Bayesian Analysis (2023) **TBA**, Number TBA, pp. 1–28

Grid-Uniform Copulas and Rectangle Exchanges: Bayesian Model and Inference for a Rich Class of Copula Functions*

Nicolás Kuschinski† and Alejandro Jara‡

Abstract. Copula-based models provide a great deal of flexibility in modelling multivariate distributions, allowing for the specifications of models for the marginal distributions separately from the dependence structure (copula) that links them to form a joint distribution. Choosing a class of copula models is not a trivial task and its misspecification can lead to wrong conclusions. We introduce a novel class of grid-uniform copula functions, which is dense in the space of all continuous copula functions in a Hellinger sense. We propose a Bayesian model based on this class which posterior distribution is strongly consistent and develop an automatic Markov chain Monte Carlo algorithm for exploring the corresponding posterior distribution. The methodology is illustrated by means of simulated data and compared to the main existing approach.

Keywords: random probability distributions, Bayesian semiparametric modelling, association modelling, multivariate density estimation.

MSC2020 subject classifications: Primary 62G07, 62C10; secondary 62H99.

1 Introduction

One of the primary interests of statistical analysis of multivariate data is the study of how random variables relate to each other. Among the various ways to express the relationship between random variables, one of the more flexible ones is the use of a marginals-copula representation, which provides a way to separate multivariate distributions into their single variate marginals and a function which represents their association structure, the copula function (Nelsen, 2007; Smith, 2013; Joe, 2014). For any d-variate distribution $H : \mathbb{R}^d \to [0, 1]$ with $d > 1$ and marginals $F_1, F_2, \ldots, F_d$, a copula is a function $C : \mathbb{R}^d \to [0, 1]$ such that

$$H(\boldsymbol{x}) = C\big(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)\big),$$

where $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$. Sklar's theorem (Sklar, 1959; Faugeras, 2013; Nelsen, 2007) is a classical result which states that this function C exists for any multivariate

*N. Kuschinski's research is supported by ANID – Millennium Science Initiative Program – NCN17_059 and Fondecyt 3210553 grant. A. Jara's research is supported by ANID – Millennium Science Initiative Program – NCN17_059 and Fondecyt 1220907 grant.

†Department of Statistics, Pontificia Universidad Católica de Chile, and Center for the Discovery of Structures in Complex Data, Casilla 306, Correo 22, Santiago, Chile

‡Department of Statistics, Pontificia Universidad Católica de Chile, and Center for the Discovery of Structures in Complex Data, Casilla 306, Correo 22, Santiago, Chile, atjara@uc.cl

 https://doi.org/10.1214/23-BA1396

distribution H, and that if H is continuous, C is unique. Copula functions are themselves multivariate probability distributions supported on the unit hyper-cube and are such that the single variate marginals are uniform (Nelsen, 2007).

Modeling phenomena with copulas involves specifying models for marginals $F_1, F_2, \ldots, F_d$, and for $C(\boldsymbol{x})$ separately. Modeling $F_i(x_i)$ is a matter of modeling single variable probability distributions, which is a well understood topic and that has been studied with great depth in both frequentist and Bayesian contexts (see, e.g., Müller et al., 2015). Our primary interest here is the modeling of the copula function $C(\boldsymbol{x})$. There is a large body of literature studying parametric copula models arising from standard multivariate distributions such as the Gaussian and multivariate-t (Nelsen, 2007; Pitt et al., 2006; Joe, 2014; Choroś et al., 2010; Smith, 2013). Some popular copula models are members of the Archimedean family, which are single-parameter copulas satisfying

$$C(u_1, \ldots, u_d \mid \phi, \theta) = \phi^{-1}\big(\phi(u_1 \mid \theta) + \cdots + \phi(u_d \mid \theta) \mid \theta\big),$$

for a function $\phi(\cdot \mid \theta)$ with a single parameter, known as the generator function of the Archimedean copula Nelsen (2007), where ϕ^{-1} is its inverse function. Archimedean copulas have been used for both frequentist and Bayesian analyses in the literature (McNeil and Nešlehová, 2009; Genest et al., 2011; Kaewsompong et al., 2020).

Choosing a class of copula models is not a trivial task and constraining the inference to parametric copula models can lead to wrong conclusions, because it reduces our ability to represent relationships between random variables. Motivated by these facts, different flexible approaches have been discussed in the literature. Non-parametric approaches for discrete data have been attempted in numerous ways, as described in Genest and Nešlehová (2007) and Yang et al. (2019). In the context of continuous data, classical nonparametric approaches can be traced back to Deheuvels (1979), and can be found in Genest et al. (1995), Kauermann et al. (2013), Oh and Patton (2013), Oh and Patton (2017), and Mukhopadhyay and Parzen (2020). The classical methods commonly rely on the use of partial- or pseudo-likelihood, and do not allow for a proper quantification of the uncertainties associated to the lack of knowledge of the marginal distributions. Furthermore, these approaches cannot be employed for modelling the association structure of latent variables in the context of hierarchical models. Approaches based on penalized splines to model the copula function can be extended to allow for a proper uncertainty quantification about unknowns and to be included in to more general modelling frameworks. However, to the best of our knowledge, there are no formal guaranties that the class of copula functions generated by this approach is sufficiently rich to ensure an appropriate estimation of any copula function.

Flexible Bayesian approaches for copula functions can be found in Guillotte and Perron (2012), Wu et al. (2013a), Grazian and Liseo (2016), Wu et al. (2013b), and Ning and Shephard (2018). Guillotte and Perron (2012) proposed an interesting semi-parametric Bayesian approach for bivariate copulas based on a finite-dimensional approximation. This approximation is structured from a partition of the unit interval based of intervals of the same length $[(i-1)/m,\ i/m]$, where m is the number of intervals and $i \in \{1, \ldots, m\}$. Their proposal is constructed using this partition and indicator functions for the corresponding intervals. The density of the copula is constructed via

a mixture of the cross products of the indicator functions, resulting in a locally uniform distribution over the unit square, parameterized by a doubly stochastic matrix. Taking advantage of the properties of doubly stochastic matrices, the authors proceed to develop both a conjugate Jeffreys prior and a Markov chain Monte Carlo (MCMC) algorithm to sample from the posterior distribution. The approach is flexible but is difficult to generalize to not equally-spaced partitions and higher dimensions because of its reliance on the properties of doubly stochastic matrices.

Wu et al. (2013a) proposed a model based on mixtures of Gaussian copulas. A Gaussian copula function is given by

$$C(u_1 \dots u_d) = \Phi_{\boldsymbol{R}}\big(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)\big),$$

where $\Phi_{\boldsymbol{R}}$ is the CDF of a d-variate normal distribution with mean zero, variance one and covariance matrix $\boldsymbol{R}$, arising from the corresponding correlation function, and Φ is the CDF of a standard normal distribution function. The authors claim that mixtures of bivariate Gaussian copulas can approximate any continuous bivariate copula function. Unfortunately, the Gaussian copula kernel is not rich enough to form a dense class. For instance, the density function of a bivariate Gaussian copula has the property that $c_{\boldsymbol{R}}(u_1, u_2) = c_{\boldsymbol{R}}(u_2, u_1)$. Therefore, the density of a mixture of arbitrarily many Gaussian copulas also has this feature and, thus, cannot approximate an asymmetrical copula function, such as the asymmetrical t copula described by Church (2012). Wu et al. (2013b) extended the idea of a mixture of copula functions to the class of skew-normal copulas. However, there is no evidence that this class of copula functions is dense in the space of all copula functions.

Grazian and Liseo (2016) proposed an estimation method for a functional of a multivariate distribution using a copula representation. The prior on the copula is constructed by eliciting the prior on the functionals of interest and it is implied for the remaining elements of the copula function by the use of the exponentially tilted empirical likelihood. The main advantage of their approach is the ease of elicitation, because it does not require the prior elicitation about all aspects of the multivariate dependence structure. However, it does not allow for posterior inferences about all aspects of the multivariate dependence structure. Finally, Ning and Shephard (2018) employed Dirichlet-based Polya trees models to propose a fully non-parametric Bayesian approach to modeling copula functions in any number of dimensions and a method for conjugate posterior simulation from the resulting posterior. This attractive result is, however, significantly marred by the flaw that the simulations from the posterior distribution are not themselves copula functions.

In this paper, we introduce a novel class of grid-uniform copula functions, which is dense in the space of all continuous copula functions. We propose a Bayesian model based on this class that has appealing theoretical properties regarding support and posterior consistency, and develop an automatic MCMC algorithm for exploring the corresponding posterior distribution, allowing for the flexible modelling of continuous joint distributions. The paper is organized as follows. In Section 2 we introduce the class of grid-uniform copula functions and state its main properties, including its ability to approximate any given continuous copula function. In Section 3 we propose a Bayesian

model based on the class of grid-uniform copulas, study the behavior of the posterior distribution under random sampling, and describe the MCMC algorithm. In Section 4, we illustrate the behavior of the model by means of analyses of simulated data. A final discussion concludes the article. The proofs of the theoretical results can be found in Supplementary Material (Kuschinski and Jara, 2023).

2 Grid-Uniform Copulas

2.1 Definition

We begin by defining the class of copulas on which we develop our proposal. Let ρ be an orthogonal grid of $[0,1]^d$. Specifically, let ρ_i be an ordered collection of points in $[0,1]$, and set $\rho = \rho_1 \times \cdots \times \rho_d$, such that $\mathbf{1}_d \in \rho$. Let ν^ρ be the collection of sets formed by ρ, which are indexed by their upper right (or higher dimensional equivalent) corner. Now, let F be a probability measure defined on an appropriate space and B a measurable set such that $F(B) > 0$. We denote by $F\,|_B$ to the restriction of F to B defined by $F_{|_B}(A) = F(A \mid B) = F(A \cap B)/F(B)$. A probability distribution F on $[0,1]^d$ is said to be *ρ-uniform*, if for each set $B \in \nu^\rho$, such that $F(B) > 0$, the restriction of F to the set B, $F_{|_B}$, is uniform on B.

Definition 1 (Grid-uniform copula)**.** Let ρ be a grid on $[0,1]^d$. A distribution C on $[0,1]^d$ is a ρ-uniform copula if it is ρ-uniform and its one-dimensional marginal distributions are uniform.

A grid-uniform copula can be completely described by specifying the grid ρ and the probabilities for every $B \in \nu^\rho$. Hence, for each grid ρ, the space of grid-uniform copulas over this grid are a compact finite dimensional domain. For a grid ρ and a distribution C, we will use C_ρ to denote the grid-uniform version of C, which assigns to each set $B \in \nu^\rho$ the probability assigned by C to that set, i.e., $C_\rho(B) = C(B)$ for every $B \in \nu^\rho$. It is easy to see that if C is a continuous copula, then C_ρ is also a grid-uniform copula and that the CDFs of C_ρ and C coincide for every $\boldsymbol{y} \in \rho$.

2.2 Richness of Grid-Uniform Copulas

We now prove that the class of grid-uniform copulas is sufficiently rich to approximate any arbitrary continuous copula function.

