I N S T I T U T D E
S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

0935

SMOOTH SEMI- AND NONPARAMETRIC BAYESIAN ESTIMATION OF BIVARIATE DENSITIES FROM BIVARIATE HISTOGRAM DATA

LAMBERT, Ph.

Smooth semi- and nonparametric Bayesian estimation of bivariate densities from bivariate histogram data

Philippe Lambert[1,2*]

[1] *Institut des sciences humaines et sociales, Université de Liège, Liège, Belgium*

2 *Institut de statistique, Université catholique de Louvain, Louvain-la-Neuve, Belgium*

Abstract

We show how penalized B-splines combined with the composite link model can be used to estimate a bivariate density from histogram data. Two strategies are proposed: the first one is semi-parametric with flexible margins modeled using B-splines and a parametric copula for the dependence structure ; the second one is nonparametric and is based on Kronecker products of the marginal B-splines bases. Frequentist and Bayesian estimations are described. A large simulation study quantifies the performances of both methods under different dependence structures and varying strengths of dependence, sample sizes and amounts of grouping. It suggests that Schwarz's BIC is a good tool for classifying the competing models. The density estimates are used to evaluate conditional quantiles in two applications in social and in medical sciences.

Key words: Grouped data, bivariate density, P-splines, composite link model, copula.

1 Introduction

Interval censored data are commonly encountered in practice. Approximate methods have long been used to deal with such data. Right or mid-point imputation were and still are convenient approaches as it enable to use familiar statistical techniques to analyze the data. Alternatively, strong parametric assumptions on the distribution of the response can also be made. Even if the so obtained results are potentially misleading (see Law and Brookmeyer, 1992, for doubly-censored data), it often provides a first series of hints for a more rigorous and detailed analysis.

Although very few dedicated procedures are available in generalist statistical softwares, there exists a vast literature on the analysis of univariate interval censored data. For a survey essentially focussed on time-to-event data, we refer to Gomez *et al.* (2004) and the recent book by Sun (2006). Extensions of the Cox proportional hazards and of the accelerated failure times models were proposed (see e.g. Goggins and Finkelstein, 2000; Komárek and Lesaffre, 2006; Zhang and

*Correspondence to: Philippe Lambert, Université de Liège, Institut des sciences humaines et sociales, Méthodes quantitatives en sciences sociales, Boulevard du Rectorat 7 (B31), B-400 Liège (Belgium). E-mail: p.lambert@ulg.ac.be Phone: +32-4-366.59.90 Fax: +32-4-366.47.51

Davidian, 2008) with, in particular, challenging and fascinating applications in dentistry (see e.g. Härkänen *et al.*, 2000; Wong *et al.*, 2005; Komárek *et al.*, 2005).

Here, we shall focus on the Bayesian estimation of a bivariate density from bivariate histogram data, thereby extending the work in a univariate setting of Eilers (1992), Kooperberg and Stone (1992), Minnotte (1998), Koo and Kooperberg (2000), Braun *et al.* (2005) and Lambert and Eilers (2009). More precisely, we assume that for each independent pair of observations, each component is only known to take a value in one of the predefined intervals partitioning the support of the corresponding variable. Thus, the setting is more restrictive than for bivariate interval-censored data.

Betensky and Finkelstein (1999) extends the maximum likelihood estimation of the survivor function from interval censored data (Peto, 1973; Turnbull, 1976) to a bivariate setting. Komárek and Lesaffre (2006) consider a mixture of bivariate normal distributions to specify the conditional density in accelerated failure times models. Yang *et al.* (2008) recently explained how mixture of Polya trees (Lavine, 1992; Hanson, 2006) can be used to estimate a bivariate density nonparametrically from interval censored data.

Our two proposals based on penalized B-splines (Eilers and Marx, 1996) and the composite link model (Thompson and Baker, 1981) are described in Section 2: the first one combines B-splines models for the marginal densities with a parametric copula, while the second is based on Kronecker products of marginal B-splines bases. The full Bayesian model and the associated conditional posteriors are detailed in Section 3. A block-wise Metropolis-Hastings algorithm is suggested to explore the joint posterior (Section 4). The merits of the method are explored in a large simulation study in Section 5. We conclude the paper with two applications and a discussion.

2 Penalized B-spline models for the bivariate density

Assume that most of the probability mass of (X^1, X^2) is contained within the rectangle $(a^1, b^1) \times (a^2, b^2)$ and that one is interested in estimating a discrete representation of the bivariate continuous density $f_{12}(\cdot, \cdot)$ on a fine rectangular lattice $\{\mathcal{I}_{i_1 i_2} = \mathcal{I}^1_{i_1} \times \mathcal{I}^2_{i_2} : i_1 = 1, \ldots, I_1; i_2 = 1, \ldots, I_2\}$ with $\mathcal{I}^d_{i_d} = (x^d_{i_d-1}, x^d_{i_d})$, $\Delta_d = x^d_{i_d} - x^d_{i_d-1}$, $a^d = x^d_0$ and $b^d = x^d_{I_d}$ $(d = 0, 1)$. Then, the probability to observe (X^1, X^2) in the small rectangular cell $\mathcal{I}_{i_1 i_2}$ is

$$\pi_{i_1 i_2} = \int_{\mathcal{I}^1_{i_1}} \int_{\mathcal{I}^2_{i_2}} f_{12}(s,t) ds dt \approx f_{12}(u^1_{i_1}, u^2_{i_2}) \Delta_1 \Delta_2,$$

where $u^d_{i_d}$ denotes the midpoint of $\mathcal{I}^d_{i_d}$.

Further assume that each marginal support is partitioned into J_d $(d = 1, 2)$ consecutive wide bins $\mathcal{J}^d_{j_d} = (L^d_{j_d}, U^d_{j_d})$ $(j_d = 1, \ldots, J_d)$. The available data take the form of bivariate frequencies $\{m_{j_1 j_2} : j_d = 1, \ldots, J_d \ (d = 1, 2)\}$ associated to the wide grid defined by the partition $\{\mathcal{J}^1_{j_1} \times \mathcal{J}^2_{j_2} : j_d = 1, \ldots, J_d \ (d = 1, 2)\}$ of the support. The marginal frequencies are denoted by $\{m_{j_1 +} : j_1 = 1, \ldots, J_1\}$ for X^1 and $\{m_{+ j_2} : j_2 = 1, \ldots, J_2\}$ for X^2.

For simplicity, suppose for now that the limits of the (wide) bins for margin d make a subset of the marginal small bins limits $\{x_0^d, \ldots, x_{I_d}^d\}$. The connection between the wide and the small bins probabilities of the dth marginal can be made through the composing matrices $C^d = \{c_{j_d i_d}^d\}$ such that $c_{j_d i_d}^d = 1$ if $\mathcal{I}_{i_d} \subset \mathcal{J}_{j_d}$ and 0 otherwise. Indeed, if $\gamma_{j_1 j_2}$ denotes the joint probability,

$$\int_{\mathcal{J}_{j_1}^1} \int_{\mathcal{J}_{j_2}^2} f_{X^1 X^2}(s,t) ds dt,$$

that (X^1, X^2) belongs to the wide rectangle $\mathcal{J}_{j_1}^1 \times \mathcal{J}_{j_2}^2$, then the marginal probability that X^1 (resp. X^2) belongs to $\mathcal{J}_{j_1}^1$ (resp. $\mathcal{J}_{j_2}^2$) is $\gamma_{j_1}^1$ (resp. $\gamma_{j_2}^2$) where

$$\gamma_{j_1}^1 = \gamma_{j_1 +} = \sum_{i_1=1}^{I_1} c_{j_1 i_1}^1 \pi_{i_1 +} \quad ; \quad \gamma_{j_2}^2 = \gamma_{+ j_2} = \sum_{i_2=1}^{I_2} c_{j_2 i_2}^2 \pi_{+ i_2}. \tag{2.1}$$

Denote by $\Pi = \{\pi_{i_1 i_2}\}$ and $\Gamma = \{\gamma_{j_1 j_2}\}$ the $I_1 \times I_2$ and $J_1 \times J_2$ matrices of probabilities of the small and wide rectangular cells, respectively. Using the lattice format of our setting, one can show that

$$\Gamma = C^1 \Pi (C^2)^T. \tag{2.2}$$

Alternatively, one can write

$$\text{vec}\,(\Gamma) = (C^2 \otimes C^1) \text{vec}\,(\Pi)\,.$$

This is a bivariate extension of the *composite link model* (CLM) (Thompson and Baker, 1981) for density estimation (Lambert and Eilers, 2009). Note that a small bin i should not necessarily be fully excluded or included in the jth wide bin. Then, the concerned element c_{ji} of the composing matrix could simply be the proportion of the small bin included in the interval.

2.1 A parametric copula model with flexible margins

The marginal distributions can be modelled using the strategy presented in Lambert and Eilers (2009). Accordingly, consider for margin d a basis $\mathcal{B}_d = \{b_{k_d}^d(\cdot) : k_d = 1, \ldots, K_d\}$ of K_d B-splines associated to equidistant knots on (a^d, b^d). Denote by B^d the $I_d \times K_d$ matrix of B-splines evaluated at the midpoints $\{u_{i_d}^d : i_d = 1, \ldots, I_d\}$ of the (fine) marginal grid induced by the fine rectangular lattice described above: one has $B^d = \{b_{i_d k_d}^d\}$ with $b_{i_d k_d}^d = b_{k_d}^d(u_{i_d}^d)$. The marginal probabilities are modelled as

$$\pi_{i_1 +} = \frac{\mathrm{e}^{\eta_{i_1}^1}}{\mathrm{e}^{\eta_1^1} + \ldots + \mathrm{e}^{\eta_{I_1}^1}} \quad ; \quad \pi_{+ i_2} = \frac{\mathrm{e}^{\eta_{i_2}^2}}{\mathrm{e}^{\eta_1^2} + \ldots + \mathrm{e}^{\eta_{I_2}^2}} \quad \text{with} \quad \boldsymbol{\eta}^d = \begin{pmatrix} \eta_1^d \\ \vdots \\ \eta_{I_d}^d \end{pmatrix} = B^d \boldsymbol{\phi}^d, \tag{2.3}$$

where $\boldsymbol{\phi}^d$ is the vector of spline parameters for margin d with identifiability constraint $\sum_{k=1}^{K_d} \phi_k^d = 0$. Remembering that $\pi_{i_1 +} = \int_{\mathcal{I}_1} f_1(s) ds \approx f_1(u_{i_1}^1) \Delta_1$ and

$\pi_{+i_2} = \int_{\mathcal{I}_2} f_2(t)dt \approx f_2(u^2_{i_2})\Delta_2$, this yields a model for the marginal densities (integrated over small predefined regions).

In the spirit of Eilers and Marx (1996), roughness penalties are forced on the B-spline coefficients to avoid overfitting. It penalizes changes in the rth order differences of the coefficients. In a Bayesian framework, this can be done by taking the following prior for $\boldsymbol{\phi}^d$ (Lang and Brezger, 2004):

$$p(\boldsymbol{\phi}^d|\tau^d) \propto \tau^{K_d/2} \exp\{-0.5\ \tau^d\ (\boldsymbol{\phi}^d)^T P^d_r \boldsymbol{\phi}^d\},$$
$$\tau^d \sim \text{Exp}(b = 0.0001),$$

where Exp(b) denotes the exponential distribution with mean b^{-1} and $P^d_r = D^{dT}_r D^d_r + \epsilon I_{K_d}$. For example, with a penalty of order $r = 3$, one has

$$D^d_3 = \begin{bmatrix} -1 & 3 & -3 & 1 & 0 & \dots & 0 \\ 0 & -1 & 3 & -3 & 1 & \dots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -1 & 3 & -3 & 1 \end{bmatrix}.$$

Then, when τ^d is very large, the conditional prior for $\boldsymbol{\phi}^d$ tends to favour a quadratic evolution of ϕ^d_k with k, and, hence, a parabola for $\{(u^d_{i_d}, [B^d\boldsymbol{\phi}_d]_{i_d}) : i_d = 1, \dots, I_d\}$. This yields a normal distribution for X^d. The addition of a multiple of the identity matrix, ϵI_d, to the matrix $D^{dT}_r D^d_r$ of rank $(K_d - r)$ just ensures that the precision matrix P^d_r is full rank and a proper conditional prior for $\boldsymbol{\phi}^d$. We refer to Lambert and Eilers (2009) for more details on the chosen specification for the marginal distributions.

