the multivariate gamma function. Then the relative correlation matrix R is given by $R = D\Sigma D$, where $D = \mathrm{diag}(1/e_1, \ldots, 1/e_d)$ and $e_j = \sqrt{\Sigma_{jj}}$.

We are now set to describe how to estimate the infinite mixture of skew–normal copula model which necessitates the implementation of an MCMC algorithm.

3 The MCMC Algorithm

First, for simplicity, we describe how to do inference for the single component skew–normal copula. Hence, we only need to demonstrate how to sample a single correlation matrix R and a skewness vector δ from the posterior, assuming all the data come from a single skew–normal copula model. After this we will adapt the algorithm to extend to the infinite mixture model.

3.1 Single Skew–Normal Copula Model

Here we describe the Metropolis–Hastings algorithm to sample the posterior of the correlation matrix and the skewness vector. The model is a single skew–normal density and we assume we observe data in 3 dimensions. So, for illustration, we take $k = 3$, $A = I_3$, and $d = 3$. At each iteration of a Metropolis–Hastings algorithm, a proposal density $q(\Sigma^*|\Sigma)$ is required which we take to be Wish$(3, \Sigma/3)$. The decision about whether we accept matrix Σ^* from this proposal density will be based on the acceptance ratio:

$$\alpha = \frac{c(\mathbf{u}|R^*(\Sigma^*)) \cdot \pi(\Sigma^*) \cdot q(\Sigma|\Sigma^*)}{c(\mathbf{u}|R(\Sigma)) \cdot \pi(\Sigma) \cdot q(\Sigma^*|\Sigma)},$$

where $c(\mathbf{u}|R(\Sigma))$ is the skew–normal copula density with the correlation matrix $R(\Sigma) = D\Sigma D$ and $\mathbf{u} = (u_1, u_2, u_3)$. Then we can construct a Metropolis–Hastings algorithm as follows:

Step 1: Choose an initial covariance matrix $\Sigma^{(0)} \sim \text{Wish}(3, I_3/3)$, then calculate the correlation matrix $R^{(0)}$.
Step 2: Sample the covariance matrix Σ^* from the proposal density $q(\Sigma^*|\Sigma^{(t)})$.
Step 3: Generate $\xi \sim \text{U}(0, 1)$.
Step 4: Set

$$R^{(t+1)} = \begin{cases} R^*, & \alpha > \xi; \\ R^{(t)}, & \alpha \le \xi. \end{cases}$$

Step 5: Increment t and repeat steps 2 through to 4.

To update $\delta = (\delta_1, \delta_2, \delta_3)$, we apply the Metropolis-Hastings algorithm on each δ_j in turn. The prior distribution of δ_j is assigned the Uniform distribution $\text{U}(-1+\eta, 1-\eta)$ for some small $\eta > 0$ and the proposal function for each δ is given by $f(\delta^\star|\delta) \sim \text{U}(\delta - \epsilon, \delta + \epsilon)$, where ϵ is a small constant. We now develop this basic algorithm to cover the infinite mixture model.

3.2 Mixture of Skew–Normal Copula Model

To be able to work on the infinite mixture model (Eq. 10), we implement a by now standard practice which is to introduce latent variables which make the infinite mixture to be a finite mixture. Following Kalli et al. (2011), we introduce a latent model which facilitates an MCMC algorithm for sampling from the posterior distribution. Two latent variables θ and κ, where $0 < \theta < 1$ and $\kappa \in \{1, 2, \ldots\}$, are introduced so that each observation is allocated to one component of the mixture model (Eq. 10). Hence, the infinite mixture model is replaced by a latent model given by

$$c(\mathbf{u}, \theta, \kappa) = \mathbf{1}(\theta < e^{-\kappa}) e^{\kappa}\, w_\kappa\, c_{R_\kappa, \delta_\kappa}(\mathbf{u}).$$

Integrating out θ and summing over κ returns the correct infinite mixture model. This effort of introducing the latent model is to create a likelihood without the infinite sum and therefore to ensure that one can sample the latent allocation variable κ, since it has, with θ, a finite selection.