Theorem 1. *Let C be an arbitrary copula which is absolutely continuous with respect to Lebesgue measure. Then for every $\epsilon > 0$, there exists a grid-uniform copula D such that the Hellinger distance between D and C is smaller than ϵ, $\mathcal{H}(D, C) < \epsilon$.*

None of the steps in the proof actually make use of the fact that C has uniform marginals, and it can easily be extended to prove that grid-uniform distributions can approximate any continuous distribution over a rectangular support.

2.3 Measures of Association

In this section we provide the expression for two important measures of dependence for grid-uniform copula models. Specifically, we provide the expression for Kendall's tau and Spearman's rho, which are considered the best alternatives to the linear correlation coefficient as a measure of dependence for non-elliptical distributions, for which the linear correlation coefficient is inappropriate and often misleading.

Let C be a ρ-uniform copula function. Let $C^{(i,j)}$ be the bivariate marginal copula of C for the variables in the coordinates i and j. Let $a_{(i,k)}$, $k = 0, \ldots, m_i$, be the kth element in ρ_i and $b_{(j,l)}$, $l = 0, \ldots, m_j$, be the lth element in ρ_j. It is straightforward to show that $C^{(i,j)}$ is a $\rho^{(i,j)}$-uniform copula function, where $\rho^{(i,j)} = (\rho_i, \rho_j)$ and

$$C^{(i,j)}((a_{(i,k-1)}, a_{(i,k)}] \times (b_{(j,l-1)}, b_{(j,l)}]) = \sum_{B \in S_{i,j}^{(a_{(i,k)}, b_{(j,l)})}} C(B),$$

where $S_{i,j}^{(a_{(i,k)}, b_{(j,l)})}$ is the collection of sets in ν^ρ such that ith coordinate of the index of the set is $a_{(i,k)}$ and the jth coordinate of the index of the set is $b_{(j,l)}$. Spearman's rho, β, and Kendall's tau, τ, for the variables in the coordinates i and j is given by

$$\beta = 3 \sum_{k=1}^{m_i} \sum_{l=1}^{m_j} \big(a_{(i,k)}^2 - a_{(i,k-1)}^2\big)\big(b_{(j,l)}^2 - b_{(j,l-1)}^2\big) c_{k,l} - 3,$$

and

$$\tau = 4 \sum_{k=1}^{m_i} \sum_{l=1}^{m_j} \big(a_{i,k}^2 - a_{(i,k-1)}^2\big)\big(b_{(j,l)}^2 - b_{(j,l-1)}^2\big) c_{k,l}^2 - 1,$$

respectively, where

$$c_{k,l} = \frac{C^{(i,j)}((a_{i,k-1}, a_{i,k}] \times (b_{i,l-1}, b_{(i,l)}])}{(a_{(i,k)} - a_{(i,k-1)})(b_{(j,l)} - b_{(j,l-1)})}.$$

2.4 Rectangle Exchanges

We now introduce a class of transformations on grid-uniform copulas referred to as rectangle exchanges, which have the following important properties: (i) a rectangle exchange on a ρ-uniform copula produces another ρ-uniform copula, and (ii) given a grid ρ, and two ρ-uniform copulas C and D, there is a finite sequence of rectangle exchanges which can transform C into D.

Definition 2. Let ρ be a grid on $[0,1]^d$ and C be a grid ρ-uniform copula function. The function C^* is the result of a rectangle exchange of C, if C^* is constructed using the following steps:

(1) Set $C^* = C$, and pick i and j in the set $\{1, \ldots, d\}$, such that $i < j$ and the cardinality of ρ_i and ρ_j is greater than or equal to 2. Also, for all $k \in \{1, \ldots, d\} \setminus \{i, j\}$ pick point $x_k \in \rho_k$.

(2) Pick $a_1, a_2 \in \rho_i$ and $b_1, b_2 \in \rho_j$.

(3) Set

$$\boldsymbol{p}_{(a_l,b_m)} = (x_1, \ldots, x_{i-1}, a_l, , x_{i+1}, \ldots, x_{j-1}, b_m, x_{j+1}, \ldots, x_d),$$

where $l, m \in \{1, 2\}$.

(4) Pick some ϵ in the interval

$$\big[\max\big\{-C\big(\nu^{\rho}_{\boldsymbol{p}_{(a_1,b_2)}}\big), -C\big(\nu^{\rho}_{\boldsymbol{p}_{(a_2,b_1)}}\big)\big\}, \min\big\{C\big(\nu^{\rho}_{\boldsymbol{p}_{(a_1,b_1)}}\big), C\big(\nu^{\rho}_{\boldsymbol{p}_{(a_2,b_2)}}\big)\big\}\big].$$

(5) Set

$$C^*\big(\nu^{\rho}_{\boldsymbol{p}_{(a_1,b_1)}}\big) = C\big(\nu^{\rho}_{\boldsymbol{p}_{(a_1,b_1)}}\big) - \epsilon, C^*\big(\nu^{\rho}_{\boldsymbol{p}_{(a_1,b_2)}}\big) = C\big(\nu^{\rho}_{\boldsymbol{p}_{(a_1,b_2)}}\big) + \epsilon,$$
$$C^*\big(\nu^{\rho}_{\boldsymbol{p}_{(a_2,b_1)}}\big) = C\big(\nu^{\rho}_{\boldsymbol{p}_{(a_2,b_1)}}\big) + \epsilon, \quad \text{and} \quad C^*\big(\nu^{\rho}_{\boldsymbol{p}_{(a_2,b_2)}}\big) = C\big(\nu^{\rho}_{\boldsymbol{p}_{(a_2,b_2)}}\big) - \epsilon.$$

The rectangle exchange operation is illustrated in two-dimensions in Figure 1. Rectangle exchanges are a closed operation on grid-uniform copulas since they conserve the uniformity of all marginals.

Lemma 1. *Let G be ρ-uniform copula and $\mathcal{G}$ the resulting rectangle exchange of G. Then $\mathcal{G}$ is also a ρ-uniform copula.*

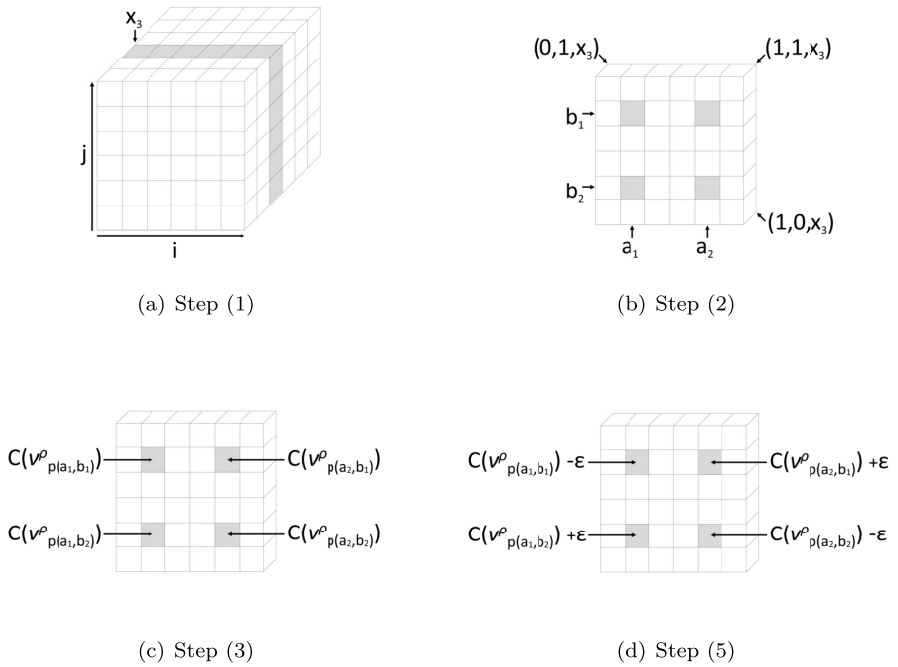
Another interesting property of grid-uniform copula functions is that starting from an uniform distribution on $[0, 1]^d$, it is possible to reach any given grid-uniform copula, say C_0, by doing certain type of operations. Specifically for a given grid-uniform copula C, we will refer as a *grid division* on C, to the addition of a division along any of the coordinates, such that the sets that are not divided retain their probabilities, and those that are divided distribute their probability in proportion to their volume. In other words, the resulting copula function arising from a grid division of C is identical to C, but is mapped onto a more refined grid.

Lemma 2. *Let U be a uniform distribution in $[0, 1]^d$, and let C be an arbitrary grid-uniform copula function. Then, there is a finite sequence of grid divisions and rectangle exchanges which will transform U into C.*

These results allow us to prove that via rectangle exchanges is possible to generate the full space of grid-uniform copula functions.

Theorem 2. *The C_1 and C_2 be two ρ-uniform copulas. There is a finite sequence of rectangle exchanges to transform C_1 into C_2.*

There is something surprising going on here. Intuition would lead us to believe that rectangle exchanges would work for 2 dimensions, whereas higher dimensions would require parallelepiped exchanges. However, the surprising fact is that the above theorem is true regardless of the number of dimensions. In essence, the apparently very complex problem of exploring the space of grid-uniform copulas is solved in any number of dimensions by repeated transformations of the sort illustrated in Figure 1.



(a) Step (1)

(b) Step (2)

(c) Step (3)

(d) Step (5)

Figure 1: Rectangle exchange – Illustration of a rectangle exchange for a ρ-uniform bivariate copula function C, where ρ_1 and ρ_2 have 7 equally-spaced points. Panel (a) illustrates step (1), where i and j are picked from $\{1, \ldots, d\}$, and x_3 is picked from ρ_3. Panel (b) illustrates step (2), where $i, j \in \{1, \ldots, d\}$ and $b_1, b_2 \in \rho_j$ are picked and where the rectangles to be exchanged are selected. Panel (c) illustrates step (3), showing the assigned sets and the mass assigned to them by C. Panel (d) illustrates step (5), where the mass of the new copula function C^* is computed.

3 Bayesian Modeling and Inference Using Grid-Uniform Copulas

Our ultimate objective is to use grid-uniform copulas to perform Bayesian statistical inference. We state the components of the Bayesian model in this section. Assume that we observe an independent and identically distributed (i.i.d.) sample of size n from a d-variate continuous distribution H, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n \mid H \overset{i.i.d.}{\sim} H$, where $H(\boldsymbol{y}) = C(F_1(y_1), F_2(y_2), \ldots, F_d(y_d))$, with $F_1, \ldots, F_n$ being the marginal distributions of H, and C is the corresponding copula function. We model C as a grid-uniform copula function. Under the grid-uniform copula model, the log-likelihood function is given by:

$$\begin{aligned}
&\ell(C, F_1, \ldots, F_d \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) \\
&\quad = \sum_{i=1}^{n} \sum_{j=1}^{d} \log\big(f_j(y_{ij})\big) + \sum_{i=1}^{n} \sum_{j=1}^{|\nu^\rho|} \log\left(\frac{C_\rho(B_j)}{\lambda(B_j)}\right) \times I_{\{(F_1(y_{i1}),\ldots,F_d(y_{id}))\in B_j\}}(\boldsymbol{y}_i), \quad (1)
\end{aligned}$$

where $B_j \in \nu^\rho$, $|\nu^\rho|$ is the cardinality of ν^ρ, $\lambda(B)$ is the Lebesgue measure of the set B, $I_A(B)$ is the indicator function that takes the value 1 if $B \in A$, and 0 otherwise.