If a parametric bivariate copula $C_{12}(\cdot, \cdot|\kappa)$ on $[0, 1]^2$ with dependence parameter $\kappa \in \mathcal{K}$ is assumed to model the dependence structure of (X^1, X^2), then one has, with respect to the upper bound of the small rectangular cells,

$$\Pr(X^1 \le x^1_{i_1}, X^2 \le x^2_{i_2}) = C_{12}(\pi^1_{\le i_1,+}, \pi^2_{+,\le i_2}|\kappa),$$

where

$$\pi^1_{\le i_1,+} = \sum_{i=1}^{i_1} \pi^1_{i,+} \quad ; \quad \pi^2_{+,\le i_2} = \sum_{i=1}^{i_2} \pi^2_{+,i}.$$

Hence, the probability that (X^1, X^2) is observed within $\mathcal{I}_{i_1} \times \mathcal{I}_{j_2}$ is

$$\begin{aligned} \pi_{i_1 i_2} =& \Pr(x^1_{i_1-1} \le X^1 \le x^1_{i_1}, x^2_{i_2-1} \le X^2 \le x^2_{i_2}) \\ =& C_{12}(\pi^1_{\le i_1,+}, \pi^2_{+,\le i_2}|\kappa) + C_{12}(\pi^1_{\le (i_1-1),+}, \pi^2_{+,\le (i_2-1)}|\kappa) \\ & - C_{12}(\pi^1_{\le (i_1-1),+}, \pi^2_{+,\le i_2}|\kappa) - C_{12}(\pi^1_{\le i_1,+}, \pi^2_{+,\le (i_2-1)}|\kappa). \end{aligned} \tag{2.4}$$

The bivariate composite link model gives, through (2.2), the wide bin probabilities.

2.2 A nonparametric model for the bivariate density

If one is not ready to make a strong parametric assumption on the dependence structure, a description of the bivariate density using Kronecker products of marginal B-splines bases is an alternative.

Using the same notation as in the previous section for the B-splines bases, one could take

$$\pi_{i_1 i_2} = \frac{\mathrm{e}^{\eta_{i_1 i_2}}}{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \mathrm{e}^{\eta_{i_1 i_2}}}, \tag{2.5}$$

where

$$\eta_{i_1 i_2} = \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} \phi_{k_1 k_2}\ b^1_{k_1}(u^1_{i_1})\ b^2_{k_2}(u^2_{i_2}). \tag{2.6}$$

This can be rewritten, using matrix notation, as

$$\{\eta_{i_1 i_2}\} = B^1 \Phi (B^2)^T \ \ \text{(or equivalently : vec}\,(\{\eta_{i_1 i_2}\}) = (B^2 \otimes B^1)\text{vec}\,(\Phi)),$$

where $\Phi = \{\phi_{k_1 k_2}\}$ is the $K_1 \times K_2$ matrix of B-spline coefficients.

As we take a large number of B-splines along each axis, one should penalize for overfitting as the proposed model would, otherwise, provide a too wiggly bivariate density. We penalize changes in the rth order differences of the B-splines along the two axes directions.

More, specifically, denote by $\boldsymbol{\phi}_{k_1:}$ and $\boldsymbol{\phi}_{:k_2}$ the k_1th row and k_2th column of Φ, respectively. Then, in a frequentist setting, roughness penalties for rth order changes in the row and column spline coefficients could be

$$Pen_1 = 0.5\ \tau^1 \sum_{k_2=1}^{K_2} \boldsymbol{\phi}^T_{:k_2} P^1_r \boldsymbol{\phi}_{:k_2} = 0.5\ \tau^1\ \text{vec}\,(\Phi)^T\,(I_{K_2} \otimes P^1_r)\text{vec}\,(\Phi)\,,$$

$$Pen_2 = 0.5\ \tau^2 \sum_{k_1=1}^{K_1} \boldsymbol{\phi}_{k_1:} P^2_r \boldsymbol{\phi}^T_{k_1:} = 0.5\ \tau^2\ \text{vec}\,(\Phi)^T\,(P^2_r \otimes I_{K_1})\text{vec}\,(\Phi)\,,$$

with the same definition for the penalty matrices as in the previous section. These penalties are subtracted from the log-likelihood for given values of the penalty coefficients τ^1 and τ^2 to define the penalized log-likelihood.

In a Bayesian setting, this translates into a joint prior for Φ:

$$\begin{aligned} p(\Phi|\tau^1,\tau^2) &\propto \exp\{-Pen_1 - Pen_2\} \\ &\propto \left|\mathcal{P}(\tau^1,\tau^2)\right|^{1/2} \exp\left\{-\frac{1}{2}\text{vec}\,(\Phi)^T\,\mathcal{P}(\tau^1,\tau^2)\text{vec}\,(\Phi)\right\}, \end{aligned} \tag{2.7}$$

where

$$\mathcal{P}(\tau^1,\tau^2) = \tau^1(I_{K_2} \otimes P^1_r) + \tau^2(P^2_r \otimes I_{K_1}).$$

Using the following notation for the eigenvalues of P^d_r,

$$\lambda^d_{[1]} > \ldots > \lambda^d_{[K_d - r]} > \lambda^d_{[K_d - r + 1]} = \ldots = \lambda^d_{[K_d]} = \epsilon,$$

one can show that the $K_1 K_2$ eigenvalues of $\mathcal{P}(\tau^1,\tau^2)$ are

$$\left\{\lambda_{k_1 k_2} = \tau^1 \lambda^1_{[k_1]} + \tau^2 \lambda^2_{[k_2]} : k_d = 1,\ldots,K_d \ \ (d = 1,2)\right\}.$$

Copula	$C_\kappa(u,v)$	$\mathcal{K}:\kappa$ range	Kendall's τ
Clayton	$\max\left\{\left(u^{-\kappa}+v^{-\kappa}-1\right)^{-1/\theta},0\right\}$	$[-1,0)\cup(0,+\infty)$	$\frac{\kappa}{2+\kappa}$
Frank	$-\frac{1}{\kappa}\log\left(1+\frac{(\mathrm{e}^{-\kappa u}-1)(\mathrm{e}^{-\kappa v}-1)}{\mathrm{e}^{-\kappa}-1}\right)$	$(-\infty,0)\cup(0,+\infty)$	No closed form
Gumbel	$\exp\left(-\left((-\log u)^\kappa+(-\log v)^\kappa\right)^{1/\kappa}\right)$	$[1,+\infty)$	$\frac{\kappa-1}{\kappa}$
Normal	$\Phi_2\left(\Phi_1^{-1}(u),\Phi_1^{-1}(v);\kappa\right)$	$(-1,1)$	$\frac{2}{\pi}\arcsin(\kappa)$

Table 1: Examples of copulas with dependence parameter κ. When no closed form is available for τ, one can evaluate $\tau = 4\int_0^1\int_0^1 C_{12}(u,v)\ dC_{12}(u,v) - 1$ numerically. Notation: $\Phi_1(\cdot)$ is the univariate normal cdf and $\Phi_2(\cdot,\cdot;\kappa)$ the standard bivariate normal cdf with correlation κ.

Therefore, the determinant of $\mathcal{P}(\tau^1,\tau^2)$ in (2.7) is

$$\left|\mathcal{P}(\tau^1,\tau^2)\right| = \prod_{k_1=1}^{K_1}\prod_{k_2=1}^{K_2}\lambda_{k_1k_2}.$$

In the special case where one assumes that $\tau^1=\tau^2=\tau$, one has

$$\left|\mathcal{P}(\tau)\right| \propto \tau^{K_1K_2}.$$

3 Bayesian model and conditional posteriors

In the preceding section, two models based on B-splines were proposed for a bivariate density. For the first one, flexible marginal densities with a parametric copula specifying the dependence structure were assumed. The smoothness of the modelled marginal densities was forced by smoothness priors for the splines coefficients. For the second model, Kronecker products of B-splines were used to specify the joint density. Again, smoothness was ensured with a suitable prior for the splines coefficients. In both cases, the smoothness priors depend on one or more penalty coefficients.

Parametric copula model with flexible margins

Large variance priors for τ^1 and τ^2 are chosen:

$$\tau^1 \sim \mathrm{Exp}(b_1=10^{-4}) \ \ ; \ \ \tau^2 \sim \mathrm{Exp}(b_2=10^{-4}),$$

where $\mathrm{Exp}(b)$ denotes the exponential distribution with mean b^{-1}.

An (possibly improper) uniform prior on $\mathcal{K}$ is taken for the dependence parameter κ of the parametric copula, see Table 1 for some examples.

Therefore, with the extra identifiability constraint that $\sum_{k_d=1}^{K_d}\phi^d_{k_d}=0$ (as ($\boldsymbol{\phi}^d$+const) and $\boldsymbol{\phi}^d$ provide exactly the same cell probabilities), the full Bayesian

model is given by

$$
\begin{aligned}
&(M_{11},\ldots,M_{J_1J_2}|\kappa,\boldsymbol{\phi}^1,\boldsymbol{\phi}^2) \sim \text{Mult}(m_{++};\gamma_{11},\ldots,\gamma_{J_1J_2}),\\
&p(\boldsymbol{\phi}^d|\tau^d) \propto \tau^{K_d/2}\exp\{-0.5\ \tau^d\ (\boldsymbol{\phi}^d)^T P_r^d \boldsymbol{\phi}^d\},\\
&\tau^d \sim \text{Exp}(b_d = 10^{-4}) \quad (d=1,2),\\
&p(\kappa) \propto 1 \times I_{\mathcal{K}}(\kappa),
\end{aligned}
$$

where $\gamma_{j_1j_2}$ is related to κ, $\boldsymbol{\phi}^1$ and $\boldsymbol{\phi}^2$ through (2.3), (2.4) and (2.2).

Denote by $L(\boldsymbol{\phi}^1,\boldsymbol{\phi}^2,\kappa|\mathcal{D})$ the likelihood

$$
L(\boldsymbol{\phi}^1,\boldsymbol{\phi}^2,\kappa|\mathcal{D}) \propto \prod_{j_1=1}^{J_1}\prod_{j_2=1}^{J_2}\gamma_{j_1j_2}^{m_{j_1j_2}}, \tag{3.1}
$$

where $\mathcal{D}$ stands for the available data. Then, the conditional posteriors are

$$
\begin{aligned}
&p(\boldsymbol{\phi}^d|\boldsymbol{\phi}^{-d},\kappa,\tau^d,\mathcal{D}) \propto L(\boldsymbol{\phi}^1,\boldsymbol{\phi}^2,\kappa|\mathcal{D}) \times (\tau^d)^{K_d/2}\exp\left(-\frac{1}{2}\ \tau^d(\boldsymbol{\phi}^d)^T P_r^d\boldsymbol{\phi}^d\right),\\
&p(\kappa|\boldsymbol{\phi}^1,\boldsymbol{\phi}^2,\mathcal{D}) \propto L(\boldsymbol{\phi}^1,\boldsymbol{\phi}^2,\kappa|\mathcal{D}) \times I_{\mathcal{K}}(\kappa),\\
&(\tau^d|\boldsymbol{\phi}^d,\mathcal{D}) \sim \mathcal{G}\left(1+K_d/2,\ b_d+(\boldsymbol{\phi}^d)^T P_r^d\boldsymbol{\phi}^d/2\right) \qquad (d=1,2),
\end{aligned} \tag{3.2}
$$

where $\boldsymbol{\phi}^{-d}$ refers to $\boldsymbol{\phi}^2$ (resp. $\boldsymbol{\phi}^1$) when $d=1$ (resp. $d=2$). These expressions will be used below to set up a Metropolis-within-Gibbs algorithm.

Nonparametric model for the bivariate density

Like in the preceding semi-parametric specification, large variance priors for τ^1 and τ^2 are chosen:

$$
\tau^1 \sim \text{Exp}(b_1 = 10^{-4}) \ \ ; \ \ \tau^2 \sim \text{Exp}(b_2 = 10^{-4}),
$$

where $\text{Exp}(b)$ denotes the exponential distribution with mean b^{-1}.