Consequently, given data $(\mathbf{u}_1, \ldots, \mathbf{u}_n)$, the full likelihood function becomes

$$\prod_{i=1}^{n} \mathbf{1}(\theta_i < e^{-\kappa_i})\, e^{\kappa_i} w_{\kappa_i}\, c_{R_{\kappa_i}, \delta_i}(\mathbf{u}_i).$$

A Gibbs sampler is implemented through sampling the variables as discussed below.

The sampling of the latent variables $(\kappa_i, \theta_i)_{i=1}^n$ is straightforward. The θ_i is simulated from a uniform distribution between 0 and $e^{-\kappa_i}$ and then

$$\Pr(\kappa_i = j | \cdots) \propto \mathbf{1}(j < \lfloor -\log \theta_i \rfloor)\, e^j\, w_j\, c_{R_j,\delta_j}(u_i), \tag{13}$$

where $\lfloor -\log X \rfloor$ defines the largest integer less than or equal to X.

The weights (w_j) are updated through the sampling of the (v_j), and for any j

$$[v_j | \cdots] = \text{beta}(a_j + n_j, b_j + m_j),$$

where $n_j = \#\{\kappa_i = j\}$ and $m_j = \#\{\kappa_i > j\}$.

The conditionals for each R_j and δ_j are sampled, using Metropolis steps, as outlined in Section 3.1, but with only the data which has been allocated to component j. To sample from the predictive copula density, which will form the basis of our analysis, at each iteration t of the MCMC algorithm, we sample the weights (w_j). At iteration t we have

$$\left(w_1^{(t)}, \ldots, w_{J_t}^{(t)}\right)$$

for some finite J_t. If $j < J_t$ is picked then a sample from the predictive is taken from the skew–normal copula with correlation matrix R_j and skew parameter δ_j. If, on the other hand, it is designated that the component should come from a $j > J_t$, then R_j is sampled from the prior, as is δ_j, and then a sample from the predictive is taken from the skew–normal copula with parameter (R_j, δ_j). Further details of how such algorithms are implemented can be found in Kalli et al. (2011).

4 Illustrations

A number of examples of the proposed methodology are presented here. We use both simulated data and real data applications to illustrate the Gaussian and skew–normal copula models described in Section 2. The use of the Gaussian copula infinite mixture model runs into trouble with a real data set which does not exhibit the symmetry property mentioned in Section 2.

In Sections 3.1 and 3.2 the prior for the correlation matrix R_j is Wish(k, A) where $k = d$ and $A = I_d$.

4.1 Single Gaussian Copula Model

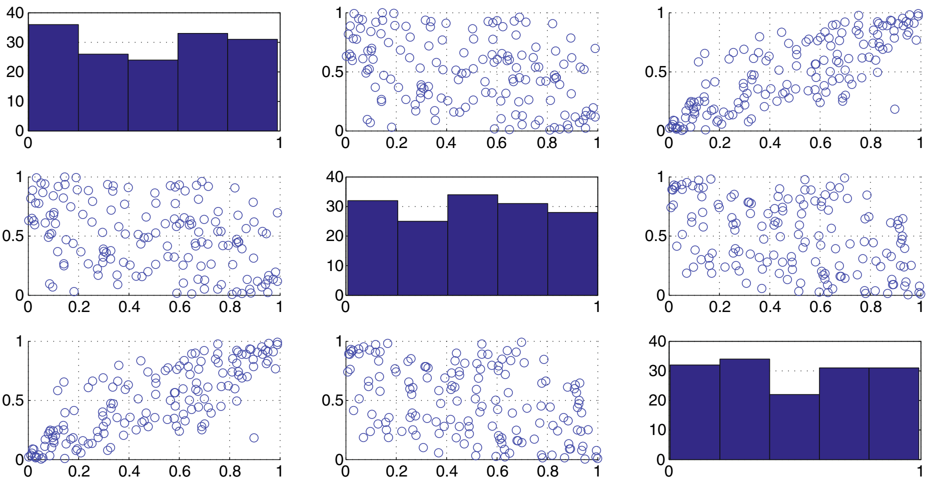
As a first example, we generated data from the Gaussian copula with the correlation matrix given by

$$R = \begin{pmatrix} 1 & -0.4 & 0.8 \\ -0.4 & 1 & -0.5 \\ 0.8 & -0.5 & 1 \end{pmatrix},$$