3.1 Grid-Uniform Prior Models

Let ρ be a grid on $[0,1]^d$. Let C_0 be an arbitrary reference copula function and $\alpha > 0$. We propose prior models for grid-uniform copula functions of the form

$$\pi(C \mid \rho, \alpha, C_0) \propto \exp\left\{-\frac{1}{2}\alpha \times \mathcal{D}(C, C_0)\right\} \times I_{\mathcal{C}_\rho}(C),$$

where $\mathcal{D}$ be a suitable distance for probability distributions, and $\mathcal{C}_\rho$ is the space of ρ-uniform copulas. Many choices for $\mathcal{D}$ could be considered. One choice that provides a simple interpretation of the hyper-parameters is the squared-L_2 distance. Let c_0 and c be densities for C_0 and C, respectively. Let $B_1, \ldots, B_p$ be the sets included in ν^ρ. Under the squared-L_2 distance the grid-uniform prior model is given by

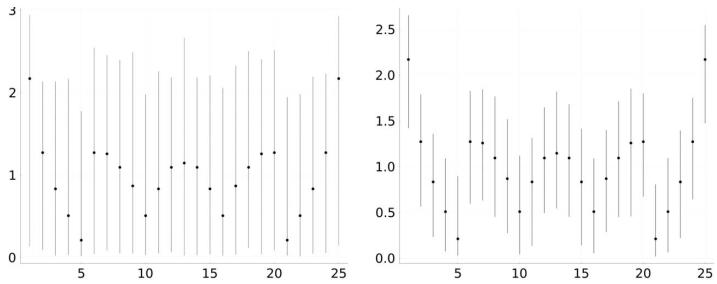
$$\begin{aligned}
\pi(C \mid \rho, \alpha, C_0) &\propto \exp\left\{-\frac{\alpha}{2} \times \int_{[0,1]^d} \big(c(\boldsymbol{x}) - c_0(\boldsymbol{x})\big)^2 d\boldsymbol{x}\right\} \times I_{\mathcal{C}_\rho}(C), \\
&= \exp\left\{-\frac{\alpha}{2} \times \sum_{l=1}^{|\nu^\rho|}\left[\int_{B_l} \big(c(\boldsymbol{x}) - c_0(\boldsymbol{x})\big)^2 d\boldsymbol{x}\right]\right\} \times I_{\mathcal{C}_\rho}(C), \\
&\propto \exp\left\{-\frac{\alpha}{2} \times \sum_{l=1}^{|\nu^\rho|}\left[-2c_l \int_{B_l} c_0(\boldsymbol{x}) d\boldsymbol{x} + c_l^2\right]\right\} \times I_{\mathcal{C}_\rho}(C), \qquad (2)
\end{aligned}$$

where $c_l = \frac{\int_{B_l} c(\boldsymbol{x}) d\boldsymbol{x}}{\lambda(B_l)}$, with $\lambda(A)$ being the Lebesgue measure of the set A. The prior takes the form of a truncated $|\nu^\rho|$-variate normal random distribution, centered at the ρ-uniform version of C_0, $C_{0,\rho}$, and precision matrix given by $\alpha \times \mathbf{I}_{|\nu^\rho|}$.

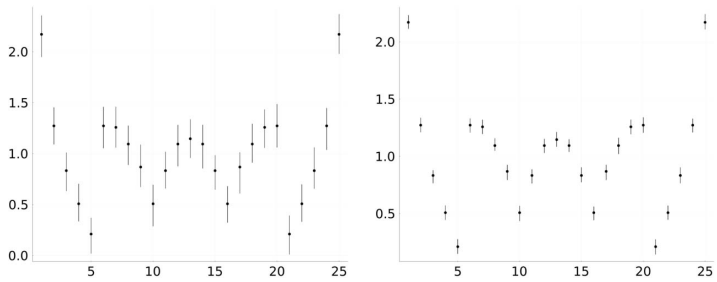
$C_{0,\rho}$ plays the role a centering parameter under the proposed prior and corresponds to the prior mode. On the other hand, α plays the role of a precision parameter, since as $\alpha \to +\infty$, the prior variance $var(\pi(C|\rho, \alpha, C_0)) \to 0$. Figure 2 illustrates the role of the parameters of the prior model. The figure displays the value of the center and a credible interval of the prior distribution of the copula density at different points of the sample space.

The impact of α scales with the size of the sets in the grid, meaning that for fine grids, α may have to be very large. To facilitate the prior elicitation process we consider the parameterization $\alpha^\star = \frac{\alpha}{|\nu^\rho|}$. For purposes of calculation, we note that using $C_{0,\rho}$ instead of C_0 as a reference copula produces the exact same prior, and calculating the prior using $C_{0,\rho}$ is much simpler, since its value is constant throughout each cell in the grid.

Assigning $\mathcal{D}$ to the squared-L_2 norm provides nicely interpretable parameters. However, it does not allow for the incorporation of prior information on the degree of smoothness of the copula function. For continuous copulas, it is often reasonable to expect that



(a) $\alpha^\star = 2$. (b) $\alpha^\star = 20$.



(c) $\alpha^\star = 200$. (d) $\alpha^\star$ =2,000.

Figure 2: Grid-uniform prior model – Center value (circle) and 95% equal-tail credible interval (vertical line) of the copula density evaluated at 25 equidistant points of the bivariate sample space. The horizontal axis is merely the index of each of the 25 equidistant points, and the vertical axis is the value of the copula density at each point. The results are shown for different values of $\alpha^\star = \frac{\alpha}{|\nu^\rho|}$. In all cases, the centering copula model, C_0, is a bivariate Gaussian copula function with correlation equals to 0.5. Panel (a), (b), (c), and (d), provides the results for $\alpha^\star = 2$, 20, 200, and 2,000, respectively.

the value of a copula density at a point is similar to the value at nearby points. Let $\mathcal{V}^\rho$ be a set containing the elements of ν^ρ in a given order. Let $\boldsymbol{W}$ be a symmetric matrix in which each entry $\boldsymbol{W}_{i,j}$ encodes information about the spatial relationship of the sets $\mathcal{V}_i^\rho$ and $\mathcal{V}_j^\rho$. Finally, let $\boldsymbol{D}_W$ be a diagonal matrix with $\boldsymbol{D}_{W_{i,i}} = \sum_{j=1}^{|\nu^\rho|} \boldsymbol{W}_{i,j}$. Borrowing ideas from models commonly used in spatial statistics and the nature of the grid-uniform model, we propose to take

$$\mathcal{D}(C, C_0) = \sum_{i=1}^{|\nu^\rho|} \sum_{j=1}^{|\nu^\rho|} (\boldsymbol{D}_W - \gamma \boldsymbol{W})_{i,j} \int_{\mathcal{V}_i^\rho} \big(c(\boldsymbol{x}) - c_0(\boldsymbol{x})\big) d\boldsymbol{x} \int_{\mathcal{V}_j^\rho} \big(c(\boldsymbol{y}) - c_0(\boldsymbol{y})\big) d\boldsymbol{y},$$

where $\gamma > 0$. Under this distance, the grid-uniform prior is given by

$$\begin{aligned} \pi(C \mid \rho, \alpha, \boldsymbol{W}, \gamma, C_0) &\propto \exp\left\{ -\frac{\alpha}{2} \mathcal{D}(C, C_0) \right\} \times I_{\mathcal{C}_\rho}(C), \\ &= \exp\left\{ -\frac{\alpha}{2} \overrightarrow{C - C_0}^T (\boldsymbol{D}_W - \gamma \boldsymbol{W}) \overrightarrow{C - C_0} \right\} \times I_{\mathcal{C}_\rho}(C), \end{aligned} \tag{3}$$

where $\overrightarrow{C - C_0}$ is the vector representation of $\{\int_B (c(\boldsymbol{x}) - c_0(\boldsymbol{x}))d\boldsymbol{x} : B \in \nu^\rho\}$ corresponding to the order induced by $\mathcal{V}^\rho$. Notice that the proposed prior corresponds to a truncated Gaussian conditional autoregressive (CAR) model, which allows for spatial correlation of nearby values, as specified by a smoothing parameter $\gamma > 0$. It is worth noting that $\mathcal{D}$ is only a distance under certain conditions on $\boldsymbol{W}$ and γ (see, e.g., Banerjee et al., 2014).

A popular option is to set $\boldsymbol{W}$ such that $\boldsymbol{W}_{i,j} = \frac{1}{d_{ij}}$, where d_{ij} is the distance between the centroids of $\mathcal{V}_i^\rho$ and $\mathcal{V}_j^\rho$ (Schmidt and Nobre, 2018). Another option is to set $\boldsymbol{W}$ as an adjacency matrix, where $\boldsymbol{W}_{i,j} = 1$ if the sets $\mathcal{V}_i^\rho$ and $\mathcal{V}_j^\rho$ are grid-neighbors (in the usual intuitive sense), and $\boldsymbol{W}_{i,j} = 0$ otherwise. For practical purposes, the sparseness of the adjacency matrix produces a prior which is faster to compute, and is a reasonable choice under most circumstances. When $\boldsymbol{W}$ is the adjacency matrix, then $(\boldsymbol{D}_W)_{i,i} = |N_{\mathcal{V}_i^\rho}|$, where N_B is the collection of grid-neighbors of the set B and the distance reduces to the following expression

$$\begin{aligned}\mathcal{D}(C, C_0) = &\sum_{B \in \mathcal{V}^\rho} |N_B| \left(\int_B \big(c(\boldsymbol{x}) - c_0(\boldsymbol{x})\big) d\boldsymbol{x} \right)^2 \\ &- \gamma \sum_{B \in \mathcal{V}^\rho} \left(\int_B \big(c(\boldsymbol{y}) - c_0(\boldsymbol{y})\big) d\boldsymbol{y} \sum_{A \in N_B} \int_A \big(c(\boldsymbol{z}) - c_0(\boldsymbol{z})\big) d\boldsymbol{z} \right). \end{aligned} \tag{4}$$

Expressions for $\mathcal{D}$ as described above are not distances for all values of γ. To force $\mathcal{D}$ to be a distance, we can pick $\gamma \in (\frac{1}{\lambda_1}, \frac{1}{\lambda_n})$, where λ_1 and λ_n are the smallest and largest eigenvalues of $\boldsymbol{D}_W^{-1/2} W \boldsymbol{D}_W^{-1/2}$. As discussed by Banerjee et al. (2014, section 6.4.3.3), the spatial correlation is low unless γ is close to 1. Because of this, a popular alternative is to consider $\gamma = 1$, which is known as the Intrinsic CAR (ICAR) model. Under the ICAR, $\mathcal{D}$ is given by

$$\mathcal{D}(C, C_0) = \sum_{i=1}^{|V^\rho|} \sum_{j=1}^{|\mathcal{V}^\rho|} W_{i,j} \left(\int_{\mathcal{V}_i^\rho} \big(c(\boldsymbol{x}) - c_0(\boldsymbol{x})\big) d\boldsymbol{x} - \int_{V_j^\rho} \big(c(\boldsymbol{y}) - c_0(\boldsymbol{y})\big) d\boldsymbol{y} \right)^2.$$

In addition, when $\boldsymbol{W}$ is the adjacency matrix, the distance reduces to

$$\mathcal{D}(C, C_0) = \sum_{B \in \nu^\rho} \sum_{A \in N_B} \left(\int_B \big(c(\boldsymbol{x}) - c_0(\boldsymbol{x})\big) d\boldsymbol{x} - \int_A \big(c(\boldsymbol{y}) - c_0(\boldsymbol{y})\big) d\boldsymbol{y} \right)^2.$$

In general, when $\gamma = 1$ the latter expression is not a distance, but only to a pseudo-metric, since adding a constant to either c or c_0 does not change the value of $\mathcal{D}(C, C_0)$. However, since C and C_0 are both restricted to the space of ρ-uniform copulas, $\mathcal{D}$ does define a distance on the corresponding domain. Finally, note that all of the prior models share the algebraic structure of a truncated Gaussian distribution centered at C_0^ρ. In fact, all but the ICAR are exactly truncated Gaussian distributions. Therefore, the interpretation of C_0 and α remains intact.