Then, with the extra identifiability constraint that $\sum_{k_1=1}^{K_1}\sum_{k_1=1}^{K_2}\phi_{k_1k_2}=0$ (as $(\boldsymbol{\phi}^d+\text{const})$ and $\boldsymbol{\phi}^d$ provide exactly the same cell probabilities), the full Bayesian model is given by

$$
\begin{aligned}
&(M_{11},\ldots,M_{J_1J_2}|\Phi) \sim \text{Mult}(m_{++};\gamma_{11},\ldots,\gamma_{J_1J_2}),\\
&p(\Phi|\tau^1,\tau^2) \propto \left|\mathcal{P}(\tau^1,\tau^2)\right|^{1/2}\exp\left\{-\frac{1}{2}\text{vec}\,(\Phi)^T\,\mathcal{P}(\tau^1,\tau^2)\text{vec}\,(\Phi)\right\},\\
&\tau^d \sim \text{Exp}(b_d = 10^{-4}) \quad (d=1,2),
\end{aligned}
$$

where $\gamma_{j_1j_2}$ is related to Φ through (2.5), (2.6) and (2.2).

If $L(\Phi|\mathcal{D})$ denotes the likelihood

$$
L(\Phi|\mathcal{D}) \propto \prod_{j_1=1}^{J_1}\prod_{j_2=1}^{J_2}\gamma_{j_1j_2}^{m_{j_1j_2}}, \tag{3.3}
$$

then the conditional posteriors are

$$p(\Phi|\tau^1,\tau^2,\mathcal{D}) \propto L(\Phi|\mathcal{D}) \times \exp\left\{-\frac{1}{2}\text{vec}\,(\Phi)^T\,\mathcal{P}(\tau^1,\tau^2)\text{vec}\,(\Phi)\right\}, \tag{3.4}$$

$$p(\tau^1|\Phi,\tau^2,\mathcal{D}) \propto \left|\mathcal{P}(\tau^1,\tau^2)\right|^{1/2} \exp\left\{-\tau^1\left(b_1+\frac{1}{2}\text{vec}\,(\Phi)^T\,(I_{K_2}\otimes P_r^1)\text{vec}\,(\Phi)\right)\right\},$$

$$p(\tau^2|\Phi,\tau^1,\mathcal{D}) \propto \left|\mathcal{P}(\tau^1,\tau^2)\right|^{1/2} \exp\left\{-\tau^2\left(b_2+\frac{1}{2}\text{vec}\,(\Phi)^T\,(P_r^2\otimes I_{K_1})\text{vec}\,(\Phi)\right)\right\}.$$

In the special case where one takes the same roughness penalty for the two axes, $\tau^1=\tau^2=\tau$ (and hyperprior parameters $b_1=b_2=b$), one has

$$p(\Phi|\tau,\mathcal{D}) \propto L(\Phi|\mathcal{D}) \times \exp\left\{-\frac{1}{2}\text{vec}\,(\Phi)^T\,\mathcal{P}(\tau,\tau)\text{vec}\,(\Phi)\right\}, \tag{3.5}$$

$$(\tau|\Phi,\mathcal{D}) \sim \mathcal{G}\left(1+\frac{1}{2}K_1K_2,\ b+\frac{1}{2}\text{vec}\,(\Phi)^T\left(I_{K_2}\otimes P_r^1+P_r^2\otimes I_{K_1}\right)\text{vec}\,(\Phi)\right).$$

These equations will be used below to set up a Metropolis-within-Gibbs algorithm.

4 Exploring the joint posterior distribution

Markov chain Monte Carlo (MCMC) will be used to generate from the joint posterior. At each iteration, the spline and the penalty parameters will be sampled from their conditional distributions using a full Metropolis or a Metropolis-within-Gibbs algorithm. As the variance-covariance matrix of some of the proposals is related to an initial frequentist estimate, we shall first discuss how one can choose the initial conditions. Then, some details will be provided on the MCMC algorithm.

4.1 Starting values

Parametric copula model with flexible margins

Frequentist estimates for the marginal parameters can be obtained for given values of the penalty parameters. Following Lambert and Eilers (2009, Section 3.1), estimates for the splines parameters can be obtained separately for each margin. Indeed, these quantities can be seen as the regression parameters of a composite link model. Then, for the dth margin, one has a polytomous logistic regression of the marginal frequencies $\mathbf{m}^d$,

$$\mathbf{m}^1=(m_{1+},\ldots,m_{J_1+})' \ ;\ \mathbf{m}^2=(m_{+1},\ldots,m_{+J_2})',$$

on a "working" design matrix $X^d=(W^d)^{-1}C^dH^dB^d$, with

$$H^d=\text{diag}(m_{++}\boldsymbol{\pi}^d(1-\boldsymbol{\pi}^d)),\ W^d=\text{diag}(m_{++}\boldsymbol{\gamma}^d(1-\boldsymbol{\gamma}^d)),$$

where

$$\boldsymbol{\pi}^1=(\pi_{1+},\ldots,\pi_{I_1+})' \text{ and } \boldsymbol{\gamma}^1=(\gamma_{1+},\ldots,\gamma_{J_1+})',$$
$$\boldsymbol{\pi}^2=(\pi_{+1},\ldots,\pi_{+I_2})' \text{ and } \boldsymbol{\gamma}^2=(\gamma_{+1},\ldots,\gamma_{+J_1})'.$$

The well-known scoring algorithm leads, at iteration $(t+1)$, to

$$\boldsymbol{\phi}^d_{t+1} = \left((X^d_t)^T W^d_t X^d_t\right)^{-1} (X^d_t)^T (\boldsymbol{m}^d - m_{++}\boldsymbol{\gamma}^d_t + W^d_t X^d_t \boldsymbol{\phi}_t),$$

the subscript t referring to the current approximation. With the roughness penalty, the scoring algorithm becomes:

$$\boldsymbol{\phi}^d_{t+1} = \left((X^d_t)^T W^d_t X^d_t + \tau^d P^d_r\right)^{-1} (X^d_t)^T (\boldsymbol{m}^d - m_{++}\boldsymbol{\gamma}^d_t + W^d_t X^d_t \boldsymbol{\phi}_t). \quad (4.1)$$

At convergence, it yields the conditional maximum penalized likelihood estimate (MPLE) $\boldsymbol{\phi}^d_\infty$ of $\boldsymbol{\phi}^d$. Then, for a given τ^d, the variance-covariance of the conditional MPLE is given by

$$\Sigma^d_{\tau^d} = \left((X^d_\infty)^T W^d_\infty X^d_\infty + \tau^d P^d_r\right)^{-1}. \quad (4.2)$$

Cross-validation or information criteria like the AIC or BIC (Eilers and Marx, 1996) can be used to select the penalty parameters τ^1 and τ^2. These estimates could be used as initial conditions for the splines parameters in an MCMC algorithm. The effective dimension of the smoother can be computed from the trace of the hat matrix,

$$\mathrm{tr}\left(X^d_\infty \left((X^d_\infty)^T W^d_\infty X^d_\infty + \tau^d P^d_r\right)^{-1} (X^d_\infty)^T\right), \quad (4.3)$$

as suggested generically by Hastie and Tibshirani (1990) and specifically in the CLM context by Eilers (2007).

Nonparametric model for the bivariate density

The same strategy can be used in the nonparametric setting as soon as one realizes that estimating Φ from the matrix of frequencies $\{m_{j_1 j_2}\}$ is equivalent to estimating $\mathrm{vec}\,(\Phi)$ from the vector of frequencies $\mathrm{vec}\,(\{m_{j_1 j_2}\})$.

Consider Kronecker versions of the composing and B-splines matrices,

$$\mathcal{C} = C^2 \otimes C^1,\ \mathcal{B} = B^2 \otimes B^1,$$

and define $\mathcal{X} = \mathcal{W}^{-1}\mathcal{C}\mathcal{H}\mathcal{B}$ where

$$\mathcal{H} = \mathrm{diag}\left\{m_{++}\mathrm{vec}\,(\Pi)\,\mathrm{vec}\,(1-\Pi)\right\},\ \mathcal{W} = \mathrm{diag}\left\{m_{++}\mathrm{vec}\,(\Gamma)\,\mathrm{vec}\,(1-\Gamma)\right\}.$$

Then, conditionally on the penalties τ^1 and τ^2, the conditional MPLE for Φ can be obtained iteratively from

$$\mathrm{vec}\,(\Phi_{t+1}) = \left(\mathcal{X}^T_t \mathcal{W}_t \mathcal{X}_t + \mathcal{P}(\tau^1, \tau^2)\right)^{-1} \mathcal{X}^T_t \left(\mathrm{vec}\,(\{m_{j_1 j_2}\}) - m_{++}\mathrm{vec}\,(\Gamma_t) + \mathcal{W}_t \mathcal{X}_t \mathrm{vec}\,(\Phi_t)\right). \quad (4.4)$$

At convergence, the variance-covariance of the conditional MPLE, $\mathrm{vec}\,(\Phi_\infty)$, is given by

$$\Sigma_{\tau^1,\tau^2} = \left(\mathcal{X}^T_\infty \mathcal{W}_\infty \mathcal{X}_\infty + \mathcal{P}(\tau^1, \tau^2)\right)^{-1}. \quad (4.5)$$

Again, cross-validation or information criteria can be used to select the penalty parameters. The effective dimension of the smoother can be computed from the trace of the hat matrix,

$$\operatorname{tr}\left(\mathcal{X}_t\left(\mathcal{X}_t^T\mathcal{W}_t\mathcal{X}_t+\mathcal{P}(\tau^1,\tau^2)\right)^{-1}\mathcal{X}_t^T\right). \tag{4.6}$$

4.2 Sampling algorithm

Whatever the selected model, a blockwise Metropolis-Hastings algorithm is proposed. Starting values for the parameters can be obtained from Section 4.1.

Parametric copula model with flexible margins

For values of the penalty parameters, τ_0^1 and τ_0^2, selected with BIC, one can use (4.1) to obtain initial values for the spline parameters, $\boldsymbol{\phi}_0^1$ and $\boldsymbol{\phi}_0^2$.

At iteration m of the MCMC algorithm, given the state of chain at the end of the previous iteration, $\boldsymbol{\theta}_{m-1}=\left(\boldsymbol{\phi}_{m-1}^1,\boldsymbol{\phi}_{m-1}^2,\kappa_{m-1},\tau_{m-1}^1,\tau_{m-1}^2\right)$:

1. *Multivariate Metropolis steps:* Each margin in turn ($d=1,2$), generate a proposal $\boldsymbol{\phi}^d$ from the multivariate normal distribution $\mathcal{N}_{K_d}\left(\boldsymbol{\phi}_{m-1}^d,\delta^d\Sigma_{\tau_0^d}^d\right)$ where $\Sigma_{\tau_0^d}^d$ is the variance-covariance matrix of the conditional MPLE of $\boldsymbol{\phi}^d$, see (4.2), and $\delta^d>0$ a tuning parameter selected to achieve the desired acceptance rate (20%, say).

   - Set $\boldsymbol{\phi}_m^1$ equal to the proposal $\boldsymbol{\phi}^1$ with probability

   $$\min\left\{1,\frac{p(\boldsymbol{\phi}^1|\boldsymbol{\phi}_{m-1}^2,\kappa_m,\tau_{m-1}^1,\mathcal{D})}{p(\boldsymbol{\phi}_{m-1}^1|\boldsymbol{\phi}_{m-1}^2,\kappa_m,\tau_{m-1}^1,\mathcal{D})}\right\},$$

     where the involved conditional posterior is given by (3.2). Let $\boldsymbol{\phi}_m^1=\boldsymbol{\phi}^1$ in case of acceptance and $\boldsymbol{\phi}_m^1=\boldsymbol{\phi}_{m-1}^1$ otherwise.
   - Set $\boldsymbol{\phi}_m^2$ equal to the proposal $\boldsymbol{\phi}^2$ with probability

   $$\min\left\{1,\frac{p(\boldsymbol{\phi}^2|\boldsymbol{\phi}_m^1,\kappa_m,\tau_{m-1}^2,\mathcal{D})}{p(\boldsymbol{\phi}_{m-1}^2|\boldsymbol{\phi}_m^1,\kappa_m,\tau_{m-1}^2,\mathcal{D})}\right\}.$$

     Let $\boldsymbol{\phi}_m^2=\boldsymbol{\phi}^2$ in case of acceptance and $\boldsymbol{\phi}_m^2=\boldsymbol{\phi}_{m-1}^2$ otherwise.