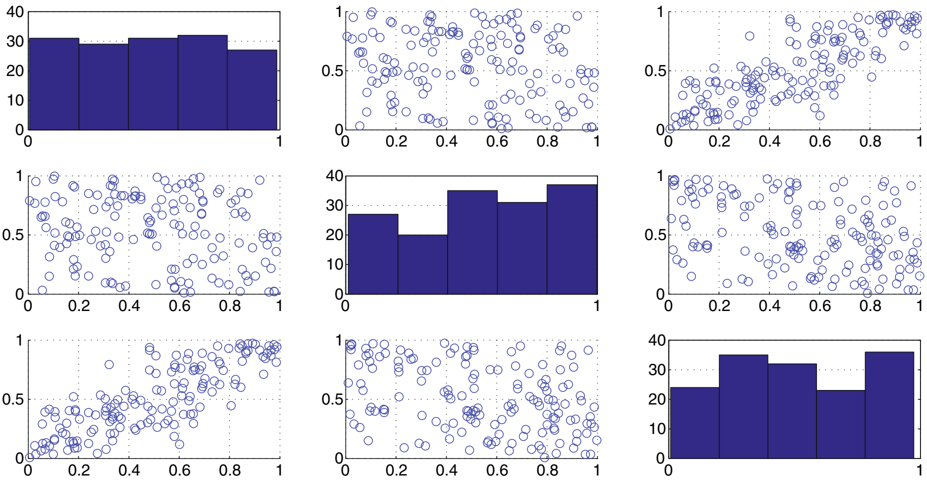
and with the sample size taken to be $n = 150$. The correlation matrices are sampled from the posterior distribution by the Metropolis–Hastings algorithm, described in Section 3.1, with δ set to 0. We subsample the chain, taking every 50th sample, to produce the output and thus 1,000 samples we collected in total based on a run length of 50,000 iterations. No burn–in was used. The Bayes estimate of the correlation matrix is the mean or sample average of the 1,000 sampled correlation matrices, evaluated as

$$\widehat{R} = \begin{pmatrix} 1 & -0.35 & 0.80 \\ -0.35 & 1 & -0.44 \\ 0.80 & -0.44 & 1 \end{pmatrix}.$$

Once we have $\widehat{R}$, we generate 150 samples from the Gaussian copula with $\widehat{R}$. Plots of the simulated true data and the data from $\widehat{R}$ are shown in Fig. 2. Figure 3 gives the



(a) simulated data



(b) predicted data

Fig. 2 Plots of data generated from single Gaussian copula model. **a** the simulated data; **b** the predictive data

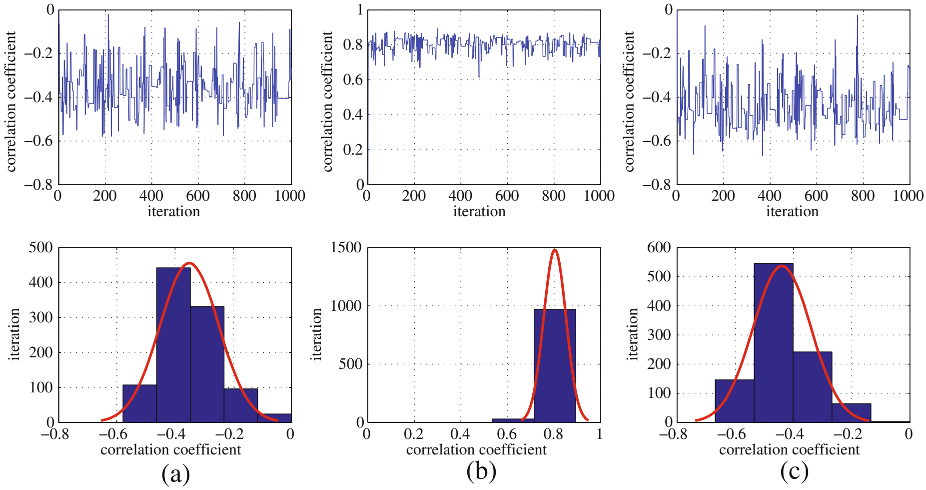


Fig. 3 *Top*: traces of every component **a** ρ_{12}, **b** ρ_{13}, **c** ρ_{23} of the correlation matrices; *bottom*: histograms of every component

trace plots and histograms of each of the components of the correlation matrix over the length of the Metropolis algorithm.