3.2 On the Choice of Hyper-Parameters

The prior depends on the choice of the grid ρ. The grid plays an equivalent role to the knots in the context of nonparametric regression based on splines. Rather than attempting to optimize the choice of a grid of reduced size and "well" located divisions, here we follow the approach proposed by Eilers and Marx (1996) in the context of penalized spline regression. Specifically, we consider an equally spaced and fine grid, along a penalization induced by an ICAR model, given in expression (3). The precise spacing of the grid can be chosen in relation to the available computational resources. We have found that on mid-range modern hardware (a 3.9 GHz processor) a good posterior estimate for a 50×50 grid (2401 free sets in the grid) can be computed in about 24 hours, whereas for a 10×10 grid (81 free sets) a posterior estimate can be computed in about 2 minutes. This is due not only to the greater computational cost of calculating the prior and likelihood functions, but also since a larger number of MCMC iterations are required.

The parameters $\alpha^\star$ and C_0 have clear interpretations, and when prior information is available, it can be used to inform their choice. For situations when such information is not readily available, we propose suitable defaults. In this setting, selecting a single default C_0 around which to center the prior is difficult because, once specified, a single centering distribution may affect inference unduly. For instance, the use of the independent copula is highly informative because the lack of dependence is itself an extreme form of association structure. Rather than selecting a single centering copula function, one option is to consider a mixture of grid-uniform copula models by allowing the parameters of the centering copula function to be random. One possible choice is to use the Gaussian copula family given by

$$C_{0,\boldsymbol{R}}(\boldsymbol{x}) = \Phi_{\boldsymbol{R}}\big(\Phi^{-1}(x_1), \Phi^{-1}(x_2), \ldots, \Phi^{-1}(x_d)\big),$$

parametrized by the correlation matrix $\boldsymbol{R}$.

Choosing a prior for $\boldsymbol{R}$ is delicate since $\pi(C|\rho, \alpha, C_0)$ is known only up to a proportionality term, and this term depends on C_0 (and hence on $\boldsymbol{R}$). We can write out the full prior for C as

$$\pi(C|\alpha, \rho, \boldsymbol{R}) = N(\boldsymbol{R}) \exp\left\{-\frac{1}{2}\alpha \times \mathcal{D}(C, C_{0,\boldsymbol{R}})\right\}$$

where $N(\boldsymbol{R})$ is a normalizing constant. By default, we consider a conditional prior for $\boldsymbol{R}$, such that $\pi(\boldsymbol{R} \mid \alpha, \rho, C_0) \propto \frac{1}{N(\boldsymbol{R})}$. This choice is computationally convenient for simulation, as described in Section 3.4. It is difficult to find a closed form expression for this prior, but it is possible to characterize its behavior by means of simulation. Figure 3 illustrates the form of the prior in the bivariate case by considering 10×10 and 20×20 grids. We observe that this prior does depend slightly on ρ and α, but the overall distribution is not greatly affected by the changing to a grid that has four times as many cells. A similar procedure could be done to observe the prior for $\boldsymbol{R}$ in a higher dimensional case, but the results of this simulation of a prior over a high dimensional correlation matrix are quite difficult to interpret. Section 3.4 describes the algorithm to

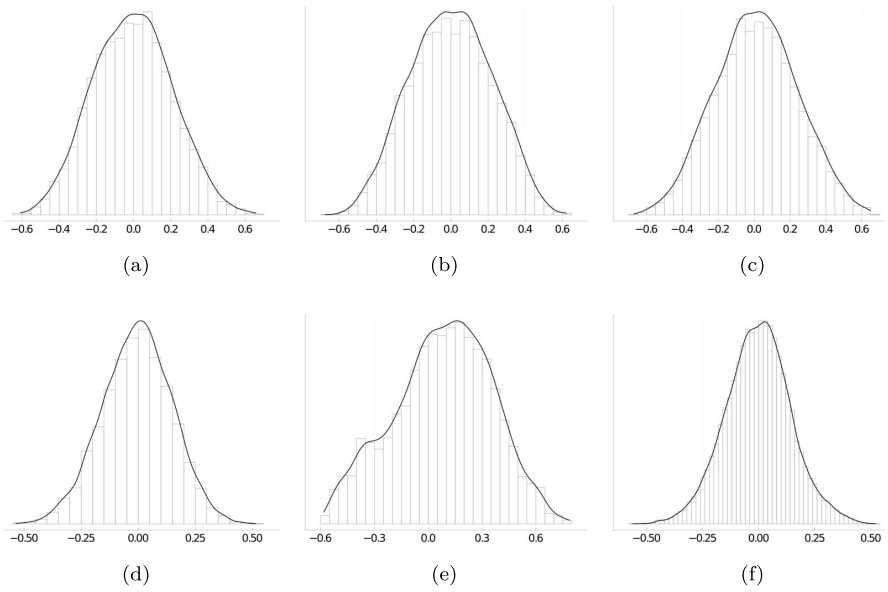


Figure 3: Prior distribution for the correlation coefficient of the centering Gaussian copula function. Panels (a), (c), and (e) display the results for an evenly spaced 10×10 grid and with $\alpha^\star = 25, 100$, and 400, respectively. Panels (b), (d), and (f) display the results for an evenly spaced 20×20 grid and with $\alpha^\star = 25, 100$, and 400, respectively.

perform this simulation in the general case, but since the prior is difficult to understand, our implementation uses a fixed independent C_0 in dimensions higher than 2.

It is also possible to use a similar structure to allow α to be random, but the proportionality term for $\pi(C|\rho, \alpha, C_0)$ also depends on α. A joint prior on $\alpha, \boldsymbol{R}$ with no closed form expression would be nearly impossible to interpret.

Finally, there is an important observation regarding the selection of the centering copula, C_0. The prior mode of the prior is not C_0 itself but rather its ρ-uniform version, $C_{0,\rho}$. Thus, if two researchers select different C_0 functions and these copulas match up to their ρ-uniform version, the corresponding priors will be identical. The information found in C_0 and $C_{0,\rho}$ is usually similar when the grid ρ is sufficiently fine, and eliciting C_0 directly is generally easier. Nonetheless, the user should inspect the ρ-uniform version of C_0 to make sure that it does provide the desired prior information. Assume, for instance, that a user select a parametric C_0 based on the existing prior information on Spearman's correlation coefficient. Depending on the grid, C_0 and its ρ-uniform version do not necessarily have the same information on Spearman's correlation coefficient. Table 1 of the Supplementary Material (Kuschinski and Jara, 2023) illustrates the differences between Spearman's correlation coefficient for the original copula C_0 and its ρ-uniform version, for different degrees of refinement, different copula models, and degrees of association. In all cases we observe that the differences are small, and that they become closer as the grid is refined. Additionally, we observe that the differences

are smaller when the correlation is lower and that the grid-uniform version of the copula generally has a slightly lower correlation than the original copula function.

3.3 Posterior Consistency

We show that under mild conditions on the prior specification of the proposed model and assuming the perfect knowledge of the marginal distributions, the posterior distribution of the proposed model is strongly consistent to the ρ-uniform version of any continuous copula function under simple random sampling.

Theorem 3. *Suppose that* $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n \mid H \overset{i.i.d.}{\sim} H$, *where*

$$H(\boldsymbol{y}) = C\big(F_1(y_1), F_2(y_2), \ldots, F_d(y_d)\big),$$

with $F_1, \ldots, F_n$ *being known marginal distributions, and* C *is the corresponding copula function of* H. *Let* $\Pi(\cdot \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ *be a version of the posterior distribution for* C *using a grid-uniform prior of the form*

$$\pi(\cdot \mid \rho, \alpha, C_0) \propto \exp\left\{-\frac{1}{2}\alpha \times \mathcal{D}(C, C_0)\right\} \times I_{\mathcal{C}_\rho}(C),$$

where $\mathcal{D}$ *be a suitable distance for probability distributions,* $\alpha > 0$, *and* $C_0 \in \mathcal{C}$, *with* $\mathcal{C}$ *being the space of all continuous copula functions and* $\mathcal{C}_\rho \subset \mathcal{C}$ *being the space of* ρ*-uniform copula functions. Let* C_ρ *be the grid-uniform version of* C. *Then, for every* $\epsilon > 0$ *and every* $C \in \mathcal{C}$,

$$\Pi(\mathcal{N}_{\epsilon, C_\rho} \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) \xrightarrow[n \longrightarrow \infty]{} 1,$$

H*-almost surely, where* $\mathcal{N}_{\epsilon, C_\rho} = \{Q \in \mathcal{C} : \mathcal{KL}(Q, C_\rho) < \epsilon\}$, *where* $\mathcal{KL}$ *denotes the Kullback-Leibler divergence given by* $\mathcal{KL}(C', C) = \int_{[0,1]^d} c(\boldsymbol{x}) \log(\frac{c(\boldsymbol{x})}{c'(\boldsymbol{x})}) d\boldsymbol{x}$, *where* c *and* c' *is a density of* C *and* C' *w.r.t. Lebesgue measure, respectively.*

The previous result implies that if the true model is ρ-uniform, the posterior distribution induced by the proposed model is strongly consistent at the truth. The result also implies that if the true model is not ρ-uniform, the posterior distribution will concentrate the mass at the ρ-uniform version of the true model and it is not consistent at the true model. The distance between the true model and its ρ-uniform version will depend on the characteristics of the true model and of the assumed grid ρ. However, it is possible to show that the thinner the grid the smaller the distance between the true model and its ρ-uniform version. This is illustrated in Table 2 of the Supplementary Material (Kuschinski and Jara, 2023), which shows the integrated squared differences between the original copula and its ρ-uniform version, for different families of copula functions and different degrees of association.

To arbitrarily increase the size of the gird, however, is not possible since the exploration of the posterior distribution becomes too demanding. To ensure the posterior consistency at any continuous copula function, a prior distribution on the grid ρ has

to be considered. By employing a prior on evenly spaced partitions, and under mild conditions on the prior specification of the proposed model, the posterior distribution of the proposed model is weakly consistent at any continuous copula function under simple random sampling.