2. *Univariate Metropolis step:* Generate a proposal κ from the truncated normal distribution $\mathcal{N}_1(\kappa^{m-1},\sigma_\kappa^2)\ I_{\mathcal{K}}(\kappa)$, where σ_κ^2 is selected to achieve the desired acceptance rate (40%, say). Set κ_m equal to the proposal κ with probability

   $$\min\left\{1,\frac{p(\kappa|\boldsymbol{\phi}_m^1,\boldsymbol{\phi}_m^2,\mathcal{D})}{p(\kappa_{m-1}|\boldsymbol{\phi}_m^1,\boldsymbol{\phi}_m^2,\mathcal{D})}\right\}.$$

   Let $\kappa_m=\kappa$ in case of acceptance and $\kappa_m=\kappa_{m-1}$ otherwise.

3. *Gibbs step:* Generate τ_m^d from $\mathcal{G}\left(1+K_d/2,\ b_d+(\boldsymbol{\phi}_m^d)^TP_r^d\boldsymbol{\phi}_m^d/2\right)$.

Nonparametric model for the bivariate density

For values of the penalty parameters, τ_0^1 and τ_0^2, selected with BIC, one can use (4.4) to obtain initial values for the spline parameters, Φ_0.

At iteration m of the MCMC algorithm, given the state of chain at the end of the previous iteration, $\boldsymbol{\theta}_{m-1} = \left(\boldsymbol{\Phi}_{m-1}, \tau_{m-1}^1, \tau_{m-1}^2\right)$:

1. *Multivariate Metropolis step:* Generate a proposal $\mathrm{vec}\,(\Phi)$ from the multivariate normal distribution $\mathcal{N}_{K_1K_2}\left(\mathrm{vec}\,(\Phi)_{m-1}, \delta\Sigma_{\tau_0^1,\tau_0^2}\right)$ where $\Sigma_{\tau_0^1,\tau_0^2}$ is the variance-covariance matrix of the conditional MPLE of $\mathrm{vec}\,(\Phi)$, see (4.5), and $\delta > 0$ a tuning parameter selected to achieve the desired acceptance rate (20%, say).

   Set Φ_m equal to the proposal Φ with probability

$$\min\left\{1, \frac{p\left(\Phi\middle|\tau_{m-1}^1, \tau_{m-1}^2, \mathcal{D}\right)}{p\left(\Phi_{m-1}\middle|\tau_{m-1}^1, \tau_{m-1}^2, \mathcal{D}\right)}\right\},$$

   where the involved conditional posterior is given by (3.4). Let $\boldsymbol{\phi}_m^1 = \boldsymbol{\phi}^1$ in case of acceptance and $\boldsymbol{\phi}_m^1 = \boldsymbol{\phi}_{m-1}^1$ otherwise.

2. *Univariate Metropolis steps:* For each penalty parameter in turn ($d = 1, 2$), generate a proposal $\log \tau^d$ from the normal distribution $\mathcal{N}_1\left(\log \tau_{m-1}^d, \sigma_{\tau^d}^2\right)$, where $\sigma_{\tau^d}^2$ is selected to achieve the desired acceptance rate (40%, say).

   - Set τ_m^1 equal to τ^1 with probability

$$\min\left\{1, \frac{\tau^1 \times p(\tau^1|\Phi_m, \tau_{m-1}^2, \mathcal{D})}{\tau_{m-1}^1 \times p(\tau_{m-1}^1|\Phi_m, \tau_{m-1}^2, \mathcal{D})}\right\},$$

     where the conditional posterior is given by (3.4). Let $\tau_m^1 = \tau^1$ in case of acceptance and $\tau_m^1 = \tau_{m-1}^1$ otherwise.

   - Set τ_m^2 equal to τ^2 with probability

$$\min\left\{1, \frac{\tau^2 \times p(\tau^2|\Phi_m, \tau_m^1, \mathcal{D})}{\tau_{m-1}^2 \times p(\tau_{m-1}^2|\Phi_m, \tau_m^1, \mathcal{D})}\right\},$$

     where the conditional posterior is given by (3.4). Let $\tau_m^2 = \tau^2$ in case of acceptance and $\tau_m^2 = \tau_{m-1}^2$ otherwise.

   In the special case where one assumes that $\tau_1 = \tau_2 = \tau$, these two Metropolis steps are replaced by a Gibbs step: given $\Phi = \Phi_m$, generate τ_m from the gamma distribution in (3.5).

For both models, after a sufficiently large number of burnin iterations, the generated sequence of $\boldsymbol{\theta}_m$ can be seen as a random sample from the joint posterior distribution. For simplicity, we shall denote the sequence after the burnin by $\{\boldsymbol{\theta}_m : m = 1, \ldots, M\}$, M being the length of the final chain.

The choice of the tuning parameters in the Metropolis steps, δ^d, $\sigma_{\tau^d}^2$ ($d = 1, 2$), δ and σ_κ^2 can be automated (Haario *et al.*, 2001) during the burnin.

5 Simulation study

A simulation study was made to evaluate the performances of these two models to estimate a bivariate density. Margin 1 was assumed to be a unimodal Beta(3,5) distribution and margin 2 to be a (bimodal) mixture of a Beta(3,10) (with weight 0.6) and of a Beta(12,8) (with weight 0.4).

$S = 100$ datasets of size $n = 200$ and $n = 1000$ were generated from bivariate distributions with the above margins and a dependence structure corresponding to a Clayton, Frank or Gumbel copula with a small ($\tau = 0.2$), medium ($\tau = 0.5$) or large ($\tau = 0.7$) Kendall's tau.

For each dataset, grouped (rectangle) data were obtained by partitioning each marginal support (naively) taken to be $(-0.15, 1.00)$ into ($J_1 = J_2 =$) 6, 8, 12 bins of equal widths, yielding ($J_1 \times J_2 =$) 36, 64 or 144 rectangles on $(-0.15, 1.00)^2$ with associated frequencies $\{n_{j_1,j_2} : j_1 = 1, \ldots, J_1; j_2 = 1, \ldots, J_2\}$.

For each of these 1800 ($= 2 \times 3 \times 3 \times S = 1800$) datasets, 2 types of models were fitted to estimate the underlying density:

- Model 1: a parametric copula (Clayton, Frank, Gumbel or normal) model with unknown dependence parameter and smooth nonparametric margins, see Section 2.1. Forty-eight small bins of equal width, 20 equidistant knots and a penalty of order 3 were considered for each margin, yielding $K_d = 23$ spline parameters $\boldsymbol{\phi}^d$ for margin d ($d = 1, 2$).

- Model 2: the nonparametric model of Section 2.2 based on the bivariate extension of the composite link model. Forty-eight small bins of equal width, 10 B-splines associated to equidistant knots and a penalty of order 3 were considered for each margin, yielding $K_1 \times K_2 = 10^2$ spline parameters $\Phi = \{\phi_{k_1 k_2}\}$.

Samples from the respective joint posteriors were obtained using the strategy described in Section 4.

- For Model 1, the length of the chain was $M = 20000$ after a burnin of 1000. Such a short burnin turned to be suitable as the initial conditions were carefully chosen and the dependence structure of the conditional posteriors was well approximated by $\Sigma^d_{\tau_0^d}$, see Section 4.1.

- For Model 2, a first chain of length 50000 following a burnin of 10000 iterations was built to improve the initial values for the two penalty parameters τ^1 and τ^2 necessary to estimate the variance-covariance matrix of the conditional MPLE. Then, that empirical variance-covariance matrix was used in the multivariate Metropolis step to build a final chain of length 200000 (after an extra burnin of 10000 iterations).

At iteration m of the MCMC algorithm:

- For Model 1: given $\boldsymbol{\theta}_m = (\boldsymbol{\phi}^1_m, \boldsymbol{\phi}^2_m, \tau^1_m, \tau^2_m)$, the corresponding marginal densities at the small bin midpoints are $f_1(u^1_{i_1}|\boldsymbol{\theta}_m)\Delta_1 = \pi_{i_1+}(\boldsymbol{\phi}^1_m)$ and

$f_2(u^2_{i_2}|\boldsymbol{\theta}_m)\Delta_2 = \pi_{+i_2}(\boldsymbol{\phi}^2_m)$, see (2.3), and the joint density at the small rectangle midpoints is $f_{12}(u^1_{i_1}, u^2_{i_2}|\boldsymbol{\theta}_m)\Delta_1\Delta_2 = \pi_{i_1 i_2}(\boldsymbol{\phi}^1_m, \boldsymbol{\phi}^2_m, \kappa_m)$, see (2.4). The estimated marginals were computed to be $\hat{f}_d(u^d_{i_d}) = \frac{1}{M}\sum_{m=1}^M f_d(u^d_{i_d}|\boldsymbol{\theta}_m)$ for margin d, while the estimated bivariate density was computed as $\hat{f}_{12}(u^1_{i_1}, u^2_{i_2}) = \frac{1}{M}\sum_{m=1}^M f_{12}(u^1_{i_1}, u^2_{i_2}|\boldsymbol{\theta}_m)$. For the copula parameter, the estimate of the posterior mean is $\hat{\kappa} = \frac{1}{M}\sum_{m=1}^M \kappa_m$.

- For Model 2: given $\boldsymbol{\theta}_m = (\Phi_m, \tau^1_m, \tau^2_m)$, the joint density at the small rectangle midpoints is $f_{12}(u^1_{i_1}, u^2_{i_2}|\boldsymbol{\theta}_m)\Delta_1\Delta_2 = \pi_{i_1 i_2}(\Phi_m)$. An estimate for the joint density is given by $\hat{f}_{12}(u^1_{i_1}, u^2_{i_2}) = \frac{1}{M}\sum_{m=1}^M f_{12}(u^1_{i_1}, u^2_{i_2}|\boldsymbol{\theta}_m)$, see (2.5) and (2.6).

For each density estimate and each simulated dataset, one can compute the integrated squared error:

$$\begin{aligned} ISE &= \int_0^1\int_0^1 \Big(f_{12}(x,y) - \hat{f}_{12}(x,y)\Big)^2 f_{12}(x,y)dxdy \\ &\approx \frac{1}{I_1 I_2}\sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2}\Big(f_{12}\left(u^1_{i_1}, u^2_{i_2}\right) - \hat{f}_{12}\left(u^1_{i_1}, u^2_{i_2}\right)\Big)^2 f_{12}\left(u^1_{i_1}, u^2_{i_2}\right)\Delta_1\Delta_2. \end{aligned}$$

The estimate with the smallest ISE is to be preferred.

The (square root of the) mean value of ISE (RMISE) over the $S = 100$ simulated datasets is reported in the left panel of Tables 2, 3 and 4 when the numbers of bigs bins, $J_1 = J_2$, are 12, 8 and 6, respectively. The copula used for data generation is indicated in the row labels (for low, medium or large Kendall's tau), whereas the assumption made for the dependence structure of the fitted density corresponds to the column label. The smallest and non significantly different values of RMISE are in bold face, while the second smallest value(s) is (are) underlined when there is only one bolded value.

Not surprisingly, the parametric model making the correct assumption for the copula is always associated to (the set of) smallest RMISE value(s). Whatever the number of big bins or sample size, the second best fit can be associated to a parametric copula or to the nonparametric estimate. There is no clear prior indication whether one should prefer a parametric or a nonparametric estimate.

Unfortunately, the RMISE is not available in practice for model selection as it requires the knowledge of the true bivariate density. A model selection tool should be computable from the observed frequencies $\{n_{j_1,j_2} : j_1 = 1,\ldots,J_1; j_2 = 1,\ldots,J_2\}$ for the $J_1 \times J_2$ big bins. Selecting the model with the smallest AIC (Akaike's information criterion, Akaike, 1974) or BIC (Bayesian information criterion, Schwarz, 1978) are workable strategies:

$$\begin{aligned} \text{AIC} &= -2\log L + 2p_{\text{eff}}, \\ \text{BIC} &= -2\log L + \log(n)p_{\text{eff}}, \end{aligned}$$

where L is the likelihood, see (3.1) and (3.3), and p_{eff} denotes the effective number of parameters. When one takes a parametric copula to describe the dependence structure, p_{eff} is one (copula parameter) plus the effective number

of parameters involved in the modelling of the two margins, see (4.3). For each margin, that quantity is computed as the trace of the smoother matrix at the posterior mean of the penalty and of the splines parameters. When a nonparametric model is used, the same formula is used with Kronecker products of the bases matrices and of the penalty matrices, see (4.6).