4.2 Mixture of Gaussian Copula Model

The simulation and real application examples are now presented to illustrate the approach for the multivariate mixture Gaussian copula model. The prior for the (v_j) is beta (a_j, b_j) where we take $a_j = 0.05$ and $b_j = 0.05$ in an attempt to be noninformative.

4.2.1 The Simulated Data

Here we consider an infinite mixture Gaussian copula model, with generated data from a 3 mixture model given by

$$c(u_1, u_2, u_3) = \sum_{j=1}^{3} w_j c_{R_j}(u_1, u_2, u_3),$$

where the weights are $w_1 = 0.25$, $w_2 = 0.55$, $w_3 = 0.2$, and the respective correlation matrices are

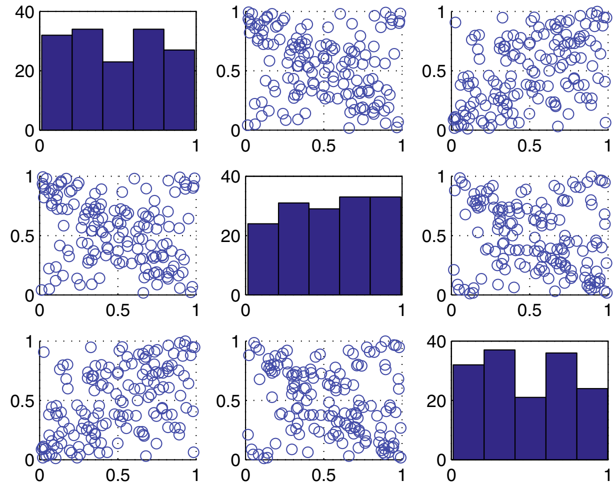
$$R_1 = \begin{pmatrix} 1 & 0.7 & 0.49 \\ 0.7 & 1 & 0.7 \\ 0.49 & 0.7 & 1 \end{pmatrix}, \quad R_2 = \begin{pmatrix} 1 & -0.9 & 0.81 \\ -0.9 & 1 & -0.9 \\ 0.81 & -0.9 & 1 \end{pmatrix},$$

and

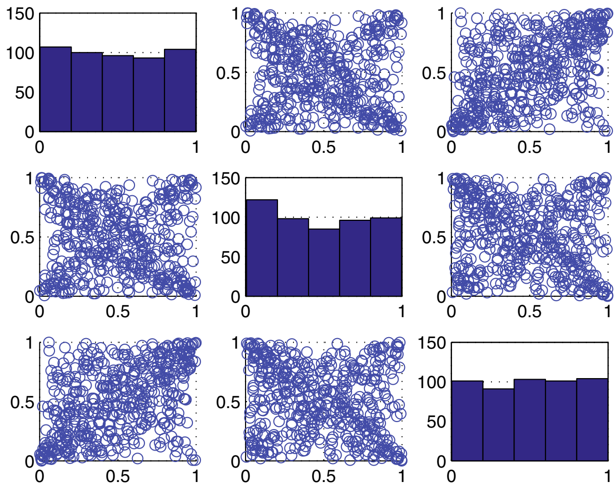
$$R_3 = \begin{pmatrix} 1 & 0.2 & 0.04 \\ 0.2 & 1 & 0.2 \\ 0.04 & 0.2 & 1 \end{pmatrix}.$$

We took a sample of size 150 and ran the chain described in Section 3.2, with δ set to 0, for 50,000 iterations. We use the last 25,000 samples, thinned to every 50th value, to provide the output–totally 500 values. Figure 4 illustrates the plots of the