Theorem 4. *Let* $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n \mid H \overset{i.i.d.}{\sim} H$*, where* $H(\boldsymbol{y}) = C(F_1(y_1), F_2(y_2), \ldots, F_d(y_d))$*, with* $F_1, \ldots, F_n$ *being known marginal distributions, and* C *is the corresponding copula function of* H*. Suppose that* C *is absolutely continuous w.r.t Lebesgue measure, with density* c*, and that fulfills the following conditions:*

1. $-\infty < \int c(x) \log(c(x)) dx < \infty$*,*
2. *There is* $B \in \mathbb{R}$ *and* $\gamma > 0$ *such that if* $c(x) > B$ *and* $L_2(x, x') < \gamma$ *then* $c(x') > 1$ *for all* x' *in the unit hypercube.*

Let $\Pi(\cdot \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ *be a version of the posterior distribution for* C *using a grid-uniform prior of the form*

$$\pi(\cdot \mid \rho, \alpha, C_0) \propto \exp\left\{ -\frac{1}{2}\alpha \times \mathcal{D}(C, C_0) \right\} \times I_{\mathcal{C}_\rho}(C),$$

and a prior $\pi(\rho)$ *for the* d*-dimensional grid of the form*

$$\pi(\rho) = P_r I_{\xi_r}(\rho),$$

where $P_r > 0, \forall r \in \mathbb{N}$ *with the property that* $\sum_r P_r = 1$*, and* ξ_r *is the* $r \times r \times r \cdots$ *evenly spaced* d*-dimensional grid,* $\mathcal{D}$ *is a suitable distance for probability distributions,* $\alpha \geq 0$*,* C_0 *a continuous copula function, and* $\mathcal{C}_\rho \subset \mathcal{C}$ *is the space of all* ρ*-uniform copula functions. Then, for every weak neighborhood* $\mathcal{N}$ *of* C*,*

$$\Pi(\mathcal{N} \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) \xrightarrow[n \longrightarrow \infty]{} 1,$$

H*-almost surely.*

Furthermore, by considering a prior on evenly spaced partitions with a particular tail behavior, the posterior distribution of the proposed model is strongly consistent at any continuous copula function under simple random sampling.

Theorem 5. *Let* $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n \mid H \overset{i.i.d.}{\sim} H$*, where* $H(\boldsymbol{y}) = C(F_1(y_1), F_2(y_2), \ldots, F_d(y_d))$*, with* $F_1, \ldots, F_n$ *being known marginal distributions, and* C *is the corresponding copula function of* H*. Suppose that* C *is absolutely continuous w.r.t Lebesgue measure, with density* c*, and that fulfills the conditions of Theorem 4. Let* $\Pi(\cdot \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ *be a version of the posterior distribution for* C *using a grid-uniform prior of the form*

$$\pi(\cdot \mid \rho, \alpha, C_0) \propto \exp\left\{ -\frac{1}{2}\alpha \times \mathcal{D}(C, C_0) \right\} \times I_{\mathcal{C}_\rho}(C),$$

and a prior $\pi(\rho)$ *for the* d*-dimensional grid of the form*

$$\pi(\rho) = P_r I_{\xi_r}(\rho),$$

where $P_r > 0, \forall r \in \mathbb{N}$ with the property that $\sum_r P_r = 1$, and ξ_r is the $r \times r \times r \cdots$ evenly spaced d-dimensional grid, $\mathcal{D}$ is a suitable distance for probability distributions, $\alpha \geq 0$, C_0 a continuous copula function, and $\mathcal{C}_\rho \subset \mathcal{C}$ is the space of all ρ-uniform copula functions. Furthermore, let $\bar{P}_l = \sum_{r=l}^{\infty} P_r$, and assume P_r satisfies $\bar{P}_l < p^{l^d+a}$ for some $0 < p < 1$, some $a > 0$ and large enough l. Then, for all $\epsilon > 0$,

$$\Pi(\mathcal{A}_\epsilon | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) \xrightarrow[n \longrightarrow \infty]{} 1,$$

H-almost surely, where $A_\epsilon = \{f : \mathcal{H}(C, f) < \epsilon\}$ and $\mathcal{H}$ is the Hellinger distance.

An example of a sequence P_r which satisfies the conditions for Hellinger convergence is $P_r \propto (1-p)^{r^d} p$ for any $0 < p < 1$, which is proportional to a geometric distribution on the number of cells in the grid.

3.4 An Automatic MCMC Algorithm for a Fixed Grid Model

We propose a Metropolis-within Gibbs algorithm for exploring the posterior distribution of the copula functions and the parameters associated with the marginal distributions.

Updating C Using a Random Rectangle Exchange as a Proposal

Rectangle exchanges provide us with a way to explore the space of ρ-uniform copulas. This movement can be used to generate proposals in the context of an Metropolis-Hastings (MH) algorithm. Let $C_\rho^{(b)}$ be the grid uniform copula which corresponds to the current state of the chain. Given, $C_\rho^{(b)}$, we propose the candidate $\widetilde{C}_\rho$ using the following random rectangle exchange:

a) Set $\widetilde{C}_\rho = C_\rho^{(b)}$ and pick d_1 and d_2 randomly from the set $\{1, \ldots, d\}$, and such that $d_1 < d_2$. The precise law of this selection does not matter so long as it is independent of $C_\rho^{(b)}$ and every pair of coordinates has a positive probability of being selected. In practice, we will select them uniformly.

b) Pick a_1 and a_2 from ρ_{d_1} and pick b_1 and b_2 from ρ_{d_2}. Also, for all $k \in \{1, \ldots, d\} \setminus \{d_1, d_2\}$, pick $x_k \in \rho_k$, and set

$$\boldsymbol{p}_{(a_l, b_m)} = (x_1, \ldots, x_{i-1}, a_l, x_{i+1}, \ldots, x_{j-1}, b_m, x_{j+1}, \ldots, x_d),$$

where $l, m \in \{1, 2\}$. The precise law of these selections does not matter so long as it is independent of $C_\rho^{(b)}$ and every rectangle along the selected coordinates has positive probability of being selected. In practice, we will select them uniformly.

c) Pick ϵ uniformly in the interval

$$\left[\max\left\{-C\left(\nu^\rho_{\boldsymbol{p}_{(a_1,b_2)}}\right), -C\left(\nu^\rho_{\boldsymbol{p}_{(a_2,b_1)}}\right)\right\}, \min\left\{C\left(\nu^\rho_{\boldsymbol{p}_{(a_1,b_1)}}\right), C\left(\nu^\rho_{\boldsymbol{p}_{(a_2,b_2)}}\right)\right\}\right].$$

d) Set

$$\widetilde{C}_\rho\big(\nu^\rho_{\boldsymbol{p}_{(a_1,b_1)}}\big) = C^{(b)}_\rho\big(\nu^\rho_{\boldsymbol{p}_{(a_1,b_1)}}\big) - \epsilon, \qquad \widetilde{C}_\rho\big(\nu^\rho_{\boldsymbol{p}_{(a_1,b_2)}}\big) = C^{(b)}_\rho\big(\nu^\rho_{\boldsymbol{p}_{(a_1,b_2)}}\big) + \epsilon,$$
$$\widetilde{C}_\rho\big(\nu^\rho_{\boldsymbol{p}_{(a_2,b_1)}}\big) = C^{(b)}_\rho\big(\nu^\rho_{\boldsymbol{p}_{(a_2,b_1)}}\big) + \epsilon, \quad \text{and} \quad \widetilde{C}_\rho\big(\nu^\rho_{\boldsymbol{p}_{(a_2,b_2)}}\big) = C^{(b)}_\rho\big(\nu^\rho_{\boldsymbol{p}_{(a_2,b_2)}}\big) - \epsilon.$$

We denote by $q(\cdot \mid C^{(b)}_\rho)$ to the candidate generating distribution induced by random rectangle exchange described by steps a)–d). An interesting property of this candidate generating distribution is that it is symmetric, which simplifies the computation of the acceptance probability. This is explained by the uniform selection of ϵ in the valid interval. The MH acceptance probability of the candidate $\widetilde{C}_\rho$ is given by

$$\max\big\{0, r\big(\widetilde{C_\rho}, C^{(b)}_\rho\big)\big\},$$

where

$$\begin{aligned}
\log\big(r\big(\widetilde{C_\rho}, C^{(b)}_\rho\big)\big) &= \log\big(\pi(\widetilde{C_\rho}|\rho,\alpha,C_0)\big) - \log\big(\pi\big(C^{(b)}_\rho|\rho,\alpha,C_0\big)\big) \\
&\quad + \ell(\widetilde{C_\rho}, F_1,\ldots,F_d|y_1\ldots y_n) - \ell\big(C^{(b)}_\rho, F_1,\ldots,F_d|y_1\ldots y_n\big), \\
&= -\frac{\alpha}{2}\big(\mathcal{D}(\widetilde{C_\rho}, C_0) - \mathcal{D}\big(C^{(b)}_\rho, C_0\big)\big) \\
&\quad + \sum_{i=1}^{n}\sum_{j=1}^{|\nu^\rho|}\log\bigg(\frac{\widetilde{C_\rho}(B_j) - C^{(b)}_\rho(B_j)}{\lambda(B_j)}\bigg) I_{\{(F_1(y_{i1}),\ldots,F_d(y_{id}))\in B_j\}}(\boldsymbol{y}_i), \\
&= -\frac{\alpha}{2}\big(\mathcal{D}(\widetilde{C_\rho}, C_0) - \mathcal{D}\big(C^{(b)}_\rho, C_0\big)\big) \\
&\quad - \sum_{i=1}^{n}\sum_{k=1}^{2}\sum_{l=1}^{2}\log\bigg(\frac{\widetilde{C_\rho}(\nu^\rho_{\boldsymbol{p}_{(a_k,b_l)}}) - C^{(b)}_\rho(\nu^\rho_{\boldsymbol{p}_{(a_k,b_l)}})}{\lambda(\nu^\rho_{\boldsymbol{p}_{(a_k,b_l)}})}\bigg) \\
&\quad \times I_{\{(F_1(y_{i1}),\ldots,F_d(y_{id}))\in\nu^\rho_{\boldsymbol{p}_{(a_k,b_l)}}\}}(\boldsymbol{y}_i).
\end{aligned}$$

Of note, if $\mathcal{D}$ is the squared-L_2 distance then the term $-\frac{\alpha}{2}(\mathcal{D}(\widetilde{C}_\rho, C_0) - \mathcal{D}(C^{(b)}_\rho, C_0))$ can also be further simplified to:

$$-\frac{\alpha}{2}\Bigg(\sum_{k=1}^{2}\sum_{l=1}^{2}\lambda\big(\nu^\rho_{\boldsymbol{p}_{(a_k,b_l)}}\big)\Big(\big(\widetilde{c_\rho}\big(\nu^\rho_{\boldsymbol{p}_{(a_k,b_l)}}\big) - c_0\big(\nu^\rho_{\boldsymbol{p}_{(a_k,b_l)}}\big)\big)^2 - \big(c^{(b)}_\rho\big(\nu^\rho_{\boldsymbol{p}_{(a_k,b_l)}}\big) - c_0\big(\nu^\rho_{\boldsymbol{p}_{(a_k,b_l)}}\big)\big)^2\Big)\Bigg).$$

In practice, we have found that this MH behaves well, with acceptance rates around 23%.