The mean ranks of AIC and BIC for the competing models are given in the middle and right parts of Tables 2, 3 and 4. Whatever the sample size, the amount of grouping and the copula used to generate the data, one can see that the smallest and second smallest BIC (see bolded and underlined values) nearly always rightly point the models with the smallest and second smallest RMISE (see bolded and underlined values for BIC). This is usually not true with AIC that tends to select too complex models.

Estimates of Kendall's tau were also produced. When a copula model is fitted, the one-to-one relationship between the copula parameter and τ directly yields an estimate for it, see Table 1. For a nonparametric density estimate, the relationship between the (fitted) bivariate density and (the population) Kendall's tau provides the desired estimation. Indeed, one has

$$\tau(X,Y) = 4 \int\int \bar{F}_{12}(x,y)\ f_{12}(x,y)\ dxdy - 1.$$

where $\bar{F}_{12}(x,y) = \Pr(X \geq x, Y \geq y)$ is the joint survival distribution. If $\hat{\pi}_{st}$ is the fitted probability to belong to the fine grid cell with midpoint (u_s^1, u_t^2), then one can estimate $\bar{F}_{12}$ at that midpoint by

$$\widehat{\bar{F}}_{12}(u_s^1, u_t^2) = \sum_{i_1=s}^{I_1}\sum_{i_2=t}^{I_2} \hat{\pi}_{i_1 i_2} - \frac{1}{2}\sum_{i_1=s}^{I_1} \hat{\pi}_{i_1 t} - \frac{1}{2}\sum_{i_2=t}^{I_2} \hat{\pi}_{s i_2} + \frac{1}{4}\hat{\pi}_{st}.$$

This suggests the following estimate for $\tau(X,Y)$:

$$\hat{\tau}(X,Y) = 4 \sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2} \widehat{\bar{F}}_{12}(u_{i_1}^1, u_{i_2}^2)\ \hat{\pi}_{i_1 i_2} - 1.$$

The (square root of the) MSE, bias and standard error of these estimators for τ, as well as the observed coverage of 90% credible intervals for τ computed over the $S = 100$ simulated datasets are reported in Tables 5, 6 and 7 when the numbers of bigs bins, $J_1 = J_2$, are 12, 8 and 6, respectively.

Not surprisingly, RMSE is smallest when the correct parametric copula is assumed to model the dependence structure. A wrong choice for the parametric copula can provide good or bad performances for the estimation of tau. A poor estimation for tau usually happens when, for example, one assumes a copula with strong left tail dependence (e.g. Clayton copula) when the dependence is strong in the right tail (e.g. Gumbel copula).

When the true Kendall's tau is small ($\tau = 0.2$), the RMSE of the nonparametric estimate is not significantly different from the RMSE for the correct parametric choice. The observed coverages for the 90% credible intervals are compatible (see percentages in bold) with the targeted nominal level.

	True copula	τ	Assumed copula: C	F	G	N	NP	C	F	G	N	NP	C	F	G	N	NP
			RMISE					AIC rank					BIC rank				
$n = 1000$	C	0.2	**0.33**	0.52	0.57	0.48	0.41	**1.0**	3.9	5.0	3.0	2.0	**1.0**	3.0	4.8	2.1	4.2
		0.5	**0.87**	1.86	2.18	1.94	1.42	**1.0**	3.6	5.0	3.4	2.0	**1.0**	3.6	5.0	3.3	2.0
		0.7	**1.69**	4.12	4.99	4.61	3.74	**1.1**	3.0	5.0	4.0	1.9	**1.0**	3.0	5.0	4.0	2.0
	F	0.2	0.39	**0.29**	0.38	**0.32**	0.35	4.6	**1.3**	4.2	2.4	2.4	3.7	**1.1**	3.3	2.0	4.8
		0.5	0.83	**0.47**	0.66	0.63	**0.51**	5.0	1.6	4.0	3.0	**1.4**	4.9	**1.0**	4.10	2.6	2.4
		0.7	1.43	**0.71**	0.93	1.01	0.95	5.0	**1.4**	4.0	3.0	1.6	5.0	**1.0**	4.00	3.0	2.0
	G	0.2	0.39	0.34	**0.30**	**0.30**	0.36	5.0	3.9	**1.1**	2.8	2.2	4.8	3.1	**1.0**	2.0	4.1
		0.5	0.95	0.57	**0.42**	0.51	0.49	5.0	4.0	**1.1**	3.0	1.9	5.0	4.0	**1.0**	2.2	2.8
		0.7	1.86	1.06	**0.72**	0.96	0.93	5.0	4.0	**1.2**	3.0	1.8	5.0	4.0	**1.0**	2.4	2.6
$n = 200$	C	0.2	**0.57**	0.67	0.75	0.66	0.71	**1.4**	3.5	4.8	2.9	2.2	**1.2**	2.8	4.1	2.2	4.7
		0.5	**1.35**	2.19	2.54	2.33	2.36	**1.0**	3.5	5.0	3.4	2.1	**1.0**	2.7	5.0	2.6	3.7
		0.7	**3.12**	5.18	5.86	5.64	5.72	**1.0**	3.1	5.0	3.8	2.1	**1.0**	2.3	5.0	3.2	3.5
	F	0.2	0.63	**0.57**	0.65	**0.60**	0.64	4.0	2.3	4.1	2.9	**1.7**	3.2	**1.6**	3.4	2.2	4.6
		0.5	1.03	**0.81**	1.04	0.96	0.91	4.5	2.2	4.2	3.0	**1.2**	4.2	**1.5**	4.0	2.4	2.9
		0.7	1.77	**1.38**	1.57	1.62	1.55	4.5	1.8	4.3	3.2	**1.3**	4.3	**1.2**	4.3	2.8	2.5
	G	0.2	0.61	**0.53**	**0.53**	**0.53**	0.59	4.8	3.6	2.0	2.9	**1.7**	4.3	2.8	**1.4**	2.0	4.4
		0.5	1.16	0.89	**0.81**	**0.82**	0.89	5.0	3.9	**1.5**	2.8	1.8	5.0	3.6	**1.1**	2.1	3.1
		0.7	2.11	1.67	**1.33**	**1.43**	1.63	5.0	4.0	**1.5**	2.6	1.9	5.0	3.8	**1.1**	2.0	3.0

Table 2: Root mean integrated squared error (**RMISE**) and mean rank for AIC and BIC over the $S = 100$ simulated datasets for different combinations for the true copula (with small, medium or large Kendall's tau) when the dependence structure is modelled using a parametric copula ('C'=Clayton, 'F'=Frank, 'G'=Gumbel or 'N'=Normal) or a nonparametric (NP) method. The first and second significantly smallest values are, respectively, bolded and underlined. Sample size n with **12 big bins** of equal width along each axis.

	True copula	τ	C	F	G	N	NP	C	F	G	N	NP	C	F	G	N	NP
								Assumed copula									
								RMISE					AIC rank									BIC rank				
$n = 1000$	C	0.2	**0.34**	0.52	0.58	0.49	0.42	**1.1**	3.8	5.0	3.2	1.9	**1.0**	2.9	4.7	2.2	4.3
		0.5	**0.97**	1.94	2.31	2.07	1.63	**1.1**	3.4	5.0	3.6	1.9	**1.0**	3.4	5.0	3.5	2.0
		0.7	**1.91**	4.65	5.50	5.13	4.45	**1.0**	3.0	5.0	4.0	2.0	**1.0**	2.9	5.0	4.0	2.1
	F	0.2	0.41	**0.30**	0.39	0.35	0.35	4.6	**1.5**	4.3	2.6	2.0	3.7	**1.1**	3.3	2.0	4.8
		0.5	0.82	**0.49**	0.69	0.66	**0.54**	4.9	1.8	4.1	3.0	**1.2**	4.9	**1.0**	4.1	2.5	2.5
		0.7	1.64	**0.78**	1.01	1.06	1.07	5.0	1.7	4.0	3.0	**1.3**	4.9	**1.0**	4.1	2.9	2.1
	G	0.2	0.41	**0.34**	**0.31**	**0.31**	0.36	5.0	3.9	**1.2**	2.8	2.0	4.8	3.1	**1.1**	2.0	4.1
		0.5	0.95	0.59	**0.46**	0.55	0.53	5.0	4.0	**1.2**	3.0	1.8	5.0	4.0	**1.0**	2.2	2.8
		0.7	2.04	1.19	**0.79**	1.08	1.08	5.0	4.0	**1.5**	3.0	1.5	5.0	4.0	**1.0**	2.2	2.8
$n = 200$	C	0.2	**0.64**	0.71	0.81	0.72	0.75	**1.8**	3.6	4.7	3.0	1.9	**1.3**	2.8	4.2	2.3	4.3
		0.5	**1.47**	2.29	2.67	2.47	2.49	**1.1**	3.3	5.0	3.5	2.1	**1.0**	2.7	5.0	2.8	3.5
		0.7	**3.50**	5.71	6.47	6.24	6.29	**1.1**	3.0	5.0	3.8	2.1	**1.0**	2.3	5.0	3.3	3.4
	F	0.2	**0.70**	**0.64**	0.73	**0.68**	**0.69**	4.0	2.6	4.0	3.0	**1.4**	3.2	**1.9**	3.4	2.2	4.2
		0.5	1.12	**0.91**	1.18	1.09	**0.98**	4.5	2.2	4.2	3.0	**1.1**	4.4	**1.6**	4.0	2.4	2.6
		0.7	1.96	**1.57**	1.92	1.89	1.76	4.3	2.1	4.3	3.1	**1.1**	4.2	**1.4**	4.4	2.8	2.3
	G	0.2	0.70	**0.60**	**0.60**	**0.60**	**0.64**	4.7	3.8	2.2	2.8	**1.5**	4.3	3.0	**1.5**	2.0	4.1
		0.5	1.21	**0.95**	**0.88**	**0.91**	**0.93**	5.0	3.9	1.8	2.8	**1.5**	5.0	3.7	**1.2**	2.1	3.0
		0.7	2.29	1.80	**1.52**	**1.57**	1.76	5.0	4.0	**1.6**	2.7	1.8	5.0	3.6	**1.2**	2.0	3.2

Table 3: Root mean integrated squared error (**RMISE**) and mean rank for AIC and BIC over the $S = 100$ simulated datasets for different combinations for the true copula (with small, medium or large Kendall's tau) when the dependence structure is modelled using a parametric copula ('C'=Clayton, 'F'=Frank, 'G'=Gumbel or 'N'=Normal) or a nonparametric (NP) method. The first and second significantly smallest values are, respectively, bolded and underlined. Sample size n with **8 big bins** of equal width along each axis.