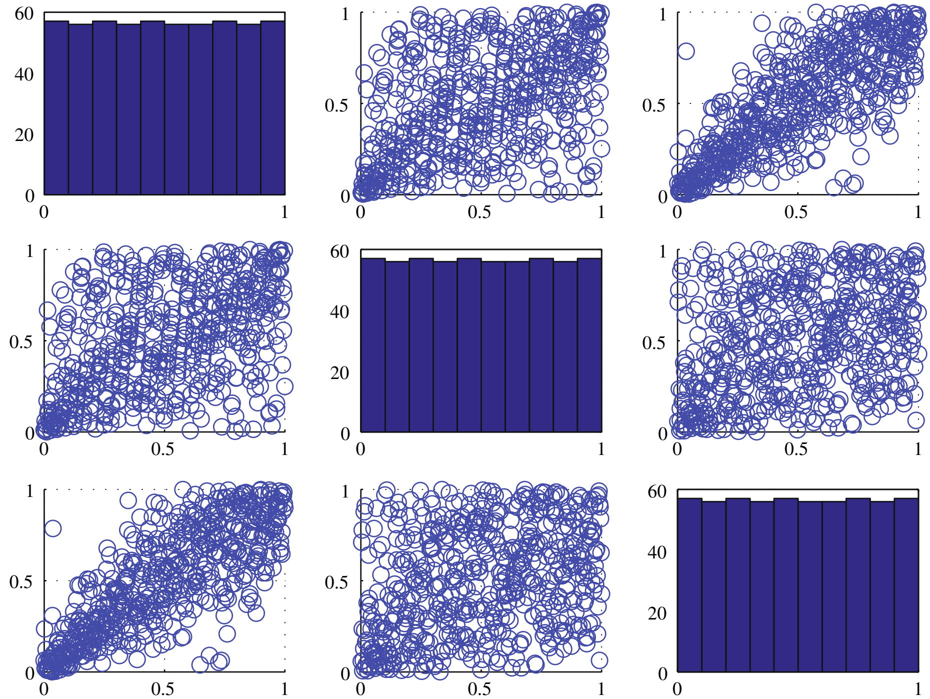
Fig. 4 Plots of data generated from the three mixture Gaussian copula model. **a** the simulated data; **b** the predictive data



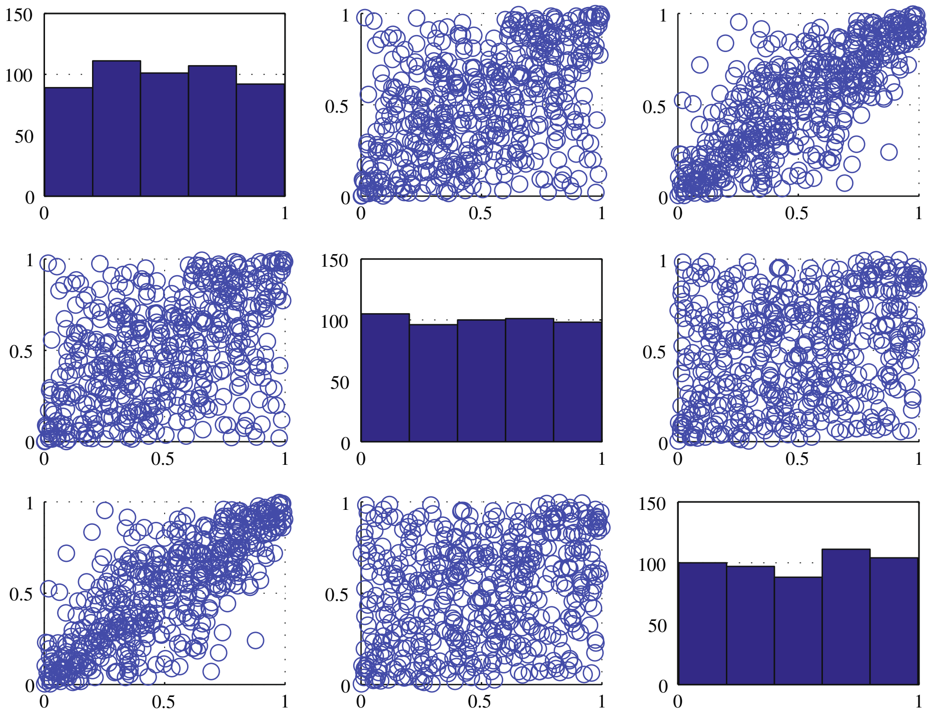
(a) simulated data



(b) predicted data



(a) real data



(b) predicted data

◀ **Fig. 5** Plots of the bike–time data. **a** the real data; **b** the predictive data using mixture of the Gaussian copula model