Updating the Marginal Distributions

There is no single technique which will work efficiently for the updating of the parameters of all possible marginal distributions, and tuning the posterior sampler may be difficult. However, there are some algorithms which are effective for a broad scope of distributions, and that can do reasonably good posterior exploration without having

to worry about tuning. A good starting point when parametric marginal distributions are considered is the t-walk (Christen et al., 2010), which is a general purpose sampler for parametric continuous distributions. The t-walk is a MH algorithm which adapts to the scale of the target distribution and can sample well from most finite dimensional continuous distributions without tuning.

The t-walk requires a support function for the parameters, an energy function, and two starting points within the support. The energy function as required by the algorithm is the negative log-density of the posterior distribution (up to a proportionality constant). We write the marginal distributions corresponding to some parametric family as $F_1(x_1, \theta_1)$, $F_2(x_2, \theta_2), \dots F_d(x_d, \theta_d)$ where θ_i is a set of parameters which defines the i^{th} marginal, with prior $\pi_i(\theta_i)$. Then the energy function is

$$\begin{aligned} E(\theta_1, \dots, \theta_d, \widetilde{C_\rho}) \propto & -\ell(\widetilde{C}_\rho, F_1, \dots, F_d | \boldsymbol{y}_1, \dots, \boldsymbol{y}_n) \\ & - \pi(\widetilde{C}_\rho | \rho, \alpha, C_0) - \sum_i \pi_i(\theta_i) - \sum_i \log\big(f_i(y_i|\theta_i)\big). \end{aligned}$$

The prior for the copula does not depend on the marginal parameters, so it can be dropped and we can use

$$E(\theta_1, \dots, \theta_d, \widetilde{C_\rho}) \propto -\ell(\widetilde{C}_\rho, F_1, \dots, F_d | \boldsymbol{y}_1, \dots, \boldsymbol{y}_n) - \sum_i \pi_i(\theta_i) - \sum_i \log\big(f_i(y_i|\theta_i)\big).$$

For the support function, we know that rectangle exchanges assure us that $\widetilde{C}$ is always within the support, so we can define a support function exclusively on $(\theta_1, \dots, \theta_n)$ to fit $\pi_1(\theta_1), \dots, \pi_n(\theta_n)$. Since the t-walk only moves $(\theta_1, \dots, \theta_d)$ it is not necessary to define two starting points for C_ρ. We only need two starting points for $(\theta_1, \dots, \theta_d)$, which can – for instance – be simulated from the prior.

If Dirichlet process mixture models (see, e.g., Müller et al., 2015) are employed for modeling the marginal distributions, the marginal algorithm of Ishwaran and James (2001), which assume a finite-dimensional approximation to the nonparametric marginals, can be easily adapted to our context.

Updating the Centering Copula Hyper-Parameter

When working with a hierarchical prior that establishes a prior distribution for C_0 which depends on $\boldsymbol{R}$, updating $\boldsymbol{R}$ can be done by adding a kernel to the MCMC chain. To update C_0, we use a variation of the metropolized hit-and-run algorithm (Chen and Dey, 1998), which makes proposals that are always valid correlation matrices.

In our specific case, we allow δ to be a pre-specified tuning parameter. We consider values between 0.3 and 1.0, as discussed by Chen and Dey (1998). To update $\boldsymbol{R}$, we propose a move from $\boldsymbol{R}^{(i)}$ to $\boldsymbol{R}^{(i+1)} = \boldsymbol{R}^{(i)} + \boldsymbol{H}$, by picking $\boldsymbol{H}$ as follows:

(1) Let $\xi^{(i)}$ be the least eigenvalue of $\boldsymbol{R}^{(i)}$.

(2) Pick a sequence of i.i.d. standard normal variables $z_{1,2}, z_{1,3}, \dots z_{d-1,d}$.

(3) Pick $\delta \sim N(0, \boldsymbol{r}^2)$ truncated to $(-\frac{\xi^{(i)}}{\sqrt{2}}, \frac{\xi^{(i)}}{\sqrt{2}})$.

(4) For $i < j$ set $h_{i,j} = \frac{\delta z_{i,j}}{\sum_{j=1}^{d-1}\sum_{l=j}^{D} z_{j,l}^2}$. Also set $h_{i,i} = 0$ for all i, and for $i > j$ set $h_{i,j} = h_{j,i}$. Set the matrix $\boldsymbol{H} = [h_{i,j}]$.

The acceptance probability is given by

$$\max\{0, r(\boldsymbol{R}^{(i+1)}, \boldsymbol{R}^{(i)})\},$$

where

$$\begin{aligned}\log\big(r(\boldsymbol{R}^{(i+1)}, \boldsymbol{R}^{(i)})\big) &= \log\big(\pi(C^{(i)}|\alpha^{(i)}, \rho, \boldsymbol{R}^{(i+1)})\big) - \log\big(\pi(C^{(i)}|\alpha^{(i)}\rho, \boldsymbol{R}^{(i)})\big), \\ &= \frac{1}{2}\alpha^{(i)}\big(\mathcal{D}(C, C_{0,R^{(i)}}) - \mathcal{D}(C, C_{0,R^{(i+1)}})\big).\end{aligned}$$

Our prior is selected so that $\log(N(\boldsymbol{R}))-\log(N(\boldsymbol{R}))$ cancels and we are not hampered by our inability to calculate it. Note that for a two dimensional copula, $\boldsymbol{R}$ depends only the single dimensional correlation coefficient, $\boldsymbol{r}$ and the hit and run algorithm reduces to a standard random walk Metropolis kernel (Robert and Casella, 2013).

4 Illustrations

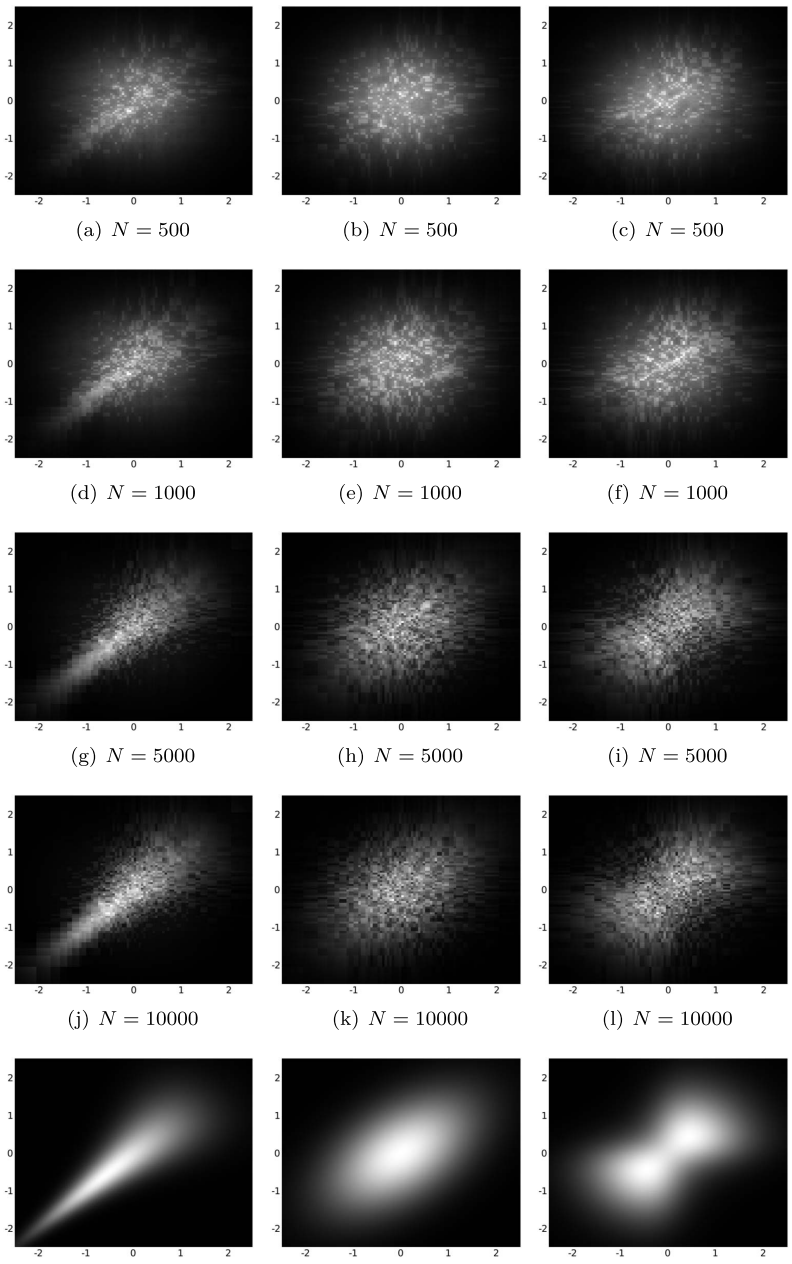
We illustrate the behavior of the proposed model by means of the analysis of simulated data. Functions implementing the MCMC algorithms employed in these analyses were written in Julia and are available upon request to the authors and in a GitHub repository.

4.1 Estimation of Parametric and Non-standard Copula Functions

To illustrate that the proposal model does not overfit the data when a parametric copula model holds and that is able to capture deviations from the standard parametric models with finite sample sizes, we consider bivariate models with Gaussian $(0, 1)$ marginals, under the following copula functions:

- **Model 1:** A Clayton copula with parameter $\theta = 3$, given by $C_\theta(x_1, x_2) = (\max\{u^{-\theta} + v^{-\theta} - 1; 0\})^{-1/\theta}$.
- **Model 2:** A Gaussian copula with correlation 0.5.
- **Model 3:** A copula arising from a mixture of two Gaussian distributions, both with identity covariance matrix, and centered at $(1, 1)$ and $(-1, -1)$, respectively.

Figure 4 display the true models under consideration. For each model, we simulate a single data set of size $N = 500, 1{,}000,$ 5,000, and 10,000. For each simulated dataset we fit our proposed model by considering a 50×50 grid, the hierarchically centered



(a) $N = 500$ (b) $N = 500$ (c) $N = 500$

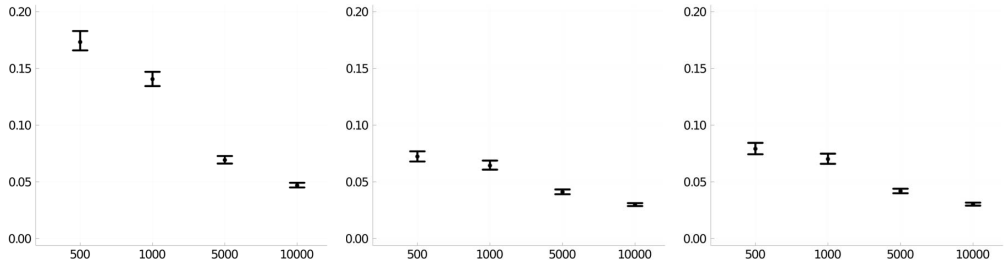
(d) $N = 1000$ (e) $N = 1000$ (f) $N = 1000$

(g) $N = 5000$ (h) $N = 5000$ (i) $N = 5000$

(j) $N = 10000$ (k) $N = 10000$ (l) $N = 10000$

(m) True model (n) True model (o) True model

Figure 4: Posterior mean of the bivariate density function. Panel (a)–(d) show the results for $N = 500, 1{,}000, 5{,}000$, and 10,000, respectively. Panel (e) displays the true model. Model 1 (Clayton copula). Model 2 (Gaussian copula) Model 3 ((Copula of a Gaussian mixture model).