	True copula	τ	Assumed copula: RMISE C	F	G	N	NP	AIC rank C	F	G	N	NP	BIC rank C	F	G	N	NP
$n = 1000$	C	0.2	**0.47**	0.64	0.67	0.60	0.66	**1.1**	3.6	5.0	3.0	2.3	**1.0**	2.7	4.4	2.3	4.6
		0.5	**1.09**	2.16	2.41	2.17	1.98	**1.1**	3.1	5.0	3.9	1.9	**1.0**	2.9	5.0	3.8	2.3
		0.7	**2.28**	5.11	5.70	5.30	5.01	**1.2**	3.0	5.0	4.0	1.8	**1.0**	2.8	5.0	4.0	2.2
	F	0.2	**0.43**	**0.44**	0.47	**0.42**	0.57	4.4	**1.3**	4.1	2.2	2.9	3.6	**1.2**	3.3	1.9	4.9
		0.5	0.95	**0.72**	0.77	**0.67**	**0.72**	4.8	1.7	4.2	3.0	**1.3**	4.7	**1.0**	4.2	2.5	2.6
		0.7	2.60	**1.14**	**1.08**	**1.03**	1.28	4.7	1.8	4.3	3.0	**1.2**	4.7	**1.1**	4.3	3.0	2.0
	G	0.2	**0.46**	**0.48**	**0.47**	**0.43**	0.68	5.0	3.7	**1.2**	2.6	2.5	4.8	3.0	**1.1**	2.0	4.1
		0.5	1.05	0.75	**0.61**	**0.61**	0.73	5.0	4.0	**1.2**	3.0	1.8	5.0	4.0	**1.0**	2.2	2.8
		0.7	3.29	1.57	**1.06**	1.26	1.44	5.0	4.0	**1.5**	2.9	1.6	5.0	4.0	**1.1**	2.3	2.7
$n = 200$	C	0.2	**0.67**	0.77	0.84	0.77	0.85	2.0	3.4	4.7	3.3	**1.6**	**1.4**	2.7	4.3	2.5	4.2
		0.5	**1.68**	2.49	2.79	2.60	2.63	**1.5**	3.1	5.0	3.7	1.7	**1.1**	2.5	4.9	3.2	3.2
		0.7	**3.94**	6.00	6.68	6.40	6.45	**1.6**	2.9	5.0	3.8	1.8	**1.1**	2.4	5.0	3.5	3.0
	F	0.2	**0.72**	**0.69**	0.79	**0.70**	0.86	3.90	2.70	3.90	3.10	**1.40**	3.20	**2.00**	3.40	2.40	4.00
		0.5	1.16	**0.98**	1.12	**1.05**	**1.05**	4.30	2.40	4.20	3.00	**1.10**	4.00	**1.90**	4.10	2.50	2.60
		0.7	2.74	**1.72**	1.96	**1.75**	**1.84**	3.90	2.30	4.50	3.30	**1.10**	3.60	**1.90**	4.50	3.10	2.00
	G	0.2	**0.70**	**0.65**	**0.67**	**0.65**	0.81	4.7	3.6	2.4	2.9	**1.3**	4.3	3.0	**1.6**	2.2	3.9
		0.5	1.27	**1.04**	**0.97**	**0.95**	**1.02**	5.0	3.8	2.0	2.9	**1.2**	5.0	3.6	**1.4**	2.3	2.7
		0.7	2.77	2.11	**1.76**	**1.77**	1.99	5.0	3.9	2.3	2.7	**1.1**	5.0	3.8	**1.8**	2.2	2.2

Table 4: Root mean integrated squared error (**RMISE**) and mean rank for AIC and BIC over the $S = 100$ simulated datasets for different combinations for the true copula (with small, medium or large Kendall's tau) when the dependence structure is modelled using a parametric copula ('C'=Clayton, 'F'=Frank, 'G'=Gumbel or 'N'=Normal) or a nonparametric (NP) method. The first and second significantly smallest values are, respectively, bolded and underlined. Sample size n with **6 big bins** of equal width along each axis.

	True copula	Assumed copula																			
		τ	C	F	G	N	NP	C	F	G	N	NP	C	F	G	N	NP	C	F	G	N	NP
			$10^2 \times$ RMSE					$10^2 \times$ BIAS					$10^2 \times$ s.e.					Coverage of 90% C.I.				
$n = 1000$	C	0.2	**2.1**	**2.3**	4.1	**2.2**	**2.4**	1	-1	-3	0	-1	2.0	2.3	2.4	2.1	2.3	**87**	**83**	56	**87**	**83**
		0.5	**1.6**	**1.7**	2.2	**1.9**	**2.0**	0	0	-1	-1	-1	1.6	1.6	1.7	1.6	1.6	**89**	**87**	**82**	**85**	**83**
		0.7	**1.5**	**1.3**	2.9	2.9	2.6	0	0	-3	-3	-2	1.4	1.2	1.2	1.3	1.3	**84**	**82**	35	33	37
	F	0.2	3.8	**2.1**	3.1	**2.3**	**2.1**	-3	0	-2	-1	0	2.0	2.1	2.1	2.0	2.1	54	**89**	71	**84**	**86**
		0.5	5.7	**1.7**	3.3	3.4	**2.0**	-5	-1	-3	-3	-1	1.7	1.6	1.6	1.6	1.6	8	**86**	47	40	73
		0.7	4.3	**0.9**	2.7	3.0	<u>2.0</u>	-4	0	-3	-3	-2	1.1	0.9	0.9	1.0	0.9	2	**93**	30	17	43
	G	0.2	4.9	**2.1**	**2.0**	**2.1**	**2.1**	-4	0	0	1	0	2.2	2.1	2.0	2.0	2.1	32	**90**	**93**	**88**	**89**
		0.5	6.4	**1.8**	**1.5**	**1.5**	**1.7**	-6	1	0	0	-1	1.7	1.6	1.5	1.5	1.6	2	**82**	**92**	**89**	**88**
		0.7	4.4	1.6	**1.0**	**1.0**	1.4	-4	1	0	0	-1	1.3	1.0	1.0	1.0	1.0	5	64	**94**	**93**	76
$n = 200$	C	0.2	**5.3**	**5.1**	7.1	**5.0**	**5.1**	3	-1	-3	0	-2	4.5	5.1	6.3	5.0	4.8	**82**	**86**	81	**84**	**83**
		0.5	**3.4**	**3.4**	**4.1**	**3.7**	**5.0**	1	0	-2	-1	-4	3.1	3.4	3.6	3.4	3.4	**94**	**93**	**90**	**92**	69
		0.7	**2.9**	**2.9**	4.3	4.7	6.8	-1	-1	-3	-4	-6	2.8	2.7	2.7	2.7	2.6	**90**	**86**	74	61	16
	F	0.2	**5.1**	**4.9**	**6.3**	**5.1**	**5.1**	-1	-1	-3	-2	-2	4.9	4.8	5.5	4.8	4.7	**86**	**88**	**83**	**84**	**86**
		0.5	5.2	**3.4**	**4.3**	**4.3**	4.9	-4	-1	-3	-3	-4	3.3	3.3	3.2	3.3	3.3	70	**93**	**87**	**82**	69
		0.7	4.0	**2.2**	**3.0**	3.4	4.9	-3	0	-2	-2	-4	2.5	2.2	2.3	2.4	2.2	73	**91**	**86**	74	33
	G	0.2	**5.9**	**4.9**	**4.6**	**4.7**	**5.1**	-3	-1	0	0	-2	4.8	4.7	4.6	4.7	4.5	80	**89**	**90**	**92**	**85**
		0.5	5.9	**3.7**	**3.6**	**3.6**	**4.6**	-5	1	1	1	-3	3.7	3.6	3.6	3.5	3.7	62	**89**	**90**	**89**	81
		0.7	5.7	**2.7**	**2.6**	**2.6**	4.7	-5	1	0	0	-4	3.2	2.7	2.6	2.6	2.7	46	**82**	**89**	**86**	50

Table 5: Root mean squared error (RMSE), bias and standard error of a point estimate for **Kendall's tau** (for $S = 100$ simulated datasets) for different combinations for the true copula (with small, medium or large Kendall's tau) when the dependence structure is modelled using a parametric copula (Clayton, Frank, Gumbel or normal) or a nonparametric (NP) method. The first and second significantly smallest values for RMSE are, respectively, bolded and underlined. The last column gives the coverage of the 90% credible interval for Kendall's tau (bold values being compatible with the nominal level) obtained from the chain for the dependence parameter under a parametric copula assumption or from the M densities generated by MCMC for a given dataset under the NP method. Sample size n with **12 big bins** of equal width along each axis.

	True copula	Assumed copula																			
		τ	C	F	G	N	NP	C	F	G	N	NP	C	F	G	N	NP	C	F	G	N	NP
			$10^2 \times$ RMSE					$10^2 \times$ BIAS					$10^2 \times$ s.e.					Coverage of 90% C.I.				
$n = 1000$	C	0.2	**2.1**	**2.4**	4.4	**2.3**	**2.5**	0	-1	-4	-1	-1	2.1	2.3	2.4	2.2	2.4	**89**	**84**	47	**86**	**84**
		0.5	**1.8**	**1.8**	3.4	2.7	2.3	0	-1	-3	-2	-2	1.8	1.7	1.7	1.7	1.7	**86**	**87**	53	67	76
		0.7	**1.5**	**2.2**	5.0	4.5	3.4	0	-2	-5	-4	-3	1.5	1.3	1.3	1.3	1.3	**85**	56	1	2	22
	F	0.2	3.5	**2.0**	3.0	**2.2**	**2.1**	-3	0	-2	-1	0	2.0	2.0	2.1	2.1	2.1	58	**86**	76	**91**	**86**
		0.5	5.0	**1.7**	3.4	3.2	**2.2**	-5	-1	-3	-3	-2	1.8	1.6	1.7	1.6	1.6	13	**89**	46	44	62
		0.7	3.2	**1.2**	3.1	3.1	2.6	-3	0	-3	-3	-2	1.4	1.2	1.2	1.3	1.1	37	**94**	30	28	37
	G	0.2	4.6	**2.2**	**2.0**	**2.2**	**2.1**	-4	0	0	1	0	2.2	2.2	2.0	2.1	2.1	41	**89**	**92**	**83**	**90**
		0.5	5.9	**1.7**	**1.6**	**1.5**	**1.8**	-6	1	0	0	-1	1.7	1.6	1.6	1.5	1.6	8	**91**	**90**	**91**	**87**
		0.7	3.7	**1.3**	**1.1**	**1.3**	2.0	-3	1	0	0	-2	1.4	1.1	1.1	1.2	1.1	18	**87**	**93**	**85**	64
$n = 200$	C	0.2	**5.5**	**5.5**	**7.6**	**5.5**	**5.5**	2	-1	-3	0	-2	4.9	5.4	6.8	5.4	5.2	**84**	**86**	80	**83**	**81**
		0.5	**3.9**	**3.9**	**4.9**	**4.3**	5.8	1	0	-3	-2	-4	3.6	3.9	3.9	3.7	3.7	**90**	**90**	**82**	**87**	62
		0.7	**3.4**	**3.8**	6.0	6.3	8.5	-1	-2	-5	-5	-8	3.3	3.2	3.1	3.1	3.0	**85**	78	52	47	8
	F	0.2	**5.0**	**4.9**	**6.4**	**5.1**	**5.2**	-1	-1	-3	-2	-2	4.9	4.8	5.8	4.9	4.7	**87**	**87**	**83**	80	**84**
		0.5	5.2	**3.7**	**4.5**	**4.5**	5.4	-4	-1	-3	-3	-4	3.7	3.6	3.6	3.6	3.5	73	**92**	**82**	80	63
		0.7	3.9	**2.5**	**3.5**	3.8	6.0	-2	0	-2	-3	-6	3.2	2.5	2.5	2.6	2.3	76	**92**	**82**	74	26
	G	0.2	**5.8**	**4.9**	**4.7**	**4.8**	**5.2**	-3	-1	0	-1	-3	4.8	4.8	4.7	4.7	4.5	77	**88**	**91**	**87**	**83**
		0.5	5.8	**3.8**	**3.8**	**3.8**	4.9	-4	1	1	0	-3	4.0	3.7	3.7	3.8	3.8	69	**87**	**87**	**87**	**84**
		0.7	5.0	**3.0**	**3.0**	**3.0**	5.4	-3	1	1	-1	-5	3.8	2.9	2.9	2.9	2.8	71	**85**	**91**	**91**	42

Table 6: Root mean squared error (RMSE), bias and standard error of a point estimate for **Kendall's tau** (for $S = 100$ simulated datasets) for different combinations for the true copula (with small, medium or large Kendall's tau) when the dependence structure is modelled using a parametric copula (Clayton, Frank, Gumbel or normal) or a nonparametric (NP) method. The first and second significantly smallest values for RMSE are, respectively, bolded and underlined. The last column gives the coverage of the 90% credible interval for Kendall's tau (bold values being compatible with the nominal level) obtained from the chain for the dependence parameter under a parametric copula assumption or from the M densities generated by MCMC for a given dataset under the NP method. Sample size n with **8 big bins** of equal width along each axis.