simulated data and the predicted data from this mixture model. It can be seen that the predictives are a good representation of the sampling density.

4.2.2 Bike Time Example

This real data set is about the number of bicycles traveling down the Main Yarra South Bank bike path in Melbourne, analyzed by Smith and Khaled (2012). There are 565 observations corresponding to the count of bicycles that have passed within each hour on the working days between 12 December 2005 and 19th June 2008, excluding weekends and special days. The first column is for the period 05:01–06:00, then hourly until the period 20:01–21:00. Smith and Khaled (2012) presented a dependence structure in the bivariate case with a parametric copula.

Triple peak times, 07:01–08:00, 09:01–10:00 and 16:01–17:00, are used here to illustrate our mixture model, can be extended to higher dimensional cases in a straightforward manner. The real data are transformed according to the strategy outlined in Section 1. We ran the MCMC as in Section 3 with the δ_j set to 0.

Figure 5 shows the scatter plots of the real data and the predictions which are in good agreement. The histogram of the number of the components in the mixture model is seen in Fig. 6. The samples are computed by determining the number of distinct κ_i at each iteration.

However, there is a clear problem with the asymmetry exhibited by one of the real data plots and the symmetry induced by the Gaussian copula. For this data we now use the skew–normal copula model to re-analyze the data. Instead of setting $\delta_j = 0$, we treat δ_j as a skewness parameter and sample it from the posterior distribution by the Metropolis–Hastings algorithm, described in Section 3.2. Figure 7 presents the scatter plot of the predictives using the mixture skew–normal copula model. We can see that the asymmetry presented by the dataset was correctly picked up and the predictives are a better representation of the real data.

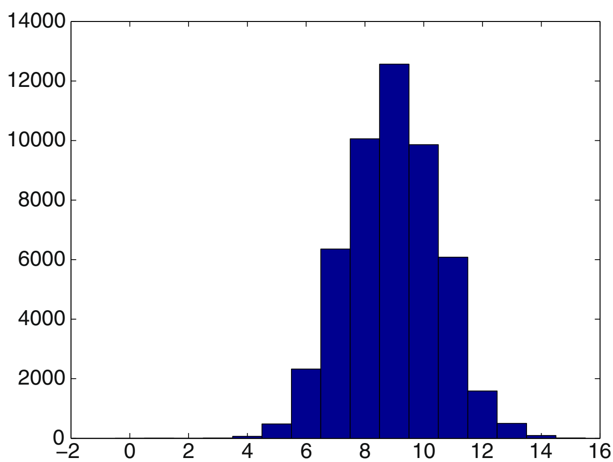


Fig. 6 Histogram of the number of the components in the mixture Gaussian copula model for the bike time data