(a) (b) (c)

Figure 5: Posterior mean (point) and 95% credibility interval (vertical bar) for the Hellinger distance to the true joint distribution, for different sample sizes (the horizontal axis is the sample size). Panel (a) displays the results for Model 1 (Clayton copula). Panel (b) displays the results for Model 2 (Gaussian copula). Finally, panel (c) displays the results for Model 3 (a copula of a mixture of Gaussian distributions).

prior, with the ICAR correlation structure described in Section 3.1, and $\alpha^\star = 400$. In these analyses we assume the marginals distributions to be known. We create a Markov chain of (conservative) of size 2,000,000 using the automatic algorithm described in Section 3.4. We considered a burn-in period of 20,000 and a thinning of 1,000. Figure 4 show the posterior mean under the different models and sample sizes.

Figure 5 displays the posterior mean and 95% credibility intervals for the Hellinger distance to the true model under the different models and sample sizes. The results show that adequate estimates for complex true models can be obtained, even for reduced sample sizes, and that when the copula model is simple, the proposed model does not overfit the data. The results also show that the posterior mean gets closer to the true model when the sample size increases and that the posterior distribution concentrates around the true model as the sample size increases.

4.2 Estimation of a Four Dimensional Copula with Unknown Marginals

We illustrate the model under a more difficult set of circumstances. Data was simulated from a four dimensional distribution with a Gaussian copula and correlation matrix reflecting different degrees of association, given by

$$\begin{pmatrix} 1.0 & 0.4 & 0.6 & 0.7 \\ 0.4 & 1.0 & 0.7 & 0.3 \\ 0.6 & 0.7 & 1.0 & 0.2 \\ 0.7 & 0.3 & 0.2 & 1.0 \end{pmatrix}.$$

We considered different supports and types for the marginals distributions. Specifically, for coordinate one, we considered mixture of Gaussian distributions given by $0.5 \times N(0,1)+0.5\times N(3,1)$. For the second coordinate we considered a log-normal distribution

with location and scale parameters set to 0 and 1, respectively. For the third coordinate we considered Gamma distribution with shape and scale parameter set to 5 and 9, respectively. Finally, for the fourth coordinate, we considered a Beta distribution with parameters 2 and 4.

We simulated a single realization of the data set and considered three sample sizes: $N = 1{,}000$, 5,000, and 10,000. For each simulated dataset, inference was performed under the ICAR prior and a $5 \times 5 \times 5 \times 5$ grid. Posterior inference was performed using the t-walk algorithm described previously, assuming the true marginal distributions with unknown parameters, and flat priors on them. We considered a burn-in period of 10,000 and the chain was run until it produced an effective posterior sample of length 1,000.

Figure 1 of the Supplementary Material (Kuschinski and Jara, 2023) shows summary statistics of the posterior distribution of the Hellinger distance to the true copula function for the different sample sizes. As expected, we observe that the posterior distribution concentrates the mass around the true copula function as the sample size increases. Figure 6 shows the posterior inferences on the marginal distributions for the different sample sizes. We observe that even under a complex marginal distribution, the marginal estimation is close to the truth in all cases and that posterior distribution concentrate the mass around the true density function as the sample size increases.

4.3 The Effect of Lack of Knowledge of Marginal Distributions

To illustrate the effect of lack of knowledge of marginal distributions on the quantification of the uncertainty regarding the copula function estimation, we generated data from bivariate Gaussian distribution. We simulated a single realization of the data set and considered three sample sizes: $N = 2{,}500$, 5,000, and 10,000. For each simulated dataset, inference was performed under the ICAR prior and a 20×20 grid. Furthermore, we considered the following three settings:

- **Setting 1:** The copula was estimated along with the marginals. The marginals were modeled as Gaussian, and estimated using a flat prior which was centered around the true marginals. In particular the prior variance was set to 1 for all parameters.
- **Setting 2:** The copula was estimated along with the marginals. The marginals were modeled as Gaussian, and estimated using a concentrated prior which was centered around the true marginal. In this case, the prior variance for the mean and variance of each marginal was 10^{-4} and 10^{-6}, respectively. This choice of prior variance is similar to the posterior variance of the parameters in the previous scenario, as estimated with sample size of 5,000.
- **Setting 3:** The marginals are assumed known and set equal to the truth.
- **Setting 4:** A two-step approach is considered. The marginals are first estimated by using the empirical distribution of the data. This distribution is then treated as the truth and the copula is estimated.

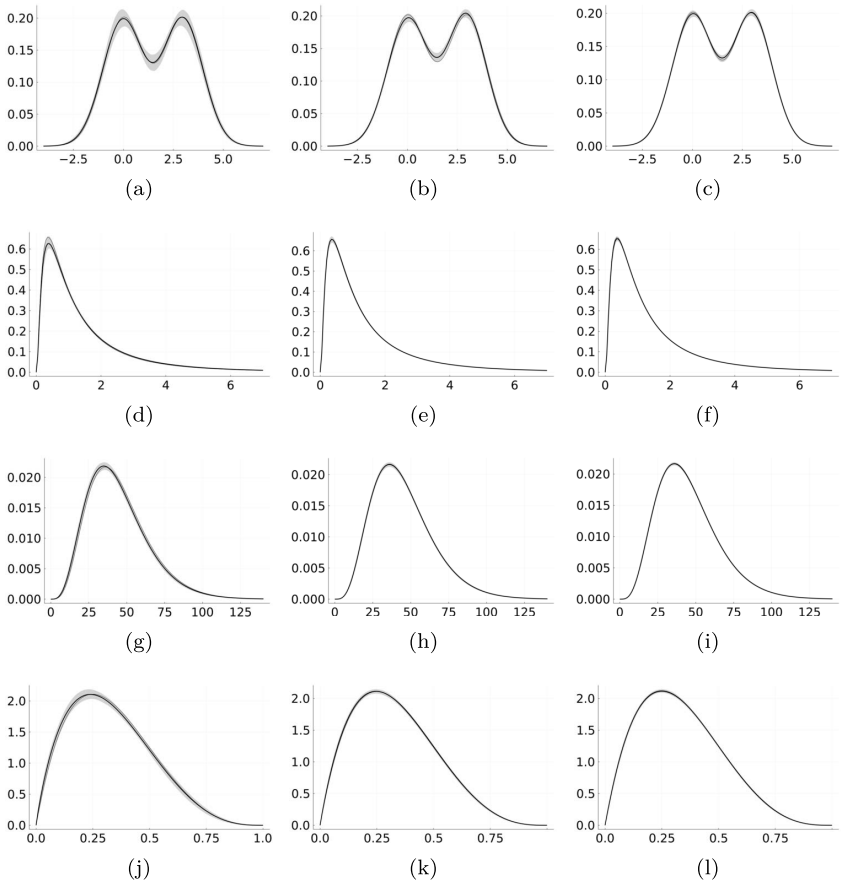


Figure 6: Four dimensional copula with unknown marginals. Posterior mean (tick line) and 95% point-wise credibility interval (grey band) for the marginal distribution. Panels (a)–(c), (d)–(f), (g)–(i), and (j)–(l) display the results for coordinate 1, 2, 3, and 4, respectively. Panels (a), (d), (g), and (j), (b), (e), (h), and (k), and (c), (f), (i), and (l) display the results for $N = 1{,}000, 5{,}000$, and 10,000, respectively. In each figure, the true marginal distribution is shown as a thin line.

Figure 7 shows summary statistics of the posterior distribution of the Hellinger distance to the true copula function under the different estimation settings and sample sizes. For sample sizes of 2,500 and 5,000, we observe that the whiskers of the box plot progressively contract as more information about the marginals is known. For a sample size of 10,000, enough marginal information is available in the sample that the four box plots might represent samples from the same distribution. Of particular interest is the fourth scenario, where the marginals are treated as known but an estimate is used in place of the true marginals. In this situation, the whiskers are just as contracted as in the oracle scenario, showing that the copula uncertainty has been under-estimated.

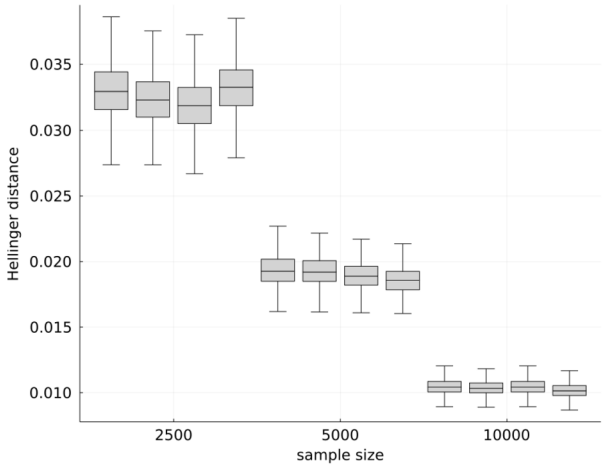


Figure 7: Two dimensional copula model. Box plots summarizing the posterior distribution of the Hellinger distance to the true copula function under different estimation settings and sample sizes. For each sample size, four box plots are presented. From left to right, the information about the posterior distribution of the Hellinger distance to the true copula function is presented under Setting 1, 2, 3, and 4, respectively.

With 2,500 observations the error is greater than for the other scenarios. In this case, it is reasonable to assume that this deviation is due to the estimation error.

We note that these results are for one random sample from one particular density, and conclusions should be drawn with care. Nonetheless, these examples do show how our model allows the user to properly account for the uncertainty associated with the lack of knowledge of the marginal distributions and that even with a large sample sizes and accurate estimation procedures for the marginal distributions, the posterior uncertainty for the copula can be strongly underestimated if full joint inference is replaced with point estimation of the marginal distributions.

4.4 A Comparison with Existing Approaches

We compared our proposal with the flat prior proposed by Guillotte and Perron (2012). There are similarities between our model and their flat proposal. The model proposed by Guillotte and Perron (2012) is a specific case of a grid-uniform copula, restricted to two dimensions and with grids that are necessarily evenly spaced. When these conditions hold, the models differs in way the prior probability mass is assigned. The work of Guillotte and Perron (2012) focused mainly on reference priors, whereas our prior is designed to share information on neighboring sets. We compare our proposal with the flat prior proposed by Guillotte and Perron (2012) here. The flat prior of Guillotte and Perron (2012) can be thought as a limiting case of our model when $\alpha \longrightarrow 0$.

We compare the models under settings considered by Guillotte and Perron (2012) in the evaluation of their proposal. We consider a two dimensional problem, a 6×6 grid, and a Gaussian, Gumbel and Clayton copula. We set the parameters of the different copula models such that they imply a similar association structure for the two variables. In particular, we set them such that they have the same Kendall's τ, and consider $\tau = 0.05$, 0.35, 0.50, and 0.64. We consider three sample sizes in the comparison, $N = 30$, $N = 100$, $N = 400$, and $N = 800$.