	True copula	Assumed copula																			
		τ	C	F	G	N	NP	C	F	G	N	NP	C	F	G	N	NP	C	F	G	N	NP
			$10^2 \times$ RMSE					$10^2 \times$ BIAS					$10^2 \times$ s.e.					Coverage of 90% C.I.				
$n = 1000$	C	0.2	**2.4**	**2.9**	5.2	**2.8**	**2.7**	0	-2	-5	-2	-1	2.4	2.4	2.5	2.4	2.4	**87**	**82**	36	**84**	**86**
		0.5	**2.0**	3.0	5.4	4.2	3.5	0	-2	-5	-4	-3	2.0	1.9	2.0	1.9	1.9	**85**	62	19	28	46
		0.7	**1.8**	3.6	6.7	5.8	4.7	0	-3	-7	-6	-4	1.8	1.6	1.6	1.5	1.6	**82**	27	0	2	11
	F	0.2	**2.8**	**2.3**	3.0	**2.2**	**2.3**	-2	0	-2	-0	0	2.3	2.3	2.3	2.2	2.2	**82**	**89**	79	**88**	**89**
		0.5	2.9	**1.9**	3.3	2.8	2.5	-2	0	-3	-2	-2	2.1	1.9	1.9	1.8	1.8	66	**86**	57	64	68
		0.7	**1.8**	**1.3**	2.9	2.8	3.2	-0	0	-3	-2	-3	1.8	1.3	1.5	1.5	1.3	**83**	**90**	46	45	25
	G	0.2	4.0	**2.4**	**2.2**	**2.5**	**2.4**	-3	0	0	1	0	2.6	2.4	2.2	2.3	2.4	59	**90**	**93**	**84**	**87**
		0.5	3.9	**1.8**	**1.8**	**1.8**	**2.2**	-3	0	0	0	-1	2.2	1.8	1.8	1.8	1.7	54	**91**	**93**	**92**	79
		0.7	1.9	**1.4**	**1.4**	**1.5**	2.9	-1	0	0	0	-3	1.8	1.4	1.4	1.4	1.4	83	**90**	**93**	**91**	41
$n = 200$	C	0.2	**5.5**	**5.4**	8.3	**5.3**	**5.5**	3	-1	-4	-1	-2	4.9	5.3	7.2	5.2	4.9	**88**	**86**	75	**88**	**82**
		0.5	**3.9**	**4.4**	6.1	**5.4**	6.9	1	-2	-4	-4	-6	3.8	3.9	4.2	3.9	3.8	**93**	**84**	74	77	58
		0.7	**3.8**	**4.7**	7.4	7.5	10.0	-1	-3	-6	-7	-10	3.8	3.4	3.7	3.6	3.1	**85**	70	40	36	4
	F	0.2	**5.4**	**5.2**	**7.2**	**5.3**	**5.5**	0	-1	-3	-2	-3	5.4	5.1	6.5	5.1	4.8	**86**	**89**	**82**	**88**	81
		0.5	**4.6**	**4.2**	**4.7**	**4.5**	5.9	-1	-1	-2	-2	-5	4.5	4.1	4.1	4.0	3.7	**85**	**88**	**85**	**82**	65
		0.7	**4.3**	**3.1**	**3.8**	**4.1**	7.0	1	1	-2	-3	-6	4.0	3.1	3.4	3.2	2.7	81	**92**	**86**	79	24
	G	0.2	**6.2**	**5.5**	**5.6**	**5.4**	**5.9**	-2	-1	0	0	-3	5.9	5.4	5.6	5.4	5.1	**85**	**86**	**88**	**88**	**84**
		0.5	**4.5**	**4.0**	**4.2**	**4.1**	**5.2**	-2	0	1	0	-4	4.3	3.9	4.1	4.0	3.9	**91**	**91**	**88**	**88**	75
		0.7	**4.3**	**3.4**	**3.4**	**3.2**	7.0	-1	0	0	-1	-6	4.1	3.4	3.4	3.0	2.9	**83**	**89**	**89**	**90**	30

Table 7: Root mean squared error (RMSE), bias and standard error of a point estimate for **Kendall's tau** (for $S = 100$ simulated datasets) for different combinations for the true copula (with small, medium or large Kendall's tau) when the dependence structure is modelled using a parametric copula (Clayton, Frank, Gumbel or normal) or a nonparametric (NP) method. The first and second significantly smallest values for RMSE are, respectively, bolded and underlined. The last column gives the coverage of the 90% credible interval for Kendall's tau (bold values being compatible with the nominal level) obtained from the chain for the dependence parameter under a parametric copula assumption or from the M densities generated by MCMC for a given dataset under the NP method. Sample size n with **6 big bins** of equal width along each axis.

For medium ($\tau = 0.5$) or large ($\tau = 0.7$) Kendall's tau, the nonparametric estimate tends to underestimate τ. Whatever the number of big bins, that bias is moderate (and at most -0.04) when the sample size is $n = 1000$. It tends to increase with τ. The same remark holds when $n = 200$ provided that the number of big bins is a least 8 and τ small or medium. For a small sample size ($n = 200$) and very wide bins, the estimation provided by a carefully chosen parametric copula seems preferable. However, note that many different shapes for a density are compatible with the observed frequencies in such a setting. The smoothness prior in the nonparametric approach penalizes quick changes in the curvature of the (log) density. Therefore, that approach favours 'slow' changes within the big rectangular cells: it tends to discard densities with a large Kendall's tau when fitted densities with a smaller dependence are perfectly compatible with the observed (big cell) frequencies.

6 Application

6.1 Association between blood pressures

The data in Table 8 give the systolic and diastolic blood pressures of 663 males aged between 45 and 62 (mean 52.5 and standard deviation 4.8) in a grouped format. The storage of these data just requires the labels and (here, 57) frequencies of the non zero cells. The practical advantages are obvious when it comes to store large databases or to ensure privacy in the manipulation of health data.

SBP \ **DBP**	(40,60]	(60,70]	(70,80]	(80,90]	(90,100]	(100,110]	(110,120]	(120,170]
(80,100]	1	3	1					
(100,110]	1	14	14	3				
(110,120]		15	41	16				
(120,130]		7	51	56	7			
(130,140]		3	38	59	24	2		
(140,150]	1	3	6	42	39	9	1	
(150,160]		1	12	22	35	17	1	
(160,170]				8	7	11	3	
(170,180]		1		4	15	14	8	
(180,190]		1		1	4	8	4	1
(190,200]				1	3	1		6
(200,210]						2	1	1
(210,280]						2	4	7

Table 8: Contigency table of the diastolic (DBP) and systolic (SBP) blood pressures (in mmHg) of 683 males. Empty cells correspond to zero frequencies.

An estimate of the bivariate density of the blood pressures was obtained

Model	Clayton	Frank	Gumbel	Normal	Nonparametric
Effective dim.	8.0	9.0	9.4	8.8	14.5
BIC	317	221	**178**	194	190

Table 9: BIC and effective dimension for the blood pressure dataset.

using the (nonparametric) bivariate composite link model described in Section 2.2. The probability mass was assumed to lie within the rectangle $(40, 170) \times (80, 280)$ in the diastolic-systolic blood pressure space. Each axis was subdivided into, respectively, 65 and 100 small bins of length 2 mmHg, yielding a partition of the rectangle into 6500 squares of equal area. A basis involving 10 B-splines associated to equally-spaced knots was taken along the two axes. Starting from a frequentist estimation of the 100 B-spline parameters with a 3rd order penalty along each direction, a first chain of length 50000 was built to get better initial values for the 2 penalty parameters, the B-spline parameters and the sphering matrix. Then, after a burnin of 10000 iterations, a final chain involving 200000 iterations was run. It took about 2 minutes on an 2.2 Ghz Apple MacBook. We refer to Section 4.2 for algorithmic details.

Parametric copula models (Clayton, Frank, Gumbel and Normal) with flexible margins, see Section 2.1, were also considered. The same rectangle support as for the nonparametric model was assumed in the diastolic-systolic blood pressure space. It was also partitioned into 6500 squares of equal area. A basis involving 20 B-splines was taken along each axis. Initial values were obtained for the 40 splines parameters using frequentist arguments, see Section 4.1. A chain of length 100000 was run after a burnin of 10000 iterations, see Section 4.2 for algorithmic details. It took about 20 seconds on the same computer.

Model selection was made using BIC, as justified by the results of Section 5. Table 9 suggests to select a Gumbel copula with flexible marginal distributions. A plot of the estimated posterior mean of the modelled bivariate density estimate is given in Fig. 1. One minus the labels of the contours give (approximately) the estimated probability that the diastolic and systolic blood pressures (of a randomly selected male from the studied population) belong to the corresponding region.

Conditional quantiles can be computed from the estimated bivariate density. The fitted conditional deciles are given on Fig. 2 together with the frequency data. By construction, the fitted curves do not cross as it sometimes happen with nonparametric estimation methods of quantile curves. Credible intervals can also be easily derived from the MCMC chain.

6.2 Age of spouses at marriage

The data of interest are the number of marriages in Belgium (in 2006) for given ages of the spouses when the husband already divorced, see Table 10.

Like in the previous example, non-parametric (see Section 2.2) and semi-parametric copula (see Section 2.1) models were used to describe the joint density of the ages of the spouses at marriage. The probability mass was assumed

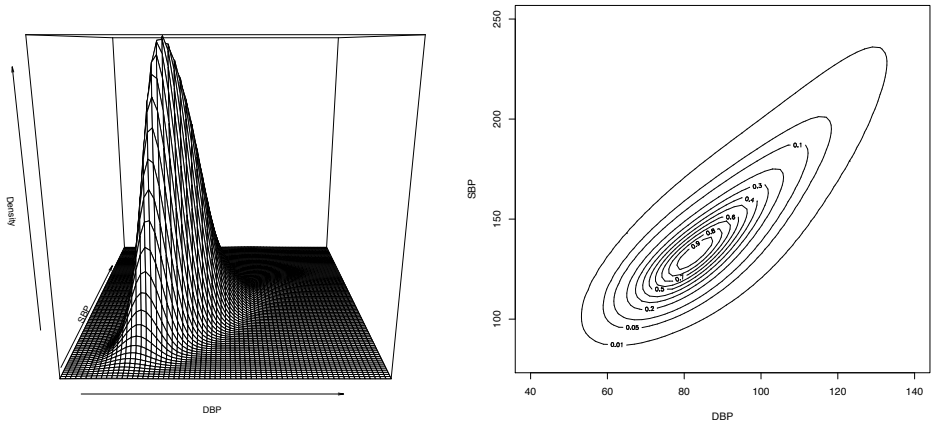

Figure 1: Fitted density based on the grouped blood pressure data.

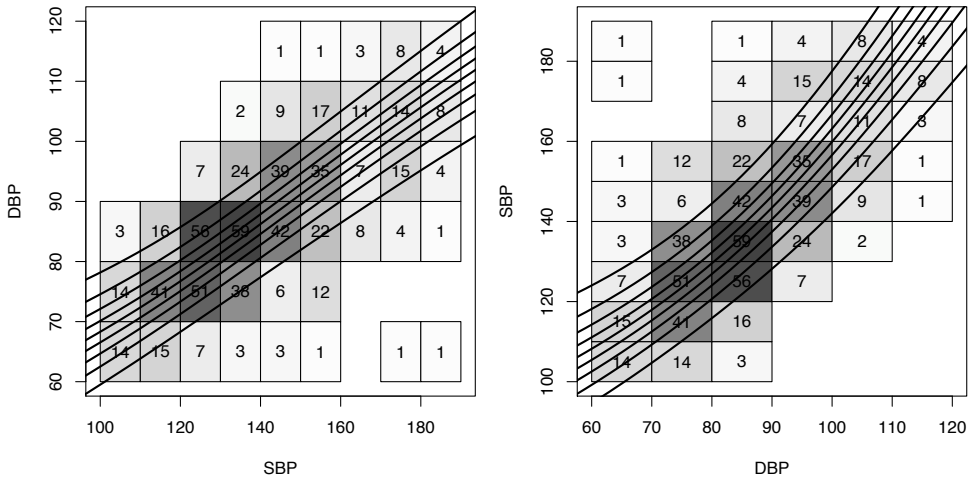

Figure 2: Conditional deciles estimated from the grouped blood pressure data.

Age of husband	Age of wife											Total	
	18-19	20-24	25-29	30-34	35-39	40-44	45-49	50-59	60-69	70-79	80+		
18-19	0	0	0	0	0	1	0	0	0	0	0	1	0.0%
20-24	8	18	10	2	0	0	0	0	0	0	0	38	0.3%
25-29	17	119	120	41	17	1	2	1	0	0	0	318	2.7%
30-34	16	175	365	333	162	47	10	2	0	0	0	1110	9.3%
35-39	18	133	394	650	519	233	69	14	1	0	0	2031	17.0%
40-44	13	79	257	465	637	556	230	83	2	0	0	2322	19.4%
45-49	1	44	125	265	438	565	526	268	5	2	0	2239	18.7%
50-59	3	27	94	158	265	552	711	969	89	1	0	2869	24.0%
60-69	0	5	12	21	45	74	136	393	167	14	2	869	7.3%
70-79	1	0	0	2	1	5	14	44	49	33	2	151	1.3%
80+	0	0	0	0	2	1	0	5	5	5	2	20	0.2%
Total	77	600	1377	1937	2086	2035	1698	1779	318	55	6	11968	
	0.6%	5.0%	11.5%	16.2%	17.4%	17.0%	14.2%	14.9%	2.7%	0.5%	0.1%		100%

Table 10: Age of spouses at marriage in Belgium (2006) when the husband already divorced (Versonnen, 2009).