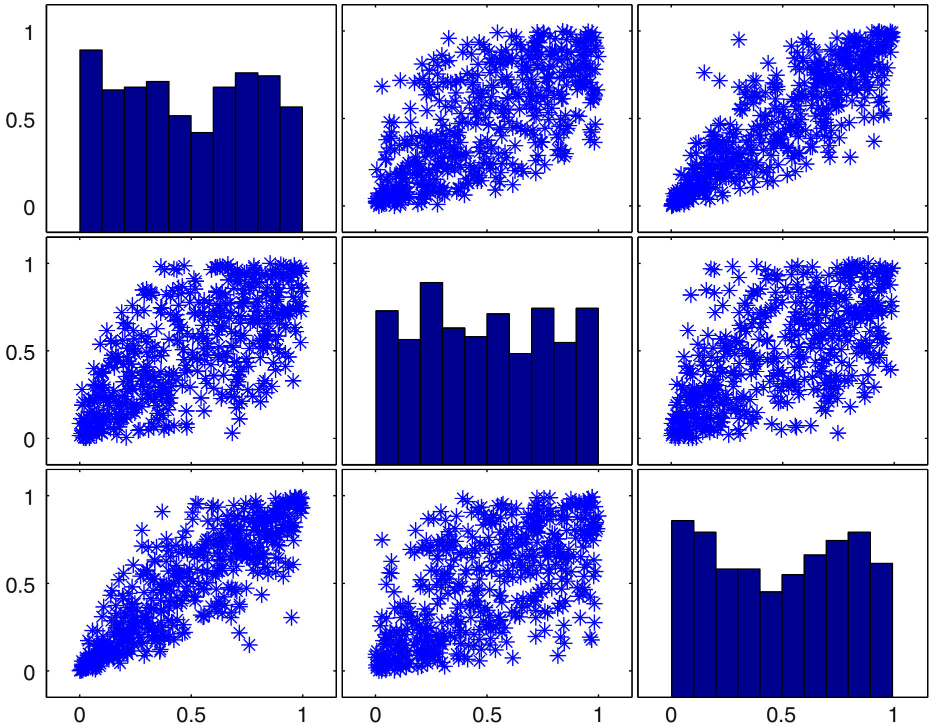


Fig. 7 Plots of the bike–time data: the predictive data using the mixture of skew–normal copula model

Acknowledgement The authors are grateful for the comments of a referee which enabled the paper to be significantly revised.

References

Azzalini A (1985) A class of distribution which includes the normal ones. Scand J Stat 12:171–178

Azzalini A, Dalla Valla A (1996) The multivariate skew-normal distribution. Biometrika 83:715–726

Demarta S, McNeil AJ (2005) The t copula and related copulas. Int Stat Rev (73):111–129

Diers D, Marek SD, Eling M (2012) Dependence modeling in non-life insurance using the bernstein copula. Insur Math Econ 50(3):430–436

Elmasri M (2012) A skew-normal copula driven generalized linear mixed model for longitudinal data. In: Master of Science in Mathematics, Dept. of Mathematics and Statistics, Concordia University, Montreal

Ferguson TS (1973) A bayesian analysis of some nonparametric problems. Ann Stat 1:209–230

Genest C, Favre AC (2007) Everything you always wanted to know about copula modeling but were afraid to ask. J Hydrol Eng 12:347–368

Genest C, Ghoudi K, Rivest LP (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika 82:543–552

Genest C, Rémillard B, Beaudoin D (2009) Goodness–of–fit tests for copulas: a review and a power study. Insur Math Econ 44(2):199–213

Ishwaran H, James LF (2001) Gibbs sampling methods for stick-breaking priors. J Am Stat Assoc 96(453):161–173

Kalli M, Griffin JE, Walker SG (2011) Slice sampling mixture models. Stat Comput 21(1):93–105

Kirshner S (2007) Learning with tree–averaged densities and distributions. In: NIPS
Nelsen R (2006) An introduction to copulas, 2nd. Springer, New York
Pitt M, Chan D, Kohn RJ (2006) Efficient Bayesian inference for Gaussian copula regression models. Biometrika 93(3):537–554
Silva R, Gramacy RB (2009) MCMC methods for Bayesian mixtures of copulas. J Mach Learn Res-Proc Track 5:512–519
Sklar A (1959) Fonctions de répartition à *n* dimensions et leurs marges. Publ Inst Stat Univ Paris 8:229–231
Smith M, Khaled M (2012) Estimation of copula models with discrete margins via data augmentation. J Am Stat Assoc 107:290–303
Tewari A, Giering MJ, Raghunathan A (2011) Parametric characterization of multimodal distributions with non–gaussian modes. In: Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on. pp 286–292
Dalla Valle L (2009) Bayesian copulae distributions, with application to operational risk management. Methodol Comput Appl Probab 11:95–115
Wu J, Wang X, Walker SG (2013) Bayesian nonparametric estimation of a copula. J Stat Comput Simul (accepted)