We performed a Monte Carlo study, considering 100 replicates for each model and sample size. For each data set we fit our proposal with the hierarchically centered prior, the ICAR correlation structure described in Section 3.1, and $\alpha^\star = 40$. The performance of the models was evaluated by computing the mean integrated squared error between the posterior mean and the true data generating copula model. The results are presented in Table 3 of the Supplementary Material (Kuschinski and Jara, 2023).

The results illustrate that the proposed model outperforms the flat prior across the board. As expected, the biggest differences between models are observed at small sample sizes; the larger the sample size, the smaller the difference between models regardless of the association structure. Furthermore, for a given sample size, our model tends to produce better results than the flat prior as the level of association increases.

5 Concluding Remarks

Flexible inference of copula functions had mainly relied on partial likelihood or pseudo-likelihood methods. This approach is useful in some cases. However, they do not allow for a proper quantification of the uncertainties associated to the lack of knowledge of the marginal distributions and cannot be employed for modelling the association structure of latent variables in the context of hierarchical models. We have proposed a novel and rich family of copula functions that can overcome these problems, the class of grid-uniform copula functions. We prove that this class is dense in the space of all continuous copula functions in a Hellinger sense.

We proposed a hierarchically centered prior distribution based on the proposed family, borrowing ideas from spatial statistics, and that have appealing support and posterior consistency properties. We have described a class of transformations on grid-uniform copulas which is closed in the space of grid-uniform copula functions and that is able to span the complete space of grid-uniform copula functions in finite number of steps, starting from any point in the space. This family of transformations, referred to as rectangle exchanges, is employed to develop an automatic MCMC algorithm for exploring the corresponding posterior distribution. We have illustrated the behavior of the proposal and compared it with the approach proposed by Guillotte and Perron (2012). By considering similar simulation settings to the ones considered by Guillotte and Perron (2012), we show that our proposal outperforms their flat model when the posterior mean is the point estimator and mean integrated squared error is considered as a model comparison criteria.

The proposed prior model can be extended in different ways. The current implementation of the proposal depends on a user-specified grid ρ. The size of the grid and the

location of the points may have an important influence in the resulting model. For instance, equally spaced grids can lead to over fitting of the copula function in areas where few data points are "observed". On the other hand, they can lead to under fitting in areas where more data points are "observed". The study of strategies for the estimation of the optimal size and location of the grid is the subject of ongoing research.

The proposed model suffers from the curse of dimensionality. For a sample of size 1,000, for a grid-uniform prior with a 10×10 grid it takes only a few seconds to generate a Markov chain of length 20,000 using the automatic MCMC algorithm and an i5 processor. We have also been able to use the proposed model in dimensions up to 10. However, the implementation of the models in high dimensions and with fine grids would result in an explosion of parameters that need to be updated, which makes the implementation of this approach practically impossible. The study of marginal versions of the model, where the copula probabilities are integrated out of the model is also subject of ongoing research.

Finally, the extension of the model to handle mixed, discrete and continuous, variables and to copula regression problems, and the extension of the consistency results to unknown marginal distributions is also subject of ongoing research.

Supplementary Material

Supplementary Material (DOI: 10.1214/23-BA1396SUPP; .pdf). In the Supplementary Material (Kuschinski and Jara, 2023) we provide: Spearman's correlation coefficient for the original copula and ρ-uniform versions of different degrees of refinement, integrated square differences between the original copula and ρ-uniform versions of it for different degrees of refinement, mean integrated squared error for the posterior mean of the copula function under our default hierarchical prior (proposal) and under the flat prior proposed by Guillotte and Perron (2012), box plots summarizing the posterior distribution of the Hellinger distance to the true copula function for different sample sizes, and proofs for Lemmas 1–2, and Theorems 1–5.

References

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press. MR3362184. 10

Chen, M.-H. and Dey, D. K. (1998). "Bayesian modeling of correlated binary responses via scale mixture of multivariate normal link functions." *Sankhyā: The Indian Journal of Statistics, Series A*, 60(3): 322–343. MR1718864. 17

Choroś, B., Ibragimov, R., and Permiakova, E. (2010). "Copula estimation." In *Copula Theory and Its Applications*, 77–91. Springer. MR3051263. doi: https://doi.org/10.1007/978-3-642-12465-5_3. 2

Christen, J. A., Fox, C., et al. (2010). "A general purpose sampling algorithm for continuous distributions (the t-walk)." *Bayesian Analysis*, 5(2): 263–281. MR2719653. doi: https://doi.org/10.1214/10-BA60. 17

Church, C. (2012). "The asymmetric t-copula with individual degrees of freedom." Ph.D. thesis, University of Oxford,. 3

Deheuvels, P. (1979). "La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance." *Bulletins de l'Académie Royale de Belgique*, 65(1): 274–292. MR0573609. 2

Eilers, P. H. C. and Marx, B. D. (1996). "Flexible smoothing with B-splines and penalties." *Statistical Science*, 11(2): 89–121. MR1435485. doi: https://doi.org/10.1214/ss/1038425655. 11

Faugeras, O. P. (2013). "Sklar's theorem derived using probabilistic continuation and two consistency results." *Journal of Multivariate Analysis*, 122: 271–277. MR3189323. doi: https://doi.org/10.1016/j.jmva.2013.07.010. 1

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). "A semiparametric estimation procedure of dependence parameters in multivariate families of distributions." *Biometrika*, 82(3): 543–552. MR1366280. doi: https://doi.org/10.1093/biomet/82.3.543. 2

Genest, C. and Nešlehová, J. (2007). "A primer on copulas for count data." *ASTIN Bulletin: The Journal of the IAA*, 37(2): 475–515. MR2422797. doi: https://doi.org/10.2143/AST.37.2.2024077. 2

Genest, C., Nešlehová, J., and Ziegel, J. (2011). "Inference in multivariate Archimedean copula models." *Test*, 20(2): 223–256. MR2834039. doi: https://doi.org/10.1007/s11749-011-0250-6. 2

Grazian, C. and Liseo, B. (2016). "Approximate Bayesian methods for multivariate and conditional copulae." In *International Conference on Soft Methods in Probability and Statistics*, 261–268. Springer. 2, 3

Guillotte, S. and Perron, F. (2012). "Bayesian estimation of a bivariate copula using the Jeffreys prior." *Bernoulli*, 18(2): 496–519. MR2922459. doi: https://doi.org/10.3150/10-BEJ345. 2, 23, 24, 25

Ishwaran, H. and James, L. F. (2001). "Gibbs sampling methods for stick-breaking priors." *Journal of the American Statistical Association*, 96(453): 161–173. MR1952729. doi: https://doi.org/10.1198/016214501750332758. 17

Joe, H. (2014). *Dependence Modeling with Copulas*. Boca Raton: CRC Press. MR3328438. 1, 2

Kaewsompong, N., Maneejuk, P., and Yamaka, W. (2020). "Bayesian estimation of Archimedean copula-based SUR quantile models." *Complexity*, 2020: 1–15. 2

Kauermann, G., Schellhase, C., and Ruppert, D. (2013). "Flexible copula density estimation with penalized hierarchical B-splines." *Scandinavian Journal of Statistics*, 40(4): 685–705. MR3145112. doi: https://doi.org/10.1111/sjos.12018. 2

Kuschinski, N. and Jara, A. (2023). "Supplementary Material for "grid-uniform copulas and rectangle exchanges: Bayesian model and inference for a rich class of copula functions"." *Bayesian Analysis*. doi: https://doi.org/10.1214/23-BA1396SUPP. 4, 12, 13, 21, 24, 25

McNeil, A. J. and Nešlehová, J. (2009). "Multivariate Archimedean copulas, d-monotone functions and $l1$-norm symmetric distributions." *The Annals of Statistics*, 37(5B): 3059–3097. MR2541455. doi: https://doi.org/10.1214/07-AOS556. 2

Mukhopadhyay, S. and Parzen, E. (2020). "Nonparametric universal copula modeling." *Applied Stochastic Models in Business and Industry*, 36(1): 77–94. MR4087740. doi: https://doi.org/10.1002/asmb.2503. 2

Müller, P., Quintana, F., Jara, A., and Hanson, T. (2015). *Bayesian Nonparametric Data Analysis*. New York: Springer. MR3309338. doi: https://doi.org/10.1007/978-3-319-18968-0. 2, 17

Nelsen, R. B. (2007). *An Introduction to Copulas*. New York: Springer Science & Business Media. MR2197664. doi: https://doi.org/10.1007/s11229-005-3715-x. 1, 2

Ning, S. and Shephard, N. (2018). "A nonparametric Bayesian approach to copula estimation." *Journal of Statistical Computation and Simulation*, 88(6): 1081–1105. MR3760400. doi: https://doi.org/10.1080/00949655.2017.1421194. 2, 3

Oh, D. H. and Patton, A. J. (2013). "Simulated method of moments estimation for copula-based multivariate models." *Journal of the American Statistical Association*, 108(502): 689–700. MR3174652. doi: https://doi.org/10.1080/01621459.2013.785952. 2

Oh, D. H. and Patton, A. J. (2017). "Modeling dependence in high dimensions with factor copulas." *Journal of Business & Economic Statistics*, 35(1): 139–154. MR3591542. doi: https://doi.org/10.1080/07350015.2015.1062384. 2

Pitt, M., Chan, D., and Kohn, R. (2006). "Efficient Bayesian inference for Gaussian copula regression models." *Biometrika*, 93(3): 537–554. MR2261441. doi: https://doi.org/10.1093/biomet/93.3.537. 2

Robert, C. and Casella, G. (2013). *Monte Carlo Statistical Methods*. New York: Springer Science & Business Media. MR1707311. doi: https://doi.org/10.1007/978-1-4757-3071-5. 18

Schmidt, A. M. and Nobre, W. S. (2018). "Conditional autoregressive (CAR) model." In *Wiley StatsRef: Statistics Reference Online*, 1–11. John Wiley & Sons, Ltd. 10

Sklar, M. (1959). "Fonctions de repartition an dimensions et leurs marges." *Publications de l'Institut Statistique de l'Université de Paris*, 8: 229–231. MR0125600. 1

Smith, M. S. (2013). "Bayesian approaches to copula modelling." In *Bayesian Theory and Applications*. Oxford University Press. MR3221171. doi: https://doi.org/10.1093/acprof:oso/9780199695607.003.0017. 1, 2

Wu, J., Wang, X., and Walker, S. G. (2013a). "Bayesian nonparametric estimation of a copula." *Journal of Statistical Computation and Simulation*, 85(1): 103–116. MR3266006. doi: https://doi.org/10.1080/00949655.2013.806508. 2, 3

Wu, J., Wang, X., and Walker, S. G. (2013b). "Bayesian nonparametric inference for a multivariate copula function." *Methodology and Computing in Applied Probability*,

16(3): 747–763. MR3239818. doi: https://doi.org/10.1007/s11009-013-9348-5. 2, 3

Yang, L., Frees, E. W., and Zhang, Z. (2019). "Nonparametric estimation of copula regression models with discrete outcomes." *Journal of the American Statistical Association*, 115(530): 707–720. MR4107674. doi: https://doi.org/10.1080/01621459.2018.1546586. 2