Model	Clayton	Frank	Gumbel	Normal	Nonparametric
Effective dim.	10.8	14.4	13.6	12.8	27.9
BIC	2485	971	696	833	**373**

Table 11: BIC and effective dimension for the age at marriage dataset.

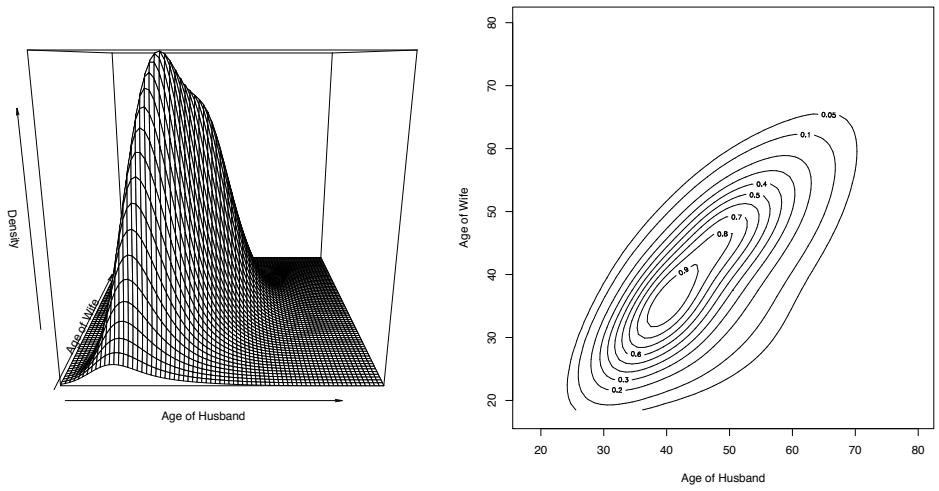

Figure 3: Fitted density for the age at marriage dataset.

to lie within the rectangle $(18, 90) \times (18, 90)$. Each axis was subdivided into 72 small bins of length 1 year, yielding a partition of the rectangle into 5184 squares of equal area. The number of B-splines along each axis and the number of MCMC iterations were taken to be 10 (resp. 20) and 200000 (resp. 100000) in the non-parametric (resp. semi-parametric) setting.

The values of BIC in Table 11 suggests that the non-parametric model is to be preferred. One explanation for the bad performances of the parametric copula models might be the implicit hypothesis of symmetry of the conditional quantile distributions on the quantile scale: these are assumed identical for men and for women.

A plot of the fitted non-parametric bivariate density is given in Fig. 3. One minus the labels of the contours give (approximately) the estimated probability that the age of the spouses belong to the corresponding region.

The fitted conditional deciles are given on Fig. 4 together with the frequency data. The dispersion of the age of the wifes markedly increase with that of the husband (see right panel), while the dispersion of the age of the husbands looks stable as soon as the wife is 35 or older (see left panel). The plot of the estimated conditional standard deviation in Fig. 5 confirms the diagnostic. It also suggests that 42 is a turning point for the age of the partner as, over it, the standard deviation of the age of the husbands is larger than that of the wifes.

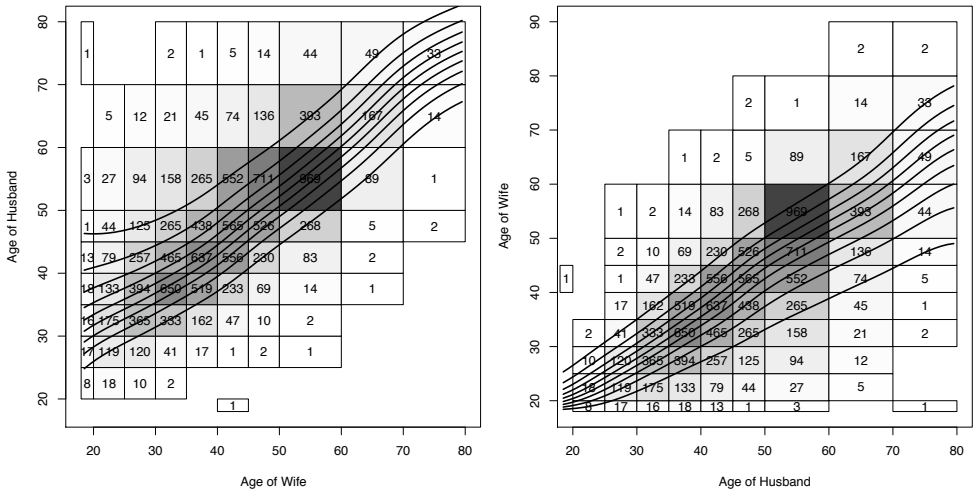

Figure 4: Estimated conditional deciles for the age at marriage dataset.

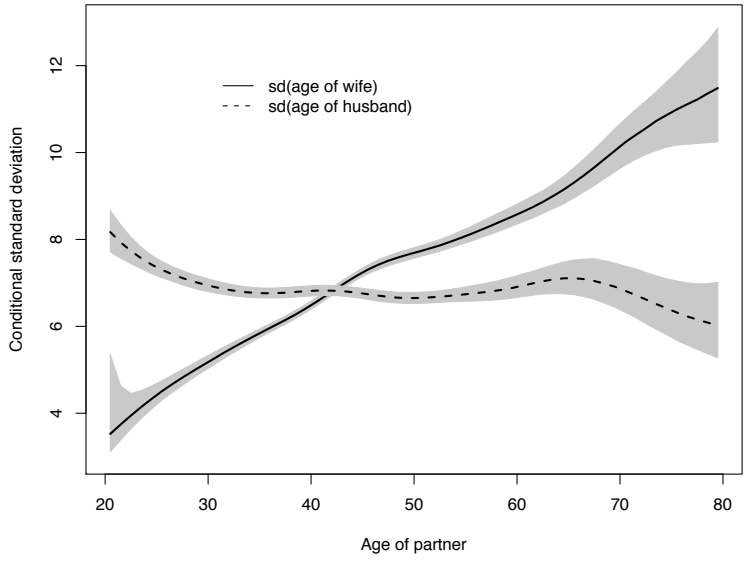

Figure 5: Conditional standard deviation of the age of the husbands (solid line) or of the wifes (dashed line) for a given age of his or her partner. Grey regions correspond to pointwise 90% credible intervals.

7 Discussion

We have proposed a semiparametric and a nonparametric model to estimate a bivariate density from bivariate histogram data. Smoothness was forced through roughness penalties or smoothness priors in both cases with the additional assumption of a parametric copula to describe the dependence structure in the first model. P-splines (Eilers and Marx, 1996) and the composite link model (Thompson and Baker, 1981) are the key components in the strategy.

Frequentist and Bayesian estimates were derived. The frequentist approaches are easier to implement and are actually used as a first step in the MCMC exploration of the joint posterior in the corresponding Bayesian implementation. The big advantages of the full Bayesian versions are the automatic account for the different sources of uncertainty (such as the choice of the roughness penalty parameters) and the ease of computation of credible regions for any derived quantity of interest such as (conditional posterior) moments, quantiles or probabilities once the Markov chains have been generated.

The simulation study suggests that BIC can be used to classify the relative merits of these semiparametric and parametric models. Relying on this, the two applications have shown that both the semiparametric and the nonparametric models are useful practice.

Extending our models to deal with bivariate interval-censored data is straightforward. Indeed, one just needs to amend the likelihood: it becomes a product over the different observed rectangular configurations formed by the pairs of intervals, each one contributing to the likelihood with the associated probability to a power equal to the number of times it was observed. Then, similarly as before, the jth row of the composing matrix for a given margin indicates which small bins are contained in the jth marginal interval configuration.

Acknowledgements

Financial supports from the FSR research grant nr FSRC-08/42 from the University of Liège and of the IAP research network nr. P6/03 of the Belgian government (Belgian Science Policy) are gratefully acknowledged.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

Betensky, R. and Finkelstein, D. (1999). A non-parametric maximum likelihood estimator for bivariate interval censored data. *Statistics in Medicine*, **18**(22), 3089–3100.

Braun, J., Duchesne, T., and Stafford, J. E. (2005). Local likelihood density estimation for interval censored data. *The Canadian Journal of Statistics*, **33**, 39–60.

Eilers, P. H. C. (1992). Nonparametric density estimation with grouped observations. *Statistica Neerlandica*, **45**, 255–269.

Eilers, P. H. C. (2007). Ill-posed problems with counts, the composite link model and penalized likelihood. *Statistical Modelling*, **7**, 239–254.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties (with discussion). *Statistical Science*, **11**, 89–121.

Goggins, W. B. and Finkelstein, D. M. (2000). A proportional hazards model for multivariate interval-censored failure time data. *Biometrics*, **56**, 940–943.

Gomez, G., Calle, M. L., and Oller, R. (2004). Frequentist and Bayesian approaches for interval-censored data. *Statistical Papers*, **45**, 139–173.

Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, **7**, 223–242.

Hanson, T. E. (2006). Inference for mixtures of finite Polya tree models. *Journal of the American Statistical Association*, **101**, 1548–1565.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*. Chapman & Hall, London.

Härkänen, T., Virtanen, J. I., and Arjas, E. (2000). Caries on permanent teeth: a non-parametric bayesian analysis. *Scandinavian Journal of Statistics*, **27**, 577–588.

Komárek, A. and Lesaffre, E. (2006). Bayesian semi-parametric accelerated failure time model for paired doubly interval-censored data. *Statistical Modelling*, **6**(1), 3–22.

Komárek, A., Lesaffre, E., Härkänen, T., Declerck, D., and Virtanen, J. I. (2005). A Bayesian analysis of multivariate doubly-interval-censored dental data. *Biostatistics*, **6**, 145–155.

Koo, J.-Y. and Kooperberg, C. (2000). Logspline density estimation for binned data. *Statistics and Probability Letters*, **46**, 133–147.

Kooperberg, C. and Stone, C. J. (1992). Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics*, **1**, 301–328.

Lambert, P. and Eilers, P. H. (2009). Bayesian density estimation from grouped continuous data. *Computational Statistics and Data Analysis*, **53**, 1388–1399.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, **13**, 183–212.

Lavine, M. (1992). Some aspects of Polya tree distributions for statistical modelling. *The Annals of Statistics*, **20**, 1222–1235.

Law, C. G. and Brookmeyer, R. (1992). Effects of mid-point imputation on the analysis of doubly censored data. *Statistics in Medicine*, **11**, 1569–1578.

Minnotte, M. (1998). Achieving higher-order convergence rates for density estimation with binned data. *Journal of the American Statistical Association*, **93**, 663–672.

Peto, R. (1973). Experimental survival curves for interval-censored data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **22**, 86–91.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.

Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data*. Springer.

Thompson, R. and Baker, R. J. (1981). Composite link functions in generalized linear models. *Applied Statistics*, **30**, 125–131.

Turnbull, B. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B (Methodological)*, **38**, 290–295.

Versonnen, A. (2009). Population et ménages, mariages et divorces. Technical report, Direction Générale Statistique et Information Economique, 44 rue de Louvain, 1000 Bruxelles (Belgique).

Wong, M. C. M., Lam, K. F., and Lo, E. C. M. (2005). Bayesian analysis of clustered interval-censored data. *Journal of dental research*, **84**(9), 817–821.

Yang, M., Hanson, T., and Christensen, R. (2008). Nonparametric Bayesian estimation of a bivariate density with interval censored data. *Computational Statistics and Data Analysis*, **52**, 5202–5214.

Zhang, M. and Davidian, M. (2008). "Smooth" semiparametric regression analysis for arbitrarily censored time-to-event data. *Biometrics*, **64**(2), 567–576.